

“十三五”国家重点出版物出版规划项目



名校名家基础学科系列

Textbooks of Base Disciplines from Top Universities and Experts

数学实验

概率论与数理统计分册

李娜 王丹龄 刘秀芹 编著



机械工业出版社
CHINA MACHINE PRESS

“十三五”国家重点出版物出版规划项目

 名校名家基础学科系列

Textbooks of Basic Disciplines from World's Top Universities and Experts

数学实验——

概率论与数理统计分册

李 娜 王丹龄 刘秀芹 编著



机械工业出版社

本书侧重于将概率论与数理统计课程中的典型问题设计成数学实验,将课程中抽象的数学公式、定理通过数学软件实现得到验证。通过实验操作,学生能够自己编程,加深对概念定理的理解,培养学生将形象思维与逻辑思维相结合的能力,让学生从问题出发,亲自动手,体验解决问题的过程。本书选择数学软件 MATLAB 和统计软件 SPSS,分两部分编写,内容包括用 MATLAB 进行关于离散型随机变量、连续型随机变量、描述性统计分析、蒙特卡罗模拟、参数估计、假设检验、方差分析、回归分析、多元回归分析的实验,以及用 SPSS 进行方差分析、回归分析、聚类分析、判别分析、因子分析和主成分分析的实验。

本书适合理、工、管理本科生学习使用,也可供工程技术人员参考。

图书在版编目(CIP)数据

数学实验. 概率论与数理统计分册/李娜, 王丹龄, 刘秀芹编著. —北京:机械工业出版社, 2018. 5

“十三五”国家重点出版物出版规划项目. 名校名家基础学科系列

ISBN 978-7-111-61183-7

I. ①数… II. ①李… ②王… ③刘… III. ①高等数学-实验-高等学校-教材 ②概率论-实验-高等学校-教材 ③数理统计-实验-高等学校-教材
IV. ①O13-33 ②O21-33

中国版本图书馆 CIP 数据核字 (2018) 第 242356 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策划编辑: 郑 玫 责任编辑: 郑 玫 汤 嘉

责任印制: 孙 炜 责任校对: 刘 岚

保定市 中画美凯印刷有限公司印刷

2019 年 2 月第 1 版第 1 次印刷

169mm × 239mm · 14.75 印张 · 276 千字

标准书号: ISBN 978-7-111-61183-7

定价: 37.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

电话服务

网络服务

服务咨询热线: 010-88379833

机工官网: www.cmpbook.com

读者购书热线: 010-88379649

机工官博: weibo.com/cmp1952

教育服务网: www.cmpedu.com

封面防伪标均为盗版

金书网: www.golden-book.com

前 言

数学实验是计算机技术与数学软件引入教学后出现的新事物,目前是理、工、管理本科生必修的一门基础课程。通过“数学实验”,学生可以更加深入地理解数学概念和理论,熟悉常用的数学软件,也有助于提高学生的动手实践能力和解决实际问题的能力。

本书侧重于将概率论与数理统计课程中的典型问题设计成数学实验,将课程中抽象的数学公式、定理通过数学软件的实现得到验证。通过实验操作,学生能够自己编程,加深对概念定理的理解,培养学生将形象思维与逻辑思维相结合的能力,让学生从问题出发,亲自动手,体验解决问题的过程,教会学生在“学”数学后,会“用”数学,有力地调动学生学习数学的积极性,加强对学生的数学知识、软件知识、计算机知识和动手能力的培养,达到“突出基础、注重实验、加强应用”的目的。通过本课程的学习,学生对概率论与数理统计课堂教学内容加以掌握与巩固,更通过数学软件的实践操作和应用为学生学习后续课程及在各个学科领域中进行数学建模和应用研究打下坚实的基础。

本书是在编者多年讲授概率论与数理统计、数学实验与 MATLAB 课程的基础上,广泛吸取了国内外相关教材的特点编写而成的。本书选择数学软件 MATLAB 和统计软件 SPSS,分两部分编写,其中李娜和王丹龄编写 MATLAB 部分,刘秀芹编写 SPSS 部分,李娜负责全书的统稿工作。

在本书的编写过程中,北京科技大学范玉妹教授,王萍、张志刚、徐尔三位副教授为本书提出了宝贵的意见和建议,北京科技大学的申婷、刘巧月和吴美琪同学在文字录入和程序调试方面给予了许多帮助,在此一并表示衷心的感谢!

