屠其璞　王俊德
丁裕国　史慧敏　编

气象应用概率统计学

气象出版社

气象应用概率统计学

屠其璞　王俊德
丁裕国　史慧敏　编

气象出版社

内 容 简 介

本书系统地阐述了概率统计学基本理论及其在气象上的应用，而重点是介绍适合气象业务工作特点的一些统计原理和方法。全书共分十一章。第一至三章为概率论基础；第四至六章为数理统计理论；第七至十一章为各种统计方法，如回归分析、判别分析、聚类分析、时间序列分析、气象要素场的正交函数展开、气候资料的序列订正等。

书中附有较多的例题，便于读者自学。本书可供广大气象业务、科研人员阅读和参考，也可作为有关大专院校的教学参考书。

气象应用概率统计学

屠其璞 王俊德
丁裕国 史慧敏 编

*

气 象 出 版 社 出 版
（北京西郊白石桥路46号）
北京印刷一厂印刷 新华书店北京发行所发行

*

开本：850×1168 1/32 印张：17.375 字数：457千字
1984年10月第1版 1984年10月第1次印刷
印数：1—10,000 统一书号：13194·0148
定价：4.20元

前　言

本世纪七十年代以来，概率统计方法应用越来越普遍。在气象工作中，“统计”这一概念已大大突破了原来仅局限于计算观测资料统计指标的范围，更多的则和“分析”联系在一起了。大量资料的积累，计算工具的改进，特别是电子计算机的推广和普及，已为概率统计方法的广泛应用提供了物质基础。

本书在阐述概率论和数理统计基本概念时，力求联系气象工作的实际而不过分强调数学的严密性，希望能有助于读者克服在学习概率统计的过程中常常感到抽象难懂的困难。在介绍各种常用数理统计方法时，除给出一些旨在说明计算方法的示意性例子外，还尽可能给出在气象各领域应用中的实例，以供读者在实际工作中参考和借鉴。

我们知道，气象工作中已引进的概率统计方法只是其中的一部分，在这方面，还有十分广阔的领域有待发展和开拓。即使在气象上现已应用的各种概率统计方法，也还在不断地改进和发展之中。为此，本书在努力联系气象工作实际的同时，注意保持概率统计基本内容的系统性和完整性，而在介绍常用统计方法时，努力吸收国内外新成果，以便使读者在阅读本书之后，对概率统计基本理论和方法及其应用能有一个比较系统的了解，适应气象应用概率统计方法日新月异的发展形势。

本书是在本院气象和气候专业统计课教材的基础上编写而成的。编写过程中翁笃鸣副教授及气候教研室的同志们给予了热情的鼓励与帮助，章基嘉教授在百忙中审阅了全部书稿并提出了许多宝贵意见，此外，王得民、施能两位同志也对本书部分内容提出宝贵意见，汪仲梅同志为本书代绘插图，编者在此向他们表示衷心的感谢。

由于编者水平所限，书中不当之处恐难避免，敬请读者批评指正。

编　　者

1982年元月于南京气象学院

目　　录

第一章　事件与概率

§1.1 引　　言

众所周知，我们观察自然现象的时候，会发现有些现象在一定条件下是必然会发生的。例如，在标准大气压下，水加热到100°C必然会沸腾；强冷空气侵入本地必然要降温；气旋过境，云量必然增多等等。这种现象称为必然现象。必然现象与其影响因素之间有比较明确的因果联系，使得它的变化具有比较确定的数学物理规律，因此，可以用经典的数学分析方法进行研究。

除必然现象外，还有另一类不同的现象，这种现象在一定条件下可能发生，也可能不发生。例如，天空有云，可能下雨，也可能不下雨；南京明年的降水，可能偏多，也可能偏少，也可能正常；掷一枚硬币，可能出现正面，也可能出现反面等等。这类现象称为随机现象或偶然现象。随机现象是大量错综复杂而又变化不定的因素影响的结果，它们之间的因果联系非常复杂，以致人们很难考查这种关系，更难准确预测它的未来状态，这就是随机现象的不确定性。

随机现象的不确定性使人们无法用经典的数学方法去研究它的规律。但是，实际经验证明，随机现象仍然是有规律可循的，不过，这种规律并不表现在现象的个别观测结果上，而是表现在大量的观测结果上。例如，表1.1是分级统计的上海1873—1970年间历年一月平均气温出现在各级范围的年数。我们知道，就个别年份来说，气温取什么值并不能预先准确无误地确定，但从表中看到，上海一月平均气温出现在各级范围内的年数却是很有规律的。即出现在中间范围的年数最多，出现高温或低温的年数都很少，年

