天文数据处理方法

丁月蓉　编著

南京大学出版社

内 容 简 介

本书从实用出发介绍了目前天文上普遍应用的几种数据处理方法。全书共分八章。第一章概率统计基础是各种方法的基础;第二章论述误差分析的基本理论,并全面系统地介绍了最小二乘法及应用。第四章至第六章是数字信号处理的内容,包括资料的平滤、滤波及各种谱估计方法;第三章、第七章属多元统计分析的范畴,介绍了几种常用的多变量数据分析方法及其在天文上的应用。除最后两章外,每章都附有习题。

本书介绍的方法不仅适用于天文观测数据,也适用于地球物理、气象、地质等学科测量数据的处理,可供这些专业的大学生和低年级硕士生使用。

天文数据处理方法

丁月蓉　编著

*

南京大学出版社出版

(南京大学校内　邮编 210008)

南京展望照排印刷有限公司排版

江苏测绘院印刷厂印刷

江苏省新华书店发行

*

开本 850×1168　1/32　印张 12.25　字数 318 千

1998 年 8 月第 1 版　　1998 年 8 月第 1 次印刷

印数 1—1000

ISBN　7-305-03166-6/P・116

定价　15.00 元

(南大版图书若有印、装错误可向承印厂退换)

前　言

天文观测数据处理是在天文观测的基础上揭示宇宙奥秘的重要手段．随着科学技术的发展，各种大型天文仪器设备的投入使用，天文学家获得的数据量正以惊人的速度迅猛增加，数据处理方法本身无论是在理论方面，还是应用方面，都在很快地发展．在这种情况下，加强这方面的教学是面向21世纪的教育义不容辞的责任．

1988年、本书作者曾与郑大伟研究员合作编著了《天文测量数据的处理方法》，受到天文界同仁的欢迎．随着教学改革的深化，教学内容不断更新，并要兼顾天体物理、天体测量和天体力学等二级学科，为此，作者结合近几年的教学和科研工作，对各个学科中使用数据处理方法的情况进行了调研，对原书进行了修改、补充，如本书的第二章到第六章是在原书的基础上进行了调整、修改和补充，第一章、第七章是新增加的．本书中保留了原书的部分内容，在此，作者对原书合作者郑大伟研究员表示由衷的感谢．

本书从实用出发介绍了目前天文上普遍应用的几种数据处理方法．作者力图在讲清方法的原理、用法的基础上，尽量结合天文实际．作为一门方法课程，我们并不过分追求数学的严谨，故而省略了一些烦杂的理论证明和冗长的数学推导．为了加深学生对所学内容的理解和掌握，在每一章的后面给出了习题或上机实习题．

本书的出版得到了南京大学教材出版基金的资助，也得到了南京大学出版社和天文系教学委员会的大力支持，作者表示衷心

的感谢．张承志教授和肖耐园副教授在百忙中分别对本书的第二、第三章和第四章至第六章进行了审阅，作者表示衷心的感谢．

由于时间匆忙，作者水平有限，错误疏漏之处在所难免，恳请读者批评指正．

作　者

1998年2月

目　录

第一章　概率统计基础

随着科学技术的迅速发展，概率统计方法已在各个科学领域里得到日益广泛的应用. 在天文数据处理中也越来越显示它的重要性，"统计"这一概念也大大突破了原来局限于计算观测资料统计指标的范围，更多地则和"分析"联系在一起了. 因此，在学习数据处理方法以前，首先应该对概率统计的基本知识要有所了解，并且作为学习数据处理方法的不可缺少的内容之一.

第一节　随机事件与概率

在概率统计中，把在一定条件下对研究对象的观测称为试验. 而在一定的试验条件下，某一现象 A 可能发生，也可能不发生，并且只有发生或不发生这两种可能性. 通常，我们把发生了的现象或试验结果 A 称为事件，并简记为事件 A.

