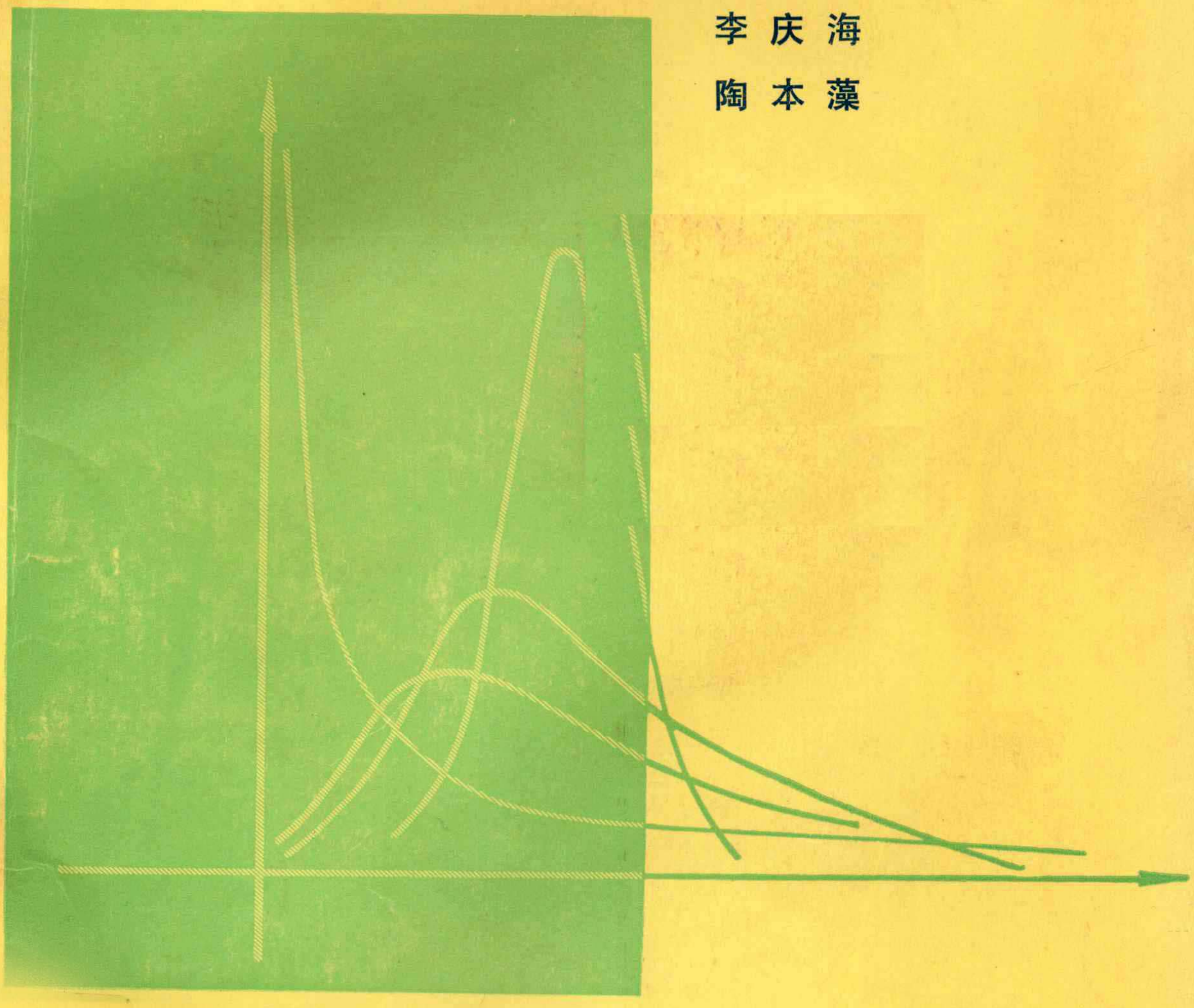
高等学校教学参考书

概率统计原理和在测量中的应用

李庆海
陶本藻



测绘出版社

高等学校教学参考书

概率统计原理和在测量中的应用

（修订本）

李庆海　陶本藻

测绘出版社

本书系统介绍了概率论和数理统计的有关理论，并详细讨论了应用于测量的参数估计和假设检验方法以及回归分析、拟合推估、卡尔曼滤波等内容。将概率统计理论与测量平差理论有机地结合起来，是本书的一个特色。

本书可作为高等院校测绘专业本科生及研究生教学用书，也适合测绘科研、工程技术人员以及其他专业有关人员参考。

高等学校教学参考书
概率统计原理和在测量中的应用
李庆海　陶本藻
*
测绘出版社出版
大厂兴源印刷厂印刷
新华书店总店科技发行所发行
*
开本 787×1092　1/16 · 印张 24 · 字数 534 千字
1982 年 6 月第一版 · 1990 年 6 月第二版 · 1990 年 6 月第三次印刷
印数 19,501—20,500册 · 定价 4.80 元
ISBN 7-5030-0322-7/P·118

