

 现代统计学系列丛书

统计分布

方开泰 许建伦 编著

Statistical Distribution

高等教育出版社

代统计学系列丛书

统计分布

Statistical Distribution

方开泰 许建伦 编著

T o n g j i F e n b u

高等教育出版社·北京

内容提要

统计分布是数理统计的基础,数理统计的理论和方法均离不开统计分布。

本书重点介绍常见的统计分布(包括离散型随机变量的分布、连续型随机变量的分布、多元分布等),详细地介绍了它们的性质、计算方法及相互之间的关系,并用大量的例子介绍这些分布的实际应用背景。

本书可供数理统计工作者、从事实际工作的技术人员、理工科师生阅读。

图书在版编目(CIP)数据

统计分布 / 方开泰, 许建伦编著. -- 北京: 高等教育出版社, 2016. 7

(现代统计学系列丛书)

ISBN 978-7-04-045501-4

I. ①统… II. ①方… ②许… III. ①统计分布 - 高等学校 - 教材 IV. ①O212.2

中国版本图书馆 CIP 数据核字 (2016) 第 103169 号

策划编辑 张晓丽
插图绘制 杜晓丹

责任编辑 张晓丽
责任校对 胡美萍

封面设计 赵 阳
责任印制 韩 刚

版式设计 张 杰

出版发行 高等教育出版社
社 址 北京市西城区德外大街4号
邮政编码 100120
印 刷 北京东君印刷有限公司
开 本 787 mm × 960 mm 1/16
印 张 18.75
字 数 340 千字
购书热线 010-58581118
咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.hepmall.com.cn>
<http://www.hepmall.com>
<http://www.hepmall.cn>

版 次 2016 年 7 月第 1 版
印 次 2016 年 7 月第 1 次印刷
定 价 33.20 元

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换
版权所有 侵权必究
物 料 号 45501-00

现代统计学系列丛书编委会

(按姓氏笔画排序)

主 编: 方开泰

副主编: 史宁中 何书元 陈 敏 耿 直

编 委: 马 洪 方开泰 史宁中 杨 虎 何书元 何晓群

张爱军 张崇岐 陈 敏 郑 明 赵彦云 耿 直

曾五一 缪柏其

总 序

统计学是一门收集、整理和分析数据的科学和艺术。这里的“数据”泛指“信息的载体”，涵盖了大千世界中的文本、图像、视频、时空数据、基因数据等。统计学是一门独立的学科，在历史上曾隶属于数学，但统计学与数学有着本质的区别，因此统计学教育有其自身的特点和要求，这些特点表现为：(1) 统计学研究的是随机现象，而数学研究的是确定性的规律；(2) 统计学是一门应用性很强的学科，许多概念和原理来自于实际的需要，不是数理逻辑的产物；(3) 数据在统计学中扮演了重要的角色。目前，统计学已被列为一级学科。

在过去的 30 年中，随着生命科学、信息科学、物质科学、资源环境、认知科学、工程技术、经济金融和人文科学等众多学科的发展，产生了许多新的统计学分支，如风险管理、数据挖掘、基因芯片分析等。此外，计算机及其有关软件在统计教育和应用中扮演了越来越重要的角色，它们提供了越来越多的图形表达和分析的方法，使得许多原来教科书中重要的内容，现在已变得无足轻重。统计教育必须要改革才能适应高速发展的形势。

大学的统计教育可分为两大类，一类是非统计学专业的课程，另一类是统计学专业的教学设计。非统计学专业的学生学习统计的目的是为了应用，在大学阶段，课程不多，主要是学习基础的统计概念和方法，学会使用统计软件，培养其解决实际问题的能力。统计学专业的课程设置十分重要，应向国际靠拢，对教师队伍的要求也较高。虽然这两类学生的教育有很多共同点，但在课程设置中必须加以区分。

我国的统计教育在过去受苏联的影响很深，把统计学作为数学的一个分支，在内容上偏理论，少应用，过于强调概率论在统计中的作用。统计学是一门应用性很强的学科，应从实际问题、从数据出发，通过统计的工具来揭示数据内部的规律。用“建模”的思路来教统计，使学生能更加容易理解统计的概念和方法，知道如何将实际问题抽象为统计模型，反过来又指导实践。对非统计学专业的学生，要强调统计的应用。学生要能熟练地使用至少一个统计软件包。对于统计学专业的学生，要培养学生对实际问题的建模能力。有些实际问题可直接应用现有的统计方法来解决，如问卷调查的统计分析。有些问题在初次接触时并不像一个统计问题，必须有坚实的统计基础和对实际问题的洞察力，才能从中发掘出统计

