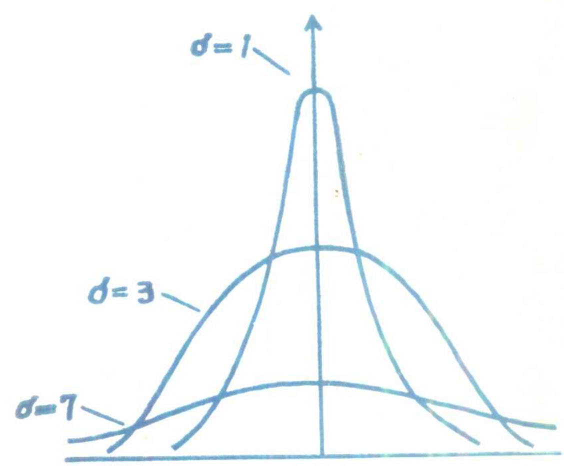
海洋科学研究中的

概率统计方法

陈敦隆 著



海洋出版社

海洋科学研究中的概率统计方法

陈 敦 隆 著

海 洋 出 版 社
1982年·北京

内 容 简 介

本书比较系统地阐述概率统计方法的基本原理及其在海洋科学研究中的广泛应用．书中首先通过许多应用实例引进概率统计的基本概念；然后重点讨论海洋调查研究中的抽样方法，“多年一遇”海洋水文气象要素的计算和水文气象、生物、地质等调查资料的多元统计分析的方法；最后一章以专题形式探索了寻求刺网相对渔获率曲线的方法，划分水团边界的概率统计方法，混合浪波高的统计分析，生物种群数量动态和空间格局的简单随机模型和评价水质的生物数学公式等．

本书可供海洋、水产、环境保护工作者、应用数学工作者以及大专院校有关专业师生参考．

海洋科学研究中的概率统计方法

陈敦隆 著

*

海洋出版社出版

(北京复兴门海贸大楼)

国防科工委印刷厂印刷

新华书店北京发行所发行　　各地新华书店经售

*

1982年12月第1版　　1982年12月第1次印刷

开本：787×1092 1/16　　印张：34¼，插页：1

字数：800,000　　印数：1—1,500

统一书号：13193·0071　　定价：7.10元

序　言

海洋科学研究包含着大量的试验和观察，这些试验和观察为我们认识海洋中各种研究对象的内在规律、揭示它们之间的本质联系和预测其未来的可能发展，提供了丰富的感性材料和观测数据．然而，如所周知，所有的海洋现象和过程，都是在各种各样的环境因素作用下产生和发展的．在每一次试验或观察中，总会有一些无法观测或意料之外的因素在起作用，从而使得人们在海上的现场试验和观察中得到的结果带有一定的随机性；即使在实验室内模拟海洋的一些“因果性”试验中，无论人们怎样严格控制预定的试验条件，但各次试验的结细也不可避免地与理论预测有所偏差．因此，要想从这样一些现场的或实验室内的试验和观察结果中找到有规律性的东西，作为可靠的结论，就必须大量重复试验和观察，并对所得到的结果进行细致深入的整理和分析．在这方面，概率论和数理统计这一数学学科，为我们提供了强有力的工具．它的一些处理技巧和计算方法可以帮助我们对大量的实验观测数据进行去伪存真、去粗取精的系统整理和科学分析，从而便于揭示大量试验和观察中所呈现出来的固有规律性（或即统计规律性）、抓住问题的关键，使问题得到更好的解决．

事实上，概率论和数理统计在海洋科学中的应用并不限于海洋试验观测数据的分析处理．近十余年来，概率论和数理统计的思想和方法已逐步渗透到海洋科学的所有分支学科中去，卓有成效地补充或取代过去常用的一些研究确定性现象的数学工具，对海洋中的一些现象、过程或系统构造概率统计模型，在电子计算机上进行模拟求解．这一发展趋势值得我们注意．

本书著者陈敦隆同志是已故著名数学家许宝騄教授和郑曾同教授的学生，多年来在山东海洋学院从事概率论和数理统计的教学和一些海洋科研工作，积累了一定的经验，这次又广泛搜集和整理有关资料，结合他自己的体会，著成这本海洋科学研究中常用概率统计方法一书．在我看来，这本书与一般的概率统计书籍比较，有它的特点，例如，本书从概率论的基本原理出发，在讲清基本知识的基础上，密切结合海洋观测中的一些具体问题和例子，比较系统而具体地介绍了常用的数理统计方法及其在海洋科学研究中某些专题的应用，应当说，这是我国广大海洋科技工作者当前急需的一本参考书，当然，也应该指出，目前以广大海洋科技工作者为读者对象撰著这样性质的一本参考书毕竟还是一种初次尝试，加以篇幅有限有些问题的研究也还不很成熟，有待提高，因此，要求本书对概率论和数理统计方法（特别是其中的随机过程论的分析方法）在海洋科学中的应用作全面的介绍是不现实的；缺点和不足之处恐亦在所难免．这些都有待我们广大的海洋科技工作者共同努力，不断予以指正和补充，作为一个海洋科学工作者，我对本书的出版表示欢迎，并寄予良好的愿望．