表 1.1

序号	1	2	3	4	5	6	7	8
分级	−1.0—0.0	0.0—1.0	1.0—2.0	2.0—3.0	3.0—4.0	4.0—5.0	5.0—6.0	6.0—7.0
年数	3	4	11	19	26	19	12	4

数的分布基本上是对称的。又如在相同条件下打靶，当射击次数较少时，靶上弹着点的分布是完全没有规律的，但当射击次数增多时，就开始出现规律，越近靶心，弹着点越多，离靶心越远，弹着点越少，而且射击次数越多，这种规律越明显。随机现象的这种规律称为统计规律。它是大量同质现象的集体规律，掌握了这种规律，就可以预言随机现象变化的基本趋势，或控制随机因素的影响，而概率统计方法就是研究这种规律的数学工具。

应当指出，把自然现象分为必然现象和随机现象只是为了研究的方便。事实上，自然界的一切现象都具有必然性和随机性两面。因为不管什么现象的发展变化，除受基本的决定性的因素控制外，总还受许多次要的微小的随机因素所干扰。所谓必然现象，只是由于对实际需要来说，这种次要因素所造成的偏差可以忽略而已。如果要求较高的精度，就应当考虑随机因素的影响。对于大气现象来说，由于影响它的天文地理因素比较稳定，因此，各地的天气和气候都有一定的特征，这就是大气现象的必然性。然而，一个地方的天文地理条件也不是一成不变的，加之大气本身的性质十分复杂多变，这就使得一个地方的天气和气候不能成为纯粹的必然现象，而具有明显的随机性。因此，概率统计方法在气象研究中得到了广泛的应用。

§ 1.2 随机事件

一、事件的定义

我们知道，不管什么自然现象的发生(或出现)总是与一定条

件相联系的，在概率统计方法中，一开始就遇到“试验”这一术语。因为这种方法是在一定条件实现之下来考察研究对象的状况的，这和物理化学中做试验相似，因此，在概率统计方法中，把在一定条件下对研究对象的观测称为试验。所谓事件，就是试验的结果，它是概率论研究的基本对象。如果一个事件在一定条件实现下必然出现，就称为必然事件；如果一个事件在一定条件实现下必然不出现，就称为不可能事件。显然，必然事件和不可能事件都是确定性现象的试验结果。对于随机现象，每次试验有多种可能的结果，而在试验之前并不能预言出现哪种结果，因此，通常称之为随机试验。随机试验的结果称为随机事件。例如，在产品质量检查时，抽查产品就是随机试验，“产品合格”或“产品不合格”就是这一试验的两个随机事件；我们也可把观察天气看作随机试验，“有雨”、“无雨”、“晴”、“阴”等都可看作随机事件。

以后用字母U表示必然事件；V表示不可能事件；$A,B,C,\cdots\cdots$表示随机事件。由于随机事件是我们研究的主要对象，因此，今后把随机试验简称为试验，把随机事件简称为事件。

一个随机试验有许多不同的可能结果，我们把这些结果称为基本事件。例如，掷一颗骰子，可出现 1,2,3,4,5,6 各种点数，每一种点数都是一个基本事件；若观测降水的形式，则其基本事件有雨、雪、雹、雾等等。以后用ω_i表示随机试验中的第i个基本事件。一个随机试验的全部基本事件组成的集合称为基本事件空间，简称基本空间，以Ω表示。弄清基本空间的组成，对于解决实际问题是十分重要的，下面举几个例子。

例 1.1　掷一枚硬币，观察出现的面，基本空间包括正面和反面两个基本事件。

例 1.2　观测一次风向，则北，北北东，东北……等 16 个方位及静风共 17 个基本事件组成基本空间。

例 1.3　观测某地年降水量，其基本空间为$\geqslant 0$的全体实数。

例 1.4　观测某地年雨日数，基本空间中共有 0，1，2，……365 这 366 个基本事件。

在实际问题中，不仅要研究基本事件，还要研究由若干基本事件组成的事件。例如，我们常说"偏东风"，它就包含了东北东，东及东南东三个不同方位的风；又如，天气预报中报"雨日 3 到 5 天"，就包括雨日为 3 天或 4 天或 5 天三个基本事件。显然，这里的"偏东风"、"雨日 3 到 5 天"也是事件，不过这不是简单事件，而是包含了若干基本事件的事件。我们称这类事件为复合事件。

