气象站数理统计预报方法

谭冠日 编

科学出版社

气象站数理统计预报方法

谭冠日 编

科学出版社
1980

内 容 简 介

本书介绍了数理统计天气预报的原理和方法，包括数理统计基本概念，选择预报因子的数理统计知识，多因子预报方法，时间序列预报方法。通俗地说明其原理，利用大量的实例讲明计算的步骤，交流使用中的经验。本书可供气象台站业务工作者在工作和学习中参考，也可供水文、地震预报基层单位参考。

气象站数理统计预报方法

谭冠日 编

*

科学出版社出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1978 年 12 月第一版　　开本：787×1092 1/32

1980 年 7 月第一次印刷　　印张：9 7/8

印数 13,361—17,230　　字数：225,000

统一书号：13031 904

本社书号：1285・13—15

定价：1.20 元

前　　言

在党的领导下，我国气象事业遵循毛主席的无产阶级革命路线，为国民经济建设和国防建设作出了积极的贡献。

近年来，数理统计方法在我国广大气象台站推广，与其他方法相结合，对提高预报准确率和预报服务质量起到了一定的作用。

数理统计是数学的一个分支，可以作为分析气象资料和使天气预报客观化的工具。借助群众看天经验和天气学原理，分析天气演变的统计规律性以及预报因子与预报对象的数量关系，建立数学模式，对未来天气作出预报，就是数理统计天气预报，简称统计预报。通过数理统计方法，可以提炼群众的天气预报经验；检验预报指标的可靠性；综合多种因子和工具的预报结论；使天气学预报思路和动力学方法作出的形势预报联系到具体天气的预报。但是，统计预报也有一定的局限性，这是由于对大气过程的内部物理联系缺乏深入的了解；对稀有天气的预报能力较差。统计预报必须和群众看天经验、天气学原理、动力学方法紧密结合，采用现代计算设备，不断研究提高，才能发挥更大的作用。

从当前气象台站的实际情况出发，本书介绍基本的统计预报方法，通俗地说明统计预报的原理，以大量的实例讲明计算的步骤，交流使用中的经验。第一部分介绍数理统计的基础概念和选择预报因子的数理统计知识（第一、二章）；第二部分讲述多因子的预报方法，即多元分析（第三至六章）；第三部分是根据天气随时间演变的规律性作预报的方法，即

时间序列分析（第七至九章）。

本书的编写得到云南大学的领导和地球物理系领导的支持和易仕明、王学仁、孙文爽、雷兆崇、李万昌同志的帮助。云南西双版纳州气象台和江苏常熟县气象站提供了宝贵的意见。由于编者水平所限，书中存在不少缺点和错误，请同志们指正。

编　者

一九七七年

目　　录

第一章 怎样度量天气发生的可能性

第一节 用概率度量天气发生的可能性

一种天气① 的发生，包含有必然性和偶然性。例如，寒潮造成的降温强度是由很多因素共同影响的，这些因素起作用的程度不同和它们互相配合的不同，使降温以或强或弱、或急或缓这样一些带有偶然性的形式表现出来。但不管每一次寒潮降温具有什么独特的形式，冷空气平流的这一共同的物理过程使降温具有必然性。恩格斯科学地论证了必然性和偶然性的关系，指出“被断定为必然的东西，是由纯粹的偶然性构成的，而所谓偶然的东西，是一种有必然性隐藏在里面的形式”。天气的发生和发展也是必然性和偶然性的辩证统一。

在统计预报中，由于天气具有偶然性的一面而把天气称作随机现象，不但不否认天气的发生发展存在必然性规律，而且正是要透过随机（偶然）现象，去探索它内部隐藏着的必然性，从而作出天气预报。随机现象不是不可测的，要预测它，首先要有一个数来度量随机现象发生的可能性，这个数量就是概率。

先解释几个术语。

随机现象的具体表现、具体结果称为随机事件，简称事件。气温的变化是随机现象，气温上升、气温下降、气温>10℃等等就是随机事件。随机事件是在一定条件下可能发生也可

---

① 为了叙述的方便，本书将气象学研究的现象，包括气象要素、天气现象、天气过程等等统称为天气。

能不发生的事件。

在一定条件下必定发生的事件称必然事件，必定不发生的事件称不可能事件。“寒潮影响气温下降”，寒潮影响是条件，气温下降是这条件下的必然事件，气温上升则是这条件下的不可能事件。

