概率与统计

（第二版）

刘婉如　徐信之　高尚华

高等教育出版社

GAODENG JIAOYU CHUBANSHE

概率与统计

（第二版）

刘婉如　徐信之　高尚华

高等教育出版社

(京)112号

内容提要

本书在第一版的基础上，参照国家教委师范司1991年12月颁发的中学教师进修高等师范专科数学专业概率与数理统计基础课程的教学大纲编写而成。本书不求多求全，但求讲解详尽，便于自学。全书分为三部分，相应为概率论、数理统计初步及应用统计方法。

本书可供卫星电视教育、教育学院、函授院校作为教材使用，也可供计算数学专业、师专数学专业及师范院校非数学专业作为教材使用。

概率与统计
(第二版)
刘婉如 徐信之 高尚华
*
高等教育出版社出版
新华书店总店科技发行所发行
中国科学院印刷厂印装
*
开本850×1168 1/32 印张 10.5 字数 270 000
1987年8月第1版
1994年10月第2版 1994年10月第1次印刷
印数 0 001—4 095
ISBN7-04-004976-7/O·1364
定价 6.10 元

第二版前言

1991年12月，国家教委师范司颁发了中学教师进修高等师范专科数学专业各课程的教学大纲。受高等教育出版社委托，我们在保留本书第一版的体系、风格的前提下，参照概率论与数理统计基础课程的教学大纲，对《概率与统计》第一版进行了修订，修订时并注意吸收了一些辅导该课程的教师的意见。

编　者

1993．10．

第一版前言

本书是根据国家教委关于通过电视教学培训中学教师的指示，受高等教育出版社的委托，按照1986年6月《概率与统计》教材编写大纲讨论会上所审定的提纲，以陈家鼎、刘婉如、汪仁官编《概率统计讲义》为基础而编写成的。

本书是针对54学时的讲课时间编写的，因此只能讲解概率论与数理统计的一些基本内容以及某些可用简单形式表示出来、能直接应用的统计方法。本书不求多求全，但求讲解详细，便于自学；本书对基本概念，重要公式和定理的直观背景和实际意义尽量多做解释，多举例子，力求通俗易懂。本书不用实变函数论与测度论等专门的数学知识，只用普通的微积分知识。因此对于涉及到较深知识的结论不给出严格的数学证明。

