林业应用数理统计

LIN YE YING YONG SHU LI TONG JI

陈华豪 丁恩统 蔡贤如
洪 伟 张忠义 编著

大连海运学院出版社

林业应用数理统计

陈华豪　丁思统　蔡贤如
洪　伟　　　　　张忠义　编著

大连海运学院出版社

内 容 提 要

本书是数理统计应用于林业科学的专著，全书共分八章。第一章概率论基础，介绍了概率论的基本概念、基本理论、随机变量的重要分布及其数字特征。第一章的内容也是全书的理论基础。第二至八章分别介绍了参数估计、统计假设检验、方差分析、回归分析、数量分类方法、试验设计、抽样技术等常用的数理统计方法。

本书中所介绍的每一种数理统计方法，都给出了结合林业实际应用的具体例子；且说理浅显，结合林业专业实际应用，通俗易懂；每一章后面均配有适量的习题；书后附有28个概率、数理统计常用表。本书可作为从事农业、林业工作的科技人员用书，也可供高校师生参考。

林业应用数理统计

陈华豪　丁思统　蔡贤如
洪　伟　　　　　张忠义　编著

大连海运学院出版社出版（大连凌水桥）
大连海运学院出版社发行
沈阳农业大学印刷厂印制

开本 787×1092 1/16　印张 17　字数 614,800
1988年8月第一版　1988年8月第一次印刷

责任编辑　杜祖缔　　封面设计　曹太文

印数001—7150　　定价6.60元
ISBN 7—5632—0032—0/D·2

前　言

定量化是生物学科、也是林业科学发展现代化的重要标志之一。林业科学发展至今日，只有定性的结论已不能满足实践的需要。

数理统计是使林业科学结论定量化的重要工具。掌握数理统计方法是今天林业科学工作者应该具备的基本素质。这是因为与林业有关的各种现象中普遍存在着随机现象，林业是大农业的重要组成部分，森林作为最大的陆地生态系统，时空宏博、变化万千、受到许多随机因素的影响，表现为各种各样的随机现象，而数理统计正是从数量方面揭示大量随机现象中存在的必然规律的学科。

在林业工作中，从测树制表到资源清查，从种苗分级到立地质量评价，从林地中生物的分布格局到群落分类排序，从生长模型到突变预测……凡此种种都可以应用数理统计方法，迄今也都积累了各种各样的成功经验。长期以来数理统计一直在各行各业直接而具体地被人们应用着，是一门在实践中应用十分活跃的数学学科分支。

我国林业科学工作者一直就非常重视数理统计这一科学研究方法与分析手段的应用。新中国成立后，老一辈林业数理统计学家谭启栋教授的《森林统计分析》开其端，全国林业教材《数理统计》继其后，近年来中国林业出版社又出版了罗鸣福先生的《林业试验设计》，唐守正先生的《林业多元分析》以及其它林业科学工作者应用数理统计的一些研究成果，都为普及与扩大数理统计方法在林业科学上的应用起了积极的推动作用，然而这一类著作毕竟太少，与我国发展着的林业科学的现状与远景极不适应，有鉴于此，编著者不揣冒昧，在林业部关切与支持下联合编写了这本《林业应用数理统计》希冀在普及数理统计方法在林业中的应用上起一点促进作用。

我们本着尽可能通俗易懂地将林业中最常用的、包括一些比较新的数理统计知识介绍给大家的愿望编写这本书，但真正要融会贯通还有待于实践，我们衷心希望读者们能够将数理统计的各种技术方法与各自的专业知识和研究工作实际有机地结合起来，让数理统计为你所用，开花结果。

本书由东北林业大学陈华豪副教授统稿（并负责第一、八章），参加编著的还有江西农业大学丁思统副教授（负责第五、六章）、沈阳农业大学蔡贤如副教授(负责第二、三章)、福建林学院洪伟副教授（负责第七章）、河南农业大学张忠义讲师（负责第四章）。出版过程中得到大连海运学院出版社、沈阳农业大学基础部以及编著者所在五所院校各级领导的支持，李世达教授百忙中为本书审定，书中插图由马云魁同志绘制，谨此一并致谢。由于编著者水平有限，编写时间又十分匆促，漏失与错误在所难免，欢迎大家不吝指正。

编著者　1988.2.