赫崇本

于青岛山东海洋学院

1980．5．10

目　　录

第一章 引 论

在海洋水文气象观测中，广泛地存在一类所谓随机事件．例如，在海面上观测风浪，由于风浪的波高和许多因素有关，譬如，产生风浪的风的大小和方向不论在空间和时间上都在不断地变化，紧贴水面的风速结构十分复杂，风浪产生后，波面的运动对水面风速又有很大影响．这些因素的影响是每时每刻变化的，因此在海面某一固定位置上所观察到的波高也就逐时逐刻不一样．即使观测时的基本条件保持不变，但由于影响波高的次要因素为数众多，有的至今还没有得到准确的了解，再加上观察时不可避免地产生测量的误差，因此，虽然每次观察后可以得到一个波高值（实际上是指在一小段时间间隔内所观测到的波高值），但是无论进行多少次观测，我们始终不能由某一时刻出现在某一位置的波高大小准确预测另一时刻出现于该点的波高大小．这就是说波高是一个随机变量，而“波高出现于某一个指定范围内”便是一个随机事件．同样，在观察海雾时，某地区4—7月份间雾日总和也是一个随机变量，而“雾日总和等于某个指定数”也是一个随机事件．这样的例子不胜枚举．

由上述随机事件的概念，人们也许会产生这样的印象，以为对随机事件无法进行研究，其实不然，任何复杂的事物总是有规律的．实践表明，尽管随机事件在单次观测中无规律可循，但是，如果观察的次数足够多，便可发现它具有某种规律性．例如，掷一个匀质的五分硬币，由于重力的作用，硬币必将落下．因此硬币下坠不是随机事件，而是必然事件．但是，如果我们关心的是“掷硬币出现国徽一面向上”这一事件（以下简称为“正面”），那末，掷出“正面”便是一个随机事件．在投掷硬币的这个随机试验中，虽然我们不能事先确切预言每次试验的结果．但是，在同一条件下反复进行很多次投掷之后，就会发现出现“正面”的相对频率近于1/2．例如把硬币投掷一万次，统计结果大概有五千次得“正面”．这就是说，尽管随机事件在个别试验中具有偶然性或称随机性，但是，大量重复的试验却具有频率的稳定性．这就是所谓的随机事件的统计规律性，或简称为统计律．利用这个统计律人们就可以预言：“下一次掷出正面”的可能性是1/2．显然，这里不是对个别随机事件作出绝对肯定或绝对否定的预测，而是根据统计律作出的一种客观估计．

海洋水文气象现象给我们提供了许多随机事件具有频率稳定性的例子．例如，于波面的固定点上观测波高，当观察的次数不多时，其所得波高值大小具备，杂乱无章，没有显著的规律性．但是，当观测的次数不断增加时，并加以适当的资料整理（见§2.4），就可发现波高出现于各个范围内有着稳定的频率；特大或特小的波所占的百分比是很小的；较大或较小的波占有一定的百分比；波高位于整个范围中央部分附近的波的百分比最大．这和我们在海上或海滨所看到的情况是相符的．在本书的第二章中将要证明波高的分布服从瑞利(Rayleigh)分布．

由上所述，大量同类的随机事件总是有统计律的．概率论与数理统计是数学的一个分支，它是一门从数量上研究随机事件统计律的学科．在海洋水文气象学中，应用概率论的方法不仅从理论上揭示了大量同类的随机事件所具有的统计律（见§4.3—§4.4），而且还为有关的实际问题的计算提供理论基础．而数理统计学的方法则是根据实测资料从数量上对随机

事件的统计律进行估计与推断. 总之，概率论（包括随机过程）和数理统计（包括多元分析）一起提供了分析和描述海洋水文气象学中的随机变量和随机过程的重要工具.

应该注意到，用概率统计方法得到的统计律在本质上只适用于大量随机事件的研究，即只能用它来预报大量性质相同的随机事件的平均结果，它不能用来预报单次观测中随机事件的结果. 例如，虽然确定了某港湾内百年一遇的波高，但是我们并不知道它是否在来年出现，或是在百年后才出现，而只能说，在长期观测序列中，这种波高平均每一百年遇上一次.

