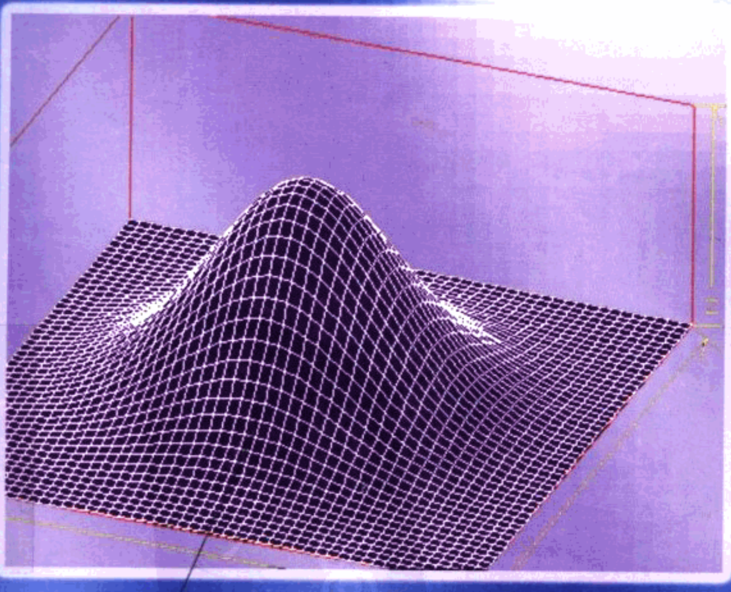
高等学校教材

概率论与数理统计

主　编：李从珠

副主编：黄伟峰　王建稳



中国商业出版社

前　　言

北方工业大学自1994年成立了“概率论与数理统计课程组”，负责全校概率论与数理统计课程的教学任务。本课程组成员在多年教学实践中积累了丰富的经验，编写了“概率论与数理统计”一书。本书共分两部分，即概率论部分（第一章至第五章）与数理统计部分（第六章至第十章），其中概率论是数理统计的理论基础，数理统计部分主要介绍统计方法及其应用，如参数估计、假设检验、方差分析和回归分析。本书作者在写作过程中，力求将概念写得直观明了、清晰易懂，尽量减少繁琐的理论证明，注重实际应用。在例题、习题的选择上，尽量体现启发性、应用性、趣味性。每章后面的习题和正文融为一体，有些习题是正文的补充。

本教材各章执笔人是：李从珠（第一章）、丁绍芳（第二章）、唐旭辉（第三章）、薛清波（第四章）、张颐（第五章、第六章）、侯峰（第七章）、黄伟峰（第八章）、王建稳（第九章、第十章）。全书由李从珠、黄伟峰和王建稳审阅，李从珠最后总纂定稿。侯志强负责全书的编排、插图的绘制以及封面设计，侯峰、崔玉杰负责全书的校对。另外，北方工业大学教务处对该教材的出版和印刷给予了大力支持，在此表示衷心的感谢。

本书可作为高等院校本科概率论与数理统计课程的教材。

由于我们水平有限，书中不足之处，在所难免，恳请读者斧正。

编者

2001年10月

目　　录

第一章　概率论的基本概念

§1.1 引言

现在人们对“概率”一词已不陌生，比如每天可在电视上看到某市（地）明天降水的概率是多少。概率是什么意思呢？直观地说，概率是度量事物发生可能性大小的一个数量指标。正像人们称重量用秤，量长度用米（或其他长度单位）一样，量一个事物发生可能性的大小用概率。

当我们观察自然界和社会上发生的多种现象时，会发现，有一类现象，在一定的条件下，必然发生。如：“太阳从东边出来”，“在一个标准大气压下，水加热到 100℃一定沸腾”，“同性电荷互相排斥”，等等，这类现象称为**必然现象**。另一类现象，在一定条件下一定不发生，如：“太阳从西边出来”，“异性电荷互相排斥”，等等，这类现象称为不可能现象。以上两类现象形式上虽然不同，但却有其共同点，那就是事先对其发生与否可以做出肯定的回答，所以这两类现象统称为**确定性现象**。然而在自然界和社会上还存在另一类与确定性现象有着本质区别的现象，它在一定的条件下可能发生也可能不发生，例如，从一群人中随便找出一个人，事先无法知道他手指中斗型纹的个数，事先也无法知道他的血型；有人买了一张奖券，在开奖前无法确知他是否中奖；对于刚怀孕的妇女，事先也无法知道她是生男还是生女等等。这类现象每次试验（或观测）的结果，都随“机会”而变，我们称这样的现象为**随机现象**。

