试验研究的数理统计方法

董德元　杨　节　苏敏文　顾关根　编著

中国计量出版社

试验研究的数理统计方法

董德元　杨　节　苏敏文　顾关根　编著

内 容 提 要

全书分为概率知识、统计检验及误差分析、回归模型分析三篇。在第一篇中，对概率的基础知识，通过四个章节作了较系统的阐明；第二篇着重介绍了试验及检测的数据处理理论及方法，涉及到数据和结论的统计检验、推理判断、误差分析等；第三篇则对过程的数学模型的建立及自适应等进行了详细说明。每一部分保持了相对完整性。因而读者可以对全书进行系统阅读，也可以对某篇独立运用。对概率基础知识及统计理论、方法的介绍，采用了统计分析母体抽样结果的模拟试验方法。这种尝试预计可以较快地帮助非数学专业读者系统深入地掌握有关知识。另外，对个别章节中有关工程数学问题，也做了必要的专门补充。

本书可作为工科专业的高等院校教材，也可供科研和技术工作者参考与自学。

试验研究的数理统计方法

董德元　杨　节　苏敏文　顾关根　编著

责任编辑　栾桂芬

中国计量出版社出版
北京和平里11区7号
中国计量出版社印刷厂印刷
新华书店北京发行所发行

开本 787×1092/16　印张 21.5　字数 510 千字
1987年 12 月第 1 版　1987年 12 月第 1 次印刷
印数 1—10 000　定价 5.10 元

ISBN 7-5026-0096-5/TB·79

前　言

本书是为研究工作者学习概率统计知识的入门书。

广大研究人员需要符合自己实际的统计知识读物。经验证明，这类著作的主要来源之一乃是研究工作在统计分析方面的成果。这种由业务研究来发展统计理论的趋势也是概率统计学的一个发展方向，本书是这样的一次尝试。

多年来，概率论的研究主要致力于随机性数学模型的更一般更抽象的本质结构，在著作中和教学上对统计理论的发展有所忽视。与此同时、各领域的研究者，则被迫进行着大量有关统计分析的探讨和发展工作。这样就形成了概率统计学的另外一个发展方向，并且它已取得了巨大的成绩（如生物、农业试验），事实证明，很多领域的业务成果与统计分析的发展是相辅相成的。但是这种发展在国内还处于初始阶段。华罗庚推广优选法、相纪珂介绍生物试验研究的统计分析成果及著作，都在这个发展方向上起了重大的推动作用。

广大研究工作者所遇到的主要困难，是在学习概率统计知识的道路上存在跳跃性很大的“沟壑”．例如，从一般高等数学到概率论缺少集合论等数学工具的“沟”，由定义、定理、分布等到实际过程有难联系的“沟”；由熟悉并习惯的演绎推理过渡到归纳推理的不习惯“沟”；在具体的较复杂的研究中，工程数学工具不足的“沟”等等．就是这些“沟壑”难越，往往使一些人的学习和应用中途止步。目前的一部分概率论专著及其通俗读物缺乏必要的统计定理和典型方法，一部分统计方法的专门性著作又往往缺少概率统计知识及基本理论。总之尚未能解决广大研究者在学习上的困难。

本书的特点之一是较全面系统地介绍了概率基础知识、统计理论和典型方法，以及有关计量检测分析；还特别收入了过程数学模型设计的试验研究的统计理论及方法。三部分内容形成了一个由基础知识、理论、方法到实际应用的整体，从而较好地解决了理论与实际之间“沟壑难越”的问题。

在第一篇中对事件、概率、随机变量的分布等都作了系统的阐述，还专门对试验结果的数值计算方法作了详细说明。这部分内容提供了必要的概率知识和试验数值的计算方法。第二篇中，对极限定理、抽样分布、假设检验、误差的估计与评定都作了详细介绍。特别是对具有基础性质的子样平均数分布、χ^2 分布、t 分布、F 分布等既给出了理论分析，还进行母体抽样试验。第三篇则是在对一元线性回归进行详尽分析的基础上对多元回归、逐步回归、非线性回归以及岭回归等作了全面说明，另外还对自适应等进行了介绍。