应该指出：必然事件与偶然事件之间并没有不可逾越的鸿沟. 事实上，在自然界中，存在着更多的同时具有必然性和偶然性的事件. 例如，在入海口处围海造田，必须首先考虑由此引起的一系列问题，如对航运和排洪等是否有影响. 为此，应该计算围海造田后港内的水位与流量以及讯期迳流量等问题. 为了讨论简单起见，仅以计算港内流量为例，人们根据流体力学的知识可以建立起描述流量的流体力学方程，并根据边界条件和初值条件求解此方程（一般要借助电子计算机），从而得到流量随时间和位置而变化的确定性规律. 特别当位置固定之后，它就是一个依赖于时间的一元函数. 以下简称为理论流量曲线. 它近似地反映了固定测点上流量随时间变化的规律，是一种抽象的数学模式. 从理论上讲，如果在绝对不变的边界条件和初值条件下，于固定点进行流量观测，则应该得到完全相同的理论流量曲线. 但是，实际观测表明：即使在相同的基本条件下，于同一测点上得到的不是同一条理论流量曲线，而是一条与理论流量曲线多少有点偏离的实际流量曲线. 这也就是说观测结果具有偶然性，但这并不否定流量的流体力学方程所反映的因果性. 因为只要对每一次观测得到的实际流量曲线加以分析研究，总是可以找到偏离的原因. 另一方面，对于同一基本条件下得到的一族实际流量曲线也具有某种规律性：实际流量曲线的分布大体上散布在理论流量曲线附近，离理论流量曲线越近越密；越远越稀. 在水文学中，一般认为流量的分布服从对数正态分布（见§4.1）.

这个例子说明，客观现实中并没有什么“无因果性的随机现象”，也没有什么“无偶然性的因果现象”. 恩格斯在《自然辩证法》中早就明确地指出：“偶然的东西是必然的，而必然的东西又是偶然的”[1]. 因此，人们在研究客观事物时，要根据具体情况，从因果律和统计律这两个侧面中的某一个侧面进行考察. 有时则要把两个侧面结合起来进行考虑. 数学作为“辩证的辅助工具和表现方式”[2]，是从数量上把客观事物的一个方面抽象化的一门科学. 微分方程等数学分支就是因果律在数学上的一种表现方式，而另一个分支概率论和数理统计则是以数量的形式对大量偶然性现象所进行的研究，它们所反映的因果律和统计律，都是对体现了必然性和偶然性辩证统一的客观事物的一种简化和近似，都只反映了对客观事物的一个侧面（必然性或偶然性）的认识. 显然，偶然性和必然性的矛盾，也存在于认识的每一层次. 因此，随着人们对客观事物认识的逐步深化，在某一范围内，可能用因果律去取代统计律，而在另一范围内，则需要求助于统计律. 或甚至在更新的方向上进行探索. 近十余年来，概率统计的思想和方法正逐渐渗透到各个不同的数学分支和其它科学技术领域中去，补充或取代原先常用的一些经典性的研究方法，对研究的对象和过程进行数学描述和分析. 概率论和数

1) 恩格斯著：《自然辩证法》，人民出版社，1971年版，第195页.
2) 恩格斯著：《自然辩证法》，人民出版社，1971年版，第3页.

理统计这个数学工具的重要性也日益为更多的人所认识．甚至，有人认为，**在一切知识领域内，广泛而有效地运用概率统计方法，是自然科学和技术科学的现阶段的特征之一**．

在海洋水文气象学的研究中，情形也是如此，例如，寒潮过程造成的海面降温是由很多因素共同影响的．由于这些因素起作用的程度不同和它们相互配合的差异，使得每次寒潮降温以或强或弱、或急或缓这样一些带有偶然性的形式表现出来．这也就是说，尽管冷平流和海面热散失等物理过程使海面降温具有必然性，但每次寒潮降温过程却有很大的偶然性．正是由于海洋水文气象过程总包含着必然性和偶然性的矛盾，在对这些过程预报实践中，相应地也就出现了基于因果律的“数值预报方法”和“统计预报方法”．而近年来，这两类方法正向着互相渗透的方向发展，并以各种方式结合成了“统计动力学方法”．这一事实表明，在海洋水文气象学的研究中，将有新的更高层次的出现．例如，在海浪研究中，虽然运用随机方法日益广泛，但是今日的主要趋势是把统计方法和流体力学方法结合起来；在研究水位时，因为任何实际水位都包含确定性部分（即所谓天文潮）和随机性部分（即所谓非周期水位）．因此，为了更准确地预报水位，客观上就要求把这两部分结合起来进行研究，事实上，今日的潮汐研究工作正朝着这个方向前进；气象学的研究也不例外，近二十年来许多气象工作者关心统计动力相结合，设计了一些新的预报方案即所谓统计动力模式或称随机预报模式，设计这些模式的基本观点主要是认为大气运动规律是由流体力学、热力学所描写的基本规律所控制的，但大气的初始状态和未来将处于何种可能状态都是随机的，预报量的最可能值是期望值．

概率论和数理统计学的历史说明了理论和应用的相互促进作用："理论的进展开辟了应用的新领域．反过来每一个新的应用又产生出新的理论问题和有成果的研究。"[1] 完全可以预料，随着我国社会主义建设事业的发展，以及海洋科学技术的现代化，概率统计方法在海洋水文气象学中将会得到愈来愈多的应用，与此同时，也必将推动概率和统计方法的进一步发展．

1) Feller, W., *An Introduction to Probability Theory & Its, Applications*, Vol. Ⅰ, 2nd ed. Wiley, N. Y., 1957.（W.费勒著，胡迪鹤、林向清译，概率论及其应用，科学出版社，1964年）．

