随 机 数 学（一）

葛余博

宋晖

清华大学数学系

2000年2月

主要符号表

凡例--

§1.6: 第一章第6节	**例 6.2**:本章第6节第二个例	(6.2):本章第6节第2式
as, ae: 几乎处处	$\Leftrightarrow$: 充要条件	:=, +:: 记为
ps: 概率空间	***df***: 分布函数	***pdf***: 概率密度函数
rv: 随机变量	$r\vec{v}$: 随机向量	*Iid*: 独立同分布

分布---

$B(n,p)$: 二项分布	$F(r,p)$: 负二项分布	$Ge(p)$: 几何分布
$P(\lambda)$: Poisson(泊松)分布	$U_{(a,b)}$: 均匀分布	$N(\mu,\sigma^2)$: 正态分布
$Ex(\lambda)$: 指数分布	$\Gamma(r,\lambda)$:Gamma(伽玛)分布	

矩---

μ_k: k阶矩（总体）	$\mu(=\mu_1)$: 数学期望（总体）	σ^2: 方差（总体）
M_k: 样本k阶矩	$\overline{X}(=M_1)$: 样本均值	S_n^2: 样本方差
S_n^2: 样本二阶中心矩		

百分位点---

z_α, $z_{1-\alpha}$	: $N(0,1)$的百分位点
$t_\alpha(n)$, $t_{1-\alpha}(n)$	: $t(n)$分布的百分位点
$\chi_\alpha^2(n)$, $\chi_{1-\alpha}^2(n)$	: $\chi^2(n)$分布的百分位点
$F_\alpha(n,m)$, $F_{1-\alpha}(n,m)$	: $F(n,m)$分布的百分位点

目　录

第一章　随机数学基础的基本概念

§1.1　引言

一、随机数学基础的对象和任务

随机数学范围很广，这里的随机数学基础主要包括传统意义上的概率论：概率论基础、数理统计和随机过程（初步），也包括概率论的应用中最活跃的一些方法和模型，例如 Monte-Carlo 方法、方差与回归分析. ARMA 模型、简单的多元统计及在信息学科广泛应用的 HMM（隐马尔可夫模型）等. 通过本书学习，能理解随机数学处理和研究随机现象的主要思想和方法，掌握重要的随机规律.

概率论是研究随机现象的数量规律的数学分支.

什么是随机现象？顾名思义，它是指一个随机的、偶然的自然现象或社会现象，它和必然现象是相对的. 北京地区冬季一定下雪，是必然现象，但降雪量多少，却是随机的；一个计算机网络上有商务信息和广告，是必然现象，而广告数量和做广告的各个企业将有多少收益，网络上访问某网站的次数，网络访问会不会遇到阻塞等等也都是随机的.

一类现象，在个别实验或观测中呈现出不确定性，在大量重复实验或观测时，又具有统计规律性，我们称它是随机现象.

"天有不测风云，人有旦夕祸福"，精彩地概述了随机现象无处不在，因此随机现象的研究便因普遍而重要. 天有不测风云，也有可测风云. 气象研究要涉及大量的随机的变量：气温、气压，气流以及降雨量等等，做气象预报就要观测和收集瞬息万变的数据，研究它们的变化规律，对明天及今后的天气形势做出预报. 说"不测"，只是因为现在对这些随机现象的规律性，把握得还不够好. "人有旦夕祸福"，正是发展各种社会保险的依据，也是要认真统计分析，进行研究的对象. 卫星发射能否成功，与发射系统的各个部件在发射过程中的性能参数以及部件间的连接协调是否合理可靠息息相关. 一个计算机网络的服务器应有怎样的配置，除物力和财力的限制外，当然要取决在网络开放时刻用户的各类需求数量，它们显然是随机的变量. 此外，某类产品的社会供求数量、股市中各上市公司的股票行情、穿过某十字路口的汽车和行人数量、一家商场在一天购买某类商品的数量及营业额、在某公共汽车站排队候车的人数与乘客的候车时间、某地区环境污染对地区流

行病的影响程度，以及对某项社会措施作计划中的民意测验会有的统计结果…都是随机变化的.

既然是随机的偶然的，有客观的数量规律么？如有，怎么找出这个数量规律？我们来看一个很著名的 Galton 钉板实验. 在一块平滑木板上如图 1.1 均匀钉上几排钉子，两侧钉有护栏，下方打上隔板，从左向右依次编号. 将此板倾斜放置，上方置一均匀小球，可使其滚下. 假设小球质量是均匀的，钉子是光滑的，并且钉子间的距离和护栏的位置，使得小球从上端落下或从上一排钉子间落下后必然碰到下一排钉子中的某一个，并且在我们假设的理想情况下，向右方和向左方落下的可能性各为一半，即 1/2. 如此滚下的小球，最后将落入哪个格子里去呢？显然小球落入哪个格子都是可能的，我们事先并不能肯定. 也就是说，结果是偶然的、随机的. 仔细分析一下，由于假定的理想条件，我们不难发现，假如小球第一次碰钉后向右落下（其可能性为 1/2），那么第二次碰钉（第二排右方的钉子）后仍然向右落下的可能性便是 $\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$，类似地或说对称地，两次碰钉都是向左落下的可能性也是 1/4. 而小球两次碰钉后从第二排中间空档落下的可能性则是 $\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$. 仿上，第三次碰钉后从 4 个空档落下的可能性，从左到右的分别为 1/8, 3/8, 3/8, 1/8. 以四排钉子为例，碰最后一排钉子后从 5 个空档落下，也即落入编号为 1 至 5 的 5 个格子的可能性则依次为 1/16，4/16，6/16，4/16 和 1/16. 可见，表面看来是偶然性起作用的地方，确实有内在的数量规律可循. 概率论的任务就是研究和发现各种随机现象中的客观规律，并掌握它们为经济建设、生产管理和科学研究服务.