人们长期的实践和研究发现，随机现象虽然每次试验的结果具有不确定性，但大量的重复试验和观测就会发现，它呈现某种

规律性。如，一孕妇生男生女虽事先难于预料，但大量观察就会发现男女大体各半。又如，任意找一个中国人，事先虽无法确知他斗型纹的手指个数，但大量观察表明中国人斗型纹约占 50%，等等。这种在大量重复试验和观测中事物所呈现出的固有规律性，我们称之为**统计规律性**。概率论和数理统计正是研究随机现象统计规律性的一门科学。

§1.2 概率的统计定义（频率）

随机试验的每一基本结果称为**基本事件**，也称**样本点**；所有基本结果构成的集合称为**样本空间**；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**，简称**事件**；在一定条件下必然发生的事件叫做**必然事件**；在一定条件下必然不发生的事件叫做**不可能事件**。以后我们分别以字母 Ω 和 Φ 表示必然事件和不可能事件；以字母 $A, B, C, \cdots$ 表示随机事件。

我们以抛掷一硬币为例，说明概率的统计定义。以 A 记“字面朝上”这一事件，显然，在上述条件实现一次时，事件 A 是否发生是不确定的，然而，大量重复试验抛掷硬币时，事件 A 发生的次数（称为**频数**），就体现出一定的规律性，即“字面朝上”出现的次数约占试验次数的一半。写成式子为：

事件 A 发生的频率=频数/试验总次数≈1/2。

历史上许多人作过这个试验，表 1.2.1 列出了他们中一些试验的记录。

从表 1.2.1 中可以看出，抛掷次数越多，频率越接近 0.5。

定义 1.2.1 在一定的条件下，重复做 n 次试验，n_A 为 n 次试验中事件 A 发生的次数，如果随着 n 逐渐增大，频率 $\frac{n_A}{n}$ 逐渐稳定在某一数值 p 附近，则数值 p 称为事件 A 在该条件下发生的概

表 1.2.1

实验者	试验次数 n	A 出现的次数 n_A	频率 $=n_A/n$
德摩尔根	2048	1061	0.5181
蒲丰	4049	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005
维尼	30000	14994	0.4998

率，记做 $P(A)=p$ 。这个定义称为概率的统计定义。

在历史上,第一个对“当试验次数 n 逐渐增大，频率 n_A 稳定在其概率 p 上”这一论断给以严格的意义和数学证明的是早期概率论史上最重要的学者雅各布伯努力(1654～1705)。

从概率的统计定义可以看到，数值 p 就是在该条件下刻画事件 A 发生可能性大小的一个数量指标。

由于频率 n_A/n 总是介于 0 和 1 之间，从定义 1.2.1 可知，对任意事件 A ，皆有 $0\le P(A)\le 1, P(\Omega)=1, P(\Phi)=0$ 。

在实际问题中，许多情况下，都是采用概率的统计定义。如：产品的合格率、某疾病的发病率、某良种的发芽率等等。所以概率的统计定义是很有用的。直观上说，概率的统计定义就是在试验次数 n 较大时，用频率代替概率。

§1.3 概率的古典定义（比率）

概率统计定义的优点是直观明了且实用性强，但用来计算事件的概率并不方便，因为用统计定义求概率，就要作试验，且试验次数 n 小了还不行，即便如此，也只能得到概率的近似值，即用频率来近似代替概率。为此引进概率的古典定义。

概率的古典定义是根据问题本身所具有的某种“对等性”，分析事物的本质，来直接计算概率的。

如果一个试验满足两条：

（1）试验只有有限个基本结果；

（2）试验的每个基本结果出现的可能性是一样的。

这样的试验，称为**古典试验**。

对于古典试验中的事件 A，它的概率定义为：

$$P(A)=\frac{m}{n} \tag{1.3.1}$$

n 表示该试验中所有可能出现基本结果的总数目。m 表示事件 A 包含的试验基本结果数。这种定义概率的方法称为**概率的古典定义**。

“古典”一词表示这一定义起源的古老，在概率论的萌芽期研究的对象全限于古典概率，它由何人何时提出，文献中并无明确记载，但可肯定的是其起源于赌博，是一种集体智慧。其中 16 世纪意大利数学家和赌博家卡丹诺（1501～1576）起了突出作用，1564 年他写的“机遇的博弈”是概率史上最早的著作。