本书的编写也参阅了许多专家、学者的论著文献,并引用部分论著中的例子,由于篇幅限制,恕不一一指明出处,在此一并向有关作者致谢!

本书的编写与出版得到了北京科技大学教材建设基金的资助。

限于编者水平,错漏之处在所难免,敬请读者不吝指正。

编 者
2018 年 4 月

目 录

前 言

概率统计——MATLAB 篇

实验一 准备实验	2	5.1 实验目的	27
1.1 实验目的	2	5.2 相关知识	27
1.2 相关知识	2	5.3 MATLAB 常用命令	29
1.3 MATLAB 常用命令	4	5.4 实验内容	30
1.4 实验内容	4	5.5 课后练习	33
1.5 课后练习	5	实验六 描述性统计分析	34
实验二 古典概型	6	6.1 实验目的	34
2.1 实验目的	6	6.2 相关知识	34
2.2 相关知识	6	6.3 MATLAB 常用命令	36
2.3 MATLAB 常用命令	7	6.4 实验内容	37
2.4 实验内容	7	6.5 课后练习	44
2.5 课后练习	13	实验七 蒙特卡罗模拟	45
实验三 条件概率	14	7.1 实验目的	45
3.1 实验目的	14	7.2 相关知识	45
3.2 相关知识	14	7.3 实验内容	45
3.3 MATLAB 常用命令	14	7.4 课后练习	51
3.4 实验内容	15	实验八 参数估计	52
3.5 课后练习	17	8.1 实验目的	52
实验四 离散型随机变量	19	8.2 相关知识	52
4.1 实验目的	19	8.3 MATLAB 常用命令	53
4.2 相关知识	19	8.4 实验内容	53
4.3 MATLAB 常用命令	20	8.5 课后练习	55
4.4 实验内容	21	实验九 假设检验	56
4.5 课后练习	26	9.1 实验目的	56
实验五 连续型随机变量	27	9.2 相关知识	56
		9.3 MATLAB 常用命令	57

9.4 实验内容	57	10.6 课后练习	70
9.5 课后练习	64	实验十一 回归分析	72
实验十 方差分析	65	11.1 实验目的	72
10.1 实验目的	65	11.2 相关知识	72
10.2 相关知识	65	11.3 MATLAB 常用命令	72
10.3 MATLAB 常用命令	65	11.4 一元回归分析	73
10.4 单因素方差分析	65	11.5 多元回归分析	80
10.5 双因素方差分析	67	11.6 课后习题	92

统计分析与 SPSS 应用

第 1 章 SPSS 统计分析软件

概述	94
1.1 SPSS 入门	94
1.1.1 软件概述	94
1.1.2 SPSS 软件的安装与激活	95
1.2 SPSS 使用基础	95
1.2.1 SPSS 的基本窗口	95
1.2.2 SPSS 软件的退出	96
1.3 SPSS 数据文件的建立和管理	97
1.3.1 建立数据文件	97
1.3.2 数据的结构和定义方法	98
1.3.3 数据的录入与编辑	100
1.3.4 数据文件的整理	104
课后练习	107

第 2 章 SPSS 基本统计分析与

统计推断	108
2.1 基本统计分析	108
2.1.1 频数分析	108
2.1.2 计算基本描述统计量	111
2.1.3 探索性数据分析	113
2.2 统计推断	115
2.2.1 由样本推断整体—— 参数估计	116
2.2.2 假设检验	120
2.2.3 检验的 p 值	122

课后练习	124
------------	-----

第 3 章 SPSS 的方差分析

3.1 方差分析概述	125
3.2 单因素方差分析	126
3.3 多因素方差分析	133
课后练习	143

第 4 章 SPSS 的相关分析与

回归分析	145
4.1 相关分析	145
4.1.1 相关分析概述	145
4.1.2 绘制散点图	145
4.1.3 计算相关系数	152
4.1.4 偏相关分析	156
4.2 回归分析	158
4.2.1 回归分析概述	159
4.2.2 线性回归分析和线性回归 模型	159
4.2.3 回归方程的统计检验	161
4.2.4 多元回归分析中的 其他问题	167
4.2.5 线性回归分析基本操作	169
4.2.6 线性回归分析的应用 举例	172
4.2.7 曲线估计	178
课后练习	185