模型。要培养学生的这种能力及统计思想(统计思想是统计文化的一部分,是用统计学的逻辑思考问题)。教师在授课中要结合较多的应用例子,要求学生做案例研究,鼓励学生参加建模比赛,参加企业的实际项目。

为满足我国统计教育发展的需要,我们计划编写一套面向高校本科生、特别是一般院校,适用于统计学专业和非统计学专业的系列教材。系列教材的编写宗旨是:突出教学内容的现代化,重视统计思想的介绍,适应现代统计教育的特点及时代发展的新要求;以统计软件为支撑,注重统计知识的应用;内容简明扼要,生动活泼,通俗易懂。编写原则为:(1)从数据出发,不是从假设、定理出发;(2)从归纳出发,不是从演绎出发;(3)强调案例分析;(4)重统计思想的阐述,弱化数学证明的推导。系列教材分为两个方向,一个面对统计学专业,另一个面对非统计学专业和应用统计工作者。

系列教材是适应形势的要求,由高等教育出版社邀请专家组成“现代统计学系列丛书编委会”负责选题、审稿,由高等教育出版社出版。

以上是我们编写这套教材的背景和理念,希望得到读者的支持,特别是高校领导和教学一线教师的支持。我们希望使用这套教材的师生和读者多提宝贵意见,使教材不断完善。

现代统计学系列丛书编委会



前 言

统计学是专门研究如何有效地对受随机影响之数据的收集和分析,并对与数据有关的问题进行解释、推断和预测的一门应用性非常强的学科。这门学科覆盖的领域很广,如医学、心理学、教育学、社会学、工程学和物理学等,但不管哪一个领域,我们都会涉及数据的统计分布,因此,统计学的基础是统计分布,它在统计学中的地位特别重要。《统计分布》一书于1987年出版后,得到广大读者的喜爱、鼓励及支持,并在1992年获得中华人民共和国国家统计局优秀统计教材奖。本书的第一作者在“概率统计”课程的多年教学中,一直把《统计分布》作为主要参考书,并发现此书对学生的课程论文的撰写有极大的帮助。本书第二作者在美国,也常把《统计分布》一书中的一些有趣例子用到他的教学中。

《统计分布》原版由科学出版社于1987年出版。由于当时计算机及统计软件在中国几乎见不到,因此原版提供了许多关于分布的近似计算方法。在过去的三十年,统计学有了突飞猛进的发展,在中国也受到了高度的重视,已被列为一门一级学科,和数学、物理学、化学等并列。同时,计算机及统计软件的使用在国内,特别是高等学校已相当普及。使用统计软件可以更方便有效地进行统计分析,如作图、计算常用分布的基本统计量和进行数据建模(即分布拟合检验)。因此,在这一版中我们增加了用统计软件R来进行有关的统计分析,删除了原书中的许多近似算法。另外,在统计分布的性质及应用方面,我们也做了适当的修订。

作为现代统计学系列丛书之一,本书的目的在于介绍各种常见的分布和它们的性质,并对其中大部分分布给出实际应用的例子,使读者不仅掌握统计分布的基本理论,而且能立即用于实际之中。本书收集的分布较多,性质也较全,兼有“统计分布手册”性质,可供教数理统计的教师、学生、工程技术人员、科学工作者等查阅。为了适应更多的读者,本书在数学上只要求读者具有初等微积分和线性代数的知识。书中个别内容需要较多的数学知识(如书中带*部分),有困难的读者可以跳过不读,并不影响对全书主要内容的理解。为了照顾不懂实变函数的读者,我们将离散型随机变量和连续型随机变量分开进行处理,而不是用分布函数的工具来统一处理。统计分布的内容极其丰富,而本书只是一本入门书,只能选择最基本的内容,所以挂一漏万在所难免,请读者原谅。