古典概率用一比率定义概率，直观上也易理解，为此我们看 10 个人分 3 张戏票的问题，为了公平，通常用抓阄的办法解决，怎么抓阄呢？通常准备一个盒（袋）子，里面放 10 个大小质地一样的球，其中红球 3 个黑球 7 个，充分混合均匀后，让每人抽出一球，凡抽出红球者得票。在抽球前存在着 10 个同等可能的结果，其中有三个有利结果，这二者的比值就是每人得票的概率。有了概率的古典定义，使得计算某些概率十分方便，如：

掷一均匀硬币，$P\{\text{字面朝上}\}=1/2$；

掷一均匀骰子，$P\{\text{出现偶数点}\}=3/6=1/2$，$P\{\text{出现点数}>4\}=2/6=1/3$

在用古典定义计算时，首先要看是否是古典试验，否则就易出错，请看一个例子。

例 1.3.1 掷两颗均匀的骰子，点数之和为 2,3,…,12 共十一个

结果，试问点数之和是 7 的概率是1/11吗？

如果把试验结果看成共十一个，而点数之和为 7 仅是 11 个结果的一个，就用古典定义算出其概率为1/11。这就错了，因为当把该试验看成 11 个结果时，这 11 个结果不是等可能的，显然点数之和为 7 要比点数之和为 2 的概率大很多，故不能用古典定义。欲用此定义，需把试验结果看成 36 个，而点数之和为 7 的试验结果为

$$(1,6), (6,1), (2,5), (5,2), (3,4), (4,3)$$

共六个，其中 (i, j) 表示第一颗骰子出现 i 点，第二颗骰子出现 j 点，$i, j=1,2,\cdots,6$。所以点数之和为 7 的概率是 $6/36=1/6$，而不是$1/11$。

计算古典概率，除首先验证试验是否是古典实验外，剩下的就是数清楚公式（1.3.1）中的分子 m 和分母 n。为帮助读者数清楚 m 和 n，对于稍复杂点的问题，需借助于排列组合知识。

§1.4 排列组合与古典概率计算

1.4.1 元素不允许重复的排列

例 1.4.1 问1,2,3,4能组成多少个无重复数字的两位数、三位数、四位数？并排出来。

解 1,2,3,4共能组成以下 $4\times3=12$ 个无重复数字出现的两位数，它们是：

12 21 31 41

13 23 32 42

14 24 34 43

这是因为，十位数有 4 种可能性，而当十位数定了之后，个位数只有三种可能性。同理，1,2,3,4共能组成以下

$$4\times3\times2=24$$

种无重复数字的三位数，它们是：

$$\begin{cases}123\\124\end{cases}\quad\begin{cases}213\\214\end{cases}\quad\begin{cases}312\\314\end{cases}\quad\begin{cases}412\\413\end{cases}$$

$$\begin{cases}132\\134\end{cases}\quad\begin{cases}231\\234\end{cases}\quad\begin{cases}321\\324\end{cases}\quad\begin{cases}421\\423\end{cases}$$

$$\begin{cases}142\\143\end{cases}\quad\begin{cases}241\\243\end{cases}\quad\begin{cases}341\\342\end{cases}\quad\begin{cases}431\\432\end{cases}$$

不难看出，组成的无重复数字的四位数是

$4\times3\times2\times1=4!=24$ 种

由上列的讨论可以看出，1,2,3,4,5 共能组成 5×4 个无重复数字的二位数；$5\times4\times3$ 个无重复数字的三位数；$5\times4\times3\times2$ 个无重复数字的四位数；5！个无重复数字的五位数。

定义 1.4.1 将 n 个不同元素按一定的次序排列叫做 n 个元素的一个**选排列**。

从 n 个元素中取出 m 个（$m<n$，不允许重复）按一定的次序排列，叫做从 n 个元素取 m 个元素的**全排列**（简称排列）。

n 个元素的全排列总数为 $n!$；

n 中选 m 的选排列总数记为 A_n^m，则：

$$A_n^m=n(n-1)(n-2)\cdots(n-m+1)$$

例 1.4.2 假定北京和天津之间有 8 个火车客运站，问要印多少种不同的火车票？

解 包括北京、天津在内共有 10 个站，任意两个站如甲到乙是一种车票，乙到甲是另一种不同的火车票，所以共要印

$$A_{10}^{2}=10\times 9=90$$

种不同的火车票。

1.4.2 元素允许重复的排列

例 1.4.3 问1,2,3,4能组成多少个二位数，三位数？(每个数码允许重复取)

