

高等院校统计学专业规划教材

概率论与数理统计

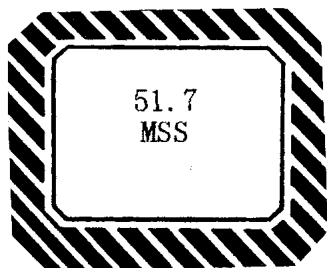
茆诗松 周纪芃 编著

中国统计出版社

高等院校统计学专业规划教材

概率论与数理统计

茆诗松 周纪芑 编著



中国统计出版社

(京)新登字 041 号

· 图书在版编目 (CIP) 数据

概率论与数理统计/茆诗松, 周纪芾编著

—北京: 中国统计出版社, 1996. 8

高等院校统计学专业规划教材

ISBN 7-5037-2025-5

I. 概…

II. ①茆… ②周…

III. ①概率论-高等学校-教材

②数理统计-高等学校-教材

IV. 021

中国统计出版社出版

(北京三里河月坛南街 38 号 100826)

新华书店经销

北京科技印刷厂印刷

*

850×1168 毫米 32 开本 20 印张 50 万字

1996 年 7 月第 1 版 1996 年 7 月北京第 1 次印刷

印数: 1—5000 册

*

定价: 18.70 元

(版权所有 不得翻印)

序 言

本书是按照全国统计教材编审委员会制定的《概率论与数理统计教学大纲》编写的，是供全国高等学校统计专业本科生学习用的教科书。全书十章分二部分，前四章是概率论部分，主要讲述概率论的基本概念和基本结论，其中心内容是随机变量及其分布，后六章是数理统计部分，主要讲述数理统计基本概念和常用统计方法，其中心内容是统计推断的三个内容：抽样分布、参数估计和假设检验。

学习这门课的读者主要是着眼于社会、经济管理领域中的应用，因此我们尽量用社会、经济、管理方面的例子讲述各种基本概念、基本理论和基本方法，努力说明其丰富的实际背景、特有的思维方式、广泛的应用范围。虽然全书有 200 个例子，为数不少，但毕竟有限，为了今后能很好应用统计方法，现在要把学习重点放在对概念、定理和方法的直观理解和数学表达上，只有理解了的东西才能更深刻地感觉它，从而才能正确地使用它，准确的数学表达是检验你是否理解了，基于这个考虑，我们在叙述上尽量启发你的思维，推理和演算上坚持严谨，能证则证，这一种严格训练对进一步学习后继的统计课程和今后的应用是十分必须的。书中部分节与段打了 * 号，在教学中可以删去，因为这都是扩大和加深知识面的内容。

本书各章后都附有大量习题，其中大部分是练习性的，以巩固本书内容为主要目的，真正有难度的题目只占少部分，独立地完成这些题目对掌握这门课程是必不可少的。如果在做习题上不肯化功夫，有畏难情绪，那今后在应用中遇到困难时，怎能有效

关的勇气和能力呢？

本书的编写自始至终得到国家统计局统计干部培训中心的关心和帮助，中国人民大学倪加勋教授耐心细致地审阅全书也使本书增色不少，中国统计出版社副总编谢鸿光先生为编辑出版此书化了很多心血，尤进红亦为本书提供大量习题，在此一并表示衷心的感谢。

由于编者水平有限，错谬之处在所难免，恳请国内同行和广大读者批评指正。

茆诗松 周纪芃

1996年2月6日

目 录

第一章 随机事件及其概率	1
§ 1.1 随机事件及其运算	1
一、随机现象	1
二、基本空间 (样本空间)	2
三、随机事件	3
四、必然事件与不可能事件	4
五、事件间的关系	5
六、事件的运算	6
七、事件的运算性质	8
§ 1.2 事件的概率	9
一、事件的概率	9
二、古典方法	10
三、频率方法	19
四、主观方法	22
§ 1.3 概率的性质	24
§ 1.4 独立性	31
一、两个事件的独立性	31
二、多个事件的独立性	33
三、试验的独立性	35
§ 1.5 条件概率	36
一、条件概率	36
二、条件概率的性质	39
三、全概率公式	42
四、贝叶斯公式	45
习题	49

第二章 随机变量及其概率分布	55
§ 2.1 随机变量	55
一、随机变量	55
二、离散随机变量的概率分布	58
三、连续随机变量的概率分布	61
四、分布函数	66
§ 2.2 随机变量的数学期望(均值)与方差	70
一、数学期望(均值)	70
二、数学期望的性质	75
三、方差	77
§ 2.3 常用离散分布	83
一、二点分布	83
二、二项分布	84
三、泊松分布	90
四、超几何分布	96
§ 2.4 常用连续分布	99
一、正态分布	99
二、 Γ 分布	112
三、 β 分布	116
§ 2.5 随机变量的其它特征数	120
一、矩	120
* 二、变异系数	121
* 三、偏度	122
* 四、峰度	124
五、中位数	125
六、分位数	126
七、众数	127
习题	128
第三章 多维随机变量	134
§ 3.1 多维随机变量及其联合分布	134
一、多维随机变量	134

