



暨南大学经济管理实验中心实验教材

统计实验 及R语言模拟

Statistical Experiment and
Simulation with R Language

侯雅文 王斌会 编著



北京大学出版社
PEKING UNIVERSITY PRESS



暨南大学出版社
JINAN UNIVERSITY PRESS

经济管理国家实验教学示范中心 共同资助
经济管理省级实验教学示范中心



暨南大学经济管理实验中心实验教材

统计实验 及R语言模拟

Statistical Experiment and
Simulation with R Language

侯雅文 王斌会 编著

 北京大学出版社
PEKING UNIVERSITY PRESS

中国·北京

 暨南大学出版社
JINAN UNIVERSITY PRESS

中国·广州

图书在版编目 (CIP) 数据

统计实验及 R 语言模拟/侯雅文, 王斌会编著. —广州: 暨南大学出版社, 2015. 8
(暨南大学经济管理实验中心实验教材)

ISBN 978 - 7 - 5668 - 1552 - 1

I. ①统… II. ①侯…②王… III. ①统计学—实验②统计分析—程序语言—程序设计 IV. ①C8 - 33②C819

中国版本图书馆 CIP 数据核字(2015)第 155554 号

出版发行: 暨南大学出版社

地 址: 中国广州暨南大学

电 话: 总编室 (8620) 85221601

营销部 (8620) 85225284 85228291 85228292 (邮购)

传 真: (8620) 85221583 (办公室) 85223774 (营销部)

邮 编: 510630

网 址: <http://www.jnupress.com> <http://press.jnu.edu.cn>

排 版: 广州市天河星辰文化发展部照排中心

印 刷: 广东广州日报传媒股份有限公司印务分公司

开 本: 787mm × 1092mm 1/16

印 张: 12

字 数: 291 千

版 次: 2015 年 8 月第 1 版

印 次: 2015 年 8 月第 1 次

定 价: 28.00 元

(暨大版图书如有印装质量问题, 请与出版社总编室联系调换)

此为试读, 需要完整PDF请访问: www.ertongbook.com

“暨南大学经济管理实验中心实验教材”

丛书编委会

编委会主任委员：胡 军

编委会副主任委员：宋献中

编委会委员(按姓氏笔画排序)：

王斌会 王惠芬 左小德 叶文晖 刘少波 汤胤

孙 彧 李广明 范 纯 郑少智 胡矗明 侯雅文

黄微平 章 牧 梁 云 谢贤芬 谭 跃

项目负责人：谭 跃

项目策划人：王斌会

总 序

百年沧桑，弦歌不辍；巍巍暨南，展焕新颜。暨南大学自1906年创办以来，始终秉承“宏教泽而系侨情”的办学宗旨，注重以中华民族优秀的传统道德文化培养造就人才。“始有暨南，便有商科”，最初设立的商科便因兼具理论与实用的“暨南特色”而享誉海内外。经过一百多年的发展，商科已分化出经济管理学科中的许多门类，各门类分工明确而细化，又交叉融合，近年来屡屡在学科发展上有突破和创新，尤为可喜的是暨南大学经济管理实验教学中心于2012年荣获国家级实验教学示范中心项目。这是我校继2008年获批媒体实验教学中心之后，再次获得国家级实验教学示范中心项目，是教育部“质量工程”重要建设项目之一，也是质量工程中含金量较高、获批难度较大的一个项目。这些项目是高等学校实验教学研究和改革的基地，引领着全国高等学校实验教学改革的方向。

暨南大学经济管理实验教学中心（以下简称“中心”）依托产业经济学和金融学2个国家级重点学科，3个一级学科博士学位点，拥有一支以珠江学者、教学名师和知名专家带头人组成的优秀教学团队，其中“会计学教学团队”被评为国家级教学团队。中心包括金融模拟、会计模拟、ERP实验、电子商务模拟、行为分析、经济统计与分析、财税管理与分析、酒店管理等18个实验室。中心师资力量雄厚，副高级以上教师占总人数的75%，承担全校22个本科专业以及研究生、博士生相关课程的实验、实训、实习等教学任务。

中心继承和发扬暨南大学经济管理教育重视实际操作、强化能力培养的优良传统，紧贴经管发展的现实需求，全面开展“虚拟仿真实验+校企合作实践”模式的实践教学形式改革，注重能力培养与社会需求相结合的教学内容改革。实验是教学不可或缺的一个重要组成部分，作为理论教学的基础和延伸，中心始终坚持“强化基础、重视实践、个性培养与创新能力紧密结合”的实验教学理念，逐步构建理论教学、实验教学、课外实践等多维互动、整体提升的创新实验教学体系，以培养未来华商领袖为己任，着力培养具有创新能力的复合型经管专业人才，为建设成为“具有浓厚华人华侨特色，享誉海内外的‘高端、优质、创新’复合型经济管理人才培养基地”而努力。在中心全体教职员工的共同努力下，中心工作取得了显著成效，比如，开设的“财务学原理”、“基础会计学”被评为国家级精品课程，工商管理类和经济类专业被评为国家级特色专业，中心申报的教学成果项目于2010年获广东省教学成果奖一等奖等。