今后，当有必要说明基本空间的成员时，采用记号 $\Omega=\{\omega_1,\omega_2,\cdots\}$或 $\Omega=\{x;x\geqslant 0\}$，这样，对 Ω 中的任意事件 A 可记为 $A=\{\omega_{i_1},\omega_{i_2}\cdots\}$ 或 $A=\{x;3\leqslant x\leqslant 5\}$。例如，观测南京 6 月中旬雨日数，基本空间可记为 $\Omega=\{0,1,2,3,4,5,6,7,8,9,10\}$，若以 A,B,C 分别表示事件"雨日不超过 3 天"，"雨日为 4 到 6 天"及"雨日不少于 7 天"，则记 $A=\{0,1,2,3\}$；$B=\{4,5,6\}$；$C=\{7,8,9,10\}$。

二、事件之间的关系

自然界的各种现象都是相互联系的，因此，我们不仅要研究个别事件的规律，还要考虑事件之间的相互关系。事件间的相互关系主要有下列几种。

1．包含关系

设 A,B 是 Ω 中的两个事件，若事件 A 出现必导致事件 B 出现，则称事件 B 包含事件 A，或称 A 是 B 的特款。并以 $A\subset B$ 或 $B\supset A$ 表示。例如，吹"东风"必导致"偏东风"出现，因为"偏东风"包含了"东风"这一特款。

如果事件 $A\subset B$，同时也有 $B\subset A$，则称事件 A 与 B 等价或相等，记为 $A=B$。例如"极大风速≥17 米/秒"与"大风"就是等价事件。

2．互不相容

设 A 与 B 是 Ω 中的两个事件，若事件 A 出现必然导致事件 B 不出现（当然，事件 B 出现也必然导致事件 A 不出现），则称 A 与 B 为互不相容事件。例如，在一次天气观测中，"碧空"和"有雨"是不能同时出现的，因此它们是互不相容事件。

3．互逆事件

若在一次试验中，事件A与B不能同时发生，但每次试验必然出现其中之一，则称A与B为互逆事件或对立事件。通常将A的逆事件记为$\overline{A}$。例如，气候的“正常”与“异常”；天气的“有雨”与“无雨”；“旬雨日在3天以下”及“旬雨日不少于3天”；射击的“命中”与“未命中”等，都是互逆事件。

4．完备群

设$A_1,A_2,\cdots,A_n$为Ω中的n个事件，若其中任意两个事件都互不相容，但每次试验能且只能出现其中之一，则称$A_1,A_2,\cdots A_n$为一互不相容事件完备群，简称完备群。显然，任一事件与它的逆事件构成一个最简单的完备群。另外，基本空间本身就是事件的完备群，其中的每一A_i只包含一个基本事件。但是，一般来说，完备群中的每一个事件A_i可能不只包含一个基本事件，也就是说，它可能是一个复合事件。例如，降水量的基本空间为$\Omega=\{x;x\geqslant 0\}$，但是无雨、小雨、中雨、大雨、暴雨都是$\Omega$中的复合事件，它们构成降水的完备群。由此看到，完备群与基本空间是既有联系又有区别的概念。

三、事件的运算

事件的运算有下列几种。

1．事件之和

设A,B为Ω中的两个事件，则由A,B至少出现一个构成的事件称为事件之和，记为$A\cup B$或$A+B$。读作“A加B”，也可读作“或A或B”。例如，“年降水异常”是“年降水偏多”及“年降水偏少”的和。

事件之和的运算可推广到多个事件的情形，若事件A是由事件$A_1,A_2,\cdots,A_n,\cdots$至少发生一个构成，则称$A$为$A_1,A_2,\cdots A_n\cdots$之和，记为$A=A_1\cup A_2\cup\cdots\cup A_n\cup\cdots=\bigcup\limits_{i=1}^{\infty}A_i$，也可记为

$$A=A_1+A_2+\cdots+A_n+\cdots=\sum_{i=1}^{\infty}A_i。$$

2．事件之交

事件之交就是由 A，B 同时发生构成的事件，记为 $A\cap B$ 或 AB。读作“A 乘 B”。例如，“雷雨”是“打雷”与“下雨”两事件之交；“雨夹雪”是“下雨”和“下雪”同时发生的事件。