第二章　概率论的基本概念

任何一门研究客观事物某一个侧面规律性的学科，都有它的一系列的基本概念．凭借这些基本概念就可以建立起各门学科的一般理论．概率论中也有这种基本概念．它们是随机事件和表征随机事件出现可能性大小的数值测度——随机事件的概率，以及随机变量和描述随机变量特征的方法——随机变量的分布律．本章的目的在于通过一些实例和简单的数学演绎，对这些基本概念作一简单的介绍．

2.1　随　机　事　件

基本事件　随机事件是概率论中的原始基本概念，是不能严格定义的，而只能加以解释．在**同样的条件下，重复实验或观测中有时出现有时不出现的事件称为随机事件**，或**简称事件**．

对于任一随机事件，总可把它和某个随机试验联系起来．也就是说，我们常常通过随机试验来观察随机事件．一个试验，如果它的结果有许多个，而且在相同条件下不断地重复做下去，就称为随机试验．或简称试验，记为 E；它的任何一个可能出现的结果，称为基本事件，记为 ω；它的全体基本事件构成的集合 $\Omega=\{\omega\}$，称为试验 E 的基本事件空间．在具体问题中，十分重要的是认清基本事件空间是由什么构成的．下面，举一些例子．

例2.1　E——掷一硬币，观察所出现的正反面．这里共有两个基本事件：ω_1——正面，ω_2——反面．于是，Ω 由两个基本事件构成，即 $\Omega=\{\omega_1,\omega_2\}$．

例 2.2　观测某海域 5 月 1 日上午 8 时的波级，如按现行海洋观测规范，海面波况分为10级（见表 2.1），则 ω_i——波况为 i 级，$i=0,1,2,\cdots,10$．于是 $\Omega=\{\omega_0,\omega_1,\omega_2,\cdots,\omega_{10}\}$．

例2.3　E——观察某海区在 9 月间台风过境的次数．如用 ω_i 表示观察结果得到的次数，$i=0,1,2,3,\cdots$，则 $\Omega=\{\omega_1,\omega_2,\omega_3,\cdots\}$．

例2.4　E——观察某测波站在 10 月间的最大波高．如用 ω_a 表示最大波高为 a 米，$0\leqslant a<\infty$，则 $\Omega=\{\omega_a;\ 0\leqslant a<\infty\}$．

值得注意的是，在例 2.4 里，我们对最大波浪的上限并不作任何限制，因为，如果我们一旦指定一个上限 H 米，那就把最大波高超过 H 米的可能性排除了．事实上，说波高能达到H米而不能达到H米加 1 毫米，这绝不会比无限波高的说法更站得住脚．其实，从概率统计计算的角度看来，承认区间$[0,\infty)$ 中的每一个值是观察波高的可能结果，不但不会影响使用，而且还可简化波高的分布律（见第三章）．

复合事件　在例 2.3 里，我们还可看出，“台风过境的次数至多是五次”，也是一个事件，它是由许多基本事件 $\omega_i(i=0,1,2,3,4,5)$ 所构成的．因此，它是一个复合事件．如果把这个基本事件记为A，则 $A=\{\omega_i;\ i=1,2,3,4,5\}$．由此可见，基本事件是事件中最简单的一种．一般的事件总是由某些基本事件共同组成的．为了说明这个问题，我们先来看一下我国现行的波级表．

表2.1

波级	名称	波高(米)	波级	名称	波高(米)
0	无浪	—	5	大浪	$2.0<H<3.5$
1	微浪	$H<0.3$	6	巨浪	$3.5<H<6.1$
2	小浪	$0.3<H<0.8$	7	狂浪	$6.1<H<8.6$
3	轻浪	$0.8<H<1.3$	8	狂涛	$8.6<H<11.0$
4	中浪	$1.3<H<2.0$	9	怒涛	$H>11.0$

在上表中我们把目测海浪资料分为十级．因此在观察海浪时，其所有的基本事件便是0，1，2，…，9．在实际工作中，我们除了关心这些基本事件以外，还要关心像波级｛至少5级｝或｛2到4级｝这样的事件．容易看出，｛至少5级｝这个事件可分解为5，6，7，8，9五个基本事件．同样地，｛2到4级｝这一事件可分解为2，3，4三个基本事件．为了区别起见，我们把这些可分解的事件称为复合事件．

今后，为了统一术语，我们把基本事件称为**样本点**，而把样本点的全体称为样本空间，**记为** Ω．

从集合论的角度看来，所谓复合事件就是包含一个样本点以上的集合，而基本事件就是仅包含一个样本点的集合．

综上所述，我们从一个随机试验 E 出发，得到它的基本事件空间，或即样本空间 $\Omega=(\omega)$；这个试验可能出现的事件（包括基本事件和复合事件），不是别的，无非是 Ω 的一些子集合而已．