目　录

第五章 回归分析

第六章 数量分类方法

第七章 试验设计

第八章 抽样技术

附表

第一章 概 率

概率论是数理统计的理论基础，是以随机现象作为基本的研究对象的。

§1.1 事件及其运算

在林业工作中处处可见受随机因素影响的现象，我们称此种现象为随机现象。它们与确定性现象相对，例如，在标准大气压下将水加温到100℃，必定出现沸腾的现象，多次重复不会变化，这是确定性现象；与此不同，随机现象在同一条件下重复试验或观察，结果可以不同。例如，同一地点，同是7月5日中午12时，每年观察一次地面温度，结果不尽相同；又如，在同一株落叶松母树上采来的种子，经同种方式育苗，栽植在同一片造林地上，十年后测定它们的高与径，结果不会一样。这些都是因为受随机因素的影响与干扰，并称为随机现象，所谓随机因素常常是指无法控制的，或者因不能完全认识也就不能切实控制的细微因素。

为方便计，今后将在一定条件下可以重复的观察或试验，统称为随机试验，简称试验。随机现象便是其试验结果不只一个，但在每次试验进行前又无法准确预言其结果究竟是哪一个的现象。不过试验告诉我们，在大量重复试验中，试验结果会表现出确定的统计规律性，这主要是指不同结果出现可能性大小不尽相同，但却十分稳定。例如，视掷一枚质地均匀的硬币为一次试验，试验结果可能有二种，一是正面朝上，一是正面朝下。这个试验可以在相同条件下多次重复，就一次试验而言，结果究竟是哪一个很难预言但在大量复重试验中，这两种结果出现的可能性大体相同。又如，在同一片林地上生长着多株同一品种的同龄红松，抽一株观测其树高视为一次试验，试验结果是得到一个树高值，那末，多次试验观察不同株的树高常常是不同的，但不会过高或过矮，而集中在某一高度范围内；另一品种的红松的林木，同样年龄在相同条件下培育，其高就会在另一范围内集中，这是相对稳定的。而假定某品种红松树高集中在[5,15](米)范围内，那末树高在[5,7)[7,9)[11,13)[13,15]（米）范围内的可能性也不会相同，各有其相对稳定的可能性，这些都是随机现象，而其中稳定的可能性大小就是概率论研究的基本内容。

概率论中将一次随机试验的各种可能的试验结果称为随机事件，简称事件，并常用A、B、C、D……来表示。不同事件在一次试验中出现的可能性大小如果能用数量表示，这个数量指标就是事件的概率。由于随机现象中一次试验无法准确预言发生什么事件，因此，人们特别对不同事件发生的可能性大小——它反映随机现象的本质特征——感兴趣。例如，上述红松树高，若在[5,15]米范围内，而树高在[11,15]内的可能性比在[5,9)内的可能性大或者小，将说明二种不同的林木生长状况。

今后将随机试验中每一可能出现的（基本而不能再分的）事件称为基本事件或样本点，而其全体称为随机试验（对应于一组指定的试验条件）的样本空间，记为Ω。样本空间包括了该随机试验的一切最基本的可能结果，例如，从5粒红松，3粒白桦，2粒杨树种子里随意抽取出一粒播种，则基本而不能再分的事件一共十个，5个是抽中红松，3个是抽中白桦，2个是抽中杨树，这十个基本事件构成了样本空间Ω。

还可以将样本空间Ω的每一基本事件理解为一个点，这些点全体形成的集合就是样本空间。概率研究表明将Ω理解成点集并将集合论的许多结果运用到概率论中来是方便的。

由基本事件构成的Ω的各种子集合一般称为复合事件。例如上面提到的例子中抽中阔叶树种子的事件可以理解为抽中白桦的三个事件与抽中杨树的二个事件形成的5个点构成的子集合，而抽中针叶树种子这种事件则直接就是抽中红松的5个基本事件构成的Ω的子集合。

最常遇见的复合事件有：

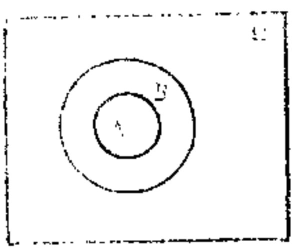
1．$A+B$或$A\cup B$，称为A，B二事件之和，它由事件A、B中至少有一个事件发生构成。集合论中，即用A、B二个点或一般地A、B二个集合的并集代表。刚刚提到的事件可以用$A_{桦}+A_{杨}$或者$A_{桦}\cup A_{杨}$表示，而如$A_{桦}$就是由三个抽中桦树种子的基本事件构成的。