有些事件在一定条件下必然会发生，称作为必然事件，常用英文字母 U 表示；有些事件在一定条件下不可能发生，称作为不可能事件，常用 V 表示. 必然事件和不可能事件都是确定性现象的试验结果. 另有一些试验，每次试验有多种可能的结果，但在试验之前不能预言出现哪种结果，通常称之为随机试验. 随机试验的结果称为随机事件，通常用 $A,B,C,\cdots\cdots$ 表示. 必然事件和不可能事件可看作随机事件的特例. 有的随机试验只有两种不同的可能结果，例如投一枚硬币，只有出现字面或徽面两种不同的可能结果. 有的随机试验有许多不同的可能结果，如果它们是不能再分的，则称它们为基本事件. 例如投一颗骰子，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. 由基本事件可以组合成复合事

件．一个随机试验的全部基本事件的集合称为基本事件空间，简称基本空间，以 Ω 表示．

随机事件是概率论中主要的研究对象．我们不仅要研究随机事件出现的规律，还要考虑事件之间的相互关系及其运算，下面介绍事件之间的关系与事件的运算．

若事件 A 出现必然导致事件 B 出现，则称事件 B 包含事件 A，并记为 $B\supset A$．

如果事件 B 包含事件 A，事件 A 也包含事件 B，则称事件 A 与事件 B **相等**，并记为 $A=B$．

若事件 A 与事件 B 不能同时发生，则称事件 A 与事件 B 为**互不相容事件**（或**互斥事件**），并有 $A\cap B=\varnothing$（$\varnothing$ 为空集）．基本事件是互不相容的．

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**对立事件**（或**互逆事件**），并记为 $A=\overline{B}$．

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．任一事件与它的对立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．

事件的运算有下列几种．

事件之和　设 A,B 为 Ω 中的两个事件，则 A,B 中至少出现一个构成的事件称为事件之和，记为 $A+B$（或 $A\cup B$）．事件之和的运算可推广到多个事件的情况，若事件 A 是由事件 $A_1,A_2,\cdots,A_n$ 中至少发生一个构成，则称 A 为 $A_1,A_2,\cdots,A_n$ 之和，记为 $A=\sum\limits_{i=1}^{n}A_i$．

事件之积　事件 A 与事件 B 同时发生构成的事件称为事件 A 与事件 B 的积，记为 AB 或 $A\cap B$，类似地，定义事件 $A_i(i=1\sim$

n)同时发生构成的事件 A 为 $A_i(i=1\sim n)$的积.

$$A=A_1A_2\cdot\cdots\cdot A_n=\prod_{i=1}^{n}A_i.$$

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 $A-B$. 如晴夜不超过 10 天与晴夜不超过 8 天的差是晴夜为 9 天或 10 天.

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,为了衡量事件出现的可能性大小的程度,需要用一个数量指标,概率这一术语正是用来刻画随机事件出现的可能性大小的一个量. 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,因为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,它出现的概率在 0 与 1 之间变化,这样规定概率的取值域不仅逻辑上是合理的,而且也是有其客观基础的. 通常把事件 A 在一定条件下出现的概率记为 $P(A)$.

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,下面我们给出常见的两种概率的定义.

古典概率 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. 若一古典型试验共有 n 种结果,其中 k 种结果是有利于事件 A 的,则事件 A 发生的概率为

$$P(A)=\frac{k}{n}, \tag{1.1}$$

这就是概率的古典定义.

下面通过一个例子说明古典型试验的概率计算.

例 1.1 在 1,2,…,9 这 9 个自然数中,求任取一个是奇数的概率.

解 设以 A 表示抽到的数为奇数的事件,有利于 A 的结果是

5，总的事件数是 9，则

$$P(A)=\frac{5}{9}.$$

在计算古典型试验的概率时，最重要的是要正确分析所有可能的结果及有利于事件 A 的可能结果，为此常需利用排列组合理论.

古典概率只适用于具有有限个等可能结果的随机试验，但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，故而引入统计概率的定义.

统计概率 在相同条件下进行了 n 次试验，其中现象 A 发生了 m 次，则记事件 A 出现的频率为 m/n，随着试验次数 n 的增加，事件 A 的频率将趋于稳定，所稳定的常数叫做理论频率，这个理论频率就作为在给定条件下事件 A 的概率的近似值，这就是统计概率的定义. 理论频率要求 n 充分大，而实际上 n 总是有限的，所以更确切地说概率是频率的极限，即

$$P(A)=\lim_{n\to\infty}\frac{m}{n}. \tag{1.2}$$

下面我们讨论概率的运算. 概率的运算是指由简单事件的概率计算较复杂事件的概率，包括概率的加法与乘法.

