全国高等林业院校试用教材

数理统计

北京林学院主编

林业专业用

中国林业出版社

全国高等林业院校试用教材

数 理 统 计

北京林学院主编

林 业 专 业 用

中 国 林 业 出 版 社

前　　言

本书是为高等农林院校林业系林业专业数理统计课程编写的试用教材．与1961年9月出版的高等林业院校数理统计交流讲义相比，有了较大的变动。主要是增加了试验设计与抽样技术等内容．这是由于，近年来这些数理统计理论和方法的发展比较迅速，并且已在林业生产和科学研究工作中得到大量的应用．同时，随着电子计算技术的发展，逐步回归，数量化方法的理论和应用也日益发展和普及．此外，在数理统计方法的数学原理方面，如所作介绍过于简略，不仅不能满足同学们的学习要求，而且也会影响到在实际工作中的正确应用，所以，在只涉及微积分与少量线性代数知识的原则下，对于公式与定理，尽可能地进行推导和证明．

全书约需讲授90—100学时．但可根据各院校的具体情况，作适当的增删．例如，试验设计与抽样技术两章，可以只选一章讲授．公式与定理的推导证明，也可不全部讲授．这样，则讲授时数约可减少20学时．

由于我们业务水平有限，时间比较仓促，错误或不当之处在所难免，诚恳地希望使用本教材的教师和同学们提出宝贵意见，以便今后改正．

南京林产工业学院廖桂宗、朱明德，福建林学院吴敬和中南林学院马新为等同志对本书的初稿提出过很好的意见，在此，我们谨向以上几位同志表示诚挚的谢意。

编　　者

1979年10月

目　录

绪　论

1. 数理统计研究的问题

数理统计所研究的问题可归结为两类，即参数的统计估计与统计假设检验．为了进一步说明，我们先从统计数字谈起．

统计数字早为人们所熟悉．在报纸，杂志及科技文献中，经常可以看到．

例如，世界人口总数、某个国家的石油年产量、某一品种小麦的平均亩产量、某工厂一批某种产品的合格率、某省在某数年内粮食产量平均每年增长速度、某林区在某时的木材蓄积量、某种合金导线的平均电阻系数、某种化合物的平均分解热、某种放射性原子的平均寿命、某地某一时期内的平均气温、某种遗传物质的平均遗传力、某种新药对于某种疾病患者的治愈率等．这些都是统计数字．

可以概括地说，统计数字是说明大量同类事物或现象的数量特征或规律性的数字．

在社会经济各部门以及科学技术各领域，人们常用统计数字说明一定时间、地点及一定条件下的某些状况，以便以此为依据，制订今后的工作计划．例如，工农业生产中，计划的执行情况，今后技术指标的制订，科学技术研究工作中所取得的成果等，常以统计数字为依据或用统计数字来说明．所以，取得统计数字的工作，在社会经济的许多部门和科学技术的许多领域中占有非常重要的地位．

取得统计数字的方法可分为两大类．第一类是采用全面调查、重点调查或典型调查等方式，搜集到研究对象的原始资料后，经过整理和分析而取得统计数字；第二类是采用抽样调查的方式，由全部研究对象中抽出一部分进行调查，取得原始资料，根据数学原理，主要根据概率论原理进行分析，最后，对于研究对象的统计数字进行估计或假设检验．

在科学技术工作中，通常采用第二类方法．这种根据抽样调查资料并应用数学原理取得统计数字的方法就是数理统计方法．用第一类取得统计数字的方法不是数理统计所研究的问题，这里不再讨论．

在数理统计中，把说明大量同类事物或现象的数量特征或规律性的统计数字称为参数．因此，数理统计所研究的问题可以归结为参数的统计估计与统计假设检验等两类问题．

关于抽样调查，参数估计，假设检验的概念、方法和原理，以后将作较详细的讨论．

2. 数理统计在林业工作中的应用

数理统计是一门应用数学，于十九世纪末开始形成为一门独立学科．当时，它的应用范

围仅限于天文测量和遗传学等方面．由于科学技术的迅速发展，特别是由于工业生产高速化、自动化以及信息科学、遥感技术和电子计算技术的发展，数理统计的应用范围不断扩大．目前，在物理、化学、天文、气象、地质、生物、水文、地震、农、林、医、药以及工农业生产和社会经济等许多科学技术领域中，都应用数理统计方法解决一些理论问题和实际问题．