为了方便起见，再将不包含任何样本点的集合（简称空集）也视为事件，并称之为不可能事件，常用的记号是 V．同样也把包含所有样本点的集合视为事件，并称之为必然事件，常用的记号是 U．例如，作一个掷骰子的试验[1]，｛骰子出现的点数大于6｝，｛掷硬币既不出现正面，又不出现反面｝等都是不可能事件．而｛掷骰子出现的点数小于或等于6｝，｛掷硬币出现正面或反面｝等都是必然事件．

必须强调指出，概率统计计算中，任何一个观测或实验总是联系着一个给定的样本空间的．因此，离开样本空间去单独地谈论事件是没有确切意义的．

事件间的关系　当我们谈到许多事件和它们之间的关系时，除非特别声明，指的都是同一个随机试验的事件，也就是说，它们都是同一样本空间的子集合．讨论事件之间的关系用逻辑符号运算（或称集合代数）是卓有成效的．下面设 $A,B,C,\cdots$ 都是事件，并且只介绍事件间的两个主要关系，即事件的和与事件的积．

1．事件的积（或简称积事件）．

由“事件 A 和 B 同时出现”所组成的是**事件的积**，记为

$$AB\ (\text{或}\ A\cap B).$$

从集合论的角度看来，事件 A 和 B 的积 AB 就是由属于 A 和 B 的样本点的全体所构成的集合．

例如，在掷骰子的试验中，样本空间 $\Omega=\{1,2,3,4,5,6\}$，令 A 表示｛得偶数点｝这一事

1）骰子是一个正立方体，六面分刻着一点到六点．

件，B 表示{至少得 3 点}这一事件，即

$$A=\{2,4,6\},\quad B=\{3,4,5,6\},$$

则

$$AB=\{4,6\}.$$

又例如在表 2.1 中，令

$$A=\{\text{1 到 5 级}\},\ B=\{\text{3 到 7 级}\},$$

则

$$AB=\{\text{3 到 5 级}\}.$$

再例如，令

$$A=\{\text{台风}\},\ B=\{\text{雷}\},\ C=\{\text{暴雨}\},$$

则

$$ABC=\{\text{台风雷暴雨}\}.$$

在事件的积中，如果考虑到事件 A 和 B 不能同时出现这一特殊情况，即 AB 的积是不可能事件，用记号表示就是

$$AB=V,$$

于是，称事件 A 和 B 为**互斥事件**.

例如，在掷骰子的实验中，令

$$A=\{1,2\},\quad B=\{3,4,5\}$$

则事件 A 和 B 是互斥事件.

又例如，在波级表中，令

$$A=\{\text{3 级}\},\quad B=\{\text{5 级}\},$$

则事件 A 和 B 也是互斥事件.

十分明显，样本空间中的任何两个基本事件都是互斥事件.

2. 事件的和（或简称和事件）.

由"事件 A 和 B 中至少有一出现"所组成的事件称为**事件的和，记为** $A+B$（或 $A\cup B$）.

从集合论的角度看来，$A+B$ 就是至少由属于 A 或 B 之一样本点的全体所组成的集合.

例如，在掷骰子的实验中，令

$$A=\{2,4,6\},\quad B=\{3,4,5,6\},$$

则

$$A+B=\{2,3,4,5,6\}.$$

又例如，在由三次风暴潮所组成的预报实验中，样本空间 $\Omega=\{A_0,A_1,A_2,A_3\}$，此处，$A_i$ 表示{报对 i 次}这一事件，$i=0,1,2,3$，

则

$$A_1+A_2+A_3=\{A_1,A_2,A_3\}.$$

表示{至少报对一次}的事件，而

$$A_0+A_1+A_2=\{A_0,A_1,A_2\}.$$

表示{至多报对二次}的事件.

在事件的和与事件的积中，如果考虑到事件 A 和 B 不能同时出现，但一定出现其中之

一的特殊情况，用事件的和与积的记号来表示，就是

$$AB=V \quad \text{且} \quad A+B=U \tag{2-1}$$

就称事件 A 与事件 B **互逆**，或者说 A 是 B（或 B 是 A）的**对立事件**. 习惯上，我们把 A的对立事件记为 $\bar{A}$.

从集合论的角度看来，$\bar{A}$ 就是 A 的补集.

例如，在掷骰子的实验中，令 $A=$｛得偶数点｝，则 $\bar{A}=$｛得奇数点｝.

又例如，在观察海雾时，｛有雾｝和｛无雾｝也是对立事件.

2.2 事件的概率

概率的概念 随机事件作为一个个别现象来说，要么发生，要么不发生，似乎是捉摸不定无规律可循. 但是，从大量的随机试验中，我们就会发现随机事件的出现量具有一定的规律性，有些事件出现的可能性较大，而另一些事件出现的可能性较小. 例如，｛青岛海面最低温度不低于-15℃｝的可能性就比｛低于-15℃｝的可能性来得大等等. 实践证明，事件发生的可能性的大小不是人的"主观信念"更不是"上帝的旨意"，而是事件本身所固有的，不随人们主观意志而改变的一种客观属性. 并且是可以通过大量重复试验来认识的. 因此，对于一个随机事件 A，人们可以用一个数 $P\{A\}$ 来描述它发生的可能性的大小，并称 $P\{A\}$ 为事件 A 的**概率**（又称**几率、或然率**）.