第5章 SPSS 的聚类分析 187

5.1 聚类分析的一般概念 187

5.2 聚类分析中“亲疏程度”的
度量方法 187

5.3 层次聚类 189

5.4 层次聚类分析中的
R 型聚类 194

5.5 K-均值聚类分析 198

课后练习 202

第6章 判别分析 204

6.1 判别分析的一般概念 204

6.2 判别分析的实现过程 204

课后练习 213

第7章 因子分析与主成分

分析 215

7.1 因子分析 215

7.1.1 因子分析概述 215

7.1.2 因子分析基本内容 216

7.1.3 因子分析的基本操作及
案例 220

7.2 主成分分析 227

7.2.1 主成分分析概述 227

7.2.2 主成分分析模型 228

课后练习 229

参考文献 230

概率统计——MATLAB篇

实验一 准备实验

1.1 实验目的

- 1) 掌握 MATLAB 的基本操作;
- 2) 学会进行矩阵和数组的输入及数组的运算;
- 3) 了解 MATLAB 中的各种函数以及数据显示格式和帮助系统等.

1.2 相关知识

MATLAB 是英文 Matrix Laboratory (矩阵实验室) 的缩写, 是一款由美国 MathWorks 公司出品的数学软件. 其将计算、可视化和编程功能集中在非常便于使用的环境中, 是一个交互式的、以矩阵计算为基础的科学和工程计算软件, 在统计分析中有广泛的应用.

视窗环境

命令窗口	命令和数据的输入输出
M 文件窗口	编辑源程序文件和调试程序
工作空间窗口	存放变量的相关信息
当前目录窗口	存放 M 或函数文件的工作目录
命令历史窗口	命令的历史纪录
起始面板窗口	工具箱组合
图形窗口	画图
帮助窗口	功能强大的帮助

矩阵基本数据操作函数

极大、极小	max, min
总和、平均	sum(x), mean(x)
中位数	median, iqr, prctile(x,p)
偏度系数	skewness
峰度系数	kurtosis
排序	sort
乘积	prod
标准差	std, std(a,l)
方差	var, var(x,l); 协方差 cov
相关系数	corrcoef(x,y), corr(x)

矩阵运算

加、减法	$a + b, a - b$
乘法	$a * b, a^2$
除法	左除 $a \backslash b$, 右除 a/b

数组运算

乘法	$a * b$
除法	$a ./b, a . \backslash b$
幂函数	$a.^2, a.^b$

函数文件

函数文件由 function 引导, 基本结构为 function [输出形参表] = 函数名 (输入形参表) 调用格式为 [输出形参表] = 函数名 (输入形参表)

二维画图

plot(x,y)	画线
plot(x,y,'ro')	画点
plot(x1,y1,x2,y2,'.....')	画多条线
plotyy(x1,y1,x2,y2)	画两条不同尺度线
stem(x,y);stem(x,y,'filled')	画竖线
text(x,y,str),(* str 必须为列向量)	在二维图形中指定位置显示字符串

plot 绘图函数的参数

参 数	意 义	参 数	意 义
r	red - 红色	-	实线
g	green - 绿色	--	虚线
b	blue - 蓝色	:	点线
y	yellow - 黄色	-.	点划线
m	magenta - 深红	o	圆圈
c	cyan - 青蓝	x	叉号
w	white - 白色	+	加号
k	black - 黑色	s	方形
*	星号	d	菱形
.	点号		

1.3 MATLAB 常用命令

调用格式	含义
<code>zeros (m, n)</code>	生成 $m \times n$ 阶零矩阵
<code>ones (m, n)</code>	生成 $m \times n$ 阶元素全为 1 的矩阵
<code>eye (m, n)</code>	生成 $m \times n$ 阶对角线元素为 1 的矩阵
<code>randn (m, n)</code>	生成 $m \times n$ 阶正态分布随机数矩阵
<code>inv (A)</code>	求矩阵 A 的逆矩阵
<code>det (A)</code>	求矩阵 A 的行列式