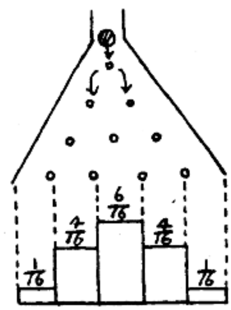

图 1.1 Galton 钉板实验

随着生产和社会经济的发展，科学研究的深入，概率论的理论和方法的研究和应用，大步向前. 有力地推动工农业生产、经济和金融管理、科学技术以及军事理论和技术的发展. 同时概率论自身也在日益丰富和深入. 概率论的理论和方法还向各个基础学科渗透，出现随机分析、随机微分方程、随机运筹和随机服务系统等等，并且随机模拟和概率统计计算也应运而生. 向工程科学渗透，出现随机信号处理、随机振动分析、与其它学科结合生长出生物统计、统计物理等边缘学科. 它也是人工智能、信息论、控制论、随机服务系统（排队论）、可靠性理论和风险分析与决策等学科的基础.

Galton 钉板试验也说明，随机现象中事件发生的可能性大小是客观存在的；因此可以对它进行量度. 量度的数量指标就是概率. 这个试验中，小球落入 5 个格子的概率就依次为 1/16，4/16，6/16，4/16 和 1/16.

二、随机数学基础的研究内容

再次回到 Galton 钉板实验，我们来速写随机数学基础的主要内容.

前面在理想条件下，就最简单的情况计算了一些事件发生的概率. 本书第一章“事件与概率”，要介绍事件、概率的概念，并介绍一些简单的概率模型（概型）及如何计算事件的概率. 介绍条件概率之后，引进几个重要公式. 在这个试验中，小球落入第 2 格，或是落入后两格，我们把这些结果都称为事件. 我们已经算出或者可以算出发生这些事件的可能性大小，或者说概率，例如小球落入第 2 格的概率是 4/16 = 0.25. 假如在 Galton 钉板实验中陆续滚下 100 个小球，在第 2 格你可能收集到 27 个小球，其频率 f 为 27/100. 但再落下 100 个小球，在第 2 格可能只收 20 个小球，频率 f 就变了. 通过试验我们会发现，落下 1 千个、1 万个，10 万个…，这个与落入总球数 n 有关的频率 $f(n)$会越来越靠近一个常数，也即当 n 趋于无穷时 $f(n)$有极限值. 这样小球落在第 2 格的概率也可用这个极限来定义. 这便是概率的统计定义，这个极限存在的事实，叫频率的稳定性，将在第四章里给出严格证明. 不难看到，利用频率稳定性，我们还可以在 Galton 钉板实验不是理想条件（光滑和均匀）下，近似求得事件的概率.

一个随机试验中，例如 Galton 钉板实验中，会有许许多多事件，这些事件的刻画可以通过叫做随机变量的取值来实现. 如令 X 是小球落入格子的序号数，则“小球落入第 2 格”可用“$X=2$”来表示. 现在假如，落下 100 个小球后，五个格子收集的小球数依次是 6，27，34，23，10. 于是我们可以画出如下的一个细线的频率直方图. 设想板上的钉子加密，增加行数，小球数量增加而质量相应减小，并且仍然假定光滑和均匀的理想情况，我们可能得到右图中的一个虚线的频率直方图. 而当钉子无限加密，小球质量小如一粒面粉. 这频率直方图将演化成右图里的一条曲线. 以粗线图示. 这条曲线就是一种叫正态密度的曲线. 在第二章“随机变量、分布与概型”，我们会看到，这条曲线是随机变量的概率密度函数，它是一种叫做正态分布的密度函数. 随机变量的引进使得我们能借助数学分析里函数的工具和知识，对一个随机现象的概率规律有一个总体性的描述. 在第二章，我们还要基于随机变量分门别类地对各种重要的概率规律进行介绍和研究. 同时把我们的视野扩大到多维的情形；引入随机变量间独立性的概念. 第三章“数字特征”则从分布里提炼出一些重要的反映概率规律特征的数量，并研究它们. 例如，刻画小球落点的按概率加权的平均位置（数学期望），以及离开这平均位置的一种波动情况（方差）等等. 第四章“极限定理”研究随机变量和的极限问题，给出“频率稳定性”和上面提到的直方图的极限为正态密度曲线的依据. 这些是概率论基础的主要内容.

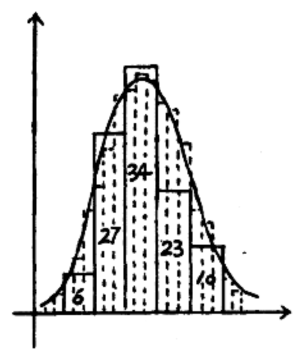

图 1.2　直方图极限

一、二章为初步，三、四章是研究的深入. 重点是第二章. 以上便是概率论基础的主要内容的一个速写. 此后三章是数理统计，介绍利用抽样得到的数据，基于概率论基础和抽样分布，作统计分析和推断：估计分布类型、参数以及对它们的检验. 以上是作为每周 3 学时一学期的随机数学基础课程的全部内容. 第八至十一章以生态环境、经济管理、系统控制、人文统计及信息科学的一两个实际问题，介绍概率统计的应用中最活跃的一些方法和模型，例如 Monte-Carlo 方法、方差与回归分析. ARMA 模型、简单的多元统计及在信息学科广泛应用的 HMM（隐马尔可夫模型）等，既可供不同专业学生从方法及模型上选用，也可作自学以扩大眼界.

§1.2 事件与概率

一、事件

研究随机现象，我们当然要考察随机现象里出现的事件. 我们先用随机试验的概念来引入事件的概念，然后用集合论知识给出事件的严谨定义（定义 2.1），初学者可先略去这一严谨定义，而理解为随机试验德结果..