解 此题与例 1.4.1 不同，例 1.4.1 中，1,2,3,4 四个数不允许重复取，而这里允许重复，所以对两位数来说十位数有四种可能的取法，个位数也同样有四种取法，所以共能组成

$$4\times 4=4^{2}$$

种两位数，同理，可以组成

$$4\times 4\times 4=4^{3}$$

种不同的三位数，可以组成

$$4\times 4\times 4\times 4=4^{4}$$

种不同的四位数。一般地有：

定义 1.4.2 从 n 个不同元素取出 m 个（允许重复）元素的排列总数为 n^{m} 种。

例 1.4.4 某城用六位电话号码，而首位不允许用 0，问该城最多能安多少台“直拨”电话机。

解 首位不允许用 0，它只能是 1 至 9 这九个数字中的一个，第二位可以是 0,1,2,⋯,9 十个数字中的任意一个，同样，第三位、四位、五位、六位都有十种可能，故该城最多能安装的“直拨”电话机数是

$$9\times 10\times 10\times 10\times 10\times 10=900000$$

1.4.3 组合

从例 1.4.2 知，北京至天津间的火车票共有 90 种，但却只有

$90/2=45$ 种票价，这是因为“北京至天津”和“天津至北京”车票不同，但票价却是一样的。这实际上是可以看成从 10 个元素（车站）中选取二个元素，有几种选法的问题。这称为从 10 个元素中选取二个元素的组合问题。

定义 1.4.3 从 n 个元素中取出 m 个不同元素，不管顺序并成一组，叫做从 n 个元素取 m 个元素的**组合**。其组合总数记为 C_n^m。

排列与次序有关，如 ab 和 ba 是不同的排列，但却是同样的组合。

再如从 30 个人中选取 3 个人照像，问共有多少种照法？因照像与次序有关，同样的三个人，站的（或坐的）次序不同，其对应的照法也不同，故这是排列问题，所以共有

$$A_{30}^3=30\times29\times28=243600$$

种照法。

若从 30 人中选 3 人组成一个学习小组，问有多少种组织法？

这与照像不同，它与次序无关，只要是三个人是一样的，就是一个学习小组，故共有

$$C_{30}^3=A_{30}^3/3!=8120$$

种组织方法。

这就是说，排列数是组合数的阶乘倍，所以一般地有

$$C_n^m=A_n^m/m!=n!/m!(n-m)!$$

可以证明：

$$C_n^m=C_n^{n-m}\text{，}\quad C_n^m+C_n^{m-1}=C_{n+1}^m$$

这两个公式。

直接由组合的表达式立即可得：$2^n=\sum_{m=0}^{n}C_n^m$

例 1.4.5 八男五女组成的学习小组，选三个组长

（1）必须是一女二男，共有多少种选法？

（2）必须是二女一男，共有多少种选法？

（3）至少有一女，共有多少种选法？

解 （1）必须是一女二男，这一女是从五个女的中间选出的，故有 C_5^1 种选法，二男是从八个男的中间选出的，故有 C_8^2 种选法，而女的任意一种选法都对应着男的 C_8^2 种，所以共有

$$C_5^1\cdot C_8^2=5\cdot 28=140$$

种选法，同理有

（2）$C_5^2\cdot C_8^1=80$

（3）“至少一女”，这句话是指正好一个女的，或正好两个女的，或正好三个女的。故共有

$$C_5^1\cdot C_8^2+C_5^2\cdot C_8^1+C_5^3\cdot C_8^0=140+80+10=230$$

种选法。

或这样解：三个组长全是男的选法有 $C_5^0\cdot C_8^3=56$，而三个组长从 13 个人中任选的选法有 $C_{13}^3=286$，所以至少一女的选法有

$$C_{13}^3-C_5^0\cdot C_8^3=286-56=230\text{。}$$

两种解法结果一样的，但后者较简便。

例 1.4.6 有 10 人的会议，开会前两两互相握手并互相赠送照片一张，问整个会议上的人共握了几次手？共送了几张照片？读者自己算算看？

（答案为：$C_{10}^2 = 45, A_{10}^2 = 90$）

我们把 n 个元素取出 m 个元素的组合看成把 n 个元素分成两堆，（第一堆 m 个元素，第二堆 $n-m$ 个元素）的分法，每堆内的元素不计次序，这样就可将组合的概念拓广为把 n 个元素分成 k 堆（计较堆与堆的次序，而不计较堆内元素的次序）第 i 堆 n_i 个，$i=1,2,\cdots,k$，$\sum_{i=1}^{k} n_i = n$，则共有 $\dfrac{n!}{n_1!n_2!\cdots n_k!}$ 种分法。