显然，这个概率的定义是直观的、定性的，它只说明了概率的作用，而没有给出概率的具体数值. 那末，我们该怎样合理地确定 $P\{A\}$ 的值呢？这个问题，需要根据实验的条件作具体的分析. 下面，我们将根据概率定义的形成和发展过程，简单地介绍几种常用的定义概率的方法.

一、古典概率的定义

有一些实验，它的各个基本事件是具有对称性的，而基于这种对称性可知，各个基本事件的发生具有等可能性. 因此，可以根据实验的条件直接算出各个基本事件的概率.

概括地说，设随机试验 E 只有有限个（n 个）基本事件

$$\Omega=(\omega_1,\omega_2,\cdots,\omega_n)$$

而且每个基本事件都处于对称（或平等）的地位，即任何一个基本事件都不会比别的基本事件更特别一些. 这时，合理地定义

$$P\{\omega_i\}=\frac{1}{n} \tag{2-2}$$

或者更一般地，定义

$$P\{A\}=\frac{k}{n}. \tag{2-3}$$

这里 k 是事件 A 中所含基本事件的个数($k\leqslant n$).

例如，作一个向上空抛一镍币的实验，由于镍币的两面具有对称性，所以，实验的所有基本事件（即｛正面｝和｛反面｝的概率是相等的，各等于$\frac{1}{2}$. 对于匀质的镍币来说，这种概率定义是经得起实践的考验的.

又例如，掷一个对称的骰子，令 $A=\{$掷骰子得偶数点$\}$，则因所有基本事件个数为6，而事件 A 所包含的基本事件个数为 3，故得$P\{A\}=\frac{3}{6}=\frac{1}{2}$.

实验的基本事件的对称性，通常只有在组织得很巧妙的实验（如随机博奕实验）上才能得到保证. 由于这种基于等可能性随机试验的概率定义产生于概率论的初期发展阶段，故有**古典概率**之称.

对于海洋水文气象学中的随机现象来说，由于不存在这种理想条件，所以，一般不采用这种古典概率的定义.

二、几何概率的定义

还在概率论开始发展的时候，人们就已经注意到建立在只考虑有限个等可能性基本事件组上的古典概率定义是不够用的. 这方面的例子很多，其中最突出的是十八世纪末和十九世纪初期所提出的射击问题. 举一个例子来说，假定我们要计算子弹命中靶面上某一环的概率，在这个问题中，即使我们可以勉强地认为子弹命中靶面上每一点的概率是相同的（即子弹打中靶面上每一点的可能性几乎是相同的），但由于射击时的所有基本事件的个数不是有限的，因此，这就使得古典概率定义无法给予计算. 为了解决像射击一类问题，就要求扩大概率的定义使其在无限多个几乎等可能性的基本事件的情形下亦适用. 不难看出，在射击中，当靶面上所指定环的面积越大时，子弹命中该环的概率亦越大. 因此，可以引进下面的几何概率定义：

如果在区域 G 中任投一质点，并且质点落在区域 G 内任何一个位置上是等可能事件，则该点落在区域 G 内某一指定小区域 g 内的概率 p 等于

$$p=\frac{g\text{的面积}}{G\text{的面积}} \tag{2-4}$$

同样，在线段 L 内任投一质点，并且质点落在线段 L 内任一位置上是等可能事件. 则该点落在线段 L 内的某一个小区间 l 内的概率 p 等于

$$p=\frac{l\text{ 的长度}}{L\text{的长度}}. \tag{2-5}$$

在上述的几何概率定义中，如果把质点的随机命中换为一个十分小的区域 $\Delta G(<g)$ 落在 G 内，或换为一个十分小的区间 $\Delta L(<l)$ 落在 L 内. 在这种情况下，ΔG 落在 g 内的概率是多少？同样，ΔL 落在 l 内的概率是多少？这类问题的计算要比前面的情形复杂得多.因为这时的概率计算不仅要取决于 G 和 g 的面积或 L 与 l 的长度，且与 ΔG 或 Δl 的大小、形状和位置有关. 今以著名的蒲丰(Buffon)问题为例.

例2.5 平面上画着一些平行线，每两条之间的距离等于$2a$，在平面上任意投一枚针，其长度为 $2l(l\leqslant a)$. 试求这针与任一直线交叉的概率.