本书的特点之二是大量地、系统地采用了母体抽样试验方法。对有关定义、定理、分布等，除给了较严谨的陈述外，还都用抽样方法进行了解释和验证。而且读者自己也可以在学习中进行抽样试验。这种理论与实践相结合的方法，不但保证了对理论形态的理解，加深了物理方面的认识，也在学习的全过程中，起了填垓“沟壑”的作用。具有一般高等数学知识

的人，自学本书是较适宜的。

本书曾以讲义形式在工科学生的选修课程中使用过五年，也曾作为工程技术人员的进修课程的教本。反应较好。

作　者

1986.10

目　　录

第一篇　概率统计知识

第一章　概　率……………………………………………………（1）
1.1　随机试验……………………………………………………（1）
1.1.1 几个模拟的母体……………………………………………（1）
1.1.2 必然性和偶然性……………………………………………（4）
1.1.3 随机试验……………………………………………………（5）
1.2　事件……………………………………………………………（6）
1.2.1 点集…………………………………………………………（6）
1.2.2 事件及子样空间……………………………………………（6）
1.2.3 事件运算……………………………………………………（7）
1.3　概率……………………………………………………………（9）
1.3.1 概率古典定义………………………………………………（9）
1.3.2 概率统计定义………………………………………………（9）
1.3.3 概率统计定义试验…………………………………………（10）
1.3.4 概率的数学定义（公理结构）……………………………（10）
1.3.5 概率的性质…………………………………………………（13）
1.3.6 边际概率……………………………………………………（14）
1.3.7 条件概率……………………………………………………（15）
1.3.8 独立性………………………………………………………（16）
1.4　随机变量………………………………………………………（17）
第二章　离散型随机变量和连续型随机变量……………………（19）
2.1　离散随机变量的概率分布……………………………………（19）
2.2　二项分布………………………………………………………（21）
2.2.1 Bernoulli 试验及二项分布…………………………………（21）
2.2.2 二项分布与相辉三角………………………………………（23）
2.2.3 二项分布的（累计）分布函数……………………………（24）
2.2.4 二项分布的试验条件问题…………………………………（25）
2.3　泊松分布………………………………………………………（26）
2.4　连续随机变量的概率分布……………………………………（31）
2.4.1 连续随机变量的密度（函数）……………………………（31）
2.4.2 分布函数……………………………………………………（33）
2.5　边际分布和独立性……………………………………………（36）
2.5.1 边际分布……………………………………………………（36）
2.5.2 条件分布和独立性…………………………………………（37）
2.5.3 随机子样……………………………………………………（38）

2.6 导出密度 …… (39)
第三章 随机变量的特征数 …… (42)
3.1 期望值 …… (42)
3.1.1 离散随机变量的期望值 …… (42)
3.1.2 连续随机变量的期望值 …… (44)
3.1.3 期望值的性质 …… (45)
3.1.4 随机变量函数的期望值 …… (45)
3.2 方差 …… (47)
3.2.1 方差的定义 …… (47)
3.2.2 方差的性质 …… (51)
3.3 随机向量的期望和方差 …… (51)
3.3.1 二维及其函数的期望值 …… (51)
3.3.2 协方差 …… (52)
3.3.3 相关系数 …… (54)
第四章 正态分布 …… (56)
4.1 正态分布的定义及其特点 …… (56)
4.1.1 定义及特点 …… (56)
4.1.2 标准正态分布 $N(0, 1)$ …… (57)
4.2 正态分布函数 …… (58)
4.2.1 正态分布与概率计算 …… (58)
4.2.2 3*母体与理论分布的比较 …… (61)
4.3 二维正态分布 …… (61)

第二篇 统计检验及误差分析

第五章 试验结果的数值计算 …… (65)
5.1 数值计算的误差分析 …… (65)
5.1.1 有效数字与舍入误差 …… (65)
5.1.2 测定值代数运算的误差 …… (65)
5.2 几个近似计算方法 …… (68)
5.2.1 插值法 …… (68)
5.2.2 数值微分 …… (69)
5.2.3 数值积分 …… (70)
5.2.4 近似公式 …… (71)
5.3 试验结果的统计整理 …… (72)
5.3.1 一个母体的子样频率分布 …… (72)
5.3.2 二个母体子样的散点图与子样相关系数 …… (75)
第六章 极限定理及抽样 …… (77)
6.1 归纳推理 …… (77)
6.1.1 母体与子样 …… (77)
6.1.2 统计量 …… (78)
6.2 极限定理 …… (78)