数理统计在林业工作中的应用可大致分为三方面：

（1）林业科学技术方面：在遗传、育种、造林、育苗、病虫害防治、化学保护等方面的科研工作中，常用到参数估计、假设检验和试验的设计与分析等方法．

（2）林业调查设计方面：在森林资源清查以及林区、林场的规划设计等生产活动中以及与此有关的森林学、测树学、航测、规划设计等学科的领域中，常用参数估计、回归分析与抽样技术等方法．

（3）森林工业方面：在森林工业产品的质量检查与质量控制等方面，常需应用参数估计，试验的设计与分析以及产品质量检查与控制等方法。

具体的应用实例，将在下面介绍。

第一章 概率论概要

概率论是一门应用范围非常广泛的数学学科，也是数理统计的理论基础. 限于条件，不准备对于概率论作严密和系统的讨论，但是，为了有助于理解数理统计方法的原理，以便能够正确地应用以下各章所介绍的数理统计方法，有必要介绍概率论中的一些基本概念和定理.

§ 1.1 概率的概念和基本性质

1. 事 件

在一定条件下进行某项试验时，常可根据试验条件和已掌握的知识预先作出判断：有些现象必定将在试验结果中出现；有些现象在试验结果中不可能出现；有些现象在试验结果中可能出现，也可能不出现.

为了便于进行以下的讨论，我们把试验结果中准备观察其是否出现的现象称为**事件**. 把试验结果中必将出现的现象称为**必然事件**；把试验结果中必定不会出现的现象称为**不可能事件**；把试验结果中可能出现，也可能不出现的现象称为**随机事件**.

举一个简单的例子：设有一小片油松，白皮松混交林，共 978 株，油松 617 株，白皮松 361 株；林木胸径最小者为 7.8 cm，最大者为 29.3 cm. 随意地由该小片林地上取一株林木进行观察，则“所观察林木为针叶树”以及“所观察林木的胸径不小于 4 cm”都是必然事件；“所观察林木为阔叶树”以及“所观察林木的胸径大于 30 cm”都是不可能事件；而“所观察林木为油松”以及“所观察林木的胸径大于 20 cm”都是随机事件.

在日常生活和工农业生产以及科学技术的许多领域中，必然事件，不可能事件以及随机事件都大量地存在着，每个人都可以举出一些实例，因此，不再举出更多的例子了. 今后，在概率论中，将以随机事件作为主要的研究对象.

2. 概率的定义

在上面所举的例中，“所观察林木为油松”与“所观察林木为白皮松”都是随机事件. 但是，如果按照所规定的试验条件进行多次试验，则将发现，“所观察林木为油松”这一事件出现的次数与“所观察林木为白皮松”这一事件出现的次数多不相等. 而且，试验次数愈多，愈明确地显示出，“所观察林木为油松”这一事件出现的次数多于“所观察林木为白皮松”这一事

件的出现次数．于是，我们很自然地认为，"所观察林木为油松"这一事件在试验结果中出现的可能性较大．

为了用一个数量客观地表明一个事件在试验结果中出现的可能性起见，我们把这样的数量称为该事件的概率，并且给予概率以如下定义：

设在一定条件下重复地进行某项试验，观察事件 A 在各次试验结果中是否出现．如果在 n 次试验结果中，事件 A 出现了 m 次，则称 $\frac{m}{n}$ 为事件 A 在 n 次试验结果中出现的频率．在继续加大试验次数 n 时，如 $\frac{m}{n}$ 逐渐稳定地在一个常数 p 的附近摆动，则事件 A 有概率．常数 p 即为事件 A 的概率．