二、联合分布函数	135
三、多维离散随机变量	137
四、多维连续随机变量	141
§ 3.2 随机变量的独立性	148
一、随机变量的独立性	148
二、独立性的应用	151
三、独立随机变量和的分布	155
四、统计上三大分布的导出	160
§ 3.3 多维随机变量的特征数	167
一、多维随机变量函数的数学期望	167
二、数学期望与方差的运算性质	169
三、协方差	174
四、相关系数	178
五、协方差阵	183
§ 3.4 条件分布与条件期望	188
一、条件分布的概念	188
二、离散随机变量的条件分布	189
三、连续随机变量的条件分布	192
* 四、构造联合分布	194
五、条件期望	196
习题	202
第四章 极限定理	212
§ 4.1 大数定律	212
一、从“频率的稳定性”说起	212
二、切比雪夫大数定律	214
三、马尔可夫大数定律	215
* 四、辛钦大数定律	216
§ 4.2 中心极限定理	217
一、问题的提出	217
二、林德贝格-列维中心极限定理	220
* 三、林德贝格条件	223

习题	227
第五章 统计量及其分布	230
§ 5.1 总体与样本	231
一、总体与个体	231
二、样本	232
三、随机数表	233
§ 5.2 统计量与抽样分布	233
一、统计量的概念与常用统计量	234
二、抽样分布	236
三、与正态总体有关的抽样分布	238
§ 5.3 次序统计量及其分布	243
一、次序统计量的概念	243
二、与次序统计量有关的常用统计量	245
三、次序统计量的分布	246
四、用随机模拟方法寻找统计量的近似分布	248
§ 5.4 数据的整理	251
一、频数、频率分布表及其图示	251
二、经验分布函数	256
三、概率纸方法	261
四、样本特征量	265
习题	267
第六章 参数估计	271
§ 6.1 矩法估计	272
§ 6.2 极大似然估计	274
§ 6.3 点估计优劣的评价标准	282
一、无偏性	282
二、有效性	284
三、均方误差准则	285
四、相合性	287
§ 6.4 区间估计	287
一、区间估计的概念	287

二、枢轴量法	288
§ 6.5 正态总体参数的置信区间	292
一、正态均值 μ 的置信区间	292
二、正态方差 σ^2 的置信区间	293
三、两个正态均值差的置信区间	294
四、两个正态方差比的置信区间	298
§ 6.6 单侧置信限	299
一、单侧置信限概念	299
二、枢轴量法	299
三、一般方法	301
§ 6.7 比率的区间估计	306
一、小样本场合下 p 的精确置信区间	306
二、大样本场合下 p 的近似置信区间	309
§ 6.8 泊松分布中参数 λ 的区间估计	311
一、小样本场合下 λ 的精确置信区间	311
二、大样本场合下 λ 的近似置信区间	313
习题	314
第七章 假设检验	321
§ 7.1 假设检验的概念与步骤	321
一、什么是假设检验	321
二、假设	325
三、两类错误	326
四、检验水平为 α 的检验	328
五、假设检验问题的类型	329
§ 7.2 正态总体参数的假设检验	332
一、关于均值的检验	332
二、关于方差的检验	339
三、关于两个正态总体方差的检验问题	341
四、关于两个正态总体均值差的检验问题	343
§ 7.3 比率 p 的检验	349
一、关于比率 p 的检验	349

二、两个比率的比较	354
§ 7.4 泊松分布参数 λ 的检验	356
§ 7.5 检验的 p 值	359
* § 7.6 广义似然比检验	363
§ 7.7 χ^2 拟合优度检验	365
一、总体可分为有限类, 且总体分布已知	366
二、总体可分为有限类, 但总体分布不完全已知	367
三、总体为连续分布的情况	370
四、列联表的独立性检验	372
* 五、中位数检验	376
§ 7.8 符号检验	378
一、单个总体的符号检验	378
二、两个总体的符号检验	381
§ 7.9 秩检验	383
一、秩和检验	384
二、符号秩和检验	388
* § 7.10 游程检验	390
一、检验两个总体是否具有相同的分布	391
二、检验序列的随机性	393
§ 7.11 正态性检验	394
一、小样本 ($3 \leq n \leq 50$) 场合的 W 检验	395
二、大样本场合 ($n > 50$) 的 D 检验	397
习题	398
第八章 方差分析	408
§ 8.1 单因子方差分析	408
一、问题的提出	408
二、单因子方差分析的统计模型	409
三、检验方法	411
四、效应与误差方差的估计	416
五、各水平下重复试验次数不等的情况	418
§ 8.2 多重比较	421