随着经济和科学技术的进步，尤其是计算机技术的飞速发展，数据、模型与实验对于当代科学乃至整个社会的影响和推动作用日益显著。“暨南大学经济管理实验中心实验教材”作为国家和广东省教学示范中心的资助教材，根据经济管理类专业、学科特点，实验教材中的数据、模型和例子全部选自经济、管理等方面的内容，形成了一个能反映经济管理类院校特色的“经济管理实验”系列教材。这一特色的形成，不仅对国内经济管理实验是一个突破与创新，而且对培养经济管理类院校的应用型、创新型、复合型人

才,有着积极意义。

本系列教材在总结过去教材建设经验的基础上,结合应用型本科教育的特点,借鉴国内外的经验做法,在经济管理各专业的课程体系、课程内容,教学方法、教材编写等方面进行进一步探索和创新。

本系列教材具有五个方面的特点:第一,创新性。从培养学生的兴趣入手,以掌握方法论和创造性思维为主线,以知识、概念和理论为基石,进行总体设计,思路新颖,写作体例风格独特。第二,前瞻性。搜集了最新的数据资料和理论研究成果,使教材内容着力体现超前性、前沿性、动态性。第三,实践性。体现了实验型本科教学的专业特点,以提高学生竞争力、综合素质和社会适应能力为最终目标。第四,系统性。基础知识、学科理论和课程体系融为一体,注重基础理论与实际应用的结合。第五,可读性。突出“以学生为中心”的思想,强调学以致用,所用语言浅显易懂,并附有一定的案例分析。

“暨南大学经济管理实验中心实验教材”的建设,改变了传统课程那种仅仅依赖“一支笔,一张纸”,由教师单向传输知识的模式。它提高了学生在教学过程中的参与程度,学生的主观能动性在实验中能得到相当充分的发挥。好的实验会引起学生学习科学知识和方法的强烈兴趣,并激发他们自己去解决相关实际问题的欲望,有助于促进学生独立思考和创新意识的培养。

教材建设是课程体系和教学内容改革的核心,是进一步加强学生教学工作,深化教学改革,提高学生教育教学质量的重要措施。暨南大学经济管理实验教学中心精心组织教材编写,通过专家组评审,分批立项,每批近十种,覆盖金融模拟、会计模拟、ERP实验、电子商务模拟、行为分析、经济统计与分析、财税管理与分析、酒店管理等实验课程。这些教材符合教育改革发展趋势,反映了经济管理学科建设的新理论、新技术、新方法、新实践,在国内同类教材中较为先进。我们期望通过几年的努力,打造出一系列特色鲜明的经济管理实验教材。



暨南大学校长、教授、博士生导师
国家重点学科产业经济学带头人

2014年6月

前 言

从 1777 年法国科学家 Buffon (蒲丰) 邀请客人参与投针试验, 得到关于 π 的估计, 到第二次世界大战期间, 美国科学家 Stanislaw Ulam (冯·诺依曼) 和 John von Neumann (乌拉姆) 为研究原子弹的裂变物质中子随机扩散问题, 提出代号为“蒙特卡罗”模拟方法。由于模拟能够较为逼真地描述事物特点和真实实验过程, 目前在经济、管理、金融、生物医学、物理学等诸多领域中都得到了应用和发展。

它与一般的数值计算方法有较大区别, 是以概率统计为理论基础的研究方法。通常, 根据问题建立概率统计模型, 设定模型分布和参数, 在一定精度要求下确立模拟量, 依照建立抽样方法产生随机数, 并计算模型的相关统计量结果作为模型估计。整个流程具体包括随机数产生、概率模型、统计量计算、模拟精度四大方面问题。

随机数是模拟的基石。本书第 1 章第 1.2 节详细介绍由一般均匀伪随机数产生, 到运用逆变换方法、接受拒绝法、复合法和广义变换方法产生服从各种分布的随机数方法。因为随机数产生的随机性, 所以并非每一次产生的随机数都符合随机性的要求。例如, 按照二项分布 $B(1, 0.5)$ 产生 0 和 1 两个随机数取值, 若计算机产生了 5 个随机数分别为 (0, 0, 0, 0, 0), 这也是可能的, 但不一定满足随机数均匀性、独立性和拟合效果要求。因而需通过增大随机数产生数量和第 1.3 节中提及的检验方法, 尽可能产生符合要求的随机数。对于产生一些难度较大的随机数, 本书第 4 章的第 4.1~4.2 节分别介绍利用构造马氏链, 产生随机数的 Metropolis-Hastings 和 Gibbs 方法。另外, 由于计算机的随机数起源是“伪”的, 蒙特卡罗方法的创立者冯·诺依曼说“任何人如果相信计算机能够产生出真正的随机的数数组都是疯子”。