Galton 钉板实验是一个随机试验. 抛一枚硬币看它落地时是否正面朝上，在一批产品中随机抽取十个记录到的正品的个数，考察某厂流水线上电视机的寿命，都是随机试验. 一般地，一个试验，如果在一定条件下可重复，试验的结果不只一个，并且每次试验时，我们不能肯定是哪一个结果出现，这样的试验称为随机试验. 随机试验里最基本的不能再分解的结果叫基本结果. 基本结果也叫基本事件. 由若干基本结果组成的，我们称之为复合事件. 基本事件和复合事件，泛称事件. 特别地，由所有基本结果组成的，我们称之为必然事件，它的反面，也认为是一个事件，就是不可能事件. 称所有事件的全体为事件体. 必然事件、不可能事件及事件体分别专记为Ω、ϕ及$\mathcal{F}$.

例 2.1 在有两排钉子的 Galton 钉板实验中基本结果只有三个：小球落入第 1 格、第 2 格及第 3 格. 它们也是基本事件. 但像“小球落入前两格”、“小球落入第奇数格”就是复合事件. 特别地，事件“小球落入第 1 至 3 格”是必然事件. 它的反面，“小球不落入第 1 至 3 格”就是不可能事件.

如果用 $\{\omega_i\}$ 表示事件“小球落入第 i 格”，i =1, 2, 3. 那么上例中必然事件就是

$\Omega=\{\omega_1,\omega_2,\omega_3\}$，而基本事件为 $\{\omega_1\}$，$\{\omega_2\}$，$\{\omega_3\}$．“小球落入前两格”这一事件现在可写为$\{\omega_1,\omega_2\}$，或写为$\Omega-\{\omega_1\}$，此时事件体为

$$\mathcal{F}:=\{\Omega,\ \phi,\ \{\omega_1\},\ \{\omega_2\},\ \{\omega_3\},\{\omega_1,\omega_2\},\ \{\omega_1,\omega_3\},\ \{\omega_2,\omega_3\}\} \tag{2.1}$$

我们看到Ω是一个非空点集，事件是Ω的一个子集，事件体是由Ω子集组成，是集合的集合．上例中Ω是一个有限的点集，事件体可以全部列出来．而在考察电视机寿命时，Ω就是一个无限的点集了，它常是一个实数区间．事件和事件体$\mathcal{F}$如何表示呢？能不能仍然可借助集合论的概念来刻画？

现在我们就来利用抽象的集合论的概念，严格定义一般的事件和事件体$\mathcal{F}$．然后说明现实中的、随机试验中的事件和事件体$\mathcal{F}$都可以用它们来解释．

定义 2.1 设$\mathcal{F}$是一个抽象的非空点集Ω的一些子集组成的集合，满足

F_1） $\Omega\in\mathcal{F}$.

F_2） 如$A\in\mathcal{F}$，则 $\overline{A}:=\Omega-A\in\mathcal{F}$.

F_3） 如$A_i\in\mathcal{F}$，$i=1,\ 2,\ \ldots$，则 $\bigcup_{i=1}^{\infty}A_i\in\mathcal{F}$.

则称$\mathcal{F}$为事件体．称$\mathcal{F}$中的每一元素（点）为事件，Ω为必然事件，事件$\overline{A}$为A的逆事件．空集合ϕ也为事件，称为不可能事件． □

今后我们一般用大写英文字母表示事件．请注意：事件是样本空间的子集而是事件体的元素（点），因此对任一事件A，有$A\subset\Omega$，而$A\in\mathcal{F}$．由下面关于事件体性质的定理，说明如上定义的事件体$\mathcal{F}$，对有限次和可列（无穷）多次（这两种情况常合称为“至多可列次”）的集合的并、交及求余运算是封闭的．

定理 2.1 1） $\phi\in\mathcal{F}$.

2）如$A_i\in\mathcal{F}$，$i=1,\ 2,\ \ldots,\ n$，则 $\bigcup_{i=1}^{n}A_i\in\mathcal{F}$.

3）如至多可列个$A_i\in\mathcal{F}$，$i=1,\ 2,\ \ldots,(n)$，则至多可列次的交集 $\bigcap_i A_i\in\mathcal{F}$.

4）如$A,\ B\in\mathcal{F}$，则$A-B:=A\cap\overline{B}\in\mathcal{F}$

证明 由F_1）及F_2）知 1)真．令 $A_{n+1}=A_{n+2}=\cdots=\phi$，则由 F_2）可推出 2)．由F_2）知$\overline{A_i}\in\mathcal{F}$，由可列并的封闭性$F_3$)、有限多次并的封闭性 2)以及集合运算的对偶原理，

$$\overline{\left(\bigcup_i \overline{A_i}\right)}=\bigcap_i A_i\in\mathcal{F}.\quad 由F_2）知\quad \bigcup_i \overline{A_i}\in\mathcal{F}$$

从而得到 3)．由F_2)及 3)，立得 4)． □

下面带有*号的注，初学者可先滑过．

注 1*：在集合论中，由集合组成的集合，叫类，满足条件F_1)至F_3)的类叫做σ-代数．因此事件体$\mathcal{F}$是Ω上的一个σ-代数．当$\Omega=R_1:=(-\infty,\infty)$时，如果定义一个类$\mathcal{L}=\{(-\infty,x):x\in R_1\}$，并将$\mathcal{L}$中所有元（它是一个半无限开区间）经过至多可列次并、交、求余所得到的全部集合记为新的类$\mathcal{B}$．容易验证它是一个σ-代数，也说它是由$\mathcal{L}$产生的σ-代

数. 我们特别称为 Borel（波雷尔）集类. 如果用$\mathcal{O}$表 R_1 中所有开区间全体，$\mathcal{G}$表 R_1 中所有闭区间全体，则可以证明，$\mathcal{O}$或$\mathcal{G}$产生的σ-代数也是$\mathcal{B}$.