通常，我们用符号 $P(A)$ 表示事件 A 的概率．因此，如果事件 A 的概率为 p，则可以写作：$P(A)=p$．

例如，设在一定的试验条件下进行落叶松种子的发芽试验时，试验结果如表 1.1 中所示：

表 1.1

累积试验次数 n	50	100	150	200	250	300	350	400	450	500	550	600	650	700
累积发芽粒数 m	27	55	85	117	152	184	201	234	269	299	334	362	391	421
发芽频率 m/n	0.54	0.55	0.57	0.59	0.61	0.61	0.57	0.59	0.60	0.60	0.61	0.60	0.60	0.60

则由于加大试验次数 n 时，落叶松种子的发芽频率逐渐稳定，并且，可以看到，稳定点在 0.60 附近(注意，表 1.1 中发芽频率多系近似值)，亦即，在所规定的试验条件下，"落叶松种子发芽"这一事件有概率，概率 p 在 0.60 附近．所以，可以用 0.60 作为这一事件的概率近似值，或用符号表示如下：

$$P(\text{落叶松种子发芽})=p\approx 0.60$$

在一般情况下，不能得到常数 p 的精确数值，只能得到其近似值．但在试验次数充分大时，近似值的误差可达充分小，并且具有充分大的可靠性．这个问题，在第三章中将作进一步的讨论．

根据概率的定义，可以看到，经常出现的事件具有较大的概率，而较少出现的事件则具有较小的概率．亦即，概率数值的大小，客观地反映了事件在试验结果中出现可能性的大小．

上面所介绍的概率定义是概率的统计定义．关于概率的定义，存在着多种不同的见解．近年来，多倾向于采用公理化的概率定义．即把概率看作是与事件相结合的并满足某些公理的一个数．但是，概率的公理化处理比较复杂，需要用到较多的数学工具，所以，在这里不

准备对于公理化的概率定义作更多的讨论．以下，关于概率的基本性质和基本定理的证明，均将以概率的统计定义为基础．此外，也准备将应用范围较为广泛但具有一定缺点的概率古典定义加以介绍．

3．概率的基本性质

由概率的定义可以得到概率的下列基本性质：

(1) 概率是不大于1的非负实数

亦即，任何一个事件A的概率$P(A)$都必定满足

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

因为，事件A在n次试验结果中出现的次数m必定满足

$$0\leqslant m\leqslant n,$$

因此，如以$\frac{1}{n}$乘以不等式的各项，可以得到，频率$\frac{m}{n}$必定满足

$$0\leqslant \frac{m}{n}\leqslant 1$$

频率的稳定点$p=P(A)$必定满足

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

(2) 必然事件的概率等于1

我们用符号U表示必然事件．亦即，U是A的特例．则这个性质可以写作：

$$P(U)=1$$

概率的这一性质是显而易见的．因为，必然事件在每次试验结果中都必定出现，因此，在n次试验结果中，必然事件出现的次数$m=n$．频率$\frac{m}{n}=1$，其稳定点$p=1$，即$P(U)=1$．

(3) 不可能事件的概率等于0

我们用符号V表示不可能事件．亦即，V也是A的特例．则这个性质可以写作：

$$P(V)=0$$

概率的这一性质也是显而易见的．因为，不可能事件在每次试验结果中都不会出现，因此，在n次试验结果中，不可能事件出现的次数$m=0$，频率$\frac{m}{n}=0$，其稳定点$p=0$，亦即，$P(V)=0$．

§ 1.2 概率的基本定理

1．概率的加法定理

(1) 事件和与事件积　如果在试验结果中同时考虑事件A与事件B等两事件的出现与否

时，则可能结果有下列四种：①，A与B同时出现，②，A出现而B不出现，③，A不出现而B出现，④，A与B都不出现.

为了便于叙述概率运算的基本定理起见，我们把“A或B出现”称为事件A与事件B之和，并用符号$A+B$表示.