广义的概率模型具体包括四个方面的内容, 分别为: ①概率; ②分布特征; ③大数定律; ④中心极限定理。本书第 2 章第 2.1~2.2 节运用大数定律、频率方法、期望方法给出关于积分蒙特卡罗模拟, 并讨论精度与模拟量之间的关系。在第 2.4、4.3 节通过优化方法获得关于模型分布特征估计量, 利用中心极限定理产生随机数、区间估计等, 上述概率模型在系统模拟中产生重要作用。本书第 5 章通过 8 个模拟过程, 分别展示概率实验 (第 5.1、5.2、5.4 节)、仿真实验 (第 5.3 节)、参数估计实验 (第 5.5 节) 和随机过程实验 (第 5.6~5.8 节)。第 5 章内容是将实际问题转化为概率统计模型, 根据已知信息, 建立对实际问题的计算机模拟过程, 最终获得问题模拟结果, 从而为研究提供简便、快捷、可靠的结论。

提高模拟精度是针对模拟方法提出的根本要求。增大模拟量是有效改善精度的重要方法之一。模拟量在受到计算机运行效率的限制时, 在有限样本容量条件下的抽样方法可以改善模拟精度。本书第 3 章详细介绍了对偶变量法、控制变量法、条件期望法、分层抽样法、重要抽样法共 5 种常用减小估计量变异抽样方法, 通过蒙特卡罗模拟, 比较各种抽样方法减少方差效率。同时第 4 章第 4.4~4.5 节分别介绍 Jackknife 和 Bootstrap 抽样方法, 前者重点在于由抽样所决定构建估计量方法与一般估计量差异, 后者则是通过重复抽样提高估计精度。

R 语言作为实现上述第 1~5 章模拟的计算机编程工具，具有举足轻重的作用。计算机编程也是蒙特卡罗模拟实现的首要条件。R 语言正如一个虚拟的实验员，运用计算机向研究者展示实验的全部过程。它的运用使理论推导与模拟达成统一。本书第 6 章介绍 R 语言基础内容，包括 R 语言的数据格式、循环和函数创建、数据基本统计量计算和图形、模型分析方法。R 语言编程与其他语言相比，属于真正意义上的统计编程语言。程序编写模式简单易懂，并且具有基于世界范围统计方法（各种统计包）的支撑。例如，常见的随机数产生仅需 R 语言中的一个命令即可实现（第 1.1 节），Bootstrap 方法可以通过一个包 boot 实现（第 4.5 节）。如果你已具备使用 R 语言和基本统计分析方法，可以忽略这一章。如果你对其他编程语言十分熟悉，也可以按照本书的要求，完成各章的模拟程序。

R 语言在创建和运行蒙特卡罗模拟方面有着卓越的表现，很多较为复杂的模拟程序仅需 10 行以内的代码就可以完成。这是因为 R 语言本身就善于生成和管控海量数据，也擅长处理大数据。随着当前的数据发展趋势呈现出规模（Volume）、速度（Velocity）和种类（Variety）的“3V”指数级发展，需要更为简单的分析工具来驾驭这些“数据洪流”，而技术世界目前的发展还无法为各行各业提供易用工具的庞大需求，巨量数据时代要求在数据管理和分析方面采用新的范式和实践。目前，Google、Twitter、Facebook、Oracle 等公司都在使用 R 语言完成对其大数据的建模和分析过程。与此同时，R 语言逐渐成为具有广泛适用性的全球通用程序语言，它所建立的这种模式也被其他软件学习和采纳。R 语言的具体优势表现为：①开放的源代码（自由且免费），可以在任何操作系统如 Windows、Linux、Mac OS X 上使用；②统计学家和前沿算法（6 797 个包，且持续增加），形成高质、广泛的统计分析与数据挖掘平台；③重复性分析工作自动生成报告，提高工作效率；④扩展与兼容性，统计软件如 SAS、SPSS、Statistica 可互相调用 R、Python、Java、C、C++ 等语言，以及强大的数据库接口。