6.2.1 切比雪夫不等式 …… (78)
6.2.2 大数定理 …… (80)
6.2.3 中心极限定理 …… (82)
6.2.4 标准化二项分布的正态近似 …… (84)
第七章 抽样分布 …… (86)
7.1 正态分布的子样平均数的分布 …… (86)
7.1.1 $\bar{X}$ 分布 …… (86)
7.1.2 $\bar{X}$ 分布试验 …… (86)
7.2 χ^2 分布 …… (87)
7.2.1 χ^2 分布密度及其概率计算 …… (87)
7.2.2 χ^2 统计量的抽样试验 …… (89)
7.2.3 χ^2 分布的一些具体形式 …… (93)
7.3 t 分布 …… (96)
7.3.1 t 分布密度及其概率计算 …… (96)
7.3.2 t 统计量的抽样试验及其各种形式 …… (97)
7.3.3 t 分布的常用形式 …… (99)
7.4 F 分布 …… (100)
7.4.1 F 分布密度及其概率计算 …… (100)
7.4.2 F 统计量的抽样试验 …… (101)
7.4.3 F 统计量的一些另外表示法 …… (102)
7.4.4 第1和第2子样 …… (103)
第八章 参数估计 …… (106)
8.1 点估计 …… (106)
8.1.1 判定函数及风险 …… (106)
8.1.2 估计量的性质 …… (108)
8.1.3 极大似然法 …… (112)
8.1.4 期望和方差的点估计 …… (115)
8.2 区间估计 …… (116)
8.2.1 已知 $D(X)$ 时期望值的区间估计 …… (116)
8.2.2 $D(X)$ 未知时期望值的区间估计 …… (119)
8.2.3 两个母体期望差 $(\mu_1-\mu_2)$ 的估计 …… (120)
8.2.4 母体方差估计（可用小子样） …… (121)
8.2.5 区间估计的抽样试验 …… (124)
第九章 假设检验 …… (128)
9.1 参数检验问题 …… (128)
9.1.1 参数检验的推理方法 …… (128)
9.1.2 最佳检验法 …… (131)
9.1.3 单边备择假设与双边备择假设 …… (135)
9.2 u 检验 …… (136)
9.2.1 单边 u 检验（σ^2 为已知） …… (136)
9.2.2 双边 u 检验（σ^2 为已知） …… (139)
9.2.3 单边对比 u 检验（σ_1^2 和 σ_2^2 为已知） …… (140)
9.2.4 双边对比 u 检验（σ_1^2 和 σ_2^2 为已知） …… (142)

9.3 t检验 …… (143)
9.3.1 单边t检验 …… (143)
9.3.2 双边t检验 …… (147)
9.3.3 对比t检验 …… (148)
9.4 χ^2检验和F检验 …… (151)
9.4.1 母体方差的χ^2检验 …… (151)
9.4.2 比较两个正态母体方差的F检验 …… (153)
9.5 χ^2的适度检验（拟合优度检验） …… (157)
第十章 误差的估计和评定 …… (160)
10.1 系统误差和随机误差的估计 …… (160)
10.1.1 系统误差与随机误差，正确度与准确度 …… (160)
10.1.2 系统误差的发现与估计 …… (161)
10.1.3 随机误差的估计与评定 …… (162)
10.2 粗大误差的评定（异常数据的取舍） …… (163)
10.2.1 t舍弃检验 …… (163)
10.2.2 F舍弃检验 …… (164)
10.3 误差的分配与传播 …… (164)
10.3.1 误差分配的基本公式 …… (164)
10.3.2 检测仪器的1/10及1/3法则 …… (165)
10.3.3 误差传播公式 …… (165)
10.4 系统误差与随机误差的综合评定方法 …… (167)
10.4.1 误差综合评定问题 …… (167)
10.4.2 随机误差（准确度）的综合评定方法 …… (168)
10.4.3 系统误差（正确度）的综合评定方法 …… (170)

