

概率统计引论

魏立力 马江洪 颜荣芳 编著



科学出版社

概 率 统 计 引 论

魏立力 马江洪 颜荣芳 编著

科 学 出 版 社

北 京

内 容 简 介

本书内容包括随机事件与概率、随机变量及其概率分布、随机变量的数字特征、大数定律与中心极限定理、统计学的基本概念、参数估计、假设检验、回归分析与相关分析、方差分析以及 Excel 在概率统计中的应用等. 全书结构体系合理, 应用背景丰富, 思想方法突出, 例题习题讲究, 并对一些问题介绍了相关的发展方向 and Excel 解决方案.

本书可作为高等学校数学专业和统计学专业的本科生教材或相关专业的研究生教学参考书, 也可作为科研人员和工程技术人员的查阅手册.

图书在版编目 (CIP) 数据

概率统计引论 / 魏立力, 马江洪, 颜荣芳编著. —北京: 科学出版社, 2012.2

ISBN 978-7-03-033404-6

I. ①概… II. ①魏… ②马… ③颜… III. ①概率论-高等学校-教材②数理统计-高等学校-教材 IV. ①O21

中国版本图书馆CIP数据核字 (2012) 第013247号

责任编辑: 杨 岭 郝玉龙/封面设计: 陈思思

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

四川煤田地质制图印刷厂印刷

科学出版社发行 各地新华书店经销

*

2012 年 2 月第 一 版 开本: 787×1092 1/16

2012 年 2 月第一次印刷 印张: 23.25

字数: 520 000

定价: 38.00 元

前 言

概率统计是研究随机现象统计规律性的学科,在自然科学、社会科学、工程技术等领域具有非常广泛的应用.概率论着眼于随机现象统计规律的演绎分析,而统计学则侧重于随机现象统计规律的归纳研究,二者互相联系、互相渗透,在科学研究、技术开发、生产管理和社会经济生活等诸多方面发挥了重要作用.现在,概率统计已经成为几乎所有科技工作者的基本工具,概率统计和其他学科相结合还产生出不少边缘学科或交叉学科,如生物统计学、统计物理学、计量化学、计量经济学和数学地质学等,极大地扩展了概率统计的应用范围.可以说,概率统计已经成为当代科学技术领域最重要的学科之一.

然而,如何根据概率统计自身的学科特点,正确理解和准确把握其基本思想和基本方法,特别是在统计软件日益普及和应用范围日益扩大的情形下避免误用或滥用概率统计知识,就显得尤为重要,这也正是撰写本书的初衷.根据多年的教学经验和科研实践,我们认为推出这样一本概率论与统计学的入门书是必要的.

本书在内容安排上,包括初等概率论(第1~4章)、基本统计方法(第5~9章)以及 Excel 在概率统计中的应用,选材尽量做到联系不同专业的实际,注重应用的广泛性;在具体行文中,注意循序渐进、深入浅出,突出对基本思想和基本概念的阐述,强调直观理解的重要性,力求融入一些增强思想性、趣味性和实用性的新内容,体现求精、求变的指导原则;在例题和习题的选择上,兼顾启发性和代表性,涵盖了很多经典的例子和近十年来的考研(数学一)题目,对于涉及的绝大部分数值计算问题,介绍了 Excel 解决方案.另外,大部分专业术语首次出现时也都附有相应的英文对照,供读者学习参考.我们希望读者通过本书的学习,能够正确理解和使用概率论与统计学的一些方法,能够对统计计算结果做出科学的分析和解释,能够准确把握概率统计的思想方法和精髓,提升运用概率统计的思想方法分析和解决有关问题的能力.

本书经过充分讨论、反复推敲、不断修改才得以完成,第1~2章由颜荣芳编写,第5~8章由马江洪编写,其余内容由魏立力编写,全书由魏立力统稿.

虽然我们竭尽全力,但仍感不尽如人意.由于水平所限,书中考虑不周或错误之处在所难免,祈望读者不吝赐教,批评指正.