此书即将出版，在这里真诚感谢出版社的朋友们，他们为本书出版进行反复沟通与论证；感谢统计学系的同仁；感谢尹居良老师、陈征老师为本书提出宝贵意见；感谢研究生吴贤坚、蔡磊和张旭的校对工作，感谢王恩丹等同学的相关案例分析。

侯雅文

2015 年 5 月于暨南园

目 录

总 序	1
前 言	1
1 随机数产生与检验	1
1.1 用 R 语言产生随机数	1
1.1.1 离散型随机数产生	1
1.1.2 连续型随机数产生	5
1.2 产生随机数的理论方法	10
1.2.1 伪随机数产生的一般方法	10
1.2.2 逆变换方法	11
1.2.3 接受拒绝法	16
1.2.4 复合法	20
1.2.5 广义变换方法	22
1.3 随机数检验	27
1.3.1 均匀性检验	27
1.3.2 独立性检验	29
1.3.3 拟合优度检验	30
2 积分模拟与优化方法	33
2.1 大数定律积分法	33
2.1.1 伯努利大数定律积分模拟	33
2.1.2 辛钦大数定律积分模拟	37
2.2 积分的精度	40
2.3 重要抽样法积分	43
2.4 优化	47
2.4.1 矩估计中的优化	48
2.4.2 极大似然估计中的优化	50
3 方差缩减方法	55
3.1 对偶变量法	55
3.2 控制变量法	62
3.3 条件期望法	68
3.4 分层抽样法	73

3.5	重要抽样法	78
4	MCMC 方法	85
4.1	Metropolis - Hastings 算法	86
4.2	Gibbs 抽样方法	91
4.3	EM 算法	94
4.4	Jackknife 方法	98
4.5	Bootstrap 方法	100
5	系统模拟仿真	106
5.1	Buffon 投针问题	106
5.2	赶火车问题	108
5.3	追逐问题	110
5.4	保险问题	112
5.5	鱼塘鱼量问题	114
5.6	收银员问题	115
5.7	储备问题	117
5.8	亚式期权定价问题	120
6	R 语言与统计基础	125
6.1	R 语言	125
6.1.1	数据类型、结构与运算	128
6.1.2	数据导入	137
6.1.3	循环与函数创建	139
6.2	统计基础	141
6.2.1	探索性数据分析	141
6.2.2	参数估计与假设检验	148
6.2.3	回归与方差分析	171
	参考文献	183

1 随机数产生与检验

随机数作为模拟真实世界发生事件的基础，由于计算机性能的提高和方法的改善，随机数产生的质量也与真实程度更为接近。例如，根据客户索赔情况，模拟某项保险业务盈亏，需要产生关于索赔人数和索赔额的随机数。实际上，在制造业、交通业、服务业、金融业等诸多领域都在使用模拟方法进行预研，以期在现实状况中获得更精确的结果和更高的收益。

本章将介绍如何使用 R 语言产生随机数，并对各类随机数产生原理进行阐述，最后展示产生随机数优劣的检验方法及应用。

1.1 用 R 语言产生随机数

R 语言产生随机数有两种方式：一是将产生方法编写成 R 程序，获得随机数；二是直接利用 R 语言自带的函数产生随机数。本节将介绍后者的产生过程。通常，随机数依据分布来源可以分为两大类，即离散型和连续型，它们由分布形式和参数两部分构成。

1.1.1 离散型随机数产生

常用的离散型随机变量包括二项分布、几何分布、超几何分布、负二项分布和泊松分布等。用 R 函数直接产生离散型随机数的方法非常便捷。一般来说，产生随机数的 R 函数由 r + 分布英文缩写 + (参数) 三部分构成，例如，产生二项分布的 R 函数为 rbinom(n , size, prob)，其中 rbinom 为产生二项分布随机数的函数，括号内参数分别表示： n ，产生随机数个数；size，二项分布试验总次数；prob，每次试验中所规定的某事件 A 发生的概率。

```
set.seed(1)           #设定随机数种子,确保结果一致
x = rbinom(100,20,0.3) #产生 100 个 n=20,p=0.3 服从二项分布随机数
x                     #显示随机数
-----
[1] 5 5 6 9 4 9...
```