第二版前言

一般认为，概率论始于十七世纪末年。当时法国贵族从事赌博，由于他们的请求，法国数学家帕斯卡（Pascal）研究了多种赌博的概率。到了十八世纪初，瑞士贝努利（Bernaulli）和法国德莫佛（De Moiver）对赌博的概率提出了统一的理论。以后概率论的潜力就逐渐为人们所认识，各国数学家研究和发展了这门学科。法国拉普拉斯（Laplace）在十九世纪初发表了他的有名的概率论著作。从二十世纪初起，英国费歇耳（Fisher）以概率论为基础，结合农业实验，发展了数理统计。其后又有多人做出贡献。从历史上看，概率论和数理统计是互相衔接的。实际上，两者之间难以划分明确的界限，故本书采用了概率统计这个词。

现在看来，概率统计已经是许多自然科学和工程技术中所必不可少的知识了。凡是需要进行实验的工作都需要用这种知识来分析其实验结果。

从十九世纪初发展起来的最小二乘法和测量平差，可以看成是概率统计侧重应用的一个分支。但把最小二乘法和概率统计结合起来进行讲解的书籍直到近年来才出现。这是一个值得鼓励的结合，可以使最小二乘法理论更为周密，内容更为丰富。本书讲到了最小二乘法的性质，所以也是这种结合讲解的书籍之一。

本版对 1982 年 6 月出版的第一版本作了适当的修改和增补，主要是突出和加强了概率统计在测量上的应用，这也是广大读者的建议。

属于概率统计原理的一些章节，如原书的一至五章、十、十一章变动极少，原书的六至八章有关假设检验的内容在本书中则归并为第八章。

作了较多的修改和增补的是本书的第六、七、九和十二、十三章。第六章是由原书分布在各章中的有关内容和增补的部分内容组成；第十二、十三章的若干节作了重写，并增补了几节新内容；这五章所述的方法应该而且有的已经用于生产。

这次修订是由陶本藻执笔的，李庆海审阅并定稿。

编著者

1985 年 3 月

目　录

第一章　数学模型和概率运算

§1-1　数学模型

用数学来分析实际问题，主要是建立一种适用的数学模型来解释实际问题，并作出结论。

在许多门自然科学中，人们从实际观察而得的事实中，取其精华，去其糟粕，提炼出其中本质的东西，用数学表达出其间的关系，将事实进行理论化，一般来讲就是建立一种数学模型。数学模型是否适用于当前的客观实际，需要在实践中验证。如果在精度许可的范围内我们经常地得到理论与实践相符合的证据，则这种数学模型是适合的。否则模型必须修改。

例如，几何学的建立过程是：从生产实践中取得了关于几何图形的经验，然后从几个简单的不加证明的公理出发，建立起一套理论，来说明点、线、面的关系，推出许多数学公式。这是一个数学模型，因为几何学上的点没有长宽厚，线没有粗度，面没有厚度，都是纯理想的、实际上没有的东西。几何学上的定理实质上是关于这个模型的定理。人们在实践中证实了这个模型可以成功地解释和分析自然世界中的事物，并可作出预断，因而这是一个与客观实际相符合的数学模型，取得了实用价值。

又例如，牛顿力学从三个运动定律和万有引力定律出发，应用数学推理的方法推出许多公式，自成一个理论体系。这也是一个数学模型，当用于解释和分析日常的、甚至一般天体的物质运动时，是与客观实际相符合的，表现了它的实用价值。用它可以预断新行星的存在，并且实际上证明了这预断是对的，例如海王星的发现。但是当人们应用这个数学模型来解释水星的运动时，便发现它不完全符合客观实际了。因而人们又建立了一种新的数学模型，即相对论力学，用它来解释速度很快的物体的运动。相对论这个模型可以看成牛顿力学的修改或推广，因为对于运动速度较慢的物体，由新旧两种数学模型所得的结果是一样的。

从这种观点来看，我们已熟悉的测量误差理论，包括误差的分类、偶然误差的特性、高斯误差定律、以及平差理论等也都是数学模型。下文中所讲的各种分布以及各种理论都需要用数学模型的观点来看它们。这些数学模型是否符合于客观实际，需要在实践中进行验证。经过实践—理论—实践的反复过程，就可以逐步提高理论，来更好地指导实践。