全书共分六章。第一章引进阅读本书所需的预备知识,第二章介绍各种离散

型统计分布,第三、四章对常用的连续型统计分布逐一进行讨论。第二、三章及第四章的各节之间有一定的独立性,不一定完全按次序去读。鉴于本书篇幅的限制,只能给出少量的统计用表,其余的表可用R软件直接计算。另外,在实际中经常需要验证某组样本是否来自一特定的分布,这需要作分布拟合检验。第五章讨论了分布拟合检验的各种方法。最后一章扼要地讨论了几个常用的多元分布,由于多元分布所需的基础知识较多,大部分性质均不予证明。

在本书的编写及修改过程中,我们参照了许多已出版的书籍,其中文献 [1]–[8] 对我们的帮助很大,本书有些内容和例子是从这些文献中吸取的,在此向这些文献的作者表示感谢。同时,本书第一作者感谢 2015 广东省“质量工程”高等学校精品教材基金 R5201504 的支持。另外,作者十分感谢高等教育出版社的大力支持,特别要感谢张晓丽编辑的仔细编辑和核对。最后,两位作者谢谢家人的支持和理解。

方开泰(中国科学院数学与系统科学研究院,北京师范大学—香港浸会大学联合国际学院)

许建伦(美国马里兰大学数学系)

2016 年 1 月

郑重声明

高等教育出版社依法对本书享有专有版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其行为人将承担相应的民事责任和行政责任；构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话 (010) 58581897 58582371 58581879

反盗版举报传真 (010) 82086060

反盗版举报邮箱 dd@hep.com.cn

通信地址 北京市西城区德外大街4号 高等教育出版社法务部

邮政编码 100120

目 录

第一章 预备知识	1
1.1 事件和概率	1
1.1.1 随机现象及其统计规律性	1
1.1.2 事件的运算	2
1.1.3 概率及其公理化定义	5
1.1.4 条件概率	8
1.1.5 独立性	11
1.2 随机变量及其分布	12
1.2.1 随机变量的定义	12
1.2.2 离散型随机变量	14
1.2.3 连续型随机变量	15
1.2.4 分布函数	16
1.3 随机变量的特征数	19
1.3.1 数学期望	19
1.3.2 方差	22
1.3.3 高阶矩	24
1.3.4 众数、分位点和中位数	26
1.4 矩母函数与特征函数	27
1.5 随机向量及其分布	30
1.5.1 随机向量	30
1.5.2 分布函数	31
1.5.3 边缘分布	32
1.5.4 独立性	33
1.5.5 条件分布	34
1.5.6 矩	36
1.5.7 特征函数	39
1.6 随机变量函数的分布	39
1.7 分布参数的估计和检验	43
1.8 计算统计分布的一些 R 函数	45

第二章 离散型随机变量的分布	48
2.1 两点分布	48
2.2 二项分布	49
2.2.1 定义	49
2.2.2 性质	50
2.2.3 应用	54
2.2.4 近似计算	59
2.3 泊松分布	62
2.3.1 定义	62
2.3.2 性质	62
2.3.3 应用	66
2.4 超几何分布	69
2.4.1 定义	69
2.4.2 性质	69
2.4.3 应用	72
2.4.4 近似计算	74
2.5 几何分布	74
2.5.1 定义	74
2.5.2 性质	75
2.5.3 应用	78
2.6 负二项分布	80
2.6.1 定义	80
2.6.2 性质	81
2.6.3 应用	85
2.7 一些其他离散分布	86
2.7.1 离散型均匀分布	86
2.7.2 负超几何分布	88
2.7.3 截塔 (Zeta) 分布	88
2.7.4 截尾泊松分布	90
2.7.5 幂级数分布	91
*2.8 缸的模型和占有问题	92
*2.9 求离散型分布矩的一种方法	95
2.9.1 有限差算子	96
2.9.2 一些离散型随机变量的矩	98
第三章 正态分布及其有关的分布	102
3.1 正态分布	102
3.1.1 定义	102