作 者
2012年1月

主要符号表

\mathbb{R}	实数集
\mathbb{R}^d	d 维实空间
Ω	样本空间或必然事件
\emptyset	空集或不可能事件
P	概率
\mathcal{F}, \mathcal{B}	σ 代数或事件域
X, Y, Z, \dots	随机变量
x, y, z, \dots	相应随机变量的观察值
$\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \dots$	随机向量或随机矩阵
$\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$	相应随机向量或随机矩阵的观察值
$E(X)$	X 的数学期望
$E(X Y)$	X 关于 Y 的条件数学期望
$D(X), \sigma_{XX}$	X 的方差
$\text{Cov}(X, Y), \sigma_{XY}$	(X, Y) 的协方差
ρ_{XY}	(X, Y) 的相关系数
$\mu_k = E(X^k)$	X 的 k 阶原点矩
$\nu_k = E(X - EX)^k$	X 的 k 阶中心矩
$\text{CV}(X)$	X 的变异系数
γ_3	偏态系数
γ_4	峰态系数
\xrightarrow{P}	随机变量序列依概率收敛
\xrightarrow{L}	随机变量序列依分布收敛
\xrightarrow{W}	分布函数序列弱收敛
$X_{(1)}, X_{(2)}, \dots, X_{(n)}$	样本 X_1, X_2, \dots, X_n 的次序统计量
\bar{X}	样本均值
S^2	样本方差
iid	独立同分布
Θ	参数 θ 的取值空间
$\hat{\theta}$	参数 θ 的估计
MLE	极大似然估计
$L(\cdot), \ln L(\cdot)$	似然函数, 对数似然函数
EMS	均方误差
UMVUE 或 BUE	一致最小方差无偏估计
H_0 和 H_1	原假设和备择假设
(a.b.c)	第 a 章第 b 节第 c 个公式
定理 (引理, 定义, 例)a.b.c	第 a 章第 b 节第 c 个定理 (引理, 定义, 例)

目 录

前 言

主要符号表

第 1 章 随机事件与概率	1
1.1 随机事件及其运算	1
1.1.1 必然现象与随机现象	1
1.1.2 随机试验和样本空间	1
1.1.3 随机事件的关系和运算	3
1.2 排列与组合	6
1.2.1 排列组合的基本模式	6
1.2.2 多项组合	7
1.3 随机事件的概率	8
1.3.1 古典概率	9
1.3.2 统计概率	14
1.3.3 几何概率	16
1.4 概率的公理化定义及概率的性质	18
1.4.1 概率的公理化定义——概率空间	18
1.4.2 概率的性质	20
1.5 条件概率	23
1.5.1 条件概率的定义和性质	23
1.5.2 有关条件概率的三个公式	24
1.6 事件的独立性	27
1.6.1 两个事件的独立性	27
1.6.2 n 个事件的相互独立性	28
1.6.3 事件独立性的应用	29
1.6.4 独立试验序列模型	31
习题一	32
第 2 章 随机变量及其概率分布	37
2.1 随机变量	37
2.1.1 直观背景及定义	37
2.1.2 随机变量的分布函数	38
2.2 离散型随机变量及其概率分布	40
2.2.1 离散型随机变量的概念及概率分布列	40

2.2.2	常见离散型随机变量及分布列	41
2.3	连续型随机变量及其概率密度函数	46
2.3.1	连续型随机变量的概念及概率密度函数	46
2.3.2	常见连续型随机变量及其概率密度函数	48
2.4	多维随机变量及其分布	54
2.4.1	二维随机变量及其分布函数	54
2.4.2	二维离散型随机变量	56
2.4.3	二维连续型随机变量	57
2.4.4	n 维随机变量	60
2.5	随机变量的独立性, 条件分布	61
2.5.1	相互独立的随机变量	61
2.5.2	条件分布	65
2.6	随机变量的变换及其分布	68
2.6.1	一个随机变量函数的分布	68
2.6.2	二维随机变量函数的分布	71
2.6.3	χ^2 分布、 t 分布、 F 分布	77
	习题二	80
第 3 章	随机变量的数字特征	88
3.1	随机变量的数学期望	88
3.1.1	离散型随机变量的数学期望	88
3.1.2	连续型随机变量的数学期望	89
3.1.3	数学期望的一般定义	90
3.1.4	随机变量函数的数学期望	91
3.1.5	数学期望的性质	92
3.2	随机变量的方差	95
3.2.1	方差的定义	95
3.2.2	方差的性质及切比雪夫不等式	96
3.3	常用概率分布的期望和方差	98
3.3.1	常用离散型随机变量的期望和方差	98
3.3.2	常用连续型变量的期望和方差	102
3.4	多维随机变量的数字特征	104
3.4.1	协方差和相关系数	104
3.4.2	多维随机变量的期望和协方差矩阵	108
3.5	其他常用数字特征	109
3.5.1	矩 (moment)	110
3.5.2	变异系数	111
3.5.3	偏态系数	111