2．$A\cdot B$或AB或$A\cap B$，称为A、B二事件之积。它由二事件同时发生所构成。用集合论的术语即A，B二个集合的交集。例如前述树高在[5,15]米内取值的红松，于其中随机抽取一株，观察其树高，树高在[5,9)内视为A事件，在[7,11)内视为B事件，则A、B二事件之积是指观察某一株树，而树高在[7,9)内取值。而A、B二事件之和即视为树高在[5,11)内取值的事件。

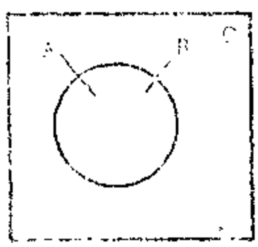
3．$\overline{A}$，称为A事件的对立事件，它表示A不发生这样的事件。例如，树高在[5,9]内取值视为A事件，则树高不在[5,9]内，或即树高在(9,15]，内取值为$\overline{A}$事件。

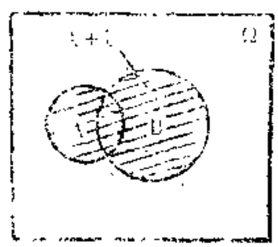
由于树高只能在[5,15]内取值，树高在[5,15]内取值的事件可以理解为在整个样本空间集合Ω上取值的事件，这是在该组试验条件下必然发生的事件，今后把事件Ω称为必然事件。根据和事件的定义，有$A+\overline{A}=\Omega$。另外如果把树高在[5,9]内取值视为事件A，把树高在[11,15]内取值视为C事件，则$A\cdot C$事件是不可能发生的。不可能发生的事件今后用Φ表示，并称$A\cdot C=\Phi$时的A，C二个事件为互不相容事件，或互斥事件。显然事件A与其对立事件$\overline{A}$间是互不相容的。

事件之间这些复合关系可以理解为在对事件进行运算，它们和集合的运算关系是一一对应的。这正是我们把Ω称为样本空间；把事件比为空间中点的集合的理由，此时基本事件便是不能再分的最基本的集合构成单位，即点。凭籍于这种理解，事件之间的运算关系可以直观地用图形表达如下：

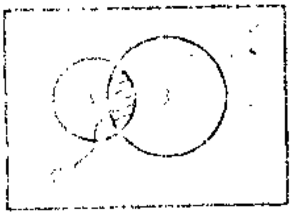


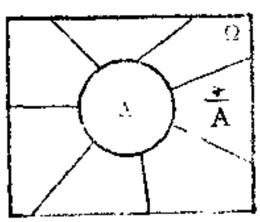
$A\subseteq B$



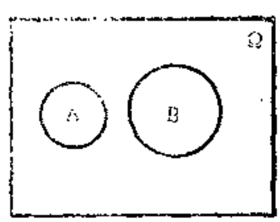


[illegible]


[illegible]



$\overline{A}$



A、B互斥

对图1.1 中事件的包含关系$A\subseteq B$，相等关系$A=B$，读者可自己按集合论的思考作出解释。

关于事件的以上各种运算关系还可以推广到多个事件的情形。

§1.2 概　　率

§1.1中指出概率是指随机事件在试验中出现的可能性大小的数量指标，出现可能性大的事件应有较大的概率，反之，应有较小的概率。概率是我们研究随机现象时十分关心的内容。那么如何来确定随机事件具体的可能性大小呢？今后将随机事件A、B、C…发生的概率用$P(A)$，$P(B)$，$P(C)$…来表示。