第三篇 回归模型分析

第十一章 一元线性回归 …… (173)
11.1 一元线性模型 …… (173)
11.1.1 固定X的模型Ⅰ及其假设 …… (173)
11.1.2 对变数的模型Ⅰ …… (174)
11.2 回归方程 …… (177)
11.2.1 配回归方程的方法 …… (177)
11.2.2 配回归方程的计算格式及其简化 …… (182)
11.2.3 模型Ⅰ的模拟及点估计 …… (184)
11.2.4 对大子样配方程的加权回归 …… (188)
11.2.5 子样相关系数γ …… (189)
11.2.6 关于相关系数γ的几种形式 …… (191)
第十二章 一元线性回归分析 …… (193)
12.1 回归问题的方差分析 …… (193)
12.1.1 总平方和L_{yy}的分解 …… (193)
12.1.2 六种均方及标准差 …… (197)

12.2 区间估计及假设检验 …… (199)
12.2.1 母体回归系数 β 的区间估计及假设检验 …… (201)
12.2.2 回归常数项 α 的区间估计及假设检验 …… (204)
12.2.3 回归值 Y 的区间估计 …… (204)
12.2.4 实测值 Y 的区间估计 …… (206)
12.3 回归的一部分应用 …… (207)
12.3.1 回归方程的稳定性问题 …… (207)
12.3.2 Y 的预报和控制 …… (207)
12.4 相关系数及其检验 …… (208)
12.4.1 子样相关系数 γ 的分布 …… (208)
12.4.2 母体相关系数 $\rho=0$ 的检验（用 γ 分布）…… (209)
12.4.3 Z 变换与 Z 分布 …… (211)
12.4.4 ρ 的假设检验 …… (211)
12.4.5 ρ 的点估计及区间估计 …… (214)
第十三章 多元线性回归 …… (216)
13.1 多元线性回归方程的求法 …… (216)
13.1.1 回归平面（二元回归问题）…… (216)
13.1.2 多元线性回归的正规方程 …… (222)
13.1.3 正规方程的解法 …… (223)
13.2 多元线性回归的矩阵格式 …… (228)
13.2.1 矩阵运算法则和正规方程的矩阵格式 …… (228)
13.2.2 正规方程的矩阵解法 …… (231)
13.3 多元线性回归分析 …… (235)
13.3.1 Y 的总离差平方和的分解及方差分析 …… (235)
13.3.2 偏相关系数 r …… (237)
13.3.3 区间估计及检验 …… (238)
13.4 标准正规方程与标准（偏）回归系数 …… (243)
13.5 每个自变量的贡献 …… (245)
13.5.1 偏回归平方和与每个自变量的贡献 …… (246)
13.5.2 关于 $b_j=0$ 的检验 …… (247)
13.6 最优回归方程的建立方法 …… (248)
第十四章 逐步回归方法 …… (255)
14.1 逐步回归方法的计算步骤 …… (255)
14.2 逐步回归计算实例 …… (258)
第十五章 岭回归 …… (269)
15.1 最小二乘估计 …… (269)
15.1.1 线性回归模型 …… (269)
15.1.2 最小二乘估计 …… (270)
15.1.3 β 的最小二乘估计(LS 估计) $\hat{\beta}$ 的性质 …… (271)
15.1.4 误差方差 σ^2 的估计 …… (271)
15.2 自变量的选择准则 …… (272)
15.2.1 全模型与选模型 …… (272)
15.2.2 基于残差平方和（RSS）的自变量选择准则 …… (274)

15.2.3 AIC信息量准则 …… (275)
15.2.4 逐步回归方法 …… (275)
15.2.5 岭回归方法 …… (276)
15.3 岭回归 …… (276)
15.3.1 岭回归定义 …… (276)
15.3.2 岭回归的简单性质 …… (276)
15.3.3 一种岭回归算法 …… (280)
15.3.4 一个计算实例 …… (280)
第十六章 非线性模型的回归 …… (282)
16.1 坐标变换——"曲线改直" …… (282)
16.2 多项式回归 …… (287)
16.2.1 非线性模型的多项式逼近 …… (287)
16.2.2 正交多项式回归 …… (292)
16.3 非线性回归 …… (302)
16.3.1 台劳级数展开法（高斯-牛顿法）[15] …… (302)
16.3.2 带阻尼的台劳级数展开法 …… (307)
16.3.3 单纯形法[17] …… (309)
第十七章 自适应算法 …… (318)
17.1 线性递推回归 …… (318)
17.1.1 增长记忆的线性递推回归算法 …… (319)
17.1.2 渐消记忆的线性加权递推回归算法 …… (321)
17.2 单参数的自适应递推算法——指数平滑法 …… (325)
17.2.1 以本项预报偏差纠正模型参数的方法 …… (326)
17.2.2 指数平滑法 …… (326)
参考文献 …… (329)