利用上面定义，现实里随机试验中的事件可以抽象为某个点集. 例如例 1 中取 $\Omega=\{\omega_1,\omega_2,\omega_3\}$，单点集$\{\omega_i\}$表示基本事件："小球落入第 i 格"，i=1,2,3. Ω的子集$\{\omega_1,\omega_2\}$表示事件"小球落入前两格"或"不落在第三格". 取事件体$\mathcal{F}$为(2.1)，就包容了这个试验中的所有事件. 而在考察电视机寿命时，可取$\Omega=[0,\ \infty)$，事件"寿命超过 400 小时而不超过 900 小时"如记为 A，则 $A=(400, 900]$. 而事件 $B=$"寿命超过 1000 小时"则可写 $B=(1000,\ \infty)$，等等，这里取小时为单位. 一般地，它是一个实数区间. 事件体$\mathcal{F}$对至多可列次的集合运算并、交及求余都是封闭的，因此事件体$\mathcal{F}$应该包括所有的非负的实数区间，以及包括由这些实数区间作至多可列次的集合运算（并、交及求余）所得到的集合. 并且事件体$\mathcal{F}$也就是这些集合就够了，就足以刻画所有的事件了. 这样，定义里给出的事件体，就确实可以看成现实中一个随机试验的所有可能的结果，也即是所有的事件的全体.

现在我们用集合间的关系和运算来刻画现实中事件间的关系和运算. A、B 集合求交的运算常略$\cap$不写，即 $AB:=A\cap B$. 在例 2.1 中，如果事件"小球落入前两格"记为 A，"小球落入第偶数格"记为 B. 那么在钉板入口处让一个小球落下，假如落入第 1 格，那么我们就可以说事件 A 出现了. 当然也可说事件 B 未出现，用集合论中的表示法分别记为 $\omega_1\in A$ 和 $\omega_1\notin B$. 当然也有 $\omega_1\in A\overline{B}=A-B$. 而如果落下的小球进入第 2 格，则 $\omega_2\in A\cap B=AB$，也即此时事件 A 和 B 同时发生了. 这样我们可以在集合间的关系与运算和事件间的关系与运算之间建立对应，列为下表 1.1.

表 1.1

集合的关系和运算	事件的关系和运算
$\omega\in A$	事件 A 发生
$A\subset B$	事件 A 发生则事件 B 必发生
$A\cup B$	事件 A 与事件 B 至少有一个发生
$\bigcup_j A_j$	事件 A_i 中至少有一个发生
$A\cap B$ 或 AB	事件 A 与事件 B 同时发生
$\bigcap_j A_j$	所有事件 A_i 都同时发生
$A\setminus B$ 或 $A-B$	事件 A 发生而事件 B 不发生

事件的关系和运算可用图表示如图 2.1，这种图叫 Ven 图. 常称 $A\cup B$ 为 A 与 B 的和

事件，而称 AB 为积事件. 如果 $AB=\phi$ 则称事件 A 与 B 互斥，或不相容，有时也说不相交. 如 $A_iA_j=\phi$，$\forall i\neq j$ 则说诸事件 A_i 两两不交，此时将 $\bigcup_i A_i$ 专记为 $\sum_i A_i$.

至此，我们已经完成利用集合论，给出概率论中事件的严谨定义，并且利用集合论中集合间的关系和运算未刻画概率论中事件间的关系和运算.

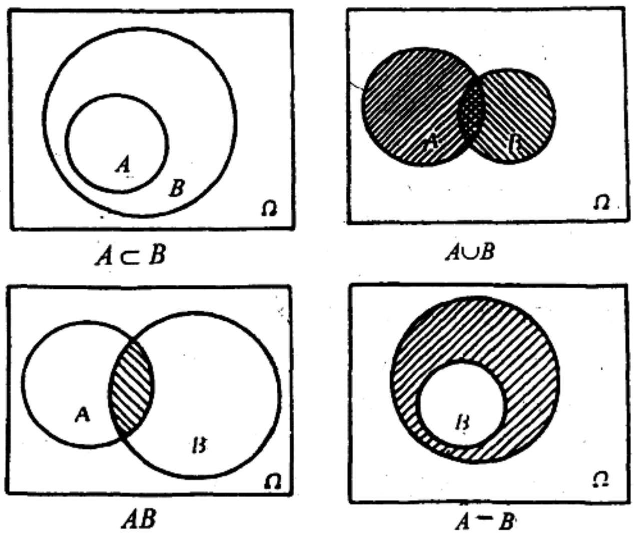

图 2.1　事件的关系和运算

二、概率

我们常说“这事有百分之百把握”、“那事有七成把握”等等，都是用 0 到 1 之间的一个实数来表示事件发生的可能性的大小的. 因此事件的概率值可以看成以事件（用集合论的话说，是集合）为自变量的一个函数值，它们在[0,1]之中. 严格的定义如下.

定义 2.2　设 P 是在事件体 $\mathcal{F}$ 上定义的实值集函数，满足

P_1）非负性　$P(A)\geq 0$，$\forall A\in\mathcal{F}$

P_2）规范性　$P(\Omega)=1$

P_3）可列可加性　设 $A_i\in\mathcal{F}$，$i=1, 2, \ldots$，且两两不交即 $A_iA_j=f$，$i\neq j$. 则

$$P\left(\sum_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i).$$

则称 P 为定义在事件体 $\Im$ 上的概率测度，简称概率. 称 $P(A)$ 是事件 A 的概率. □

如 $P(A)=0$，称 A 为几乎不可能事件；如 $P(A)=1$，称 A 为几乎必然事件. 由于我们关心

的是概率，因此今后对几乎必然事件与Ω及几乎不可能事件与ϕ不作区分.