1.4 实验内容

实验 1 产生一个 4 阶随机矩阵.

```
rand(4,4)
ans =
    0.6324    0.9575    0.9572    0.4218
    0.0975    0.9649    0.4854    0.9157
    0.2785    0.1576    0.8003    0.7922
    0.5469    0.9706    0.1419    0.9595
```

实验 2 编写函数文件求半径为 r 的圆的面积和周长.

```
function [s,p]=fcircle(r)
% r 圆半径
% s 圆面积
% p 圆周长
s=pi*r*r;
p=2*pi*r;
```

实验 3 在区间 $0 \leq x \leq 2$ 上, 绘制曲线 $y = 2e^{-0.5x} \cos(4\pi x)$.

```
x=0:pi/100:2;
y=2*exp(-0.5*x).*cos(4*pi*x);
plot(x,y)
结果见图 1.
```

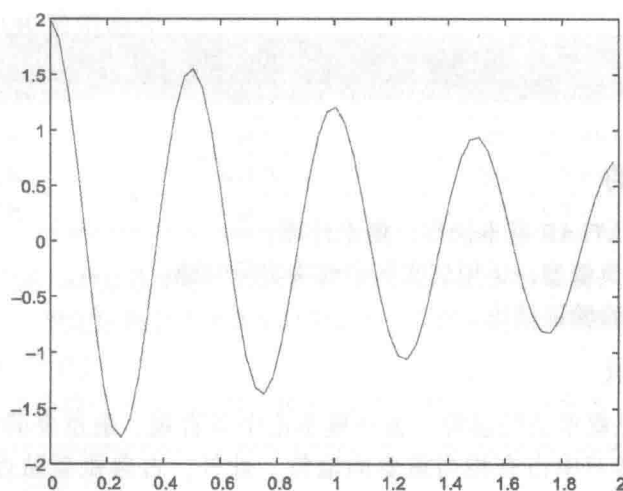


图 1

1.5 课后练习

1. 自行输入一些简单的矩阵，并对矩阵进行四则运算，体会左除和右除的区别，以及矩阵的乘法、除法、乘方与数组相应运算的区别。
2. 利用函数的递归调用，求 $n!$ 。
3. 画出蔓叶线 $y^2 = \frac{x^3}{4-x}$ 的图像。

实验二 古典概型

2.1 实验目的

- 1) 掌握 MATLAB 基本操作, 概率计算;
- 2) 理解古典概型, 运用到实例中解决实际问题;
- 3) 模拟实验验证结论.

2.2 相关知识

古典概型是概率论的起源, 也是概率论中最直观、最重要的模型之一, 在概率论的整个学习中占有相当重要的地位. 此外, 古典概型贴近于实际生活, 在密码学、经济学、管理学等学科中也有着重要的应用.

1. 古典概型的定义

- 1) 试验的样本空间只包含有限个元素 $S = \{e_1, e_2, \dots, e_n\}$;
- 2) 试验中每个基本事件发生的可能性相同 $P(\{e_1\}) = P(\{e_2\}) = \dots = P(\{e_n\})$.
具有以上两个特点的试验称为等可能概型或古典概型.

2. 古典概型的计算公式

1) 古典概型中因为

- ① $\{e_1\}, \{e_2\}, \dots, \{e_n\}$ 两两互不相容;

② $\bigcup_{i=1}^n \{e_i\} = S$;

③ $P(S) = 1$.

由有限可加性:

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

故

$$P(\{e_i\}) = \frac{1}{n}, \quad i = 1, 2, \dots, n.$$

2) 若 A 是 S 中任一事件, A 的概率如何计算?