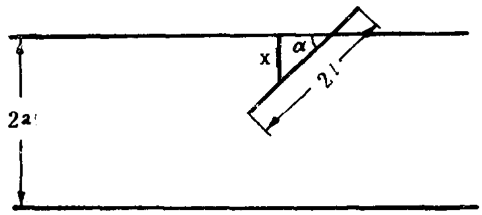


图 2.1

以 x 表示由针的中点到最近一条平行线的距离并且以 φ 表示针与此平行线所成的角（见图2.1）x 与 φ 两个数完全决定针的位置. 针的所有可能位置由以 a 和 π 为边的矩形的点子

所决定．由图 2.1 可以看出，要针与平行线相交，其必要而充分的条件是

$$x \leqslant l \sin\varphi$$

所求的概率由所作假设应等于图 2.2 中阴影区域的面积与矩形面积之比：

$$p=\frac{1}{a\pi}\int_0^{\pi} l\sin\varphi d\varphi=\frac{2l}{a\pi}. \tag{2-6}$$

蒲丰问题为人们提供了一个意味深长的模拟试验．这个实验就是画出间隔为 $2a$ 的两条平行线，让长度为 $2l$ 的针随意地落下，然后从这枚针与平行线中任一条相交的次数来计算 π 值．如果在 n 次投针实验中，有 m 次与平行线之一相交，则根据统计概率定义可以从下式求出 π 的近似值

$$\frac{m}{n}=\frac{2l}{a\pi}. \tag{2-7}$$

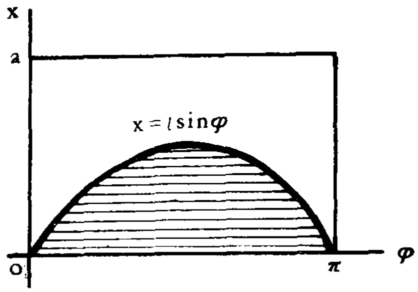


图 2.2

特别当 $l=a$ 时，有

$$\frac{m}{n}=\frac{2}{\pi}$$

或

$$\pi=\frac{2n}{m}.$$

1850 年，沃尔夫(Wolf)通过实验，曾得出 $\pi=3.1596$．

这样的实验，随着运筹学的发展以及快速电子计算机的诞生而逐渐被人们所接受．时至今日通过试验求数学解的这种模拟法（也称蒙特卡罗（Mont Carlo)法）已成为主流了．

在几何概率定义中，因其对“任意”或“随机”之辞没有做过解释，因此，即使是同一问题，也可能因任意的含义不同而有所不同．其中表现得最鲜明的应推十八世纪的贝特朗(Bertrand)氏奇论．今举一例如下：

例2.6 在半径为 r 的圆内任意作一弦，试求这条弦的长度大于该圆的内接正三角形边长的概率是多少？

对于这个问题可以有以下三种解法：

1． 由对称的道理可以预先指定弦的方向．我们画一条直径垂直于这方向．显然，只有交直径于四分之一分点及四分之三分点之间的弦之长才超过正三角形的边长，所以

$$P\{\text{弦的长度大于内接正三角形边长}\}=\frac{1}{2}.$$

2． 由于对称的道理可以预先把弦的一端固定在圆周上．在这点上的切线与以这点为顶点的正三角形的两边各成 60° 的角．显然只有落在中间角里的弦的长度才大于该圆内接正三角形的边长，所以

$$P\{\text{弦的长度大于内接正三角形边长}\}=\frac{1}{3}.$$

3． 要决定一条弦的位置，只要指定它的中点就够了．要使一条弦的长度大于内接正三

角形的边长必须是其中点落在与所给的圆同心而且半径为其一半的圆内，这个圆的面积等于所给半径为 r 的圆的面积的$\frac{1}{4}$. 所以

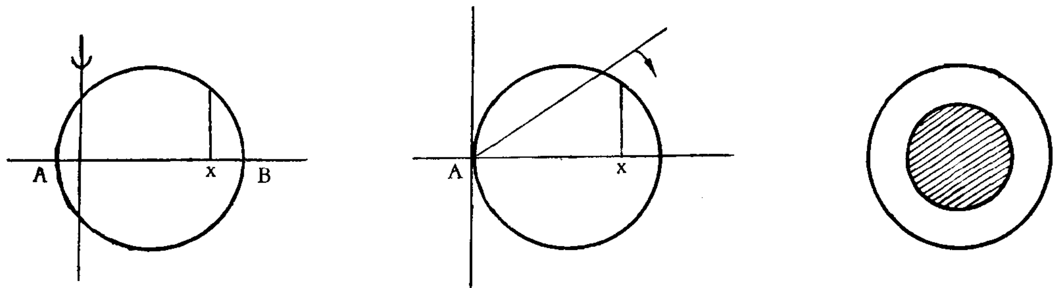
$$P\{\text{弦的长度大于内接正三角形边长}\}=\frac{1}{4}.$$

为什么同一个问题的解不统一呢？原因在于问题的条件中未给“任意作弦”这概念下定义，因此把三个不同问题的解混充作同一问题的解了. 事实上，这三种解法有根本的差别.