若事件 A 是事件 $A_1, A_2\cdots A_n\cdots$ 同时发生构成的事件，则记为 $A=A_1A_2\cdots=\prod\limits_{i=1}^{\infty}A_i$。

3．事件之差

设 A，B 为 Ω 中的两个事件，由 A 发生而 B 不发生构成的事件，称为事件 A 与 B 的差，记为 $A-B$。例如，“雨日不超过 5 天”与“雨日不超过 3 天”的差是“雨日为 4 天或 5 天”。

应当指出，在事件之差的运算中，关系式 $A-B+B=A$ 一般不成立。例如，设 A 表示“打雷”，B 表示“下雨”，则 $A-B$ 表示“只打雷不下雨”，而 $(A-B)+B$ 表示“或打雷或下雨”，显然，它与“打雷”不是等价事件。

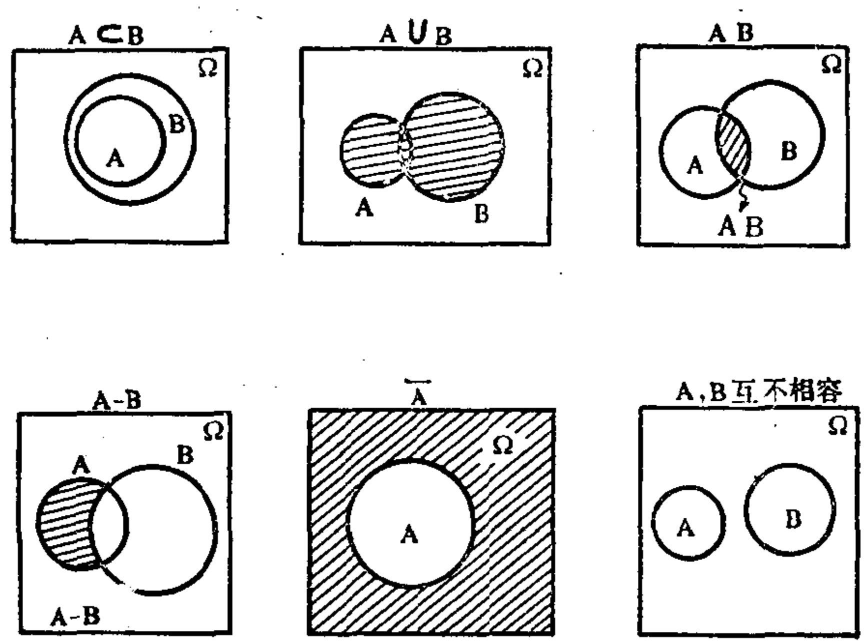

图 1.1

事件之间的关系和运算，可用图 1.1 形象地说明，图中 Ω 表示基本空间，除 $A\subset B$ 及互不相容的情形外，其他关系用阴影部分

表示。

§ 1.3 概率的定义

一、概率的意义

随机事件的发生具有不确定性。在实际需要和理论研究中，人们常要定量地说明随机事件发生的可能性，以便对试验结果进行预测。例如，在分析旱涝规律时，要估计各种大小的降水出现的可能性；在作建筑物风荷载计算或风能利用规划时，要估计各种量级的风力出现的机会等等。为了衡量事件出现的可能性程度，就要用一种数量指标来测量，使出现可能性大的事件有较大的数值，出现可能性小的事件有较小的数值，这样的指标就称为概率。通常把事件 A 在一定条件下出现的概率记为 $P(A)$。

那么应该用怎样的数量作为事件的概率呢？显然，必须有一个测量的单位。我们知道，必然事件在一定条件下必然会出现，因此它出现的可能性最大，可令其概率为 1；而不可能事件在一定条件下一定不会出现，因此它出现的可能性最小，可令其概率为 0；随机事件在一定条件下可能出现也可能不出现，因此，它出现的概率变化于 0 与 1 之间。以后我们会看到，这样规定概率的取值域，不仅逻辑上是合理的，而且也是有其客观实践基础的。