3.1.2	性质	103
3.1.3	应用	110
3.2	对数正态分布	113
3.2.1	定义	113
3.2.2	性质	114
3.2.3	应用	116
3.3	χ^2 分布和 χ 分布	117
3.3.1	定义	117
3.3.2	性质	118
3.3.3	应用	122
3.3.4	χ 分布	124
3.4	t 分布	125
3.4.1	定义	125
3.4.2	性质	126
3.4.3	应用	128
3.5	F 分布	132
3.5.1	定义	132
3.5.2	性质	132
3.5.3	应用	136
*3.6	χ^2 分布、t 分布和 F 分布密度的推导	138
3.6.1	χ^2 分布密度的推导	138
3.6.2	t 分布密度的推导	139
3.6.3	F 分布密度的推导	141
*3.7	非中心 χ^2 分布	142
3.7.1	定义	143
3.7.2	性质	143
3.7.3	应用和计算	146
3.8	非中心 t 分布	147
3.8.1	定义	147
3.8.2	性质	147
3.8.3	应用和分布函数的计算	148
3.9	非中心 F 分布	149
3.9.1	定义	149
3.9.2	性质	150
3.9.3	应用和分布函数的计算	151

第四章 连续型随机变量的分布	152
4.1 有限区间上的均匀分布	152
4.1.1 定义	152
4.1.2 性质	152
4.1.3 应用	155
4.2 韦布尔分布	159
4.2.1 定义	159
4.2.2 性质	160
4.2.3 应用	164
4.3 伽玛分布	169
4.3.1 定义	169
4.3.2 性质	170
4.3.3 应用	175
4.4 贝塔分布	177
4.4.1 定义	177
4.4.2 性质	178
4.4.3 应用	184
4.5 幂函数分布	187
4.5.1 定义	187
4.5.2 性质	187
4.6 柯西分布	189
4.6.1 定义	189
4.6.2 性质	190
4.6.3 应用	192
4.7 逻辑斯谛分布	193
4.7.1 定义	193
4.7.2 性质	194
4.7.3 应用	197
4.8 极值分布	197
4.8.1 定义	197
4.8.2 性质	198
4.8.3 应用	200
4.9 拉普拉斯分布	201
4.9.1 定义	201
4.9.2 性质	201
4.9.3 应用	204
4.10 最大似然估计	205

第五章 分布拟合检验	208
5.1 皮尔逊的 χ^2 检验	208
5.2 经验分布函数	215
5.3 柯尔莫哥洛夫检验、斯米尔诺夫检验	217
5.3.1 柯尔莫哥洛夫检验	217
5.3.2 斯米尔诺夫检验	221
*5.4 A^2 和 W^2 检验	224
5.4.1 A^2 和 W^2 统计量	224
5.4.2 计算公式	225
*5.5 参数未知的 A^2 和 W^2 检验	227
5.5.1 正态分布	227
5.5.2 韦布尔分布	228
5.5.3 极值分布	230
5.5.4 逻辑斯谛分布	231
5.5.5 柯西分布	232
5.6 正态性检验	238
5.6.1 正态概率纸的构造	238
5.6.2 正态概率纸的使用方法	239
5.6.3 偏斜系数和峰态系数法	245
第六章 多元分布	247
6.1 多项分布	247
6.1.1 定义	247
6.1.2 性质	248
6.1.3 应用	250
*6.2 多元超几何分布	251
6.2.1 定义	251
6.2.2 性质	252
6.2.3 应用	254
*6.3 多元负二项分布	254
6.3.1 定义	254
6.3.2 性质	255
6.4 多元正态分布	257
6.4.1 定义	257
6.4.2 性质	258
6.4.3 应用	262
*6.5 狄利克雷分布	265
6.5.1 定义	265

6.5.2	性质	266
6.5.3	应用	267
附表 1	正态分布表	269
附表 2	卡方 (χ^2) 分布表	270
附表 3	t 分布表	272
附表 4	F 分布表	273
索引		279
参考文献		283

第一章

预备知识

本书的目的是介绍各种常见的分布及其应用,为此需要一些概率论和数理统计的预备知识,这些知识对理解全书非常重要.本章将介绍:事件、概率、随机变量、分布函数、分布密度、矩和各种母函数等.将上述这些内容推广至多个随机变量——随机向量的情形(见1.5节),也是应随后几章的需要而设立的.1.5节数学形式稍微复杂一些,数学基础不够的读者可跳过该节,先读第二章,待今后需要时再回过头来读它.那时有了许多具体分布的背景,再读这一节就不会有太多的困难.在今后几章中涉及一些统计应用,这就需要知道总体、样本等概念,为此在1.7节我们简单地介绍这些概念,以及参数估计和假设检验的提法.