3.5.4 峰态系数	112
3.5.5 分位数	112
3.5.6 中位数	113
3.6 条件数学期望	114
习题三	117
第 4 章 大数定律与中心极限定理	123
4.1 特征函数	123
4.1.1 特征函数的概念	123
4.1.2 特征函数的性质	124
4.1.3 特征函数和分布函数之间的关系	127
4.2 随机变量序列的两种收敛性	130
4.2.1 依概率收敛	130
4.2.2 依分布收敛、弱收敛	131
4.3 大数定律	133
4.3.1 大数定律的一般形式	133
4.3.2 切比雪夫大数定律	134
4.3.3 辛钦大数定律	136
4.4 中心极限定理	137
4.4.1 中心极限定理的一般概念	137
4.4.2 独立同分布情形的中心极限定理	138
4.4.3 独立不同分布情形的中心极限定理	140
习题四	141
第 5 章 统计学的基本概念	144
5.1 导 言	144
5.1.1 统计学的任务	144
5.1.2 统计学的应用	145
5.1.3 学习建议	146
5.2 总体与样本	147
5.2.1 总体与样本的概念	147
5.2.2 无限总体与有限总体	148
5.2.3 样本的二重性和样本分布	148
5.3 样本数据及其分布的描述	150
5.3.1 数据的类型	150
5.3.2 频数与频率	152
5.3.3 累加频数和累加频率	154
5.3.4 直方图	155

5.3.5	茎叶图	157
5.3.6	经验分布函数	159
5.4	统计量和抽样分布	160
5.4.1	统计量	160
5.4.2	正态总体抽样分布	162
5.4.3	次序统计量及其分布	166
	习题五	170
第 6 章	参数估计	173
6.1	点估计	173
6.1.1	矩估计	174
6.1.2	极大似然估计	175
6.2	评价估计量的准则	179
6.2.1	无偏性	179
6.2.2	有效性	180
6.2.3	均方误差	181
6.2.4	相合性	182
6.2.5	稳健性 (robustness)	183
6.3	最小方差无偏估计量	184
6.4	贝叶斯 (Bayes) 估计	188
6.4.1	贝叶斯统计的基本思想	188
6.4.2	贝叶斯公式的概率函数形式	188
6.4.3	贝叶斯估计	189
6.5	区间估计 (置信区间)	192
6.5.1	区间估计的概念	192
6.5.2	区间估计的求法 —— 枢轴量法	193
6.5.3	单个正态总体参数的区间估计	194
6.5.4	两个正态总体参数的区间估计	196
6.5.5	非正态总体参数的区间估计	198
	习题六	198
第 7 章	假设检验	203
7.1	基本概念	203
7.1.1	统计假设与检验法则	203
7.1.2	两类错误	205
7.1.3	检验的功效和显著性水平	209
7.2	单个正态总体均值与方差的假设检验	211
7.2.1	已知 σ^2 , 检验关于 μ 的假设	211

7.2.2	σ^2 未知, 检验关于 μ 的假设	213
7.2.3	检验关于 σ^2 的假设	215
7.3	两个正态总体均值与方差的假设检验	216
7.3.1	方差已知时均值的检验	217
7.3.2	方差未知但相等时均值的检验	218
7.3.3	方差未知 (且不假定相等) 时均值的检验	218
7.3.4	方差的检验	218
7.4	成对数据比较检验法	221
7.5	其他分布参数的假设检验	225
7.5.1	指数分布参数的假设检验	225
7.5.2	比例参数的假设检验 (小样本检验 —— 基于二项分布)	226
7.5.3	比例参数假设的大样本检验 (基于正态分布)	228
7.5.4	两个比例参数的比较检验	229
7.6	分布拟合检验	230
7.6.1	分类数据的 χ^2 检验法	230
7.6.2	总体分布的假设检验	231
7.6.3	列联表和独立性检验	234
7.7	两个重要的非参数检验 —— 符号检验与秩和检验	236
7.7.1	符号检验	237
7.7.2	秩和检验	239
7.8	检验的 p 值	241
	习题七	244
第 8 章	回归分析与相关分析	248
8.1	相关关系	248
8.2	一元线性回归	249
8.2.1	基本概念	249
8.2.2	a 、 b 的估计 —— 最小二乘法	250
8.2.3	随机误差的方差 σ^2 的估计	252
8.2.4	关于 a 、 b 的统计推断	253
8.3	多元线性回归	261
8.3.1	多元线性回归模型	261
8.3.2	最小二乘估计	262
8.3.3	回归方程与回归系数的显著性检验	266
8.4	可线性化的回归方程	267
8.4.1	常用的可化为线性函数的回归函数	267
8.4.2	多项式回归	268
8.5	相关分析	269