1. 概率的统计定义

大量工作、生活经验告诉我们，若在一定条件下重复某种试验，并考察我们所关心的某随机事件A（即某种试验结果）发生与否的话，当试验次数足够多时，事件A发生的频率，即事件A发生次数m对试验总次数n之比常常会逐渐稳定于某个与试验总次数n无关的常数p。表1.1.1提供的实验数据就是一个例证，今后我们把这个常数p（如果它存在的话）定义为事件A发生的率概，记作

$$P(A)=p$$

表1.1.1　　　　落叶松种子发芽试验的频率记录

n累计试验次数	50	100	150	200	250	300	350	400	450	500	550	600	650	700
m累计发芽粒数	25	55	85	117	152	154	201	234	269	299	334	362	391	421
发芽频率$\frac{m}{n}$	0.54	0.55	0.57	0.59	0.61	0.57	0.59	0.60	0.61	0.60	0.61	0.60	0.60	0.60

在一般情况下，未必能得到p的精确值，但当试验次数充分大时，误差常常很小，按此方式定义的概率称为事件的统计概率。例如上述种子发芽的统计概率p约为0.60。

2. 概率的古典定义

实践中常常遇到一类简单的随机试验，它具有这样的特性：（1）试验的可能（不可再分）的结果，只有有限个，即只有有限个基本事件$\omega_1\omega_2\cdots\omega_n$；（2）所有基本事件的发生都有相等的可能性。

具有这种特性的试验称为试验的古典概型。在古典概型中$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$。若我们关心的事件$A$包含$m$个基本事件，则事件$A$的概率定义为$P(A)=\frac{m}{n}$。这就是概率的古典定义。这里不可能事件$\phi$的概率直接规定为0，$P(\phi)=0$。

例如，欲从10株林木中用抽签方式抽取一株进行某项生理试验，我们没有理由认为某一株比另一株有更大的可能性被抽中，换言之，10株林木在地位上是平等的，又是不能再分的基本抽签单位，此时若其中有7株红松、1株冷杉、2株椴树的话，自然抽中红松的可能性应该大些，而按古典概率的定义，该试验中基本事件数$n=10$，有利于红松的基本事件数

$m=7$，将抽中红松事件记为A，抽中冷杉事件记为B，抽中椴树事件记为C，则

$$P(A)=\frac{7}{10} \qquad P(B)=\frac{1}{10} \qquad P(C)=\frac{2}{10}$$

例1.1.1，共5株林木，其中有3株病腐，随意抽取2株，问抽中都是病腐木的概率P_1有多大？抽中1株病腐木、1株健康木的概率P_2有多大？

解：随意抽取意味着抽取每株林木（基本事件）的等可能性，符合古典概型的要求。

由排列组合知识知道，在5株中抽取2株，总共有C_5^2种不同的抽取方式，而2株均为病腐木意味着这2株必须从3株病腐木中抽取出来，一共有C_3^2种取法，从而

$$p_1=\frac{C_3^2}{C_5^2}=\frac{3\times 2/2!}{5\times 4/2!}=\frac{3}{10}$$

类似地，抽中1株病腐木，1株健康木分别有C_3^1和C_2^1种取法，从而取出一株健康木，1株病腐木的可能方式数为$C_3^1\times C_2^1$，故

$$p_2=\frac{C_3^1\times C_2^1}{C_5^2}=\frac{6}{10}$$

一般地从N件产品，其中含M件正品，$N-M$件次品，抽取n件，问恰恰抽中m件正品的概率有多大？记该事件为A，则

$$P(A)=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n} \qquad (1.2.1)$$

这是一个典型问题，或者说是一类很基本的试验概型，称为超几何概型。实践中有许多问题可以归结为这种概型。

3. 几何概率

作为古典概率的推广，我们介绍几何概率的概念。设Ω是二维平面上可求面积的点集，面积为$S(\Omega)$，考虑随机试验：在Ω上随机掷一点M。这种随机试验称为几何概型，随机投掷意指每次投掷，M落在这可求面积的集合$A(A\subseteq\Omega)$中的可能性与集A的形状，位置无关，只与A的面积$S(A)$大小成正比。在几何概型中，Ω中的点是基本事件，因此基本事件个数无限且不可列，$P(A)$因此不能用$\frac{m}{n}$表示。