一、重复数相等场合的 T 法	421
二、重复数不等场合的 S 法	423
§ 8.3 方差齐性检验	425
一、样本容量相等的场合	425
二、样本容量不等的场合	427
§ 8.4 两因子方差分析	429
一、交互作用	430
二、两因子方差分析的统计模型	432
三、有交互作用的情况	435
四、无交互作用的情况	441
习题	447
第九章 回归分析	451
§ 9.1 一元线性回归	452
一、一元线性回归模型	452
二、回归系数的最小二乘估计	454
三、回归方程的显著性检验	458
四、利用回归方程作预测	464
* 五、利用回归方程作控制	468
* 六、重复观测 (试验) 的情况	470
§ 9.2 可化为一元线性回归的 <u>曲线回归</u>	474
一、模型的确定	474
二、参数估计	478
三、回归曲线的比较	479
§ 9.3 多元线性回归	481
一、多元线性回归模型	481
二、回归系数的最小二乘估计	482
三、回归方程的显著性检验	485
四、回归系数的显著性检验	488
五、预测	492
§ 9.4 回归诊断	494
一、残差及其性质	495

二、模型假定的诊断	497
习题	505
第十章 贝叶斯统计初步	512
§ 10.1 先验分布与后验分布	512
一、贝叶斯统计所利用的三种信息	512
二、贝叶斯公式	515
三、共轭先验分布	520
* 四、确定先验分布的方法	523
§ 10.2 贝叶斯估计	531
一、损失函数	531
二、贝叶斯估计	533
三、例	536
* 四、贝叶斯估计的误差	541
§ 10.3 贝叶斯区间估计	543
一、可信区间	543
二、最大后验密度可信区间 (HPD 可信区间)	546
习题	550
参考文献	557
附表	
附表 1 二项分布表 $P(X \leq x)$	558
附表 2 泊松分布表 $P(X \leq x)$	568
附表 3 正态分布表 $\Phi(u)$	573
附表 4 t 分布分位数 $t_{1-\alpha}(n)$ 表	574
附表 5 χ^2 分布分位数 $\chi^2_{1-\alpha}(n)$ 表	575
附表 6 F 分布分位数 $F_{1-\alpha}(f_1, f_2)$ 表	577
附表 7 柯尔莫哥洛夫 D_n 检验的临界值 $D_{n,1-\alpha}$ 表	585
附表 8 正态分布容许区间 $\bar{x} \pm \lambda s$ 中系数 $\lambda(n, \beta, \gamma)$ 值表	586
附表 9 正态分布容许限 $\bar{x} \pm ks$ 或 $\bar{x} - ks$ 中系数 $k(n, \beta, \gamma)$ 值表	587
附表 10 秩和检验临界值	588

附表 11	符号秩和检验临界值	589
附表 12	游程总数检验临界值	590
附表 13	正态性检验统计量 W 的系数 $a_i(n)$ 的值	592
附表 14	正态性检验统计量 W 的系数 α 分位数 W_α	594
附表 15	正态性检验统计量 Y 的 α 分位数 Y_α	595
附表 16	多重比较的 $q_{1-\alpha}(r, f)$ 表	596
附表 17	$F_{\max}$ 的分位数表	599
附表 18	$G_{\max}$ 的分位数表	600
附表 19	检验相关系数 $\rho=0$ 的临界值表	602
附表 20	$D-W$ 检验临界值表	603
附表 21	随机数表	605
习题答案		606

第一章 随机事件及其概率

§ 1.1 随机事件及其运算

一、随机现象

随机现象是概率论与数理统计的研究对象。

在一定条件下,并不总是出现相同结果的现象称为随机现象。从这个定义可见,随机现象的结果至少要有二个,至于那一个出现,人们事先并不知道,这些都是随机性的特征。随机现象在人们的生活、生产和经济交往中经常出现。

例 1.1 随机现象的例子

1. 抛一枚硬币,可能出现正面朝上,也可能反面朝上。
2. 掷一颗骰子,朝上的一面可能是1点、2点、 $\cdots$, 6点中的某一个。
3. 一颗麦穗上长着麦粒的个数可能是50粒,但也可能多于50粒,也可能低于50粒。
4. 我国农户的年收入可能是1万元,但也可能高于1万元,也可能低于1万元。
5. 下一个交易日的上海证券交易所的综合指数(简称上证指数)可能是1000点,也可能高于或低于1000点。
6. 某产品的不合格品率可能是3%,但也可能高于3%,也可能低于3%。