8.5.1	相关与回归的区别	269
8.5.2	相关系数	270
8.5.3	散点图	271
8.6	积差相关系数——数值型变量间相关性度量	273
8.6.1	积差相关系数的概念	273
8.6.2	相关性检验	274
8.6.3	多个变量的情况	275
8.7	秩相关系数	276
8.7.1	Spearman 秩相关系数	276
8.7.2	Kendall τ 相关系数	278
8.7.3	多变量 Kendall 协和系数	280
	习题八	281
第 9 章	方差分析	283
9.1	单因素试验的方差分析	283
9.1.1	单因素方差分析的统计模型	283
9.1.2	统计分析	285
9.1.3	应用举例	288
9.2	双因素试验的方差分析	290
9.2.1	双因素方差分析的统计模型	291
9.2.2	统计分析	292
9.2.3	无重复试验的方差分析	297
	习题九	299
第 10 章	Excel 在概率统计中的应用	302
10.1	Excel 简介	302
10.1.1	Excel 与统计软件	302
10.1.2	Excel 的分析工具库	302
10.2	常见概率分布的计算	304
10.2.1	二项分布	304
10.2.2	负二项分布	305
10.2.3	几何分布	305
10.2.4	超几何分布	306
10.2.5	泊松分布	306
10.2.6	正态分布	307
10.2.7	对数正态分布	308
10.2.8	指数分布	308
10.2.9	贝塔分布 (均匀分布)	309

10.2.10 Γ 分布 (χ^2 分布)	310
10.3 在假设检验中使用 Excel 软件	311
10.3.1 Z 检验 —— 单样本情形	311
10.3.2 Z 检验 —— 双样本情形	311
10.3.3 t 检验 —— 单样本情形	313
10.3.4 t 检验 —— 两个样本的情形	313
10.3.5 F 检验 —— 两总体方差的假设检验	315
10.3.6 χ^2 检验 —— 单个总体方差的假设检验	316
10.3.7 χ^2 检验 —— 独立性假设检验	316
10.4 方差分析	317
10.4.1 单因素方差分析	317
10.4.2 双因素方差分析 —— 无交互作用	319
10.4.3 双因素方差分析 —— 有交互作用	320
10.5 相关分析与回归分析	321
10.5.1 相关分析	321
10.5.2 一元线性回归分析与预测	322
10.5.3 多元线性回归分析与预测	324
习题答案	327
附表	337
附表 1 二项分布的数值表	337
附表 2 泊松分布表	341
附表 3 标准正态分布函数表	342
附表 4 χ^2 分布的分位数表	343
附表 5 F 分布的分位数表	344
附表 6 t 分布的分位数表	350
附表 7 两样本秩和双边检验的临界值表	350
附表 8 相关系数临界值表	351
参考文献	353

表格目录

1.1	种子发芽统计数据	14
1.2	英文字母的使用频率	14
1.3	英文字母使用统计数据	15
3.1	常用离散型分布的期望和方差	102
3.2	常用连续型分布的期望和方差	104
4.1	常用分布的特征函数	124
5.1	不光滑焊点频数、频率分布表	152
5.2	螺钉次品数的频数、频率分布表	153
5.3	某月城市 A、B 交通事故的经济损失频数和频率分布表	154
5.4	某月城市 A、B 交通事故的经济损失累加频率分布	155
5.5	100 个零件的重量 (单位: 克)	156
5.6	100 个零件的重量变换后的数据	156
5.7	频数和频率分布表	157
5.8	两个车间各 40 名工人的产量	159
5.9	容量为 3 的一个样本和次序统计量的所有取值	166
7.1	检验的两类错误	205
7.2	单个正态总体均值的假设检验	213
7.3	单个正态总体方差的假设检验	216
7.4	两个正态总体的均值的假设检验	217
7.5	两正态总体方差的假设检验	219
7.6	成对纪录的数据结构	222
7.7	例 7.4.4 成对数据的纪录	223
7.8	例 7.4.5 成对数据的纪录	224
7.9	例 7.4.6 成对数据的纪录	224
7.10	指数分布参数的假设检验	226
7.11	关于比例参数 p 的假设检验	228
7.12	例 7.6.1 中 χ^2 统计量值的计算	231
7.13	例 7.6.2 中 χ^2 统计量值的计算	233
7.14	某种铀所放射到达计数器上的 α 粒子数	233
7.15	某种铀所放射到达计数器上的 α 粒子数及其计算结果	234
7.16	2×3 列联表	234
7.17	$r \times s$ 列联表	235