全书分为三部分。第一部分是概率论（一、二、三章）；第二部分是数理统计初步（四章）；第三部分是应用统计方法（五、六章）。

在编写过程中我们得到汪仁官、徐文培、林鸿洲、周赛花等先生的帮助，谨在此表示衷心感谢。由于编者水平有限，时间仓促，书中的缺点、错误在所难免，恳请读者批评指正。

编　者

一九八七年冬于北京大学

目　录

第一章　随机事件与概率 …………………………………………………… 1

§1　随机事件及其概率 ………………………………………………… 1

§2　古典概型 ……………………………………………………………… 4

2.1　数数——排列与组合 ……………………………………………… 4

习题一 ………………………………………………………………… 10

2.2　古典概型 …………………………………………………………… 11

习题二 ………………………………………………………………… 18

§3　事件的运算及概率的加法公式 …………………………………… 19

3.1　事件的包含与相等 ………………………………………………… 19

3.2　事件的和与积 ……………………………………………………… 19

3.3　对立事件与事件的差 ……………………………………………… 20

3.4　事件的运算规律 …………………………………………………… 21

3.5　事件的互不相容性 ………………………………………………… 23

3.6　概率的加法公式 …………………………………………………… 23

习题三 ………………………………………………………………… 29

*§4　集合与事件 …………………………………………………………… 29

§5　条件概率·乘法公式·独立性 ……………………………………… 34

5.1　条件概率 …………………………………………………………… 34

5.2　乘法公式 …………………………………………………………… 35

5.3　独立性 ……………………………………………………………… 38

习题四 ………………………………………………………………… 43

§6　独立试验序列概型 ………………………………………………… 44

习题五 ………………………………………………………………… 49

§7　全概公式与逆概公式 ……………………………………………… 50

7.1　全概公式 …………………………………………………………… 50

7.2 逆概公式……………………………………………… 53
习题六……………………………………………… 55
第二章 随机变量的概率分布与数字特征 ……………… 56
§1 随机变量 ……………………………………………… 56
§2 离散型随机变量……………………………………… 59
2.1 概率分布列……………………………………………… 59
2.2 几类常用的概率分布列………………………………… 61
习题七……………………………………………… 70
§3 连续型随机变量……………………………………… 71
3.1 概率密度函数…………………………………………… 71
3.2 几种常用的连续型随机变量…………………………… 78
§4 分布函数与随机变量函数的分布…………………… 89
4.1 分布函数………………………………………………… 89
4.2 随机变量函数的分布…………………………………… 96
习题八………………………………………………101
§5 期望……………………………………………………103
5.1 离散型随机变量的期望………………………………104
5.2 几个常用的离散型随机变量的期望…………………105
5.3 连续型随机变量的期望………………………………108
5.4 几个常用的连续型随机变量的期望…………………110
5.5 随机变量函数的期望…………………………………112
5.6 期望的简单性质………………………………………114
习题九………………………………………………115
§6 方差……………………………………………………117
6.1 方差的概念……………………………………………117
6.2 常用分布的方差………………………………………120
6.3 方差的简单性质………………………………………124
6.4 切比雪夫不等式………………………………………124
习题十………………………………………………127
第三章 随机向量……………………………………………… 129
§1 随机向量的(联合)分布与边缘分布…………………130

1.1 二维离散型随机向量……130
1.2 边缘分布及其与联合分布的关系……135
1.3 二维连续型随机向量的分布密度……139
1.4 随机变量的独立性……147
习题十一……153
§2 两个随机变量的函数的分布……154
习题十二……161
§3 随机向量的数字特征……163
3.1 两个随机变量的函数的均值公式……163
3.2 均值与方差的性质……165
3.3 协方差……168
3.4 相关系数……171
习题十三……173
§4 n维随机向量……174
4.1 联合密度与边缘密度……175
4.2 独立性……177
4.3 n 个随机变量的函数的分布……177
4.4 数字特征……177
§5 大数定律和中心极限定理……182
习题十四……187
第四章 参数估计与假设检验……189
§1 总体与样本……189
§2 点估计……193
2.1 期望的点估计……193
2.2 方差的点估计……197
2.3 标准差的估计……199
2.4 样本平均值$\bar{x}$及样本方差s^2的简化算法……200
§3 最大似然估计……202
§4 矩估计……206
§5 介绍几种分布……207
5.1 χ^2分布……207

5.2 t 分布 ……209
5.3 F 分布……209
§6 期望的置信区间……210
6.1 已知方差，对期望作区间估计……211
6.2 未知方差，对期望作区间估计……214
§7 方差的置信区间……217
习题十五 …… 219
§8 假设检验 ……220
8.1 一个正态总体的假设检验……221
8.2 两个正态总体的假设检验……234
习题十六……239
第五章 回归分析方法……242
§1 一元线性回归 ……243
1.1 经验公式与最小二乘法……243
1.2 平方和分解公式与线性相关关系……249
1.3 数学模型与相关性检验……251
1.4 预报与控制……257
1.5 曲线改直……258
*§2 多元线性回归 ……262
2.1 模型……262
2.2 最小二乘估计与正规方程……263
2.3 平方和分解公式……263
2.4 相关性检验……264
2.5 偏回归平方和与因素主次的判别……264
习题十七……266
第六章 正交设计……268
§1 正交表简介……268
§2 正交表的应用……271
§3 正交设计的实施步骤 ……289
3.1 一般步骤……289
3.2 因素的挑选……290

3.3 位级的确定……………………………………………………290
3.4 正交表的选择……………………………………………………291
3.5 因素位级表的制定………………………………………………291
3.6 均衡分散性与整齐可比性………………………………………291
习题十八…………………………………………………………292
附表1 正态分布数值表………………………………………………294
附表2 t分布临界值表………………………………………………295
附表3 χ^2分布临界值表……………………………………………296
附表4 F分布临界值表($\alpha=0.05$)…………………………………297
附表5 F分布临界值表($\alpha=0.025$)…………………………………299
附表6 F分布临界值表($\alpha=0.01$)…………………………………301
常用正交表……………………………………………………………303
习题答案………………………………………………………………315