第一章 概 率

1.1 随 机 试 验

1.1.1 几个模拟的母体

科学试验的目讨探是的被研究对象．在这里把被研究的对象叫作母体（或叫总体），从母体中抽取一部分个体，这一部分个体就叫子样．

为了理解课程内谷和学者自已动手做一些试验，在这里提供几组数据，且把这几组数据当作母体．

1# 母体：如果投一枚硬币，出现可能是正（1）面或反（0）面，一次一次投下去，并按次序纪录下来，比如 0，1，0，0，1，1，0，1，1，1，0，0，…把“正”和“反”推广一下，比如，“正品”“次品”，这就成了工厂里产品的质量等级．

下面有 100 支晶体管，“1”代表正品，“0”代表次品，堆放在货架上，如表 1.1 把这堆数据当作 1# 母体．

100 支晶体管的等级（1# 母体）

表 1.1

	0	1	2	3	4	5	6	7	8	9
0	1	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	0	1	1	1
2	1	1	1	0	1	1	1	1	1	1
3	0	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	0	1	1
5	1	1	1	1	0	1	1	1	1	1
6	1	0	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	0	1
8	1	1	1	1	1	0	1	1	1	1
9	1	1	0	1	1	1	1	1	1	1

2# 母体：用碳化钨粉压制，烧结成的合金模具，按尺寸及性能分为“1”“2”和“3”，表中 1，2，3 是等级数．把第二组数据当作 2# 母体．

1# 和 2# 母体的数据都是整数，是可数的．

100块模具等级表(2$^{\#}$母体)

表 1.2

	0	1	2	3	4	5	6	7	8	9
0	1	2	1	1	1	2	1	1	1	3
1	1	1	2	1	1	1	3	1	2	1
2	1	1	1	3	1	2	1	1	1	2
3	3	1	2	1	1	1	2	1	1	1
4	1	1	1	2	1	1	1	3	1	2
5	2	1	1	1	3	1	2	1	1	1
6	1	3	1	2	1	1	1	2	1	1
7	2	1	1	1	2	1	1	1	3	1
8	1	2	1	1	1	3	1	2	1	1
9	1	1	3	1	2	1	1	1	2	1

3$^{\#}$母体：第三组数据是每班在初轧机上用固定工艺规程轧制名义尺寸为 250×1300 的板坯。表中数据为“宽差”，即板坯尺寸与名义尺寸之差，比如，板坯的实际尺寸为1301，则其宽差为 1301－1300＝1，当然，这里是为了简化取了整数，因为 1 mm 可能是 1.1 mm，也可能是 1.105 mm 等。

100个板坯宽差尺寸(3$^{\#}$母体)

表 1.3

	0	1	2	3	4	5	6	7	8	9
0	−3	−7	5	3	−1	6	−5	−6	2	1
1	−1	1	−5	9	−3	7	−7	3	5	−9
2	1	−3	0	−9	11	−2	3	−5	−6	10
3	10	−3	1	−5	−11	5	−7	−2	4	−1
4	−13	1	−1	−2	1	2	4	−4	8	0
5	8	−4	−6	0	4	−2	0	2	−10	6
6	−4	−1	−8	2	−6	0	6	11	−3	3
7	6	−8	4	6	8	−4	3	0	0	−2
8	0	7	1	5	−5	3	−2	2	−7	−3
9	−3	0	9	4	2	−10	7	−9	6	−5