设 $A = \{e_{i_1}, e_{i_2}, \dots, e_{i_k}\} \subset S$, 则 $P(A) = ?$

$$P(A) = \frac{k}{n} = \frac{\text{A 中包含样本点的个数}}{\text{样本点总数}}$$

有利数
可能数

2.3 MATLAB 常用命令

函 数 名	调用格式	注 释
0-1 随机数	rand(N)	返回一个 $N \times N$ 的随机矩阵
0-1 随机数	rand(N,M)	返回一个 $N \times M$ 的随机矩阵
0-1 随机数	rand	产生一个 0~1 之间的随机数

例 1 连续掷 100000 次硬币, 记录重复 10 次, 100 次, 1000 次, 10000 次试验模拟出现正面的频率. 规定随机数小于 0.5 时为正面, 否则为反面, 求解正面朝上的概率.

【MATLAB 程序】

MATLAB 命令存于 ex11.m 内, 现给出算法程序, 仅供参考:

1. 定义一个数组 aa
2. for 循环: $i = 1: 6$
 - a) $a(i)$ 是随机产生 10^i 次方个随机数将其四舍五入后求和
 - b) aa 数组由 $a(i)$ 组成
3. 跳出循环, 列表显示结果

【运行结果】

掷硬币模拟实验

试验次数	10	100	1000	10000	100000	1000000
正面朝上概率	0.700000	0.450000	0.484000	0.497500	0.500720	0.499835

2.4 实验内容

实验 1

一个口袋装有 8 个球, 其中白球 3 个, 红球 5 个. 从袋中取出 3 个球, 每次随机取一个. 考虑两种取球方式:

有放回抽样: 第一次取一个球, 观察其颜色后放回, 搅匀后再取一个球;

无放回抽样: 第一次取一个球不放回, 第二次从剩余的球中再取一个球.

分别就上述两种方式求:

- 1) 第三次才摸到红球的概率;
- 2) 取到的三个球中至少有一个是白球的概率.

【求解过程】

从袋中取两个球, 每一种取法就是一个基本事件. 假设 $A =$ “第三次才摸到

红球”， $B =$ “取到的三个球中至少有一个是白球”， $\bar{B} =$ “没有取到白球”。

有放回抽样，则放回摸球的三次试验互不影响，因此三次摸球相互独立有 $P(\text{白}) = 3/8$ 和 $P(\text{红}) = 5/8$ 。

$$1) P(A) = (3^2 \times 5) / 8^3 \approx 0.0879,$$

$$2) P(B) = 1 - P(\bar{B}) = 1 - 5^3 / 8^3 \approx 0.7559,$$

故第三次才摸到红球的概率 $P(A) \approx 0.0879$ ；取到的三个球中至少有一个是白球的概率 $P(B) \approx 0.7559$ 。

无放回抽样

$$1) P(A) = \frac{C_3^1 C_2^1 C_5^1}{C_8^1 C_7^1 C_6^1} \approx 0.0893,$$

$$2) P(B) = 1 - P(\bar{B}) = 1 - \frac{C_5^3}{C_8^3} \approx 0.8214,$$

故第三次才摸到红球的概率 $P(A) \approx 0.0893$ ；取到的三个球中至少有一个是白球的概率 $P(B) \approx 0.8214$ 。

【MATLAB 程序】

MATLAB 命令存于 ex21.m 内，现给出算法程序，仅供参考：

1. 定义白球、红球个数

2. 计算事件 A, B 发生的概率

%A = “第三次才摸到红球” %B = “取到的三个球中至少有一个是白球”

3. 代入 MATLAB 命令公式，列表显示结果

【运行结果】

摸球实验理论值

试验	事件A发生的概率: $P(A)$	事件B发生的概率: $P(B)$
有放回抽样	0.087891	0.755859
无放回抽样	0.089286	0.821429

【MATLAB 模拟】

设有 m 个球，其中白球 w 个，红球 r 个。取出 n 个球，利用二项分布产生随机数模拟摸球实验，摸出红球事件为“0”，概率为 $\frac{w}{m}$ ，摸出白球事件为“1”，概率为 $\frac{r}{m}$ 。