应说明，“A或B出现”是指A与B两事件中的任一事件出现，或两事件同时出现．也就是说，“A或B出现”包含“A与B同时出现”，“A出现B不出现”以及“A不出现B出现”等三种情况在内.

如果在试验结果中同时考虑$A_1, A_2\cdots, A_k$等k个事件的出现与否时，则该k个事件中的任一事件出现，或某几个事件出现，即称为该k个事件之和，并用符号$A_1+A_2+\cdots+A_k$或用符号$\sum_{i=1}^{k} A_i$表示.

此外，我们把“A与B同时出现”称为事件A与事件B之积，并用符号AB表示．同样，我们把“A_1及A_2及$\cdots$及A_k等k个事件同时出现”称为该k个事件之积，并用符号$A_1A_2\cdots A_k$或用$\prod_{i=1}^{k} A_i$表示.

（2）**概率的加法定理** 首先，我们来讨论概率加法定理的最简单情况，即概率加法定理应用于两事件的情况．两事件和的概率，可按照下式进行计算：

$$P(A+B)=P(A)+P(B)-P(AB) \tag{1.2.1}$$

亦即，“A或B出现”的概率等于“A出现”的概率加“B出现”的概率减“A与B同时出现”的概率.

证：设在n次试验结果中，事件A出现m_1次，事件B出现m_2次，但在A出现的m_1次中以及B出现的m_2次中都包含了A与B同时出现的次数．设n次试验中，A与B同时出现的次数为m_3次，则$A+B$在n次试验中出现的次数为：

$$m=(m_1-m_3)+(m_2-m_3)+m_3=m_1+m_2-m_3$$

因此，$A+B$的频率为：

$$\frac{m}{n}=\frac{m_1+m_2-m_3}{n}=\frac{m_1}{n}+\frac{m_2}{n}-\frac{m_3}{n}$$

无论试验次数n为何，上式恒成立．当n充分大时，$\frac{m}{n}$稳定于$P(A+B)$，且$\frac{m_1}{n}$，$\frac{m_2}{n}$，$\frac{m_3}{n}$分别稳定于$P(A)$，$P(B)$与$P(AB)$．因此，可以得到(1.2.1).

由(1.2.1)式，用递推法，可以得到计算三事件和的概率时，有

$$P(A_1+A_2+A_3)=P(A_1)+P(A_2)+P(A_3)-P(A_1A_2)-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3) \tag{1.2.2}$$

因为：

$$
\begin{aligned}
P(A_1+A_2+A_3)&=P[(A_1+A_2)+A_3]=P(A_1+A_2)+P(A_3)-P[(A_1+A_2)A_3]=\\
&=P(A_1)+P(A_2)-P(A_1A_2)+P(A_3)-P(A_1A_3+A_2A_3)\\
&=P(A_1)+P(A_2)+P(A_3)-P(A_1A_2)-[P(A_1A_3)+P(A_2A_3)-P(A_1A_2A_3)]\\
&=P(A_1)+P(A_2)+P(A_3)-P(A_1A_2)-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3)
\end{aligned}
$$

仿此，并以此为基础，用数学归纳法，可以证明：

$$
P\left(\sum_{i=1}^{k}A_i\right)=\sum_{i=1}^{k}P(A_i)-\sum_{i<j}^{k}P(A_iA_j)+\sum_{i<j<l}^{k}P(A_iA_jA_l)-
$$

$$
-\sum_{i<j<l<m}^{k}P(A_iA_jA_lA_m)+\cdots+(-1)^{k+1}P\left(\prod_{i=1}^{k}A_i\right) \tag{1.2.3}
$$

这就是概率的加法定理．显然，(1.2.1)与(1.2.2)各为(1.2.3)在$k=2$与$k=3$时的情况．

由概率的加法定理(1.2.3)可以得到下面一些推论．在叙述这些推论之前，需要先介绍互斥事件的定义如下：

如果在试验结果中，事件A与事件B不可能同时出现，亦即，$P(AB)=P(V)=0$．则称A与B为互斥事件．或者说，A与B互斥或互不相容．

如果在$A_1,A_2,\cdots,A_k$等k个事件中的任何两个事件都不可能在试验结果中同时出现，则称$A_1,A_2,\cdots,A_k$等k个事件两两互斥或两两互不相容．