7.18	例 7.6.4 理论频数 E_{ij} 的计算	236
7.19	甲、乙两班生产棉条均匀度数据	238
7.20	当 $n_1 = 3, n_2 = 4$ 时, 第一样本秩的所有可能取值	240
7.21	当 $n_1 = 3, n_2 = 4$ 时, 第一样本秩和的概率分布	240
7.22	不同显著性水平下的拒绝域和检验结论	242
8.1	5 名学生四项测验的成绩	271
9.1	一个单因素试验的结果	283
9.2	单因素试验的样本数据结构	284
9.3	单因素方差分析表	287
9.4	某种纤维的缩水率	288
9.5	纤维缩水率方差分析表	289
9.6	不同方案生产的产品的废品率	290
9.7	废品率的方差分析表	290
9.8	两组双因素试验结果	291
9.9	两因素试验的样本数据结构	291
9.10	双因素试验方差分析表	295
9.11	不同生产条件制造的硬橡胶的抗牵强度	296
9.12	硬橡胶抗牵强度的方差分析表	296
9.13	双因素方差分析表 (不考虑交互作用)	298
9.14	纤维的缩水率 (%)	298
9.15	纤维的缩水率变换后的数据和计算	299
9.16	纤维缩水率的双因素方差分析表	299

插图目录

1.1 韦恩图 ($A \cup B, A \cap B, A - B, A^c$ 分别为图中阴影部分)	4
1.2 几何概率示意图	16
1.3 例 1.3.12 示意图	16
1.4 蒲丰投针示意图	17
1.5 蒲丰投针概率计算示意图	17
1.6 串联系统示意图	30
1.7 并联系统示意图	30
1.8 一个开关电路图	30
2.1 例 2.3.1 中密度函数	47
2.2 例 2.3.1 中分布函数	47
2.3 均匀分布的密度函数	48
2.4 均匀分布的分布函数	48
2.5 指数分布的密度函数	49
2.6 指数分布的分布函数	49
2.7 正态分布的密度函数	50
2.8 正态分布的分布函数	50
2.9 不同 σ 的正态密度函数	51
2.10 不同 μ 的正态密度函数	51
2.11 标准正态密度函数	51
2.12 标准正态分布函数	51
2.13 Γ 分布密度函数 ($\beta = 1, \alpha = 1, 2, 4$)	53
2.14 贝塔分布概率密度函数	53
2.15 正态和柯西分布的密度函数	54
2.16 柯西分布和均匀分布的联系	54
2.17 联合分布函数的几何意义	55
2.18 随机向量在矩形区域取值的概率	55
3.1 具有三种不同偏度的密度函数	111
3.2 分位数和上分位数	112
3.3 连续型随机变量的中位数	113
5.1 统计数据的分类	151
5.2 电路板不光滑焊点的频率分布	152
5.3 螺钉次品数的频率分布	153

5.4	某月城市 A、B 交通事故的经济损失频率分布	154
5.5	某月城市 A、B 交通事故的经济损失累积频率分布	155
5.6	直方图示意图	155
5.7	直方图的例子	157
5.8	某电脑公司销售量数据的茎叶图	158
5.9	某电脑公司销售量数据的扩展茎叶图	159
5.10	某两车间产量的背靠背茎叶图	160
5.11	标准正态分布函数和两个经验分布函数	160
6.1	标准正态分位数点	194
6.2	t 分布分位数点	195
6.3	χ^2 分布分位点	196
6.4	F 分布分位点	197
7.1	检验的拒绝域和接受域	207
7.2	参数的真值越接近原假设下的值时, β 值 (阴影部分面积) 就越大	208
7.3	例 7.1.8 中两个检验的功效函数	210
7.4	例 7.1.9 中一个检验的功效函数	210
7.5	假设 (7.2.8) 的水平为 α 的接受域	216
7.6	假设 (7.3.9) 的水平为 α 的接受域和拒绝域	219
7.7	假设检验的 p 值示意图	243
8.1	一元回归函数示意图	248
8.2	散点图	250
8.3	回归直线示意图	255
8.4	回归函数置信区间示意图	258
8.5	控制求解图示	261
8.6	利用数据的一致性说明相关关系图示	271
8.7	完全正相关和完全负相关图示	272
8.8	不同形状的散点图表示的相关度	272
8.9	标准化数据的散点图	272
8.10	相关系数为 0 时两个变量的散点图	274
10.1	Excel 的数据分析工具库	302
10.2	ZTEST 函数做双边检验	312
10.3	【Z 检验: 双样本平均差检验】对话框	312
10.4	【Z 检验: 双样本平均差检验】结果	313
10.5	t 检验 —— 成对样本的均值的检验结果	314
10.6	F 检验 —— 两总体方差的假设检验	315