1.4.4 古典概率计算

例 1.4.7 设有 N 件产品，其中有 D 件不合格品，今从中任取 n 件，试问这 n 件产品中恰有 $k(k \le D)$ 件不合格品的概率是多少？

解 在 N 件产品中任取 n 件（这里指不放回抽样）共有 C_N^n 种取法，从 D 件不合格品中取 k 件的可能取法为 C_D^k，在 $N-D$ 件合格品中取 $n-k$ 件的可能取法为 C_{N-D}^{n-k}，故所求概率为 $p=\dfrac{C_D^k C_{N-D}^{n-k}}{C_N^n}$

这个概率称为**超几何概率**。

例 1.4.8 将 n 个球任意地放入 $N(N \ge n)$ 个盒子中，试求每个盒子至多有一只球的概率（设盒子的容量不限）。

解 将 n 只球放入 N 个盒子中去，每一种放法是一基本事件。亦知，这是古典概率问题。因每一只球都可以放入 N 个盒子中的任一个盒子。故共有 $N \cdot N \cdots N = N^n$ 种不同的放法，而每个盒子

中至多放一只球共有 $N(N-1)\cdots[N-(n-1)]$ 种不同放法。因而所求的概率为

$$P=[N(N-1)\cdots(N-n+1)]/N^n=A_N^n/N^n$$

有许多问题和本例具有相同的数学模型。例如，假设每人生日在一年 365 天中的任一天是等可能的，即都等于 $1/365$，那么随机选取 $n(n\leq 365)$ 个人，他们的生日各不相同的概率为

$$[365\cdot 364\cdots(365-n+1)]/365^n$$

因而，n 个人中至少有两人生日相同的概率为

$$P=1-[365\cdot 364\cdots(365-n+1)]/365^n$$

经计算可得如下结果：

n	20	23	30	40	50	64	100
p	0.411	0.507	0.706	0.891	0.970	0.997	0.9999997

从上表可看出，在仅有 64 人的群体中，至少有两人同生日的概率很接近 1。

例 1.4.9 将 15 名乒乓球手随机地平均分配到三个组（组编号），这 15 名球手中有 3 名国手，试问：

（1）每组各分配 1 名国手的概率是多少？

（2）3 名国手分配在同一组的概率是多少？

解 15 名球手平均分配到 3 个组的分法总数 n 是 $15!/(5!5!5!)$。

（1）将 3 名国手分到 3 组每组 1 名国手的分法共 $3!$ 种，对这每一种分法，其余 12 名球手平均分到 3 个组的分法有 $12!/4!4!4!$ 种。故每组恰有一名国手的分法共有 $m=(3!12!)/(4!4!4!)$ 种。于是所求概率 P_1 为

$$P_1=m/n=0.2747$$

（2）将 3 名国手分到同一组内的分法有 3 种，对这每 1 种分

法，其余 12 名球手的分法（一组 2 名，另 2 组各 5 名）有 $12!/(2!5!5!)$ 种，因此 3 名国手分在同一组的分法共有 $(3\times12!)/(2!5!5!)$ 种，于是，所求概率 P_2 为

$P_2=0.0659$

有了排列组合知识，给计算古典概率带来了方便，但还不够，请看下例。

例 1.4.10 从数字 $1,2,\cdots,9$ 中每次任取一个，共取 n 次，试问这 n 个数的积被 10 除尽的概率是多少？

像这种类型的题即使用排列组合知识也不好算。我们试图把求一个较复杂事件的概率问题，转化为一些求较简单事件的概率。为此我们引进事件的关系与运算。

§1.5 事件的关系与运算，加法公理

1.5.1 包含关系

若事件 A 发生必导致事件 B 发生，则称 B **包含** A（见图 1－5－1），记作 $A\subset B$。若 $A\subset B$ 且 $B\subset A$，则 $A=B$，称 A **与** B **相等**。

1.5.2 事件的并与交

事件 A,B 的并（见图1－5－2）表示，A,B 两个事件中至少有一个发生，记为 $A\cup B$。事件 A,B 的交（见图1－5－3）表示 A,B 两事件中同时发生，记为 $A\cap B$，$A\cdot B$ 或 AB。

同理可定义 n 个事件的并与交。

例如，投掷两枚匀称的硬币，A 表示，“正好一个字面朝上”，B 表示“正好两个字面朝上”，C 表示“至少一个字面朝上”，于是