定义 $P(A)=\frac{S(A)}{S(\Omega)}$，这就是几何概率。这种定义很直观。

4. 概率的基本性质

总结以上三种概率定义，概率应具备以下基本性质：

（1）不可能事件概率为0　　$P(\phi)=0$

（2）必然事件概率为1　　$P(\Omega)=1$

（3）对任意随机事件A　　$0\leqslant P(A)\leqslant 1$

我们指出，不同概率定义所定义的事件的概率性质是一致的，今后将根据方便和需要用任意一种方式说明各种有关概率的结果。

§1.3 概率的基本定理

1. 加法定理

对任意事件A，B均有

$$P(A+B)=P(A)+P(B)-P(A\cdot B) \qquad (1.3.1)$$

定理可按图1.3.1用几何概率理解。$A+B$事件发生相当于在Ω中随机掷一点落入A或B集合范围内，而这个集合的面积为$S(A)+S(B)-S(A\cdot B)$，其中$S(A\cdot B)$是A，B公共部分所占面积。

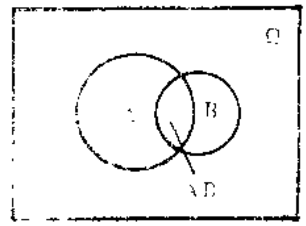


图 1.3.1

特别应当指出，若A、B互斥，则

$$P(A+B)=P(A)+P(B) \qquad (1.3.2)$$

因为此时$P(A\cdot B)=P(\phi)=0$

这个结论易于推广到更多个事件的情形，如当A，B，C两两互斥时

$$P(A+B+C)=P(A)+P(B)+P(C)$$

这种推广对下面将讨论的许多定理都是类似的，今后将不一一叙述。

2. 条件概率

在实际问题中除了讨论随机事件A的概率外，往往要研究在B事件发生条件下A事件发生的概率问题。把这种概率称为事件B发生条件下事件A发生的条件概率，记作$P(A|B)$。

根据图形，用几何概率来理解，我们定义

$$P(A|B)=\frac{P(A\cdot B)}{P(B)} \quad (P(B)>0) \qquad (1.3.3)$$

若$P(B)=0$，规定$P(A|B)=0$，今后论证中将忽略这种特定情形。

条件概率也具备概率的各种性质。

3. 乘法定理

由条件概率定义

$$P(A|B)=\frac{P(A\cdot B)}{P(B)} \qquad P(B|A)=\frac{P(A\cdot B)}{P(A)}$$

立即可得乘法定理

$$P(A\cdot B)=P(A)\cdot P(B|A) \qquad (1.3.4)$$

$$P(A\cdot B)=P(B)\cdot P(A|B) \qquad (1.3.5)$$

4. 独立性概念

若$P(A|B)=P(A)$，称A独立于B，这意味着B事件发生与否不影响A事件发生的概率。

类似，若$P(B|A)=P(B)$称B独立于A。

可以证明，若$P(A|B)=P(A)$则$P(B|A)=P(B)$。事实上，若$P(A|B)=P(A)$，由乘

法定理

$$P(A\cdot B)=P(A)\cdot P(B|A)=P(B)\cdot P(A|B)=P(A)\cdot P(B)$$

故$P(A)[P(B|A)-P(B)]=0$

在$P(A)>0$条件下，必有$P(B|A)=P(B)$

今后若$P(A|B)=P(A)$或$P(B|A)=P(B)$，则将一般地称A，B相互独立。

实践中往往直接由直观作出事件A、B相互独立的种种判断。例如，从一批出芽率为P的红松种子中随意取出二粒，播种于同一穴内，则它们各自出芽的事件被视为是相互独立的。

5. 乘法定理的重要推论

（1）若A，B相互独立，则

$$P(A\cdot B)=P(A)\cdot P(B) \qquad (1.3.6)$$

（2）若A，B相互独立，则，A，$\overline{B}$；$\overline{A}$，B；$\overline{A}$，$\overline{B}$也分别相互独立。

这里仅给出$\overline{A}$，B相互独立的证明，其它留作练习。

$\because\quad B=A\cdot B+\overline{A}\cdot B$，此处 $A\cdot B$ 与$\overline{A}B$互斥

$$P(B)=P(A\cdot B)+P(\overline{A}\cdot B)$$

$$\therefore\quad P(\overline{A}\cdot B)=P(B)-P(A\cdot B)=P(B)-P(A)\cdot P(B)$$

$$=P(B)(1-P(A))=P(B)\cdot P(\overline{A})$$

但 $P(B|\overline{A})=\dfrac{P(\overline{A}\cdot B)}{P(\overline{A})}$ 故 $P(B|\overline{A})=P(B)$