在本书中我们首先建立一些数学模型，说明它们的性质和可能的用处，然后举出一些在测量问题上应用的方法和实例。

我们所用到的数学模型是以事物出现的概率（即或然率、或是率）为基础的，因而在本章中需要讲述有关概率的定义和运算规则。

§1-2　随机事件的统计规律性

在自然界中有一类现象，对它我们可作精确的预断，例如在天文学知识的基础上，我们可以预断某地某年月日可以见到日蚀。这一类现象很多，都是因为我们已经掌握了这些现象的因果性规律，所以可以预知。这类事件叫做**因果性事件**。

另有一类现象，对其进行观察或实验时我们无法对结果作出精确的预断，这主要是因为影响这些现象出现的因素过多，并且每一因素所起的作用难以精确地知道，更难以精确地计算出来。纵令已尽力使实验的条件保持不变，而每次实验的结果仍然有不规则的变化，这一类现象叫做**随机事件**，实验叫做**随机实验**，随机是“偶然”的意思。测量工作中这种事件很多，例如每一偶然误差的出现就是随机事件。

但是，如果我们考虑问题的另一面，即考虑一系列观测的总的结果，情况就完全不同了：虽然单独一次的观测结果是不规则的，但多次随机实验的平均结果就呈现了规律性。

例如，我们听到过的投掷硬币的实验，当投掷次数较少时，正面在上的**频率**，即正面在上的**频数** x 除以投掷总次数 n 所得的商，变化较大；当 n 逐渐增大时，此频率的变动就渐渐小下来。

有人做过这样的实验：投掷硬币最初以 10 次作为一组，计算正面在上的频率，共进行 5 组投掷，求出最大和最小的频率。再以 100 次作为一组，也进行 5 组投掷，最后以 1000 次作为一组，也进行 5 组投掷，皆进行同样的计算，结果如下表所示。从表中可以看出，当每组的次数增多时，5 组中最大最小频率之差逐渐减小，最大和最小频率逐渐接近于 $\frac{1}{2}$。

每组中投掷次数	频率		
	最大	最小	差数
10	0.600	0.300	0.300
100	0.550	0.480	0.070
1000	0.507	0.496	0.011

这种现象指出，如果实验在相同的条件下无限制地继续下去，这个频率将趋于一个固定的、很近于 $\frac{1}{2}$ 的极限值。

测量上，随着三角形个数的增多，三角形闭合差的算术平均值的波动就逐渐减小，如此等等，都是这一类型的例子。

这就是所谓的统计规律性。由此可以给出概率的定义。

§1-3 概率的定义

正如几何学中有若干个不加证明的公理作为几何学中数学模型的基础一样，我们给概率下的定义也是属于这种性质的。

正如上节所述，如某随机实验共进行n次，其中事件A出现x次，设该实验具有统计规律性，那么，当n很大时，则频率$\frac{x}{n}$的波动很小。此时，自然会使人想到：有一个数P，可以作为频率$\frac{x}{n}$的理论“极限”值。此处的“理论”二字指的是：我们无法进行无穷多次实验，因而无法求出其可能具有的极限值。这种理论化，有些象从纸上的一条具体的直线使人想到有一条数学上没有粗度的直线。

数字P就定义为该实验中事件A的概率。观测而得的频率$\frac{x}{n}$就认为是概率P的观测值。根据前述的统计规律性，当n很大时，观测而得的频率$\frac{x}{n}$就近似地等于该事件的概率P。以下用记号P（A）来表示事件A的概率。

现在我们来讨论，在某种实验中如何求得某一事件A的概率P（A）这个问题。

在有些问题中，如果实验结果只可能为有限的n个，每一结果出现的可能性相等，并且这些结果是互斥的，即每次实验只能出现一个结果，其中能使事件A发生的结果有m个，则在这种情况下事件A的概率P（A）可用下式定义：

$$P(A)=\frac{m}{n} \tag{1}$$

通常称之为概率的**古典定义**。

例1：从0、1、2、…、9十个数字中任取一个数字，求取得小于6的数字的概率。

取一个数全部等可能并互斥的结果$n=10$，其中小于6的数字为$m=6$，按（1）得所求概率为

$$P(A)=\frac{6}{10}=\frac{3}{5}$$

例2：测量中如果不考虑误差恰为零这种少有的情况，求一次测量结果误差为正的概率。

一次测量结果的误差只可能为正或负，如果正和负的误差出现的可能性相等，而且出现了正就不会出现负，则$n=2$，$m=1$，所以正误差出现的概率为$\frac{1}{2}$。