注 2*：我们来看熟知的“长度”的概念．“长度”是非负的集函数.

令 $\Omega=[0,\ 1)$，$A_1=\left[0,\ \frac{1}{2}\right)$，对$n>1$ $A_n=\left[\frac{1}{2^{n-1}},\ \frac{1}{2^n}\right)$.

易知$L(A_n)=\frac{1}{2^n}$，注意诸A_n不交，且$\sum_{n=1}^{\infty}A_n=[0,\ 1)$. 又

$$\sum_{n=1}^{\infty}L(A_n)=\sum_{n=1}^{\infty}\frac{1}{2^n}=1=L[0,\ 1),$$

因此$L\left(\sum_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}L(A_n)$，这就是可列可加性. 可见作为一个量度的尺度的概念，应该有可列可加性. 更一般地，“测度”定义为非负的有可列可加性的R_1中的集函数. 从而$\Omega=[0,1]$上的长度(一点的长度为 0，因此[0,1)和[0,1]在讨论长度时不作区别)，也就可看成规范化（$L(\Omega)=1$）的测度，它也可视为概率测度.

我们知道概率应该是刻画事件发生的可能性大小的数量指标. 现在这样定义的概率，确实能够担当起这个角色么？比如说，不可能事件的概率应该为 0，又比如说，事件 B 如果包容了事件 A，即事件 A 发生必然导致事件 B 发生，那么 $P(A)$ 应该不大于 $P(B)$，等等.. 我们可以证明，这样定义的概率确实能够保证这些事实仍然是正确的，作为刻画事件发生的可能性大小的数量指标的概率，所有应有的结论都得到了保证.

概率的性质：

定理 2.2（概率的性质） 设 P是事件体$\mathcal{F}$上的概率，则

1）$P(\phi)=0$;

2）有限可加性：设$A_i\in\mathcal{F}$, $i=1,\ 2,\ \dots,\ n$且两两不交，则$P\left(\sum_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)$.

3）设$A\in\Im$ 则$P(\overline{A})=1-P(A)$.

4）单调性：如果$A\subset B$则$P(A)\le P(B)$.

5）连续性：设$A_i(\in\Im)$单调，即$A_i\subset A_{i+1}$或$A_i\supset A_{i+1}$，$i=1,\ 2,\ \dots$，此时分别定义$\lim_{n\to\infty}A_n=\bigcup_{n=1}^{\infty}A_n$ 或 $\lim_{n\to\infty}A_n=\bigcap_{n=1}^{\infty}A_n$，则

$$P(\lim_{n\to\infty}A_n)=\lim_{n\to\infty}P(A_n).$$

证明 1）由可列可加性及规范性

$$1=P(\Omega)=P(\Omega+\phi+\phi+\cdots)=P(\Omega)+P(\phi)+P(\phi)+\cdots=1+P(\phi)+P(\phi)+\cdots$$

故$P(\phi)+P(\phi)+\cdots=0$，注意$P(\phi)\ge 0$，这样只能有$P(\phi)=0$;

2)仿定理 2.1(2)的证明，令 $A_{n+1}=A_{n+2}=\cdots=\phi$，则由可列可加性$P_2$)及刚证得的$P(\phi)=0$可推出 2).

3）设$A\in\Im$. 由$A+\overline{A}=\Omega$、有限可加性及规范性P_2)，易得 $P(\overline{A})=1-P(A)$.

4）由$A\subset B$知可写$B=A+B\overline{A}$，则由有限可加性及$P(B\overline{A})\ge 0$得 $P(A)\le P(B)$.

5）仅对单调非降列证明 P 的连续性，单调非升时类似.

设 $A_i \subset A_{i+1}$，$i=1, 2, ,\ldots$，此时定义

$$B_1 = A_1,\ B_2 = A_2 - A_1,\ B_{i+1} = A_{i+1} - A_i,\ i=1,\ 2,\ ,\ldots\ ,$$

则诸 B_i 两两不交，且 $\bigcup_{i=1}^{n} A_i = A_n = \sum_{i=1}^{n} B_i$．从而 $\bigcup_{n=1}^{\infty} A_n = \lim_{x\to\infty} A_n = \sum_{i=1}^{\infty} B_i$．

由可列可加性和有限可加性

$$P(\lim_{n\to\infty} A_n) = P(\sum_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i)$$
$$= \lim_{n\to\infty} P(\sum_{i=1}^{n} B_i) = \lim_{n\to\infty} P(A_n). \qquad \square$$

可见定义 2.1 给出的概率的简洁定义，确实能够保证它作为刻画事件发生可能性大小的数量指标．至此，对我们所观测的对象 Ω，定义了建基于集合论的事件体 $\mathcal{F}$ 和建基于测渡论的概率（测度）P．我们称三元体 $(\Omega,\ \mathcal{F},\ P)$ 为概率空间，简记为 *ps.*

对 n 个不相交事件的和事件，其概率计算有如下一般加法公式．

定理 2.3（一般加法公式） 设 $A_i \in \Im$，$i=1, 2, ,\ldots,n$，则

$$P(\bigcup_{i=1}^{n} A_i) = s_1 - s_2 + s_3 + \cdots + (-1)^{n+1} s_n \tag{2.1}$$

其中 $s_1 = \sum_{i=1}^{n} P(A_i)$，$s_2 = \sum_{1\le i<j\le n} P(A_iA_j)$，$s_3 = \sum_{1\le i<j<k\le n} P(A_iA_jA_k)$，$\ldots$，

$s_n = P(A_1A_2\cdots A_n)$．

我们将特别地对 $n=2$ 时证明这个定理，在 $n=3$ 时图解这个公式，而在一般的情形，不去详细写出定理的证明，只在这里指出，将相交的和化为不相交的和，从而利用归纳法和有限可加性可以证得．

$n=2$ 时 $\quad P(A_1 \cup A_2) = P(A_1 + A_2\overline{A}_1) = P(A_1) + P(A_2\overline{A}_1)$， (2.2)

另一方面，由 $P(A_2) = P(A_2A_1) + P(A_2\overline{A}_1)$ 知 $P(A_2\overline{A}_1) = P(A_2) - P(A_1A_2)$．

代入(2.2),

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1A_2) \tag{2.3}$$

此即 $n=2$ 时的 (2.1).