6. 全概率公式

全概率公式是把加法定理和乘法定理相结合而产生的。

设B_1 $B_2\cdots B_K$两两互不相容，$B_1+B_2+\cdots+B_K=\Omega$，称$B_1$，$B_2$，…，$B_K$构成互斥事件完备群，在此条件下对任意事件$A$，有全概率公式

$$P(A)=\sum_{i=1}^{k}P(B_i)P(A|B_i) \qquad (1.3.7)$$

证：$\because\quad A=AB_1+AB_2+\cdots+AB_k$，诸$AB_i$间两两互不相容($i=1,2,\cdots k$)(由$B_1,B_2,\cdots B_k$的互斥，完备性)

$$\therefore\quad P(A)=P(AB_1+AB_2+\cdots+AB_k)=P(AB_1)+P(AB_2)+\cdots+P(AB_k)$$

$$=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+\cdots+P(B_k)P(A|B_k)$$

$$=\sum_{i=1}^{k}P(B_i)P(A|B_i)$$

7. Bayes公式 贝叶斯公式

若$B_1B_2\cdots B_k$构成互斥事件完备群，则对任意事件A，($P(A)>0$)有

$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{k}P(B_i)P(A|B_i)} \qquad (1.3.8)$$

证：$P(B_i|A)=\frac{P(B_iA)}{P(A)}=\frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{k}P(B_i)P(A|B_i)}$

例1.3.1 前面提到7株红松、1株冷杉，2株椴木中任抽一株可形成A,B,C三个互斥事件

$$P(A+B)=P(A)+P(B)=\frac{7}{10}+\frac{1}{10}=\frac{8}{10}$$

若这10株林木中有3株红松病腐，1株椴树病腐，则从中任抽一株，在抽中红松条件下它还是病腐木（抽中病腐木记为D事件）的概率为

$$P(D|A)=\frac{P(DA)}{P(A)}=\frac{\frac{3}{10}}{\frac{7}{10}}=\frac{3}{7}$$

例1.3.2 在造林地上作红松直播造林，已知红松种子出芽率为90%，若每穴播一粒种子，将约有10%造林地上缺苗，为提高出苗率，决定每穴播2粒种子，问这时出苗率提高到了多少？

解：从每个穴来看，播一粒种子，90%可能出苗，10%可能不出苗，播二粒种子时，以A，B分别表示第一、第二粒种子出苗的事件，显然只要有一粒种子出苗，该穴便不缺苗，故所求出苗率为

$$P(A+B)=P(A)+P(B)-P(A\cdot B)$$

现已知 $P(A)=P(B)=0.90$，另一方面A、B两个事件根据经验可以判断它们相互独立。

$$\therefore\quad P(A+B)=P(A)+P(B)-P(A)\cdot P(B)$$
$$=0.9+0.9-0.9^2=0.99$$

概率的基本定理使人们能根据已知的某些事件的概率进一步计算出由这些事件构成的复合事件的概率来，而这些复合事件的概率正是实践中人们最关心的。

另解：播二粒种子后，每个穴缺苗意味着穴中二粒种子都不出苗，而这一事件为$\overline{A}\cdot\overline{B}$，已经知道，若$A$，$B$相互独立，则$\overline{A}$，$\overline{B}$也独立，从而由独立事件的乘法定理（1.3.6)知

$$P(\overline{A}\cdot\overline{B})=P(\overline{A})\cdot P(\overline{B})=[1-P(A)][1-P(B)]$$
$$=0.1\times0.1=0.01$$

$$故\quad P(A+B)=1-P(\overline{A}\cdot\overline{B})=1-0.01=0.99$$

这种考虑问题的思路在概率论中是常用的，值得注意。

例1.3.3 某林区根据气象资料知道防火期内，该区阴、晴、雨天的概率分别为15%，60%，25%，根据长期积累的火情资料，知道在阴、晴、雨条件下，发生森林火灾的概率分别为

$$P(\text{火}\mid\text{阴})=0.3\% \qquad P(\text{火}\mid\text{晴})=0.5\% \qquad P(\text{火}\mid\text{雨})=0.2\%$$