二、古典概率

虽然概率的意义是简单而明确的，但是若要回答某一事件出现的概率是多大，却并不是很容易的。下面我们先考虑古典型的概率问题。

若一个随机试验只有有限种等可能的结果，则这一试验就称为古典型的。例如，掷一枚硬币；掷一颗骰子等，就是这样的试验。

若一古典型试验共有 n 种结果，其中有 m 种结果具有性质 A（或称有利于 A），则 A 出现的概率为

$$P(A)=\frac{m}{n} \tag{1.3.1}$$

这就是概率的古典定义。

这里应该注意“等可能”的涵义，对此，我们无法下一个确切的定义，只能根据人们的实践经验来理解。它是以客观事物的对称性为基础的，例如，掷一枚构造均匀的硬币，经验证明，出现正面或反面两种结果的可能性是相等的。

在计算古典型试验的概率时，最重要的是要正确分析试验的所有可能结果（即基本空间）及有利于事件 A 的那些结果(即 A 由哪些基本事件组成)。为此，常需应用排列组合理论。现在举几个例子。

例 1.5 一袋里装有 4 个白球，6 个黑球，它们的形状大小相同，随机地从中取两球，求其中恰有一个白球和一个黑球的概率。

解：设以 A 表示取到一个白球及一个黑球的事件，从 10 个球中任意取 2 个球的取法，按组合理论应为 $C_{10}^2=\frac{10!}{2!8!}$ 种。因球的形状大小都相同，故可认为每一种取法都是等可能的。当 A 发生时，意味着有一个球取自 4 个白球中，显然从 4 个中任取一个的取法共有 4 种，而取另一个黑球的取法有 6 种，两种颜色的球搭配，共有 4×6 种取法有利于 A，所以 A 出现的概率为

$$P(A)=\frac{C_4^1\times C_6^1}{C_{10}^2}=\frac{8}{15}\approx0.53$$

如果把抽球比作抽查产品或仪器，“白球”、“黑球”比作“正品”、“次品”等等，就可以看到，这类问题是有实际意义的。

例 1.6 一只箱子里装有标号分别为 0，1，2，3，4，5，6，7，8，9 的 10 个同样的球，从中任取五个球（每次取一个，记下号码后放回），求五个球号码都不同的概率。

解：以 A 表示五个球号码都不同，我们把从 10 个数字中任取五个的一切取法(每次一个，取后放回)作为基本空间，则基本空间中的基本事件共有 10^5 个，要使五个球号码都不同，则第一次取得的球可以是 0 至 9 号中的任意一个，共 10 种取法，第二次所取球的号码必须是其他 9 个球中的一个，依次类推，最后一个球应从剩

下的 6 种号码球中任取一个，这些取法相配合，所以有利于 A 的取法共有 $10\times9\times8\times7\times6$，于是

$$P(A)=\frac{10\times9\times8\times7\times6}{10^5}\approx0.3024$$

三、频率

从上面的讨论中看到，古典概率只适用于具有有限个等可能结果的随机试验，但是，就大多数实际随机现象来说，其可能的试验结果往往不是有限的，而且实际上也无法判断各种结果是不是等可能的，因而就不能用古典概率的方法来计算概率。此时，一般用频率来代替概率。

设对某随机现象在同样条件下进行了 n 次重复试验，其中事件 A 出现了 m 次，则比值

$$P^*(A)=\frac{m}{n} \tag{1.3.2}$$

称为事件 A 在 n 次试验中出现的频率，而 m 称为频数。

经验证明，若 A 出现的概率比较大，则在多次重复试验中，它出现的次数也就多，因此，频率 $P^*(A)$ 也就愈大。反之，若 A 在每次试验中出现的概率比较小，则在多次重复试验中，它出现的次数也就少，因此，频率 $P^*(A)$ 也就愈小。所以，频率与概率之间应该有密切的关系。我们用表 1.2 中的实例来说明这种关系。表中给出按不同年数统计的上海一月逐日平均气温正距平频率。我们看到，当统计年数较少时，频率的变化比较大，可是当统计年数很多时，频率的变化愈来愈稳定地在某未知常数 p 附近摆动。这一事实就称为频率的稳定性。尽管我们并不知道频率稳定所趋的那个常数是多少，但是我们不难设想，它就是事件的概率。而且，可以

表 1.2

年数	5	10	20	30	40	50	60	70	80	90	100
频率	0.335	0.394	0.429	0.452	0.487	0.489	0.485	0.490	0.507	0.503	0.501