第一章　随机事件与概率

§1　随机事件及其概率

粗略地说，在一定的条件下，可能发生也可能不发生的事件叫做随机事件①．通常用大写的英文字母 A，B，C 等表示随机事件．

例1.1.1　抛掷一枚分币，落下后它的正面朝上(常把有币值的一面称为正面)，这是一个随机事件．记作 A =“正面朝上”；落下后正面朝下也是一个随机事件，记作 B =“正面朝下”．

例1.1.2　抛掷两枚分币，则

A =“两个都是正面朝上”，B =“两个都是正面朝下”，

C =“一个正面朝上，一个正面朝下”

都是随机事件．不难看出

D =“至少有一个正面朝上”

也是一个随机事件．

例1.1.3　掷一个匀称的骰子，则

A =“出现一点”，　B =“出现两点”，……，

F =“出现六点”

都是随机事件；

G =“出现奇数点”，　H =“出现偶数点”

也都是随机事件．

例1.1.4　在 10 个同类的产品中，有 8 个正品、2 个次品．从中任取 3 个，那么

① 对随机事件确切的叙述见定义1.1.5．

$$A=\text{“三个都是正品”},\quad B=\text{“至少一个是次品”}$$

均为随机事件；而事件

“三个都是次品”与“至少一个是正品”，

前者是不可能发生的；后者是必定要发生的。我们称不可能发生的事件为不可能事件，记作 V；称必定要发生的事件为必然事件，记作 U。如在例 1.1.3 中，“点数大于 6”是不可能事件；“点数不大于 6”是必然事件。不可能事件与必然事件是否发生都是确定的。但是为了今后讨论问题方便起见，将它们也都当作随机事件。

虽然在一次试验中对某随机事件是否发生不能预先知道，但是对它发生的可能性的大小是可以知道的。比如，在例1.1.1中，如果被抛掷的分币是匀称的，那么随机事件 A(=“正面朝上”)和随机事件 B(=“正面朝下”)发生的可能性是一样的；在例 1.1.2 中，如果两枚分币都是匀称的，那么随机事件 A(=“两个都是正面朝上”)和随机事件 B(=“两个都是正面朝下”)发生的可能性也是一样的，并且它们比随机事件 C(=“一个朝上，一个朝下”)发生的可能性要小。不仅如此，由我们的直觉还可以说，例 1.1.1 中的随机事件 A(=抛掷一枚分币出现“正面朝上”)发生的可能性比例 1.1.2 中的随机事件 A(=抛掷两枚分币出现“两个都是正面朝上”)发生的可能性要大。以上只是对事件发生的可能性的一种定性的描述。只停留在这样的描述上是很不够的，我们希望对它给出一种客观的定量的描述。

考虑例 1.1.1 中抛掷一枚分币的试验。该试验是在一定的条件下作的。比如说我们规定：“分币是匀称的，放在手心上，用一定的动作向上抛，让分币自由地落在有弹性的桌面上等等”。我们称这些条件为条件组 S。在条件组 S 的一次实现下，事件 A(=“正面朝上”)是否发生是不确定的。这仅是问题的一面。另一方面，当条件组 S 大量重复实现时，事件 A 发生的次数(频数)，能体现出一定的规律性，它约占总试验次数的一半。不难看出

$$A\text{ 发生的频率}\left(=\frac{\text{频数}}{\text{试验次数}}\right)\text{接近于}\frac{1}{2}$$

根据长期积累的经验，我们心目中的所谓某事件发生的可能性的大小，不就是这个“频率的稳定值”吗？

历史上，有些人作过成千上万次抛掷钱币的试验，其试验结果见下表。

试 验 者	抛 掷 次 数 n	出现“正面朝上”的次数 μ（频数）	频率 $\left(\frac{\mu}{n}\right)$
德摩根 (De Morgan)	2048	1061	0.518
蒲 丰 (Buffon)	4048	2048	0.5069
费 勒 (Feller)	10000	4979	0.4979
皮尔逊 (Pearson)	12000	6019	0.5016
皮尔逊 (Pearson)	24000	12012	0.5005

由表中的数字容易看出，抛掷次数越多，频率越接近0.5.