4$^{\#}$母体：表 1.4 是误差的记录，把它当做 4$^{\#}$ 母体，它与 3$^{\#}$ 都是占据了实数轴上的可能取值。

如果研究对象的范围是某种工艺对晶体管、模具、宽差、误差的影响，那么 100 个数也

是一部分数据，也就是一个子样，因为，生产还要继续下去。如果研究的范围就是限定在100个数据之内，那么就可以认为它们都是母体了。

表 1.4

	0	1	2	3	4	5	6	7	8	9	
0	0.2	-2.0	0.6	-0.4	1.6	-0.2	0.4	-0.6	1.2	-0.6	0
1	-1.2	0.6	-0.4	0.2	-1.4	1.4	-0.2	0	0.4	0.6	1
2	-0.2	1.6	0.4	-2.4	0.6	-0.4	0	0.2	-1.0	1.2	2
3	0.6	-0.6	-2.0	1.8	-0.6	0.6	1.0	-0.6	0	0.2	3
4	0.6	0	-0.2	1.2	0.4	1.0	-1.2	-1.4	0.2	-0.6	4
5	0	-0.4	1.8	-1.0	0.6	-0.6	0.6	-1.6	-1.2	1.4	5
6	-1.0	1.0	0	-0.4	-0.4	0.2	-0.6	0.6	-0.2	0.6	6
7	-0.4	-0.6	1.0	0	1.2	-1.0	2.4	0.4	-1.2	-1.8	7
8	1.4	0.2	-0.2	0.6	-0.6	0.4	-1.8	1.0	-0.2	-1.0	8
9	-0.2	0.4	-0.6	0.2	-1.4	-1.6	-0.6	2.0	2.0	0	9
	0	1	2	3	4	5	6	7	8	9	

母体的实际情况有时为人们所熟知，有时人们不了解它，而且大多数母体是我们不清楚的。为了今后使用这些母体，下面对它们加以说明：

把1*母体中所有个体直接列数（按0，1）就可以知道"0"是10个，"1"是90个，就是说，在1*母体中"0"占10%，"1"占90%，或者 $p=0.9$，$q=0.1$。

$$p+q=1 \tag{1.1}$$

可以设想，投掷一枚均匀硬币，将有 $p=0.5$，$q=0.5$，$p+q=1$。

2*母体也可以很容易用列数方法知道：$p_1=0.70$，$p_2=0.20$，$p_3=0.1$，即

等级	1	2	3	
p	0.7	0.2	0.1	(1.2)

3*和4*母体也可以用上述列数的办法求得各种"宽差"和"误差"所占的比率。但有时会遇到困难，特别是数据少时，数据将过分分散，因为它们可以在全实数轴取值。较好的办法是采用分组的办法。比如，$-13.5\sim-11.5$ 算一组，并且用该组的上下限的平均数 x_i 当作该组的代表值——叫作组中值，比如1用

$$[-13.5+(-11.5)]/2=-12.5$$

作为落在 $-13.5\sim-11.5$ 之间所有数据的代表。

在这里原数据为 -13.0，用的是提高一位的 -13.5，这是为了避免正好是 -13.0，-11.0 等数据分组时不好决定归属的问题。

表1.5是3*母体落在各区间的数字频数（n_i）和比率（n_i/n），n 是母体内的总数，用 $\frac{n_i}{n}\Big/\Delta x\approx f(x)$ 作为纵坐标，x_i 作为横坐标画出母体的分布图形（图1.1）。图中

表 1.5

(事件)组号	区间 $a_i \sim a_{i+1}$	组中值 x_i	频数	频数 n_i	比率 n_i/n
1	$-13.5 \sim -11.5$	-12.5	一	1	0.01
2	$-11.5 \sim -9.5$	-10.5	下	3	0.03
3	$-9.5 \sim -7.5$	-8.5	正	5	0.05
4	$-7.5 \sim -5.5$	-6.5	正下	8	0.08
5	$-5.5 \sim -3.5$	-4.5	正正	10	0.10
6	$-3.5 \sim -1.5$	-2.5	正正下	13	0.13
7	$-1.5 \sim 0.5$	-0.5	正正正	15	0.15
8	$0.5 \sim 2.5$	1.5	正正下	13	0.13
9	$2.5 \sim 4.5$	3.5	正正一	11	0.11
10	$4.5 \sim 6.5$	5.5	正止	9	0.09
11	$6.5 \sim 8.5$	7.5	正一	6	0.06
12	$8.5 \sim 10.5$	9.5	止	4	0.04
13	$10.5 \sim 12.5$	11.5	丅	2	0.02
				100	1.00

Δx 是组距，它是每个方块的底边.

$\frac{n_i}{n}$是方块的面积，它代表了 x_i 在母体中所占比率.