读者还可列举很多有趣的随机现象。

很多随机现象是可以大量重复的,如抛一枚硬币可以无限次

重复，不同麦穗上的麦粒数可以大量观察等，这种可重复的随机现象又称为随机试验，简称试验。以后常把检查一件产品看作做一次试验，观察一颗麦穗上的麦粒数也看作一次试验。也有一些随机试验是不能重复的，明年世界经济的增长率是不能重复的，本书主要研究能大量重复的随机现象，但也十分注意研究不能重复的随机现象。因为后者在我们经济生活中占有重要地位。

二、基本空间（样本空间）

认识一个随机现象的关键就看你能否罗列出它的一切可能发生的基本结果。这里的“基本结果”是指随机现象的最简单的结果。如在检查产品中，若只关心合格与否，那其基本结果只有二个：合格品与不合格品；若为了满足顾客的不同需求，那就要把合格品再分为三组，这时其基本结果就有四个：一级品，二级品，三级品和不合格品。可见，基本结果是相对的，是随人们研究需要而划分的。

随机现象所有基本结果的全体称为这个随机现象的基本空间，常用集合 $\Omega = \{\omega\}$ 表示。其中元素 ω 就是基本结果。在数理统计中，基本空间又称为样本空间，基本结果又称为样本点。

例 1.2 例 1.1 中的六个随机现象的基本空间分别是：

1. 抛一枚硬币的基本空间 $\Omega = \{\text{正面}, \text{反面}\}$ 。
2. 掷一颗骰子的基本空间 $\Omega = \{1, 2, 3, 4, 5, 6\}$ 。
3. 一颗麦穗上的麦粒数的基本空间 $\Omega = \{0, 1, 2, \dots\}$ 。当然，一颗麦穗上长 1 万个麦粒是不可想象的，但我们仍把它作为一个基本结果放入基本空间是为了以后数学上处理方便，但又不会失真。类似地，一匹布上的疵点数、一小时内总机收到的电话呼叫次数都可用此基本空间描述。

4. 我国农户年收入用 x 表示，那末其基本空间 $\Omega = \{x : x \geq 0\}$ 。

5. 上证指数（用 x 表示的基本空间 $\Omega = \{x : x \geq 100\}$ ），因为上证指数 x 的起点是 100 点。

6. 产品的不合格品率 θ 的基本空间 $\Omega = \{\theta : 0 \leq \theta \leq 1\}$ 。

三、随机事件

随机现象的某种结果称为随机事件，简称事件，常用大写字母 A, B, C 等表示。如“产品是合格品”是检查产品这一随机现象的一个事件，“电视机寿命超过 1 万小时”是电视机寿命试验的一个事件。“我国农户年收入超过 1 万元”是一个事件。

任一个事件都可看成是由若干个基本结果组成的集合。它是基本空间 Ω 的某个子集。该子集中任一个基本结果发生了就认为该事件发生了。

例 1.3 掷一颗骰子，“出现 6 点”、“出现偶数点”、“出现点数不超过 2”、“出现点数不等于 3”都是事件，若依次记为 A, B, C, D ，那它们都可以用其基本空间 $\Omega = \{1, 2, 3, 4, 5, 6\}$ 的某个子集表示。

$$\begin{aligned} A &= \{6\}, & B &= \{2, 4, 6\} \\ C &= \{1, 2\} & D &= \{1, 2, 4, 5, 6\} \end{aligned}$$

可以设想你处在这样一种情况，在掷骰子前约定，出现偶数点（事件 B ）就中奖。若掷的结果出现 4 点，你就中奖了（事件 B 发生了），若掷出 2 点或 6 点，你也中奖了，这说明，事件 B 虽由三个基本结果组成，但只要其中任一个出现就说事件 B 发生了。

假如掷二颗骰子，这时基本结果可用一个数对 (x, y) 表示，其中 x 表示第一颗骰子出现的点数， y 表示第二颗骰子出现的点数，这一随机现象的基本空间为

$$\Omega_1 = \{(x, y) : x, y = 1, 2, 3, 4, 5, 6\}.$$

它含有 36 个基本结果。下列事件都可看作 Ω 的某个子集。

$$A_1 = \text{“点数之和等于 2”} = \{(1, 1)\}.$$

$$B_1 = \text{“点数之和等于 5”} = \{(1, 4), (2, 3), (3, 2), (4, 1)\}.$$

$$C_1 = \text{“点数之和超过 9”} = \{(4, 6), (5, 5), (6, 4), (5, 6), (6, 5), (6, 6)\}.$$