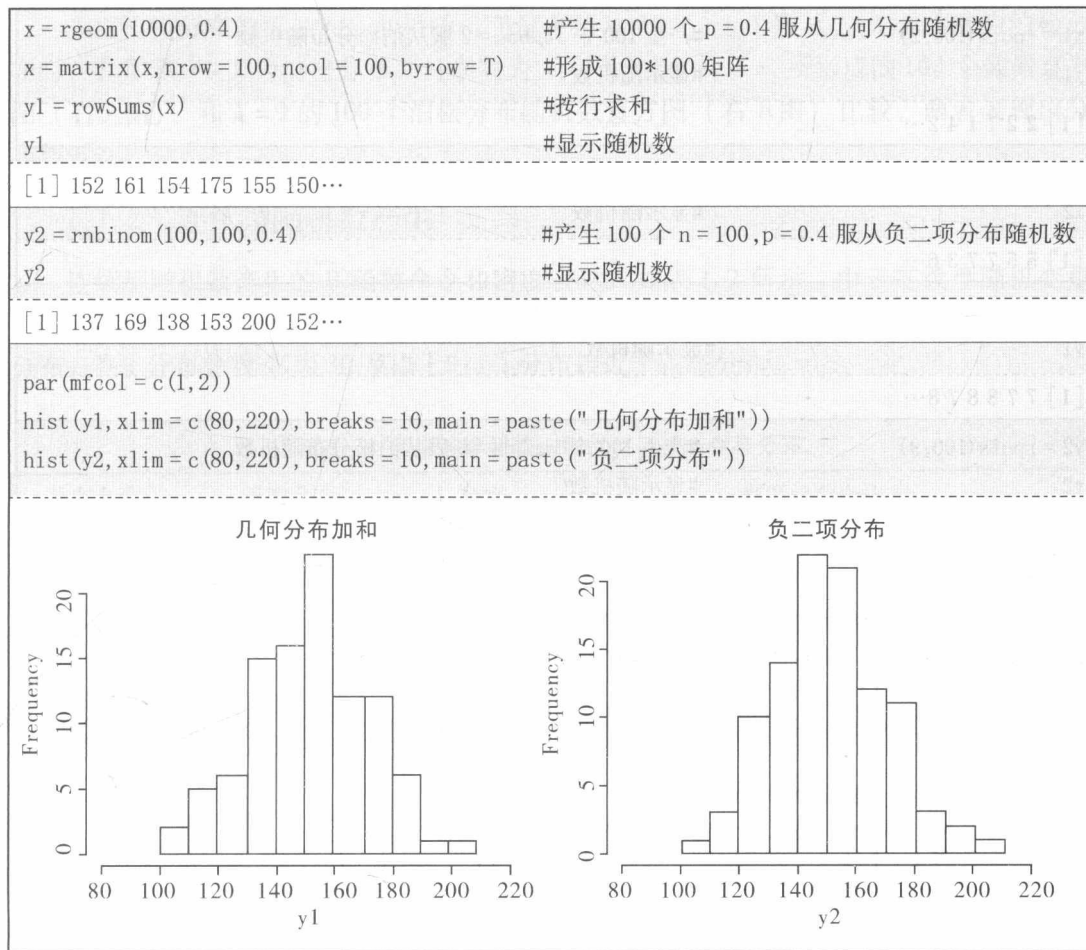
在 R 语言中，常用的离散型随机数都可以通过函数命令迅速产生，具体如表 1.1 所示，由于随机变量的分布表达形式是多样化的，因而 R 语言中离散型随机数的产生在参数含义与常见含义上可能稍有区别。例如，负二项分布在很多书籍、文献中，多用总试验次数表示随机变量 X ，在 R 语言中， X 则表示在所有试验中，某事件 A 不发生的次数。

表 1.1 常用离散型随机数的 R 函数命令与分布

分布名称	函数命令	参数	分布表达式
二项分布 Binomial	rbinom (n, size, prob)	size = n, 试验总次数 prob = p, 每次试验 A 事件发生概率	$P(X = x) = c_n^x p^x (1 - p)^{n-x}$ $x = 0, 1, \dots, n$
几何分布 Geometric	rgeom (n, prob)	prob = p, 每次试验 A 事件发生概率	$P(X = x) = (1 - p)^x p$ $x = 0, 1, 2, \dots$ X 表示 A 事件首次出现前, A 事件 不发生的试验次数
超几何分布 Hypergeometric	rhyper (nn, m, n, k)	m = 不合格数 n = 合格数 k = 抽取数	$P(X = x) = \frac{c_m^x c_n^{k-x}}{c_{m+n}^k}$ $x = 0, 1, \dots, m$
负二项分布 Negative binomial	rbinom (nn, size, prob)	size = n, A 事件发生 的次数 prob = p, 每次试验 A 事件发生概率	$P(X = x) = \frac{\Gamma(x + n)}{\Gamma(n)x!} p^n (1 - p)^x$ $= c_{x+n-1}^{n-1} p^n (1 - p)^x$ $x = 0, 1, 2, \dots$ X 表示 n 次 A 事件发生时, 所有试 验中 A 事件不发生的试验次数
泊松分布 Poisson	rpois (n, lambda)	lambda = λ	$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$ $x = 0, 1, 2, \dots, \lambda \geq 0$
多项分布 Multinomial	rmultinom (n, size, prob)	size = N prob = c(p ₁ , p ₂ , ..., p _k)	$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$ $= \frac{N!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$ $N = \sum_{i=1}^k x_i$