$\frac{n_i}{n}\Big/\Delta x$ 是方块的高度. 当 $\Delta x \to 0$ 时，图形将趋于一条圆滑曲线，即

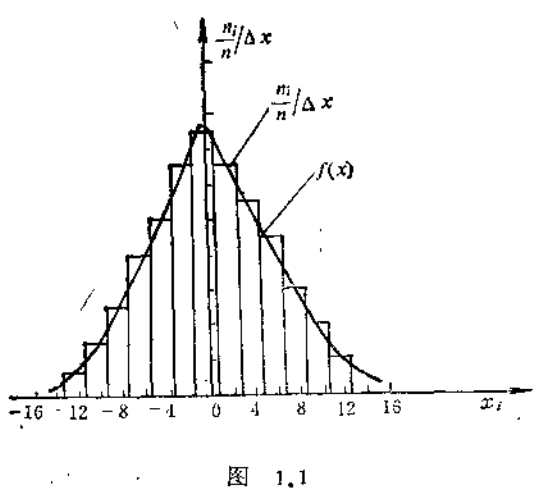

图 1.1

$$P=\lim_{\Delta x\to 0}\frac{n_i}{n}\Big/\Delta x = f(x) \tag{1.3}$$

很明显，当 $\Delta x \to 0$ 时，

$$\frac{\left(\frac{n_i}{n}\right)}{\Delta x}\cdot\Delta x\to 0 \tag{1.4}$$

所以，在 x_i 处的比率也就趋于 0，意思是说，当数据在实数轴上取值时，某一个定值（比如，$x_i=4.5$）在母体中所占的比率是零. 比率是由面积给定的，而不是由高度 $f(x)$ 给定的.

1.1.2 必然性和偶然性

我们做试验，比如从 1# 母体中随便取一数. 这个数是“0”，是“1”，并不一定，或

者说，结果是偶然的。

科学主要是研究在重复试验中的不变性质．因此，在这里必须把偶然性和必然性的辩证关系给予说明。

恩格斯深刻地指出："凡是断定为必然的东西，都是用纯粹偶然构成的．而凡被认为偶然的东西，则是一种有必然性隐藏在里面的形式。"这个科学的论断揭示了必然和偶然的辩证关系，它是认识事物的有力武器。

世界上的事物按其"发生的可能性"来看，可以分为：必然发生的和偶然发生的两大类。

如果，在某一定条件下，某个事件在试验中一定会发生，就称之为"必然事件"．如水在一个标准大气压力下，100℃时会沸腾，这就是一个必然事件．还有一些是在一定条件下不可能发生的事件，其实也是属于必然事件范畴的．比如"等速直线运动在无外力作用下，改变运动状态"是一个不可能事件，其实与下列说法是等价的："在无外力作用下，不可能改变等速直线运动状态"是一个必然事件．"人不可能长生不老"也是一个必然事件．

但是在自然界中，偶然性是永远伴随着一切事物的．但在偶然性中又蕴藏着必然的规律．这是因为母体中情况所决定的．投一枚硬币出现"正面"或"反面"是不能事先预料的。但是投掷次数很多时，则将发现"正面"和反面，将是近乎相等。

我们可认定，偶然事物的规律性主要表现在各种事物都具有自己出现的"可能性"上．比如，1# 母体中"1"占90%，"0"占10%，那么，当由里面随便地，不受主观愿望左右地抽取一个样，虽然"1"或"0"出现是偶然的，但是可确知，在同一条件下，重复抽取子样，在偶然中会发现必然性的规律，就是出现"1"的比率约占90%，出现"0"的比率约占10%。

我们过去曾大量接触过"确定性规律"，即"动态规律"．例如：欧姆定律 $U=IR$，牛顿第二定律 $F=ma$，在这些定律中，各参数之间呈现出一定的函数关系．另外，也存在着很多的"统计规律"，例如，由无数气体分子所组成的总体中，尽管每个分子的运动轨迹、速度等都是带有偶然性的，但从总体来观察，将发现气体的压力，则是稳定在一定值上。