试求该地区发生森林火灾的总概率多大？

解：由全概率公式，视阴、晴、雨天分别为B_1，B_2，B_3构成互斥事件完备群

$$\begin{aligned}P(A)&=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)\\&=15\%\times0.3\%+60\%\times0.5\%+25\%\times0.2\%=0.00395=0.395\%\end{aligned}$$

例1.3.4 在上例条件下，若发生森林火灾，问火是在阴天发生的概率多大？

解：问题是要求$P(B_1|A)$，由$Bayes$公式

$$P(B_1|A)=\frac{P(B_1)P(A|B_1)}{P(A)}=\frac{0.15\times0.3\%}{0.00395}=0.114=11.4\%$$

8． 重复独立试验概型（贝努里概型）

在概率论及其应用中经常要遇到一个很基本的试验概型，即重复独立试验概型或称贝努里概型。它表达在一组条件下独立地重复一种试验，每次试验关心A事件发生或不发生这二个对立试验结果，一次试验中A事件发生概率已知为$P(A)=p$，希望知道在n次独立试验中事件A恰恰发生k次的概率。

例1.3.5 设有20株林木，其中有5株病腐木，用重复抽样方式随机抽取4株，试求

（1） 4株均为健康木的概率

（2） 4株中恰有3株健康木的概率。

解：重复抽样是指一次抽样后，抽中样木依旧放回，下次抽样时它仍有被抽中的可能，意即每一抽样单元都可以重复参与抽样。由此可知，每一次抽取，抽中病腐木的概率都是5/20，而抽中健康木的概率都是15/20。抽取4株可以理解为4次独立的抽取。于是问题归结为贝努里概型。

（1） 4株均为健康木这一事件是由每次抽中健康木的4次重复试验构成的，由独立事件的概率乘法定理，4个抽中健康木的事件同时发生的概率为$\left(\frac{15}{20}\right)\left(\frac{15}{20}\right)\left(\frac{15}{20}\right)\left(\frac{15}{20}\right)=\left(\frac{3}{4}\right)^4$

（2）恰有3株健康木这一事件由以下诸事件之一发生所构成，即

A_1：病、健、健、健
A_2：健、病、健、健
A_3：健、健、病、健
A_4：健、健、健、病

显然A_1，A_2，A_3，A_4间两两互不相容，且每个事件发生的概率由独立事件概率乘法定理知它们均为 $\frac{1}{4}\times\left(\frac{3}{4}\right)^3$，故所求概率为

$$P=P(A_1)+P(A_2)+P(A_3)+P(A_4)=4\times\frac{1}{4}\times\left(\frac{3}{4}\right)^3$$

等式右端$\frac{1}{4}\times\left(\frac{3}{4}\right)^4$系数4可以理解为$C_4^1$，它等于4个位置按1病3健4个元素填入的一切可能方式数，由于3株健康木填入3个位置时与它们之间的先后顺序并不相关，故采用组合公式。又病腐木位置确定，剩余三个位置就一定是3株健康木，故问题决定于病腐木放到什么位置上，这正相当于1株病腐木每次放在哪个位置（共4个位置）上，故有$C_4^1=4$，种不同放法。

（3）实际上上述计算可以推广到4株林木中恰有2病2健；3病1健；4病0健等各种可能情形中去，相应概率用与这里完全一样的原则可以算出为

$$P(2健2病)=C_4^2\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^2$$

$$P(3病1健)=C_4^3\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^1$$

$$P(4病0健)=C_4^4\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^0$$

$$它们与P(1病3健)=C_4^1\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^3$$

$$P(0病4健)=C_4^0\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^4$$

合在一起给出该独立试验概型的一切可能结果及相应概率。这5个概率计算式可以归结为

$$P_i=C_4^i\left(\frac{1}{4}\right)^i\left(\frac{3}{4}\right)^{4-i}\qquad i=0,1,2,3,4$$