注: 函数命令中第一个参数 n 或 nn 均表示随机数产生的数量。

【例 1.1】分别产生 $p=0.4$ 的几何分布随机数 10000 个, 形成 $100 * 100$ 的矩阵, 并取行加和, 形成 $100 * 1$ 向量, 作出直方图。同时, 产生 $n=100, p=0.4$ 的负二项分布随机数 100 个, 作出直方图, 比较两者的相似性。



左侧直方图和右侧直方图有很大的相似性，右侧直方图为使用 R 函数产生的 100 个参数 $n=100$ ， $p=0.4$ 的负二项分布随机数， $n=100$ 意味着在试验中共有 100 次 A 事件的发生，每次 A 事件发生的概率为参数 $p=0.4$ 。若假设该参数条件下的负二项分布随机变量为 y_2 ，那么 y_2 可以看成是由 100 个服从参数为 $p=0.4$ 的几何分布随机变量（用 y_1 表示）加和构成，即 $y_2 = \sum_{i=1}^{100} y_{1i}$ 。根据直方图中 y_1 的产生方法，首先产生 10000 个服从参数 $p=0.4$ 的几何分布随机数，转化为 100×100 的矩阵，按照行取加和，每个加和可以看成是一个来自参数 $n=100$ ， $p=0.4$ 的负二项分布随机数，共 100 个。

由此，离散型分布随机数可以通过 R 语言自带函数产生，也可以利用 R 函数和分布之间的关系产生。

【例 1.2】分别产生参数 $\lambda=2$ 和 $\lambda=7$ 的泊松分布随机数各 100 个，取两两加和，形成 100 个数据。再产生参数 $\lambda=9$ 的随机数 100 个，与加和结果进行比较。

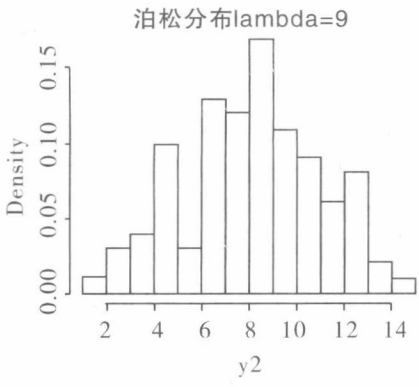
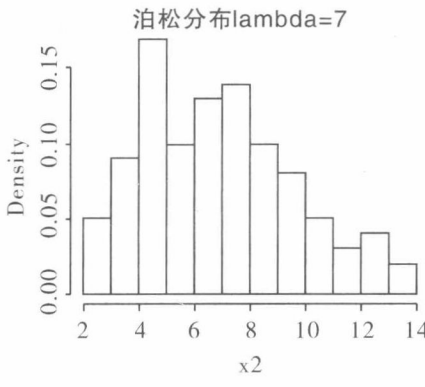
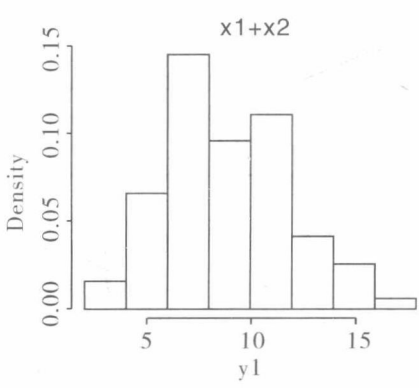
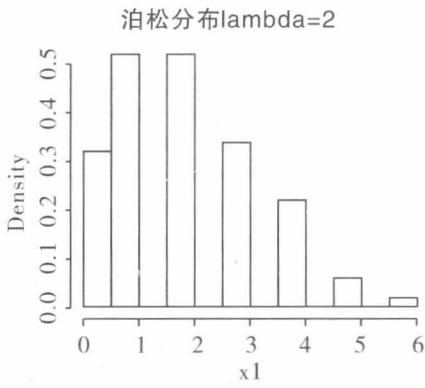
```
x1 = rpois(100,2)          #产生 100 个 lambda = 2 服从泊松分布随机数
x1                          #显示随机数
-----
[1] 2 2 1 1 4 2...

x2 = rpois(100,7)          #产生 100 个 lambda = 7 服从泊松分布随机数
x2                          #显示随机数
-----
[1] 5 5 7 7 3 6...

y1 = x1 + x2
y1                          #显示随机数
-----
[1] 7 7 8 8 7 8...

y2 = rpois(100,9)          #产生 100 个 lambda = 9 服从泊松分布随机数
y2                          #显示随机数
-----
[1] 15 11 15 6 5 9...

par(mfrow = c(2,2))
hist(x1,breaks = 10, prob = T, main = paste("泊松分布 lambda = 2"))
hist(x2,breaks = 10, prob = T, main = paste("泊松分布 lambda = 7"))
hist(y1,breaks = 10, prob = T, main = paste("x1 + x2"))
hist(y2,breaks = 10, prob = T, main = paste("泊松分布 lambda = 9"))
```