概率加法定理 若 A,B 为互不相容事件，则

$$P(A+B)=P(A)+P(B). \tag{1.3}$$

这个定理表达了概率的重要特性，即可加性. 它从大量的实践中概括出来，又成了我们研究概率的基础. 从概率的定义出发，这个定理的证明是很容易的，这里从略. 用归纳法不难证明，若事件 A_1，A_2，…，A_n 互不相容，则

$$P\Big(\sum_{i=1}^{n}A_i\Big)=\sum_{i=1}^{n}P(A_i). \tag{1.4}$$

必须强调，上面的概率加法定理仅适用于互不相容事件，对于一般的事件，则有

$$P(A+B)=P(A)+P(B)-P(AB). \tag{1.5}$$

下面我们加以具体证明：

由于 $A+B=A+B\overline{A}$，A 与 $B\overline{A}$为互不相容事件，所以

$$P(A+B)=P(A)+P(B\overline{A}),$$

又∵ $B=BA+B\overline{A}$，BA 与 $B\overline{A}$也为互不相容事件，故

$$P(B)=P(BA)+P(B\overline{A}).$$

$$P(A+B)=P(A)+P(B)-P(AB).$$

(1.5)式也可推广到有限多个事件的情形，若 $A_1, A_2, \cdots A_n$ 为某试验中的 n 个事件，则

$$P\left(\sum_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)-\sum_{1\leqslant i\leqslant j\leqslant n} P(A_iA_j)+\sum_{1\leqslant i\leqslant j\leqslant k\leqslant n} P(A_iA_jA_k)-\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n), \tag{1.6}$$

若诸事件互不相容，则(1.6)式变成(1.4)式.

例 1.2 箱内有 10 个灯泡，其中 3 个是废品，7 个是正品，从中任取 4 个，求全是正品或只有一个废品的概率.

解 以 A 表示全是正品的事件，B 表示只有一个废品的事件，显然，A 与 B 互不相容，而 4 个全是正品或只有一个废品的事件为 $A+B$，按(1.5)式有

$$P(A+B)=P(A)+P(B)=\frac{C_3^0\times C_7^4}{C_{10}^4}+\frac{C_3^1C_7^3}{C_{10}^4}\approx 0.677.$$

下面介绍一下利用对立事件计算概率的方法. 因为 A 与 $\overline{A}$ 构成互不相容事件完备群，因此，$A+\overline{A}=U$，故有

$$P(A+\overline{A})=P(A)+P(\overline{A})=1,$$

于是得到

$$P(A)=1-P(\overline{A}).$$

许多情况下 $P(\overline{A})$比 $P(A)$容易计算，这时利用这一关系就比较方便.

概率乘法定理 在讨论概率乘法定理之前，先介绍条件概率的概念.

在事件 B 发生的条件下，事件 A 发生的概率，称作事件 A 对于事件 B 的条件概率，记作 $P(A|B)$. 下面通过一个例子说明条件概率与一般概率的区别.

例 1.3 箱内有 3 个白球，2 个红球，甲乙两人依次从中取一球，甲先取，若已知甲取得(a)白球，(b)红球，试在甲取后放还及不放还两种情况下，求乙取得红球的概率.

解 以 A 表示乙取得红球的事件，B_1 表示甲取得白球的事件，B_2 表示甲取得红球的事件. 在甲取后放还的情况下，不管甲取的是红球还是白球，在乙取球时，箱内球的成分没有变化，因此，乙取到红球的概率 $P(A)=2/5$.

若甲取后不放回，当甲取的是白球时，箱中还有 2 个白球，2 个红球，因此，乙取到红球的概率是 $P(A|B_1)=2/4$；但若甲取的是红球，则在乙取球时，箱中有 3 个白球，1 个红球，于是他取得红球的概率是 $P(A|B_2)=1/4$.

应该注意，不管什么事件的概率，总是与"一定条件"相联系的，从这个意义上说，凡概率都是有条件的，这里所说的条件概率，指的是在一般条件之外，另加的附加条件下的概率.

有了条件概率的基础，下面讨论概率的乘法定理.

如果在 N 次试验中，事件 B 发生 N_B 次，事件 A，事件 B 同时发生 N_{AB}次，则有

$$P(A|B)=\lim_{N\to\infty}\frac{N_{AB}}{N_B},$$

而

$$\frac{N_{AB}}{N_B}=\frac{N_{AB/N}}{N_{B/N}}\xrightarrow[N\to\infty]{}\frac{P(AB)}{P(B)},$$

所以

$$P(A|B)=\frac{P(AB)}{P(B)}. \tag{1.7}$$

并称它为事件 B 发生的条件下事件 A 发生的**条件概率**.