$n=3$ 时 (2.1)变为

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3) \end{aligned} \tag{2.4}$$

从图 2.2 可以直观地得到上式的证明. 事实上, 如用 p_k 表示图 2.2 中对应的彼此不相交的第 k 个事件的概率，容易看到

$P(A_1 \cup A_2 \cup A_3) = \sum_{i=1}^{7} p_i$；而(2.4)式右方的计算列表如下:

		p_1	p_2	p_3	p_4	p_5	p_6	p_7
$+s_1$	$+P(A_1)=$	+			+		+	+
	$+P(A_2)=$		+		+	+		+
	$+P(A_3)=$			+		+	+	+
$-s_2$	$-P(A_1A_2)=$				−			−
	$-P(A_1A_3)=$						−	−
	$-P(A_2A_3)=$					−		−
$+s_3$	$+P(A_1A_2A_3)=$							+

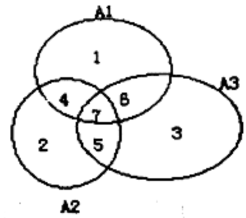

图 2.2　图解加法公式

表中“＋”表示加上在第一行上此列对应的概率值，而“－”表示减去这个概率值. 容易确认(2.4)式成立.

例 2.2 设 A、$B\in\mathcal{F}$，$P(A)=0.5$，$P(B)=0.7$，求 $P(A\cup B)$ 最大值与最小值.

解 由(2.3)知 $P(A\cup B)$ 最大值与最小值应该分别是 $P(AB)$ 取最小值与最大值的情况，也即前者为 $A\subset B$ 和 $A\supset B$ 后者为 AB 最大的情况，因此 $P(A\cup B)$ 最大值与最小值分别是 1 和 0.2. 注意 0.5+0.7>1，此时 $AB\neq\phi$.

§1.3　古典概型

本节介绍我们常常碰到、也最为简单的一类随机现象的概率模型，其间概率的计算用到初等数学中的排列和组合.

定义 3.1　基本事件个数有限且等可能的概率模型，称为古典概型.

例如 Galton 钉板试验中小球每一次碰钉后有向左和向右落下两种可能的结果，由于钉板的设计和假定小球均匀及钉子足够光滑，这两种可能结果的概率各为 1/2; 在红色、黑色、白色三种小球数量相等的袋子中任取一球，取出的球的颜色构成三个等可能的基本事件；闰年里一个人的生日有 366 个等可能的基本事件（人为因素除外）. 所有这些概率模型，都是古典概型. 但是请注意，§1.1 中有四排钉子的 Galton 钉板试验，小球落向下方哪个格子，结果虽有限，但非等可能，因此不是古典概型. 它是连续四次碰钉后的结果，或者说是连续四次古典概型试验的结果.

古典概型中可记空间 $\Omega=\{\omega_1,\omega_2,\ldots,\omega_N\}$，每一个 ω_i 为一个样本点或基本场合，基本

事件的概率 $P(\omega_i):=P(\{\omega_i\})=1/N$, i=1,2,...,N, 而$\mathfrak{I}$由Ω的一切子集所组成. 故如事件 A 中由n_A个样本点，则 $P(A)=n_A/N$. 有时也记 N 为含义更清楚的n_Ω. 这样

$$P(A)=n_A/n_\Omega \tag{3.1}$$

例 3.1 包括 a 和 b 两人在内共 n 人排队，问 a, b 间恰有 r 人的概率 p.

解 事件"a, b 间恰有 r 人"记为 A. 所做的试验是让 n 人排队，所有可能的排队结果是有限并且等可能的. 人与人间是有区别的，因此 $n_\Omega=n!$. 下面来求n_A.

如果 a 排在 b 的左边且占在第 1 位，则符合 A 要求的排队结果，一定是 b 占在第 r+2 个位置上. 而其余 n-2 个人没有限制，不论是在 a, b 之间还是之外，只要 $0\le r\le n\text{-}1$. 此时应有 $(n\text{-}2)!$ 种可能. a 可以退到第 2 位、第 3 位、直到第 $n\text{-}r\text{-}1$ 位（这时 b 占最右边了）. 考虑到 b 也可在 a 的左边，因此当 $0\le r\le n\text{-}1$ 时，$n_A=2(n\text{-}r\text{-}1)(n\text{-}2)!$. 故

当 $0\le r\le n\text{-}1$ 时， $p=2(n\text{-}r\text{-}1)(n\text{-}2)!/n!=2(n\text{-}r\text{-}1)/n(n\text{-}1)$ ，否则 $p=0$. □

下面一个例子有典型意义，它还引出两类重要的概率分布规律.

例 3.2 设有 a 件正品 b 件次品，从中按有放回和无放回两种方式逐一随机抽 n 次，求恰抽出 k 件正品（记此事件为 A）的概率 p_k.