(1)在第一个解法里，可以设想为一条圆柱形轴沿一条直径滚动（见图 2.3a）. 这根轴的所有可能停止位置的全体就是直径上的点子的全体，而轴停在直径内任何位置上是等可能事件；

(2)在第二个解里，我们可以设想把轴用铰链钉在圆周一点上而使它摆动，其摆幅不大于180°（见图 2.3b）在此假设轴的停止在长度为 h 的圆弧内这一事件的概率只与弧长有关而与其位置无关. 如此，轴的停止在任何同长的圆弧内被看作是等可能事件；

(3)在第三个解法里，我们在圆内任投一个点，而所要求的是该点落在半径为$\frac{r}{2}$的同心圆内的概率（见图 2.3c）.



(a) (b) (c)

图 2.3

在水文气象学中，应用几何概率可以解决许多实际问题. 例如，计算一个地区被冰雹打中的概率；在指定寒潮的期间内某港口结冰的概率，又例如，计算某港口由风暴引起的最大水位的概率分布[2].

三、统计概率定义

随着生产的发展，自然科学和工程技术中所提出的问题，使建立在等可能性基础上的古典概率定义遭到了不可克服的困难. 这方面的例子是很多的. 譬如决定生男生女的概率和计算港口内最高水位超过警戒限的概率等等，都遇到区分等可能性的困难. 为了解决这一类问题，就要求大幅度地推广概率的定义. 人们在实际工作中（例如人口性别统计和水文气象资料的整理中）总结出事件的一个重要的属性——随着试验或观察次数的增多，事件出现的频率越来越接近于某一个固定常数 p（这里 $0\leqslant p<1$）. 根据随机事件的这种属性，自然可以引入下述概率定义：

设 E 为任一随机试验，A 为其中任一事件. 在同样的条件下，把 E 独立地重复作 n 次（这里，可重复性是随机试验的定义中规定的）. 设事件 A 在这 n 次试验中出现了 m 次，称比值 m/n 为 A 出现的频率，而 m 为出现的频数. 显然，$m<n$. 在一般情形下，如果 n 充分大，

频率 m/n 稳定在某个数 $P\{A\}$ 的周围，
即

$$\frac{m}{n}\approx P\{A\}, \tag{2-8}$$

这时，称 $P\{A\}$ 为 A 的概率.

上述的概率定义称为统计概率定义.

由此可见，在水文气象学的许多实际问题中，当事件的概率不易求出时，我们总可以通过频率来找出这一事件客观存在的概率. 在这里，引入频率的重大意义在于一方面它能适当地反映事件A出现可能性的大小，另一方面，频率的概念比较直观，容易掌握，我们常常可以根据频率的性质去推想概率的性质. 因此，概率的统计定义给我们提供了一个广泛应用的工具.

根据频率的定义，显然可知，它是依赖于试验次数 n 的，我们把它记为 $f_n\{A\}=\frac{m}{n}$，并且，它具有如下性质：

1. $0\leqslant f_n\{A\}\leqslant 1$； (2-9)
2. $f_n\{\Omega\}=1$； (2-10)
3. 若 $A_1,A_2,\cdots$是两两互斥的事件，

则

$$f_n\left\{\bigcup_{i=1}^{\infty}A_i\right\}=\sum_{i=1}^{\infty}f_n\{A_i\}, \tag{2-11}$$

这里，$\bigcup\limits_{i=1}^{\infty}A_i=A_1+A_2+\cdots$，即事件 $A_1,A_2,\cdots$中至少有一出现频率.

容易想像，如将上列各式中的频率 $f_n\{A\}$置换成概率 $P\{A\}$，便可以得出统计概率所具有的性质. 此外，可以证明，对于古典概率定义来说，也具有上列统计概率相同的性质.

四、公理化概率定义

统计概率定义推广了古典概率定义和几何概率定义，使概率定义的应用范围大幅度地扩大. 但如果仔细分析，仍可以发现统计概率定义的提法存在不少问题. 主要的问题集中在建立统计概率定义时所依据的事件的属性上. 即随着试验次数的增加，事件出现的频率呈现某种稳定性. 定义中对试验次数应大到什么程度和所谓“稳定性”应如何理解都没有明确的说明，它在很大程度上是凭直观的. 因此，从逻辑上讲它是不严格的.

于是，建立一个严格的概率定义，以便满足现代自然科学对概率定义所提出的更高要求就显得十分必要了.

在引进公理化概率定义之前，必须首先严格定义事件的概念. 前面已指出任何一个随机试验 E，总可以得到一个样本空间 Ω，而试验中所有可能出现的事件只不过是 Ω 的一些子集而已，我们把它的全体记为$\mathscr{F}$. 在具体的随机试验中，$\mathscr{F}$ 由 Ω 的哪些子集构成是可以确定的. 但是，由于公理化概率定义是摆脱具体的随机试验的，所以我们所关心的不是$\mathscr{F}$是由 Ω 的哪些子集来构成，而是 $\mathscr{F}$ 应该满足的实际和理论上所需要的共同要求. 概括地说，$\mathscr{F}$ 必须满足下属的三条事件运算的规则.