实际上，在做研究"动态规律"的试验时，必然伴有偶然现象．比如，按欧姆定律做试验观测时，每次试验结果也必有一定的波动，这是因为在试验过程中，不可避免地伴有其它因素影响的缘故．不过，在误差不大时，一般是牺牲了偶然性误差所影响的精度而已．但必须指出，有时这种牺牲是不能允许的，有时要求我们把这种误差分离出来，并作必要的估计。

1.1.3 随机试验

通过前面的介绍，可以对母体、随机试验、随机事件等作如下的直观陈述．

母体 母体是指研究对象的全体，它所包含的各种个体，是按一定机会而出现的。

随机试验 在同一组条件实现的情况下，所得到的结果，不一定是相同的，然而，每一个可能的实验结果都有一定的出现机会，并假设这种试验在相同条件下可重复进行，这种试验叫作"随机试验"．

随机事件 随机试验的结果叫作"随机事件"．

比如，从2# 母体中，任意抽取一个子样，它可能是"2"或"3"，其结果是不定的．但

是，在母体中的“1”，“2”，“3”都有各自的比率，因而它们的出现必然是按一定机会的，因此，从 2# 母体中，随机地抽取子样，便是随机试验.

从 2# 母体中，每次抽一个数（子样含量为 1）则出现结果是“1”，“2”，“3”等都是随机事件，如果每次抽 2 个数，那么它的结果必然是（1，2）、（2，3），（3，1）、（1，1）、（2，2）、（3，3）六者之一，（1，2）、（2，3）或（3，1）等都是随机事件（在这里没有要求出现先后的次序）.

还有一点需要说明，就是试验在相同条件下是可重复进行，第一次试验不影响第二次试验的结果（称为随机独立的）.

实际上，为了模型化，我们是把上述随机试验理想化了. 在现实生活中，真正的重复试验是难做到的.

1.2 事　　件

为了描述概率的数学模型，下面对事件进一步加以说明，并且使它抽象化. 说明事件的根本原则是设法将一切可能的结果都包括到一个空间里去.

1.2.1 点　集

基本事件是指在一定研究范围内不能再分的随机试验结果. 从数学上说，可以把它当作空间一个点 ω，或者叫作一个元素.

基本事件空间是指基本事件所构成的集合. 抽象来讲，它是全部点 ω 所构成的集合. 记作 $\Omega=\{\omega\}$.

例 1.1 投掷一枚硬币，它有两个结果，一个是“正”，一个“反”. “正”和“反”都是基本事件，即 $\omega_1=$ 正，$\omega_2=$ 反. 由“正和反”两个点组成一个集合，就是投一枚硬币的基本事件空间，记作 $\Omega=\{$正，反$\}$或者写成 $\Omega=\{\omega$：正，反$\}$.

例 1.2 从 2# 母体中抽取一个数，它的不可能再分的结果只可能是“1”，“2”或者“3”. “1”，“2”，“3”都是基本事件，并且由这三个点集合成一个基本事件空间 $\Omega=\{1, 2, 3\}$.

例 1.3 设有 n 个试样，每个试样上都标号，即 $1, 2, \cdots, n$ 等. 任取其一，其结果必为 $1, 2, \cdots n$ 中一个数，这 n 个数分别都是基本事件，而且 $1, 2, \cdots, n$ 个点就形成一个集合，基本事件空间 $\Omega=\{1, 2, 3, \cdots, n\}$.

例 1.4 上述的例子的基本事件空间都是由有限数目的点所组成的集合. 下面是一个由无限个点组成的例子. 我们要合计一个季度内某自动生产线上仪表发生故障的次数，它可能是 0 次，1 次，2 次. 实际上这个基本事件空间将由与正整数的数目相同的无限个点所组成.

例 1.5 下面这个例子是连续性的. 从生产线上抽取一块板坯，测量它的宽差，其结果可能是实数轴上的任何一个点，即实数轴上任何一个点都是一个基本事件，而实数轴上的全部点的集合便是基本事件空间，$\Omega=\{\omega: (-\infty, \infty)\}$.

例 1.6 某随机试验的基本事件空间必须包括代表所有基本事件的点. $\Omega=\{\omega: 1, 2\}$，它不是从 2# 母体中抽取子样的基本事件空间，因为它没有把点“3”包括进去.

1.2.2 事件及子样空间

基本事件空间是由全部基本事件所组成的集合. 那么事件是由基本事件所组成的某个集