解 [有放回] 由于每次抽完后放回去，所以每一次抽取都是在 a+b 个产品中任意抽取, 并且这 a+b 个产品都是等可能地被抽到，这样，任意两次抽取应有$(a+b)^2$种等可能的结果，因此有放回抽取 n 次时，空间的样本点个数 $n_\Omega=(a+b)^n$. 为求n_A，先假定前 k 次都抽到正品，那么后 n-k 件就只能抽取到次品了. 仿n_Ω的计算，k 件正品的抽取应有a^k种等可能的情况，而 n-k 件次品的抽取应有b^{n-k}种等可能的情况，由于要抽 n 次，从而符合事件 A 要求的样本点个数应该为$a^k b^{n-k}$. 由于 n 次抽取中究竟哪 k 次抽取到正品（另外 n-k 次应该抽出次品）是没有限制的，因此$n_A=C_n^k a^k b^{n-k}$. 于是

$$p_k=n_A/n_\Omega=C_n^k a^k b^{n-k}/(a+b)^n,\quad k=1,\ 2,\ \ldots,\ n. \tag{3.2}$$

现在我们将(3.2)改写为如下形式

$$p_k=C_n^k\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^{n-k},$$

并令$p=a/(a+b)$, $q=1-p=b/(a+b)$, 则

$$p_k=C_n^k p^k q^{n-k},\quad k=1,\ 2,\ \ldots,\ n. \tag{3.3}$$

容易看到，每一次抽取也都是一个古典概型，p 实际是一次抽取中抽到正品的概率，q 则是抽到次品的概率，因此(3.3)式的概率意义就十分明显了. 因为(3.3)右方是$(p+q)^n$二项展开式中有p^k的项，$\sum_{k=0}^n C_n^k p^k q^{n-k}=(p+q)^n=1$，因此这一类有放回抽取的概率模型，叫做**二项概型**. 由(3.3)决定的数列$\{p_k\}$叫**二项分布**.

注意，每一次抽取都是一个古典概型，空间的元数（元素个数）为 a+b. 现在的最终

试验是抽取 n 次，实际上是个复合随机试验，它也是一个古典概型，空间的元数为 $(a+b)^n$.

[不放回] 由于抽取不放回，此时每一次抽取虽然仍是一个古典概型，但每次抽取时产品数已经比上次少一个了. 既然每次抽取不放回，因此逐一抽取 n 次也可以看成是从 $a+b$ 个产品中一次抽走了 n 个，因此 $n_\Omega = C_{a+b}^n$. k 件正品取自总的 a 个正品，可能的取法有 C_a^k 种，同理 $n\text{-}k$ 件次品的取法有 C_b^{n-k} 种，从而符合 A 的抽取的应有 $C_a^k\ C_b^{n-k}$ 种可能，故

$$p_k = C_a^k\ C_b^{n-k} / C_{a+b}^n\ , \quad k = 1,\ 2,\ \dots,\ n. \tag{3.4}$$

此时还应有 $k \le a$， $n-k \le b$. 由(3.4)决定的这一类概率模型，叫做超几何概型. □

逐一不放回抽样，在第二次抽取时，可以抽取的产品数少了一个. 但是当产品数很大，抽样数量远小于产品数，抽出的正品数也远小于全部产品中的正品数时，第二次抽取比第一次抽取只是少了一个产品，因此和第一次抽取几乎没有什么不同，从而不放回抽样可以用有放回抽样近似，按二项分布(3.3)计算. 这样就可少算两个组合数，并且有专门编制的二项分布表可查. 在工厂企业及社会经济问题调查中进行的抽样几乎都是这种情况.

上述事实是说，当 $a+b$ 很大且 $k \ll a$ 、$n-k \ll b$ 时，(3.4)右方与(3.3)右方近似，超几何概型问题化为二项概型问题,. 现在从数学上我们证明*:

当 $N := a+b \to \infty$ 时，设 $\lim\limits_{n\to\infty} \dfrac{a}{N} = p > 0$， $\lim\limits_{N\to\infty} \dfrac{b}{N} = q > 0$ 则对任一正整数 n, 有

$$\lim_{n\to\infty} \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, \cdots, n \tag{3.5}$$

实际上,

$$\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{N}{n}} = \binom{n}{k}\left(\frac{a}{N}\right)^k \left(\frac{b}{N}\right)^{n-k} \times$$

$$\times \frac{\left(1-\frac{1}{a}\right)\left(1-\frac{2}{a}\right)\cdots\left(1-\frac{k-1}{a}\right)\left(1-\frac{1}{b}\right)\cdots\left(1-\frac{n-k-1}{b}\right)}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}.$$

由条件知: 当 $N \to \infty$ 时，$a \to \infty$；故于上式中令 $N \to \infty$ 即得证(3.5). □

例 3.3 一袋装有 r 个红球、b 个黑球，除去颜色外不可辨别. 今随机逐一取球，不放回，求第 k 次取出红球的概率 p_n.

解 I 球不可辨，将 $r+b$ 个球随机逐一全数取出，依次一线排列，占 $r+b$ 个位置，共有 C_{r+b}^r 种可能. 第 k 个位置固定为红球，则只有 C_{r+b-1}^{r-1} 种变化，故

$$p_k = C_{r+b-1}^{r-1} / C_{r+b}^r = r/(r+b).$$

解 II 将原本不可辨的球编号从而变成可辨的，按排列处理同样可解此问题. 仍然将 $r+b$ 个球随机逐一全数取出一线排列，共有（$r+b$）！种可能. 第 k 个位置固定为一红球，有 r 种

可能，而其余（$r+b-1$）个位置却有（$r+b-1$）！种变化，故也有

$$p_k = r(r+b-1)!/(r+b)! = r/(r+b). \qquad \square$$

我们看到，不可辨的组合问题可以化为可辨的排列处理，“条条大路通罗马”. 不过一般情况下，建议还是按原来的题设处理为好：不可辨的用组合而可辨时用排列处理. 因为改变后的处理，稍有不慎，便容易出错. 不可辨的组合问题可以化为可辨的排列处理,同样可解此问题. 仍然将 $r+b$ 结果是一致的.