定义1.1.5① 在一组不变的条件S下，重复作n次试验.记n次试验中事件A发生的次数为μ. 当试验次数n很大时，如果频率μ/n稳定地在某一数值p的附近摆动，并且一般说来，摆动的幅度随着试验次数的增多而愈变愈小，则称A为随机事件，称数值p为随机事件A在条件组S下发生的概率，记作

$$P(A)=p$$

① 我们这里只给概率一个直观的朴素的描述，通常称为概率的统计定义.精确的数学定义，即所谓公理化定义，在本课程中将不叙述. 至于概率$P(A)$的实际计算法，定义本身也给出了一种近似求法，即作大量的试验，计算事件A发生的频率.虽然得到的是近似值，但我们相信读者不至于因为现实生活中某一数值的获得只是些近似值而感到不实在. 事实上，我们周围许多量的测量完全是近似的，如长度的概念并不会因为每次实测数值都是近似值而建立不起来，也不会因为温度计读数都是近似值而怀疑起“温度”的客观存在性.

显然，数值p是从数量上刻画了在条件组S下随机事件A发生的可能性的大小。

换言之，上述定义可以简述为："频率具有稳定性的事件叫做随机事件；频率的稳定值叫做该随机事件的概率。"

需要强调的是，人类的大量实践证明，实际上遇到的事件一般都是随机事件，也就是说它们都有确定的概率。以下我们常简称随机事件为事件，

由于事件A的频数大于等于0，并且不会超过试验次数，所以频率μ/n一定介于0与1之间，即

$$0 \leqslant \mu/n \leqslant 1$$

于是由定义1.1.5可知，对任何事件A都有

$$0 \leqslant P(A) \leqslant 1$$

对于必然事件U，它在每次试验中都发生，即U的频数μ等于试验次数n。而对于不可能事件V，不管做多少次试验，其频数总等于0。所以由定义1.1.5可知

$$P(U)=1, \qquad P(V)=0$$

§2 古典概型

2.1 数数——排列与组合

1. 排列

例1.2.1 问由1,2,3,4四个数字能组成多少无重复数字出现的两位数？

这里次序是很要紧的，例如13与31是不同的数。这种有次序的一排东西叫做一个排列。由列举法可知四个数字能组成12个无重复数字出现的两位数，即

12，13，14；21，23，24；31，32，34；41，42，43

换言之，十位数上的数字共有4种可能，即1，或2，或3，或4。取定十位数字后，由于要求无重复数字，所以剩下来只有3个数字可以放

在个位上.因此总共有 $4\times 3=12$ 个无重复数字的两位数.

问由 1,2,3,4 四个数字能组成多少无重复数字出现的三位数呢？首先考虑百位数上的数字共有 4 种可能；在百位数上取定一个数字后，十位数上的数字只有 3 种可能；十位数上再取定一个数字后，个位数上的数字只有 2 种可能。这样一来，总共有 $4\times 3\times 2=24$ 个无重复数字的三位数.

这个问题的一般提法是：从 n 个不同的东西 $a_1,a_2,\cdots,a_n$ 中，无放回地任取 $m\,(1\leqslant m\leqslant n)$ 个排成一列.问这样的排列共有多少种？

我们称这个问题为非重复的排列问题，简称排列问题．排列总数记为 P_n^m.下面来导出 P_n^m 的计算公式.注意 m 是不超过 n 的正整数．每个排列都是有次序的一排东西

$$a_{i_1}\quad a_{i_2}\quad \cdots\quad a_{i_m}$$

不难看出 a_{i_1} 有 n 种可能，选定 a_{i_1} 后，a_{i_2} 有 $n-1$ 种可能，a_{i_1}，a_{i_2} 选定后，a_{i_3} 有 $n-2$ 种可能，……，a_{i_1}，a_{i_2}，$\cdots$，$a_{i_{m-1}}$ 选定后，a_{i_m} 有 $n-(m-1)$ 种可能(因为这是非重复的排列)．于是

$$P_n^m=n(n-1)\cdots(n-m+1)$$

这是计算排列的一般公式．注意一共有 m 个因子相乘．特别地，当 $m=n$ 时，称为全排列．全排列的公式为

$$P_n^n=n(n-1)(n-2)\cdot\cdots\cdot 2\cdot 1$$

这个乘积在数学中经常出现．把它叫做 n 的阶乘，记作 $n!$，即

$$n!=n(n-1)(n-2)\cdot\cdots\cdot 2\cdot 1$$

以后为了方便，规定 $0!=1$．易知

$$P_n^m=\frac{n!}{(n-m)!}\qquad (1\leqslant m\leqslant n)$$

例1.2.2 有 3 个球，它们分别染着红、黄、蓝 3 种颜色.从这 3 个不同颜色的球中有放回地任取两次，每次取一个球，记下球的颜色和它出现的次序．问最多能出现多少种花样？