P_i即4株林木中出现i株病腐木的概率。

（4）推广到更一般的情形，若抽取n株林木，其中病腐木出现概率为p，健康木出现概率为$q=1-p$，则

$$P_i=C_n^i p^i q^{n-i}\qquad i=0,\ 1,\ 2,\ \cdots n\qquad (1.39)$$

这是贝努里概型中计算概率的通用公式。

§1.4　随机变量及其分布

实践中有许多随机现象的结果本身具有数量特征，例如某林场共十台集材拖拉机，采伐季节每天能上山的拖拉机台数通常是不确定的，但总不外乎0，1，2，…，9，10这11种情形，每天对应一种情形，可以理解为一次随机试验的结果，这里每一结果对应于一个数。某林业局统计每年的火灾事件数，这也是不确定的，但不外乎是在0，1，2，…内，即各种非负整数范围内的一个数，每年统计一次视为一次观测或试验，其结果是这种数。当然还有一些随机现象的结果不具有数量特征，如取一粒种子做发芽试验，结果可能是发芽、不发芽这二种情形之一，但即使如此，让发芽结果对应于1，不发芽结果对应于0，就可以使结果实现了数量化，这种处理对于人们是方便的，因为可以使用数学工具。因此现在开始，设想所研究

的随机现象的每个结果都可以用数值表示。

假定一个随机现象，它的每一试验结果都可以用变量ξ取得某确定的值x（或y，z，……等）表示，而每次试验结果，ξ究竟取哪一个值是随机的，不确定的，我们就称ξ是一个随机变量，ξ可能取到的全部数值，即x构成的集合，亦即ξ的取值范围，自然是研究随机现象的人们所关心的，不同的随机现象或者说不同的随机变量，它们的取值范围一般是不同的，这种不同正是区别不同随机变量的一种标志，利用我们已经习惯的术语，可以认为不同的随机现象（随机变量）对应于不同的Ω。

随机变量有各种不同类型，主要分两大类，一是离散型的，一是连续型的，由于Ω是基本事件构成，两类不同的基本事件形成了两类不同的随机变量。若基本事件是可以一一列举的，由有限个或可列无限个基本事件构成Ω时，称相应的随机变量是离散型的；若基本事件不能一一列举而连续地在数轴上或数轴上的一个或几个区间上取值时，称相应的随机变量是连续型的。前面提到林場集材拖拉机每天的上山台数是一个随机变量，林业局每年的森林火灾次数也是一个随机变量，它们都是离散型的。而在一片较大范围的林子里随机抽取林木测定其胸径时我们也遇到了随机变量，其取值只能认为是某个区间内，例如，〔4,80〕(cm)内，这个区间就是Ω，基本事件是林木胸径取该区间内的各种各样值，但它们是无法一一列举的，这种随机变量是连续型的。总之，随机变量ξ不同于高等数学中的变量，它在基本事件ω构成的集合Ω上取值，今后恒用ξ或η或ζ等希腊字母表示随机变量，而用x，y，z，…表示它们的取值。随机变量ξ取值1用$\xi=1$表示，ξ取值x用$\xi=x$表示，而随机变量在$[a, b]$中取值用$a\leqslant\xi\leqslant b$表示。

前面的讨论还指出，仅仅说明一个随机变量可能取哪些值，即随机现象可能取得哪些结果是不够的。为了揭示随机现象中蕴含着的统计规律性，需要掌握与各种可能结果相应的概率即随机变量取各种值的概率。

1． 离散型随机变量

对于离散型随机变量，由于其基本事件的可列性，而各种各样事件无非是基本事件复合构成，因此若能将各基本事件取值及对应的概率列出在一张表上，则随机变量所能取的各种各样值及相应概率就可以由这张表推算出来，这张表显示了随机变量取各种可能值对应的概率分配情况称为ξ的概率表，由于ξ的取值可以一一列出，亦称为概率分布列，简称分布列。随机变量的取值及其概率的对应关系称为概率分布，概率分布完整地刻画了随机变量。

（1）例1.4.1.设一批种子发芽率为$P=0.8$，从中随机抽取一粒做发芽试验，记发芽为1，不发芽为0，我们遇到一个离散的，只取二种结果的随机变量ξ，其概率分布列如右表所示，而这个概率分布常称为0—1分布。

ξ取值x	0	1
$P(\xi=x)$	0.2	0.8

一般的离散型随机变量的概率分布列如表1.4.1

表1.4.1　　离散型随机变量的概率分布列

ξ 取值 x	x_1	x_2 ……	x_n	……	总和
$P(\xi=x)$	p_1	p_2 ……	p_n	……	1

由于全部基本事件合在一起构成样本空间Ω，所以全部与基本事件对应的概率之和必定