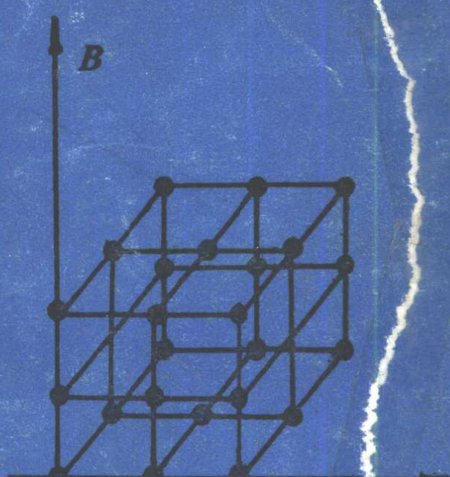
全国高等林业院校教材

# 数理统计

(第2版)

贾乃光 主编



(1)

中国林业出

全国高等林业院校教材

# 数 理 统 计

（第 2 版）

贾乃光 主编

中国林业出版社

（京）新登字033号

全国高等林业院校教材
数 理 统 计
（第 2 版）
贾乃光 主编

中国林业出版社出版（北京西城区刘海胡同7号）
新华书店北京发行所发行 通县振兴印刷厂印刷

787×1092毫米 16开本 25.75印张 591千字
1993年6月第2版 1993年6月第1次印刷
（1981年第1版，共印11次，印数：1-51550册）
印数1-6100册 定价：11.80元
ISBN7-5038-0987-6/S·0538

# 第1版 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

# 第2版 前　言

本书第1版由符伍儒教授主编，1980年10月出版，至今已有10年了。在这10年期间，许多院校的老师曾提出许多宝贵的意见。此外，本课程的教学大纲也在林业部高教处的组织领导下进行了多次修订。在此基础上，我们受林业部高教处的委托，对第1版进行修改补充。

我们可以设想，一位林学或水保系的本科毕业生在实际工作中遇到数理统计的问题时，他恐怕总会首先查阅他曾学过的教材。因为教材一般是他最熟悉、印象最深也是读得最仔细的书。他希望在教材中能找到解决他所遇到的问题的方法，或者进一步详查文献的线索。由于他所学的专业并非数理统计，所以他期望于数理统计教材的，不但有基本的原理和方法，还最好有进一步广泛的内容，和深入一步研究的途径。

然而，教材毕竟是针对课堂教学的讲义，受教学时数的限制，应以基本理论和方法为主，不能也不应该涉及得过广过深。

以上两方面是矛盾的，也是本书编者试图解决的重要问题。本书采用加注和加附录的方法加强内容的深度和广度，而这些注和附录无需在课堂讲授，属参考读物。比如，两总体均值的差异显著性检验和方差分析都要求总体是正态的，那么，细心的同学必然要问，对非正态总体如何解决这些问题呢？我们在附录中给出了解答，即使有些同学来不及细读附录，然而他知道，教材中有这方面的内容，一旦需要可以查阅。附录的内容不可能很详尽，仅仅是启发性地讲一部分，如还需要深入，则要参考另外的文献了。附录尽量做到简单易懂，希望起到过渡到参考文献的桥梁作用。

各章的注放在各章习题之后，因为注与正文关系密切，常常是一些较复杂的定理的证明，如果将这些证明放在正文中，占讲课学时过多，而这些证明的难度又是本科生可以接受的，所以作为注，供参考。有些定理的证明难度较大或占篇幅过大，则只在本文中指出参考文献，不再作注。至于附录，放在本书的最后，它们与正文的关系不像注那么密切，如复合分布等是为一部分学习优秀的同学扩大眼界的。

第一版的第一章很完整，可惜占用学时过多，故加以精简，并将一部分证明放在注里不在课堂讲授。

假设检验、方差分析和回归分析都采用从模型出发的方法，这在问题的提法上更简明。

如果学时数为80，建议只讲前6章；如果为100学时，可讲前7章；120学时，则可以8章都讲。

本版第一章由贾乃光、黄用廉编写，第二至七章由贾乃光编写，第八章仍用第一版的全部内容，未作变动。

由于编者水平所限，错误、疏漏、重点不当、叙述不妥之处一定不少，希望同行们和同学们批评指正。如能在内容的取舍上，深度、广度的安排上得到多数同行的赞同，编者将深感欣慰。

贾乃光　1991年9月于北京林业大学

# 目　录

# 绪　论

## 1. 数理统计研究的问题

统计数字如世界人口总数、某国的石油年产量、某一品种的小麦的平均亩产量、某工厂一批产品的产量或合格率、某省在某些年内粮食产量平均每年的增长速度、某林区在某时的木材蓄积量、某地某一时期内的平均气温等，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。

取得统计数字的方法可分为两大类。第一类采用全面调查、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的统计数字；第二类是采用抽样调查的方式，从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。用第一类方法取得统计数字的不属于数理统计的范畴。

## 2. 数理统计在林业中的应用

数理统计于19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。由于科学技术的飞速发展，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机的发展，数理统计的应用范围不断扩大，可以说，它可以应用于工农业生产、研究及社会经济的所有领域。

数理统计在林业工作中的应用可大致分为三个方面：(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，常用到参数估计、回归分析及抽样技术等方法。(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、试验设计以及产品质量检查与控制等方法。具体应用实例将在各章中介绍。

# 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，这里不准备对概率论作严密的和系统的讨论。但是，为了理解数理统计方法的原理，以便今后能够正确地应用数理统计方法，必须介绍概率论中一些基本概念和定理。

## §1.1　概率的概念和基本性质

### §1.1.1　概率的概念

**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：有些现象在试验结果中必然出现；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。