下一例子在统计物理中起重要作用，它也告诉我们不同的假定，使得等可能的基本事件个数不同，也即空间的点数不同. 因此我们着手解决问题之前，要弄清 Ω 是什么，哪些是基本事件，他们是否为等可能的.

例 3.4 设有 n 个质点落入 N（$>n$）个合子. 事件 A：指定某 n 个合各含 1 点. 求 $P(A)$.

解 I (Maxwell-Boltzmann) 设质点可辨，合容不限，即每个合子可容纳的质点数无限制.

因为每个质点可落入 N 个盒子中的任意一个，故 n 个质点落入合子的所有可能情况有 N^n 种， 即 $n_\Omega = N^n$. 并且每一种可能情况都认为是等可能的出现. 指定的 n 个合子中各含 1 个质点，质点可辨别，因此 $n_A = n!$，从而 $p_1 := P(A) = n!/N^n$. 这适于空间

II (Fermi-Dirac) 设质点不可辨，每个合子至多可容纳一个质点. 容易算得 $p_2 := P(A) = 1/C_N^n$. 这适用于空间基本粒子费米子.

III (Bose-Einstein)设质点不可辨，合容不限. 我们用*占位法*来解.

题设试验的一个结果可如图 3.1(*a*)所示. 这里合子依次连续排开，竖线表示合壁，O 表质点. 图中示 2 只质点落向 3 个合子的一种情况. 现在将合子的每个内壁和每个质点都平等地看成一个位置，这样一共有 $N\text{-}1+n$＝3-1+2＝4 个位置. 如图 3.1(*b*)所示. 如果从这 $N\text{-}1+n$=4 个位置（记以×）任选 n=2 个解释为质点，而其余解释为合子的内壁，例如图 3.1(c)选前两个为质点，则对这一选法现在的解释是第一个合子中有两个质点，第二合空，第三合空，这正好描述 n 个质点落入 N 个合子的一种情况(*a*). 这样质点的每一个落法，对应从 $N\text{-}1+n$ 个位置选 n 的任意一个选法；反之，从 $N\text{-}1+n$ 个位置任选 n 的一个选法对应质点的一个落法. 因此，$n_\Omega = C_{N+n-1}^n$. 假如现在的每一个选法都是等可能的（物理学中基本粒子玻色子，就是这情况），那么所求的概率 $p_3 := P(A) = 1/C_{N+n-1}^n$. 本例适用物理学中的基本粒子玻色子. $\square$

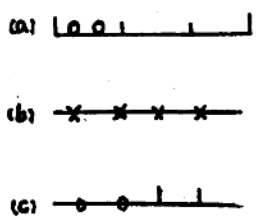

图 3.1　占位法图示

注 上例若以 2 个质点 a 和 b 落向 3 个合子为例，解 1 的空间元数为 3^2＝9，而解 3 因不可辨，图 3.1 中 a 与 b 相同，空间元数只有 $C_{3+2-1}^2 = C_4^2 = 4\times3/2 = 6$ 个，它们是不同的. 但一般说这 6 种情况不是等可能的.

1.(ab, –, –)
2.(–, ab, –)
3.(–, –, ab)
4.(a, b, –)
5.(b, a, –)
6.(a, –, b)
7.(b, –, a)
8.(–, a, b)
9.(–, b, a)

再看一个例子，著名的 Bertrand(贝特朗)奇论，说明对空

间等可能性的假定是至关重要的. 这是一个几何概率问题. 几何概率在现代概率概念的发展中曾经起过重大作用. 十九世纪时，不少人相信，只要找到适当的等可能性描述，就可以给概率问题以唯一的解答. 然而有人却构造出这样的例子，它包含着几种似乎都同样有理但却相互矛盾的答案，下面就是一个著名的例子.

例 3.5 (Bertrand 奇论) 在半径为 1 的圆内随机地取一条弦，求其长超过该圆内接等边三角形的边长 $\sqrt{3}$ 的概率 p.

这个几何概率问题, 在基于对术语 "随机地"含义的不同解释，却存在多种不同答案.

解 I 任何弦交圆周二点，不失一般性，先固定其中一点于圆周上，以此点为顶点做一等边三角形. 显然只有落入此三角形内的弦才满足要求，这种弦的另一端跑过的弧长为整个圆周的$\frac{1}{3}$，故所求概率等于$\frac{1}{3}$（见图 3.2(a)）.

解 II 弦长只跟它与圆心的距离有关，而与方向无关，因此可以假定它垂直于某一直径，而且当且仅当它与园心的距离小于$\frac{1}{2}$时, 其长才大于$\sqrt{3}$, 因此所求的概率为$\frac{1}{2}$.（见图 3.2(b)）

解 III 弦被其中点唯一确定,. 当且仅当其中点属于半径为$\frac{1}{2}$的同心圆内时, 弦长大于 $\sqrt{3}$. 此小圆面积为大圆面积的$\frac{1}{4}$, 故所求概率等于$\frac{1}{4}$.(见图 3.2(c)) □

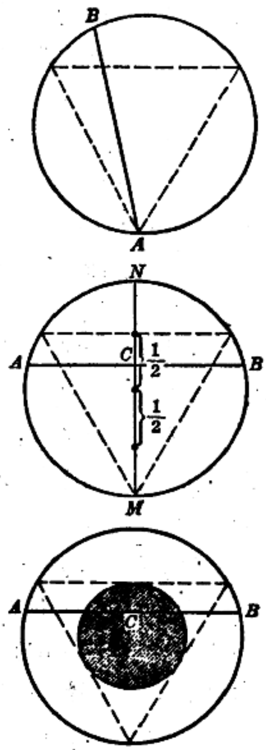

图 3.2 "Bertrand 奇论"

古典概型的例子和解法很多、有些也很精彩. 例如随机取数问题、递推法等等，有兴趣的读者可看[1]或[2].

§1.4 几何概型

一、引例与定义