【MATLAB 程序】

MATLAB 命令存于 ex21.m 内，现给出算法程序，仅供参考：

1. 定义红、白球个数以及取出球个数
2. 设摸到红球为随机数 1
3. $a = \text{binornd}(1, pp, n, 3)$
4. for $i = 1 : \text{size}(a, 1)$
5. for $j = 1 : \text{size}(a, 2)$
 - a) if $a(i, j) == 0$ 画红圈
 - b) else 画白圈
6. 跳出循环
- %模拟第三次才摸到红球的概率
7. for $i = 1 : n$
8. if $(r(i, 1) == 0) \& (r(i, 2) == 0) \& (r(i, 3) == 1)$ 计数加一
9. 用频率值估计概率值
10. 跳出循环, 画图
- %模拟取到的三个球中至少有一个是白球的概率.
11. for $i = 1 : n$
12. if $(r(i, 1) == 1) \& (r(i, 2) == 1) \& (r(i, 3) == 1)$ 计数加一
13. 用频率值估计概率值
14. 跳出循环, 画图

【运行结果】见图 2.

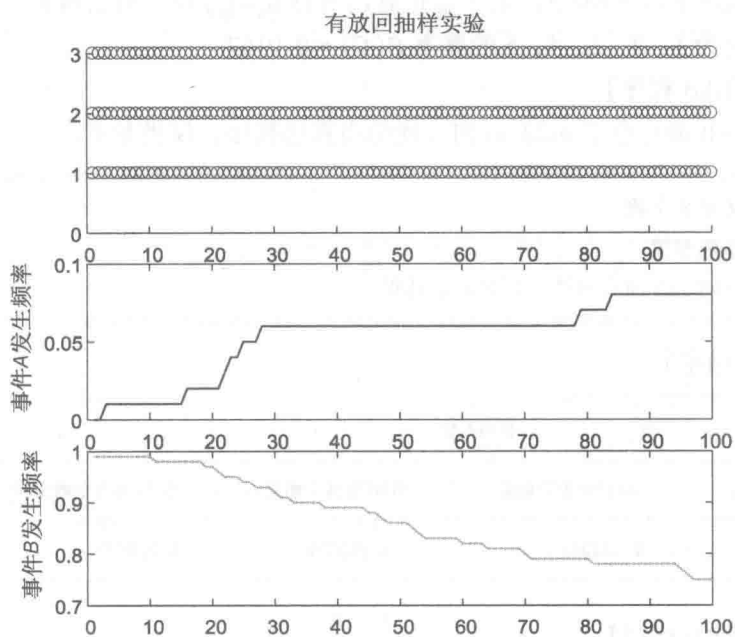


图 2

实验 2

一部五卷的选集，按任意的次序放在书架上，试求下列事件的概率：

- 1) 第一卷及第五卷出现在两端；
- 2) 第一卷及第五卷不出现在两端；
- 3) 第三卷正好出现在正中，自左向右或自右向左的卷号顺序恰好为 1, 2, 3, 4, 5.

【求解过程】

设 $A =$ “第一卷及第五卷出现在两端”， $B =$ “第一卷及第五卷不出现在两端”， $C =$ “第三卷正好出现在正中，自左向右或自右向左的卷号顺序恰好为 1, 2, 3, 4, 5”。

$$1) P(A) = \frac{2! \times 3!}{5!} = \frac{1}{10} = 0.1,$$

$$2) P(B) = \frac{P_3^2 \times 3!}{5!} = \frac{3}{10} = 0.3,$$

$$3) P(C) = \frac{2}{5!} = \frac{1}{60} \approx 0.0167,$$

故第一卷及第五卷出现在两端的概率 $P(A) = 0.1$ ；第一卷及第五卷不出现在两端的概率 $P(B) = 0.3$ ；第三卷正好出现在正中，自左向右或自右向左的卷号顺序恰好为 1, 2, 3, 4, 5 的概率 $P(C) \approx 0.0167$ 。

【MATLAB 程序】

MATLAB 命令存于 ex22.m 内，现给出算法程序，仅供参考：

1. 定义书本个数
2. 计算概率值
3. 用 MATLAB 命令计算，列表显示结果

【运行结果】

排书实验

试验	事件A发生的概率:P(A)	事件B发生的概率:P(B)	事件C发生的概率:P(C)
概率	0.100000	0.300000	0.016667

【MATLAB 模拟】

采用 randperm() 函数产生 1 - n 之间不重复的随机整数模拟排书实验。