我们把试验结果中准备观察其是否出现的现象称为事件。把试验结果中必然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。

例如，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。随意地从中取一株林木进行观察，则“所观察林木为针叶树”以及“所观察林木胸径不小于4 cm”都是必然事件、“所观察林木为阔叶树”及“所观察林木胸径大于30 cm”都是不可能事件；而“所观察林木为油松”以及“所观察林木胸径大于20 cm”都是随机事件。

类似以上的例子，在日常生活、工农业生产和科学技术的许多领域中大量存在，读者不难自己举出一些。今后，在概率论中将以随机事件作为主要研究对象。事实上，我们今后把必然事件和不可能事件也都看作随机事件，是随机事件的两种极端的情形，正如在高等数学中把常量也看作是变量的特殊情况一样。

**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”出现的次数不相等，而且，试验的次数愈多，愈明显地显示出前者出现的次数多于后者出现的次数。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，我们把这样的数量称为该事件的概率，概率的定义如下：

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 $n$ 次试验中，事件 $A$ 出现了 $m$ 次，则称 $m/n$ 为事件 $A$ 在 $n$ 次试验中出现的频率。在继续加大试验次数 $n$ 时，如 $m/n$ 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，用 $P(A)$ 表示事件 $A$ 的概率。因此，如果事件 $A$ 的概率为 $p$，则记为

$$P(A)=p$$

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，试验结果为表 1.1 所示：

表 1.1 落叶松种子发芽试验结果

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

当加大试验次数 $n$ 时，落叶松种子的发芽频率逐渐稳定，稳定点在0.60附近，可以用0.60作为这一事件的概率的近似值，即

$$P(\text{落叶松种子发芽})=p\approx 0.60$$

在一般情况下，不能得到常数 $p$ 的精确数值，只能得到它的近似值。当试验次数 $n$ 充分大时，以频率值 $m/n$ 作为 $p$ 的近似值的误差可以达到充分小，并且具有充分大的可靠性。这个问题将在第三章作进一步讨论。

### § 1.1.2 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，任何事件 $A$ 的概率 $P(A)$ 都满足

$$0\leqslant P(A)\leqslant 1 \tag{1.1.1}$$

这是因为，任何事件 $A$ 在 $n$ 次试验结果中出现的次数 $m$ 必定满足 $0\leqslant m\leqslant n$，从而有 $0\leqslant m/n\leqslant 1$。作为频率 $m/n$ 的稳定点 $p=P(A)$ 也必然满足(1.1.1)式。

(2) 必然事件的概率等于 1　我们把每次试验都必定出现的事件称之为**必然事件**，并记为 $U$，则有

$$P(U)=1 \tag{1.1.2}$$

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 $m/n=1$，其稳定点 $p=1$，即(1.1.2)式成立。

(3) 不可能事件的概率等于零　我们把每次试验都不可能出现的事件称之为**不可能事件**，并记为 $V$，则有

$$P(V)=0 \tag{1.1.3}$$

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，$V$ 出现的次数 $m=0$，$V$ 的频率 $m/n=0/n=0$，其稳定点 $p=0$　即(1.1.3) 式成立。

## § 1.2　概率的基本定理

### § 1.2.1　概率加法定理

**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$B$两事件的出现与否时，则可能结果有四种：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。

我们把"$A$或$B$出现"称为事件$A$与事件$B$之和（或并），记为$A+B$，事件$A+B$出现包括以上的①②③三种情况在内，即$A+B$的出现是$A$或$B$至少有一事件出现。

如果在试验中同时考虑$A_1$，$A_2$，$\cdots$，$A_k(k\geqslant 2)$共$k$个事件的出现与否，则该$k$个事件之和记为$A_1+A_2+\cdots+A_k$。事件$A_1+A_2+\cdots+A_k$的出现是表示$A_1$，$A_2$，$\cdots$，$A_k$中至少有一个出现。$A_1+A_2+\cdots+A_k$常用符号$\sum_{i=1}^{k}A_i$表示。

此外，我们把"$A$与$B$同时出现"（即以上四结果中的①）称为事件$A$与$B$的积(或交，记为$AB$。同样，"$A_1$及$A_2$及$\cdots$及$A_k$等$k$个事件同时出现"的事件称为该$k$个事件之积，记为$A_1A_2\cdots A_k$或$\prod_{i=1}^{k}A_i$。

$p(A)+p(B)-p(AB)$

**2. 概率加法定理**

两事件$A$、$B$之和$A+B$的概率为：

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

亦即，"$A$或$B$"出现的概率等于$A$出现的概率加$B$出现的概率减去$A$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$m_1$次中有$m_3$次$B$也出现了，$m_2-m_3$次为$B$出现但$A$不出现的次数，$m_3$为$A$、$B$同时出现的次数。事件$A+B$(即$A$、$B$至少一个出现）的次数$m$可分解为① $A$、$B$同时出现，② $A$出现但$B$不出现，③ $B$出现但$A$不出现。即

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

因此，$A+B$出现的频率为

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

当$n$充分大时，$m/n$稳定于$P(A+B)$，而$m_1/n,m_2/n,m_3/n$分别稳定于$P(A),P(B)$，$P(AB)$，故得(1.2.1)。

如果用图形示意，如图1.1所示，设矩形表示必然事件$U$，左边的圆表示事件$A$，右边的圆表示事件$B$，两圆相交部分表示事件$AB$，而事件$A+B$为两圆外围的粗线条所围的部分，即图中有阴影的全体。若事件$A$的概率由$A$的面积表示（长方形$U$的面积为1)，