例3：设袋中有8个球，其中5个白球3个红球，从中任意抽取4个，求恰好抽到3个白球的概率。

从8个球中任取4个的组合方法共有C_8^4种，其中每一种组合取到的可能性相等而且

互斥，抽 4 个球包含 3 个白球的可能结果是 $C_5^3 \cdot C_3^1$，故所求概率为

$$P(A)=\frac{m}{n}=\frac{C_5^3C_3^1}{C_8^4}=\frac{3}{7}$$

这种所谓概率古典定义与根据频率而下的定义是不相违背的，因为既然每一结果出现的机会相等，则根据统计规律性原理在很多次实验下，事件A出现的频率必然接近于使事件A出现各个结果的数目与结果总数之比。根据频率而下的定义称为概率的**统计定义**。

当实验结果不能只分为有限个，并且每一结果出现的机会不相等时，就不能用上述方法确定概率的数值了。测量中的问题多属于这一类，例如三角形闭合差落在（0″，1″）和（1″，2″）这两个区间内的机会就不相等。这时就需要根据多次实验中该事件出现的频率来确定（更确切地说是估计）其概率的数值，方法见后。在此只说明这个方法是比较好的方法。譬如，根据古典概率的说法，投掷硬币正面在上的概率是$\frac{1}{2}$。但这个$\frac{1}{2}$只能认为是近似的，因为任何一个硬币不能说正面和反面是完全对称的，由于两面的花纹不同以及制造上的其他原因，两面很可能有轻有重。对于某一硬币来讲，投后正面在上的概率可认为是该币固有的常数，这常数很可能与$\frac{1}{2}$稍有区别，其数值只能用多次投掷所得的频率来估计。

三角测量中三角形闭合差为负值这个事件，按古典概率定义的说法，应该是$\frac{1}{2}$。但是在国家一等三角测量各个地区的成果中负值出现的较多，如果经常发生这种现象，则古典定义只能造成与客观实际不相符合的结果。

§1–4　概率的性质和运算定则

(一)

先讨论各种事件的性质，以备应用。

1.　在某项实验中每次一定发生的事件称为**必然事件**。

2.　实验中永不会出现的事件称为**不可能事件**。

3.　设有A和B两事件，在每次实验中A和B可能出现，也可能不出现。A和B至少出现一个（也可能两个皆出现）这样一个综合事件称为A和B的**事件和**，记为$A+B$。

4.　A和B两事件皆出现这个综合事件称为A和B的**事件积**，记为AB。

5.　如两事件A和B不可能同时出现，则称A和B为**互斥事件**或**互不相容事件**。

6.　如$A+B$为必然事件，并且A和B互斥，这就是说A出现则B必不出现，且A和B必出现其一，则称A和B互为**逆事件**。一般用$\overline{A}$表示A的逆事件。

如将事件A、$\overline{A}$、$A+B$、AB用图形表示出来，则如图 1-1。在图中A和B分别用圆圈表示，长方形表示必然的事件，画有斜线的部分为欲表示之事件。

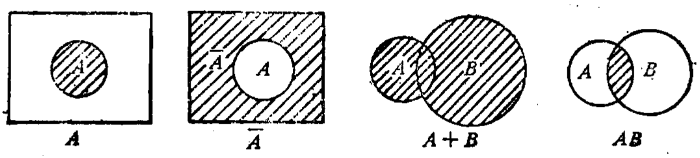


图 1-1

(二)

根据上节中概率的定义，可指出概率的几个性质如下：

1. 设在n次实验中事件A共出现x次，则显然$0\leqslant x\leqslant n$，由此得$0\leqslant\frac{x}{n}\leqslant 1$。既然频率$\frac{x}{n}$的理论极限值为事件$A$的概率$P(A)$，自然应要求$P(A)$满足下列条件：

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

2. 如A为必然事件，则$\frac{x}{n}$必等于1。因此可知：必然事件的概率永为1。

3. 如A为不可能事件，则$\frac{x}{n}$必恒为零。由此可知：不可能事件的概率恒为零。

4. 设将某项实验重复进行n次，现在考察两事件A和B出现的次数。每次实验中只能产生下列四种情况之一：

A出现而B不出现，设共有a次；

B出现而A不出现，设共有b次；

A和B同时出现，设共有c次；

A和B皆不出现，设共有d次。

自然，$a+b+c+d=n$。由此得出：

事件A出现的频率$f_A=\frac{1}{n}(a+c)$；

事件B出现的频率$f_B=\frac{1}{n}(b+c)$；

事件$A+B$出现的频率$f_{A+B}=\frac{1}{n}(a+b+c)$；

事件AB出现的频率$f_{AB}=\frac{c}{n}$。

由此可以得下列关系：

$$f_{A+B}=f_A+f_B-f_{AB}$$

再由此可得相应概率的关系式：

$$P(A+B)=P(A)+P(B)-P(AB) \tag{1}$$

如果A和B两事件是互斥的，则它们不可能同时出现，此时f_{AB}和$P(AB)$皆为零，