根据泊松分布的可加性, 如果随机变量 X 服从 $P(\lambda_1)$, Y 服从 $P(\lambda_2)$, 都为泊松分布时, 其加和 $X+Y$ 为泊松分布, 且参数为 $\lambda_1 + \lambda_2$ 。由 $x_1 + x_2$ 所形成的 100 个数据直方图 (右上图), 和 $\lambda=9$ 的 100 个泊松分布随机数直方图 (右下图) 比较, 两者表现出分布相近性。

1.1.2 连续型随机数产生

连续型随机数产生的 R 函数命令和密度表达式如表 1.2 所示, 由于连续型随机变量密度函数随参数变化而发生改变, 各种不同连续分布之间存在关联。例如, 正态分布和 t 分布, 在 t 分布参数 df 为 30 及以上时, t 分布接近于正态分布。

表 1.2 常用连续型随机数的 R 函数命令与分布

分布名称	函数命令	参数	密度函数表达式
贝塔分布 Beta	rbeta (n , shape1, shape2)	shape1 = a shape2 = b	$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}$ $a > 0, b > 0, 0 \leq x \leq 1$
柯西分布 Cauchy	rcauchy (n , location, scale)	location = l scale = s	$f(x) = \frac{1}{\pi s (1 + (\frac{x-l}{s})^2)}$
卡方分布 Chi-squared	rchisq (n , df, ncp)	df = n ncp = λ	$f(x, n) = \frac{1}{2^{n/2} \Gamma(n/2)} x^{(n/2)-1} e^{-x/2}$ $g(x, n, \lambda) = e^{-\frac{\lambda}{2}} \sum_{r=0}^{\infty} \frac{(\lambda/2)^r}{r!} f(x, n+2r)$ (非中心化)
指数分布 Exponential	rexp (n , rate)	rate = λ	$f(x) = \lambda e^{-\lambda x}, x \geq 0$
F 分布 F	rf (n , df1, df2)	df1 = n_1 df2 = n_2	$f(x) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} (\frac{n_1}{n_2})^{n_1/2} x^{\frac{n_1}{2}-1} (1 + \frac{n_1}{n_2}x)^{-\frac{n_1+n_2}{2}}$ $x > 0$
伽马分布 Gamma	rgamma (n , shape, rate, scale = 1/rate)	shape = a rate = λ scale = $1/\lambda$	$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, x \geq 0, a > 0, \lambda > 0$
对数正态分布 Lognormal	rlnorm (n , meanlog, sdlog)	meanlog = μ sdlog = σ	$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, x > 0$
逻辑斯特分布 Logistic	rlogis (n , location, scale)	location = m scale = s	$f(x) = \frac{1}{s} e^{-\frac{x-m}{s}} (1 + e^{-\frac{x-m}{s}})^{-2}$

(续上表)

分布名称	函数命令	参数	密度函数表达式
正态分布 Normal	rnorm (n, mean, sd)	mean = μ sd = σ	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
t 分布 t	rt (n, df)	df = n	$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(n/2)} (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$
均匀分布 Uniform	runif (n, min, max)	min max	$f(x) = \frac{1}{\max - \min}, \min \leq x \leq \max$
威尔科克森 符号秩分布 Wilcoxon Signed Rank	rsignrank (nn, n)	n	随机数: 来自一组数据经过绝对值排序取秩后, 将所有数据值为正对应的秩求和结果。 n: 随机数取值范围在 0 到 $\frac{n(n+1)}{2}$
威布尔分布 Weibull	rweibull (n, shape, scale)	shape = a scale = b	$f(x) = \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a}, x > 0$
威尔科克森 秩和分布 Wilcoxon Rank Sum	rwilcox (nn, m, n)	m n	随机数: 来自样本量分别为 m 和 n 的随机、独立数据 X 和 Y, 计算所有满足 $X_i \geq Y_j$ 的个数结果。 随机数取值范围: 0 到 mn
多元正态分布 Multivariate Normal	mvrnorm (n, mu, Sigma)	mu = μ Sigma = Σ	$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \left \Sigma \right ^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)' \Sigma^{-1}(x-\mu)}$