1.1 事件和概率

1.1.1 随机现象及其统计规律性

在自然界和人类社会中,有许多现象,我们完全可以预言它们在一定条件下是否出现.例如“同性的电互相排斥”,“在没有外力作用的条件下,作等速直线运动的物体必然继续作等速直线运动”等等是一定会出现的;而上述现象的反面,即“同性的电互相吸引”,“在不受外力作用的条件下,作等速直线运动的物体改变其等速直线运动状态”等等是必然不会出现的.这种在一定条件下一定出现的现象叫必然事件,在一定条件下必然不出现的现象叫不可能事件.显然,必然事件的反面是不可能事件.

然而,在自然界和人类社会中,还有许多现象,它们在一定条件下可能出现也可能不出现,这种现象称为随机事件,或简称事件.例如“掷一枚质地均匀的硬币,数字面(正面)向上”,这件事可能发生,但也可能发生背面向上的情形.“新生的婴儿是男孩”,这件事可能发生,但也可能发生新生的婴儿是女孩的情形.

我们常常通过随机试验来观察随机事件的统计规律性,例如:事件“正面向上”是随机试验“掷一枚质地均匀硬币”的一个可能结果,而事件“婴儿是男孩”是随机试验“婴儿出生”的一个可能结果.

一般地, 设 E 为一试验, 如果不能事先准确地预言它的结果, 而且在相同条件下可以重复进行, 就称 E 为随机试验. 以 ω 表示它的一个可能的结果, 称 ω 为 E 的一基本事件. 全体基本事件的集合 $\Omega = \{\omega\}$ 称为基本事件空间或样本空间.

例 1.1 E 表示掷一枚质地均匀的硬币而观察所出现的面, ω_1 表示正面 (数字面), ω_2 表示背面, 于是 $\Omega = \{\omega_1, \omega_2\}$.

例 1.2 E 表示在一个箱子里有 10 个球, 上面分别标以 $1, 2, \dots, 10$, 若从箱子里随机地取一个球, 令 ω_i 表示球上的数字是 i , 则 $\Omega = \{\omega_1, \dots, \omega_{10}\}$. 如果简记 ω_i 为 i , 则得 $\Omega = \{1, \dots, 10\}$.

例 1.3 E 表示计算某电话交换台在某段时间内所接到的呼唤次数, 若 ω_i 表示正好接到 i 次呼唤, 则 $\Omega = \{\omega_0, \omega_1, \dots, \omega_n, \dots\}$. 如果简记 ω_i 为 i , 则 $\Omega = \{0, 1, \dots, n, \dots\}$.

基本事件是事件中的一种, 一般的事件总是由若干基本事件共同组成的, 因而是 Ω 的一个子集. 例如, 在例 1.2 中, 事件 A : “球上的数字不大于 3” 是由三个基本事件组成, 即 $A = \{\omega_1, \omega_2, \omega_3\}$. 同样, 对于事件 B : “球上的数字为偶数”, 即有 $B = \{\omega_2, \omega_4, \omega_6, \omega_8, \omega_{10}\}$. 由此可见, 每一事件都对应于 Ω 的一个子集, 显然 Ω 就是必然发生的事件, 以后用 $\emptyset$ 表示不可能发生的事件.

1.1.2 事件的运算

进行一个试验, 可能发生的事件往往很多, 这些事件各有不同的特点, 但彼此之间又有一定的联系, 下面我们介绍一些事件之间的重要关系和运算.

(1) 如果事件 A 发生必然导致事件 B 发生, 就说事件 A 是事件 B 的特款, 或说 B 包含 A , 记作 $A \subset B$. 如果 $A \subset B$, 且 $B \subset A$, 就说 A 与 B 相等, 记作 $A = B$.

在例 1.2 中, 若令 $A = \{2\} = \{\text{球上的数字是 } 2\}$, $B = \{2, 4, 6, 8, 10\} = \{\text{球上的数字是偶数}\}$, 显然 $A \subset B$. 图 1.1 形象地表示 B 包含 A .

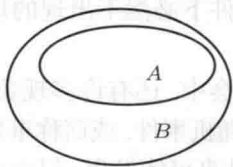


图 1.1 $A \subset B$

(2) “两事件 A 和 B 中至少有一个发生” 也是一个事件, 称此事件为 A 与 B 的和, 记作 $A \cup B$, 见图 1.2 中阴影部分.