推论 1　如果$A_1,A_2,\cdots,A_k$等k个事件两两互斥，则

$$
P\left(\sum_{i=1}^{k}A_i\right)=\sum_{i=1}^{k}P(A_i) \tag{1.2.4}
$$

证：显然，如$A_1,A_2,\cdots,A_k$两两互斥，则(1.2.3)中的$P(A_iA_j)=P(A_iA_jA_l)=P(A_iA_jA_lA_m)=\cdots=P\left(\prod_{i=1}^{k}A_i\right)=0$．由此即可得到(1.2.4)．

这个推论也称为互斥事件的概率加法定理．

通常，两事件是否互斥是不难根据试验条件进行判断的．例如，在§1.1所举的例中，如油松与白皮松两树种的林木中都有胸径大于20 cm的林木，则“所观察林木为油松”与“所观察林木胸径>20 cm”就不是互斥的．同样，“所观察林木为白皮松”与“所观察林木胸径>20 cm”也不是互斥的．但是，“所观察林木为油松”与“所观察林木为白皮松”则是互斥的．同样，“所观察林木胸径>20 cm”与“所观察林木胸径<15 cm”也是互斥的．

请注意，如果我们用符号ξ表示“所观察林木胸径”，则“12 cm$\leqslant\xi<$16 cm”，“16 cm$\leqslant\xi<$20 cm”以及“20 cm$\leqslant\xi<$24 cm”等三个事件是两两互斥的．但是，“12 cm$\leqslant\xi\leqslant$16 cm”，“16 cm$\leqslant\xi\leqslant$20 cm”以及“20 cm$\leqslant\xi\leqslant$24 cm”等三个事件则不一定两两互斥，除非已知该小片林地上的978株林木中，没有胸径恰等于16 cm与20 cm的林木．

推论 2　事件的概率与该事件之对立事件的概率之和等于 1.

我们把试验结果中"A不出现"称为"A出现"的对立事件. 并且，把"A出现"简单地以符号A表示，而把其对立事件"A不出现"简单地用符号$\overline{A}$表示. 则推论 2 可以写作：

$$P(A)+P(\overline{A})=1 \tag{1.2.5}$$

证：显然，$A+\overline{A}=U$，$A\overline{A}=V$. 因此，$P(A+\overline{A})=P(U)=1$. 同时，由于$P(A\overline{A})=P(V)=0$，所以，由(1.2.4)可以得到：

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

于是，可以得到：

$$P(A)+P(\overline{A})=1$$

由(1.2.5)容易得到：

$$P(A)=1-P(\overline{A}) \tag{1.2.6}$$

以及

$$P(\overline{A})=1-P(A) \tag{1.2.7}$$

推论 3　若 $A_1,A_2,\cdots,A_k$ 等 k 个事件两两互斥，且在试验结果中必定出现其中之一，则

$$\sum_{i=1}^{k}P(A_i)=1 \tag{1.2.8}$$

证：由于 k 个事件两两互斥，所以，根据(1.2.4)，有：

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

又由于试验结果中必定有该 k 个事件中的一个事件出现，所以：

$$P\left(\sum_{i=1}^{k}A_i\right)=P(U)=1$$

因此，容易得到：

$$\sum_{i=1}^{k}P(A_i)=1$$

今后，我们把两两互斥且必出现其中之一的 k 个事件 $A_1,A_2,\cdots,A_k$ 称为**互斥事件的完备群**.

推论 4　如果 $A_1,A_2,\cdots,A_k$ 等概，且为互斥事件的完备群，则"其中某 m 个事件之一出现"的概率等于 $\frac{m}{k}$.

证：由于 $A_1,A_2,\cdots,A_k$ 等概，即

$$P(A_1)=P(A_2)=\cdots=P(A_k)=p$$