注: mvrnorm 函数需调用 MASS 包, 函数命令中第一参数 n 或 nn 均表示随机数产生数量。

【例 1.3】分别产生标准正态分布随机数 100 个, $df=2, 10, 30$ 的 t 分布随机数各 100 个。作出随机数直方图和密度函数曲线, 进行比较。

options(digits=3)	
x = rnorm(100)	#产生 100 个 mean=0, sd=1 服从标准正态分布随机数
x	#显示随机数

[1]	1.020 0.788 0.418 1.611 -0.828 0.351...
y1 = rt(100,2)	
y1	#产生 100 个 df=2 服从 t 分布随机数
	#显示随机数

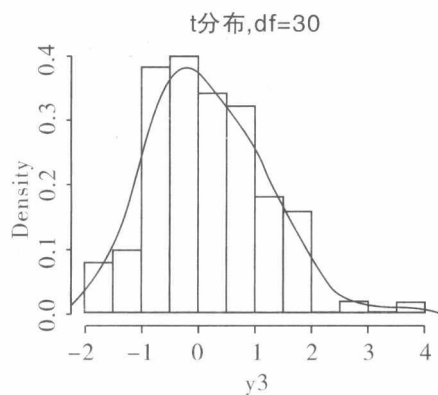
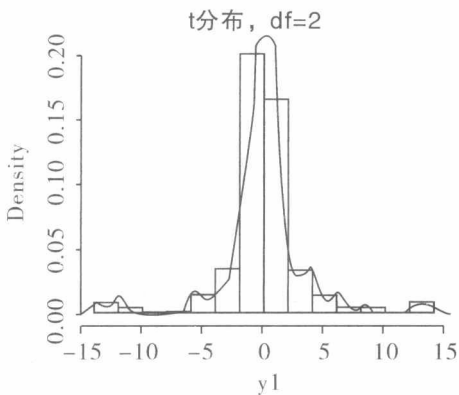
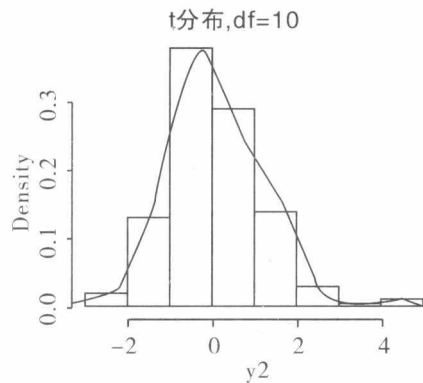
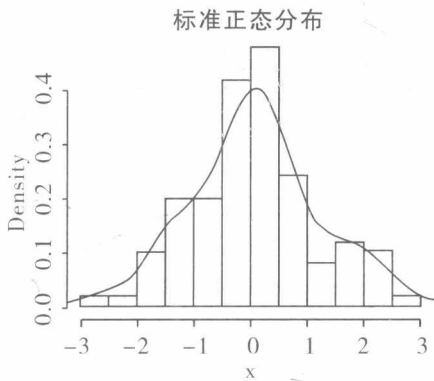
[1]	0.00847 0.46177 1.17985 -0.42883 -1.03086 4.13947...
y2 = rt(100,10)	
y2	#产生 100 个 df=10 服从 t 分布随机数
	#显示随机数

[1]	0.657 1.267 1.294 1.747 -1.333 1.015...

```
y3 = rt(100, 30)      #产生 100 个 df = 30 服从 t 分布随机数
y3                    #显示随机数
```

```
[1] -0.2384 -0.5865 -1.5909 2.0114 -1.5195 0.0616...
```

```
par(mfcol = c(2, 2))
hist(x, breaks = 10, prob = T, main = paste("标准正态分布"))
lines(density(x), col = "red", lwd = 2)
hist(y1, breaks = 10, prob = T, main = paste("t 分布, df = 2"))
lines(density(y1), col = "red", lwd = 2)
hist(y2, breaks = 10, prob = T, main = paste("t 分布, df = 10"))
lines(density(y2), col = "red", lwd = 2)
hist(y3, breaks = 10, prob = T, main = paste("t 分布, df = 30"))
lines(density(y3), col = "red", lwd = 2)
```



$df=2$ 时 (左下图), 呈现 t 分布典型的厚尾分布特征, 随着 df 的增加, $df=10$ 时 (右上图), 横轴取值范围缩小到 $[-3, 3]$ 之间, 稍有厚尾, $df=30$ 时 (右下图), 厚尾进一步减小, 密度函数特征 (实线) 与标准正态分布 (左上图) 更为接近。

实际上, 上述直方图和密度函数是基于随机数而产生的, 图形结果都会有所差异。下面以各分布的理论密度作出密度函数曲线图进行比较。