与例1.2.1不同，此例的排列是允许有重复的．因为按序抽两次，每次抽出的球在下一次抽取之前都要放回去，所以第二次抽

到的球可能与第一次抽到的相同.不难看出,有放回地任取两次,每次抽到的球的颜色都有 3 种可能,两次共有 $3\times3=9$ 种可能的花样。

这个问题的一般提法是:从n个不相同的东西$a_1,a_2,\cdots,a_n$中,有放回地任取 m 次,每次取一个,所得到不同的序列共有多少种?

显然m次有放回地抽取就得到由$a_1,a_2,\cdots,a_n$组成的可重复的排列。这种排列共有

$$\underbrace{n\times n\times\cdots\times n}_{m\text{个}}=n^m$$

种。这就是可重复排列的公式。

现在考虑稍微复杂点的一种排列问题。它所排列的 n 个东西不是各不相同的,其中有些东西彼此不能区别。为了讲清计算方法,我们借助一个例子。

试问用字母$PEPPER$能组成多少种不同的字母排列?

首先把 3 个P、2 个E之间彼此加以区别,记成

$$P_1E_1P_2P_3E_2R$$

由全排列公式,它可以组成$6!$种排列.但是,考虑这$6!$个排列中的任何一个,例如$P_1P_2E_1P_3E_2R$,若在 3 个P之间以及 2 个 E 之间交换位置,得到的排列仍是$PPEPER$型的。就是说$3!\,2!$个排列

$$\begin{array}{lll}
P_1P_2E_1P_3E_2R, & P_1P_2E_2P_3E_1R, & P_1P_3E_1P_2E_2R \\
P_1P_3E_2P_2E_1R, & P_2P_1E_1P_3E_2R, & P_2P_1E_2P_3E_1R \\
P_2P_3E_1P_1E_2R, & P_2P_3E_2P_1E_1R, & P_3P_1E_1P_2E_2R \\
P_3P_1E_2P_1E_1R, & P_3P_2E_1P_1E_2R, & P_3P_2E_2P_1E_1R
\end{array}$$

都是$PPEPER$型的。因此,所求的不同排列数等于

$$\frac{6!}{3!\;2!}=60$$

一般地,运用类似上述的推理可得:

如果n个东西中有某n_1个彼此不能区别,另 n_2 个彼此不能区别,……另n_r 个彼此不能区别($n_1+n_2+\cdots+n_r=n$),则其全排

列数等于

$$\frac{n!}{n_1!\ n_2!\ \cdots n_r!}$$

2. 组合

例1.2.3 将4本不同的书分给甲、乙两学生，如果甲得1本，乙得3本，问共有多少种分法？

设这4本不同的书为A,B,C,D。从中任取3本，不同的取法是

$$A,B,C;\qquad A,B,D;\qquad A,C,D;\qquad B,C,D$$

因此一共只有4种分法。容易看出这个问题与次序是无关的。

不讲次序的问题就是所谓的"组合问题"。组合问题的一般提法是：从n个各不相同的东西里，任取出m个(不管顺序)，问共有多少种取法？

每一种取法称为一个组合。不同的组合总数通常用符号C_n^m表示，或者用符号$\binom{n}{m}$表示。

注意，排列与组合的不同之处在于：从n个东西中取出m个后，排列是要考虑这m个东西的次序，而组合是不考虑这m个东西的次序。

下面推导$C_n^m(1\leqslant m\leqslant n)$的计算公式。设有$n$个东西：$a_1,a_2,\cdots,a_n$。当$m=1$时，任取其中一个，显然有$n$种取法。因为一个东西无所谓次序，因此排列与组合的种数是相同的，即

$$C_n^1=P_n^1=n$$

当$m=2$时，从n个东西里任取两个，每种取法(组合)可以排成两种次序.例如取到a_1与a_2,则可以排成a_1a_2与a_2a_1。由此可见排列数比组合数多一倍。故$P_n^2=2C_n^2$。于是

$$C_n^2=\frac{n(n-1)}{2}$$

当$m=3$时，从n个东西里任取三个，每种取法(组合)可以排成6种次序.例如取到a_1,a_2,a_3,则这三个东西可以排成$a_1a_2a_3$,