统计分析的对象的全体称为总体或母体。总体中每一个别事物称为个体。在许多情况下，只能研究总体中的一部分个体，这部分个体称为样本或子样。抽样时总体中每一个个体都有同等的机会被抽到，称为随机抽样。随机抽样时总体中个体的成份不变，所得样本称简单随机样本，它的统计特征能较好地代表总体。样本中个体的数目称为样本容量或样本大小、样本数，习惯以 n 表示。例如，要研究某地 6 月降水量，该地古往今来以至未来的无数个 6 月降水量是总体，实有的 n 年的 6 月降水量是样本，样本容量是 n。又如研究 4 月每天 14 时 ΔP_{24} 和 $(e-T)$ 同次日晴雨的关系，每年有 30 个个例，10 年资料的样本容量 $n=300$。在天气分析研究中，每一次天气过程叫做个例，样本容量 n 也就是个例数目。天气的总体是无限的，总体的统计特征是未知的，只能由样本去推断它。样本的统计特征往往与总体的统计特征有偏离，这种偏离是由于样本的随机性造成的，称为随机波动。一般说来，随机样本的样本容量越大，其统计特征的随机波动越小，越可能接近总体。

毛主席教导我们说：“我们有许多同志至今不懂得注意事物的数量方面，不懂得注意基本的统计、主要的百分比，不懂得注意决定事物质量的数量界限，一切都是胸中无‘数’，结果就不能不犯错误。”为了对随机事件作出预报，我们应对随机事件发生的可能性做到胸中有“数”，这个数就是概率。

气象报表上要统计风向频率。某一种风向的频率就是观

测到该种风向的次数(又称频数)占观测总次数的百分率。频率的意义是:在观测的 n 次中,若观测到事件 A 出现 m 次,事件 A 的频率 $f_A=\frac{m}{n}$。

由频率可以引伸出概率。气象员都有这样的经验:短年代资料的统计结果不大可靠,长年代资料统计的结果比较可靠。取云南某地 1901 年至 1960 年间实有的 50 年资料,先计算每五年 7 月的平均雨日(将五年 7 月雨日总数除以 5)和雨日频率(平均雨日除以 31),见表 1.1。我们看到每五年的雨日频率随机波动很大,变化在 0.57—0.74 之间。再计算 10 年、15 年、20 年……50 年的 7 月雨日频率,随着年数的增加,雨日频率越来越稳定在 0.64 上下。

随着观测次数的增加,事件 A 的频率 f_A 的波动越来越小,稳定在常数 p 附近,事件 A 的概率 $P(A)$ 就等于 p。即

表 1.1

纪录年代	每五年		纪录年代	实有年数	平均雨日	雨日频率
	平均雨日	雨日频率				
1901—1907	17.6	0.57	1901—1907	5	17.6	0.57
1908—1918	17.8	0.57	1901—1918	10	17.7	0.57
1919—1923	21.2	0.68	1901—1923	15	18.9	0.61
1924—1928	21.4	0.69	1901—1928	20	19.5	0.63
1929—1933	19.0	0.61	1901—1933	25	19.4	0.63
1934—1938	19.4	0.63	1901—1938	30	19.4	0.63
1939—1943	21.4	0.69	1901—1943	35	19.7	0.64
1944—1950	17.6	0.57	1901—1950	40	19.4	0.63
1951—1955	21.2	0.68	1901—1955	45	19.6	0.63
1956—1960	23.0	0.74	1901—1960	50	19.9	0.64

$$P(A)=p$$

规定把事件的符号 A 写在 P 后面的括号中，表示事件 A 的概率。

当观测次数 n 足够大，频率可以大体上代表概率。即当 n 足够大，频率 f_A 近似于概率 $P(A)$，$\frac{m}{n}$ 近似于 p。

概率必在 0—1 之间。不可能事件的概率是 0，必然事件的概率是 1。

频率和概率的关系是个性和共性的关系，可对照如下：

表 1.2

频　　率	概　　率
1.由样本求得	1.对总体而言。当总体无限，用大样本的频率代表
2.随着样本的不同，有随机波动，是随机变量	2.是常数
3.是个性，其共性(概率)寓于其中	3.是共性，通过个性(频率)表现出来

频率和概率的关系可以总结如下：当观测次数无限增加，事件的频率无限接近概率的可能性趋于百分之百。这称为伯努利(Bernoulli)定理。

气象总体是无限的，我们只能用较大的样本计算频率来代表概率。气象台站目前习惯把小样本的频率称为概率或机率，本书在讨论概率时也常常通过一些小样本的频率来讨论。但我们应该认识到，样本越大，频率可能越接近概率；样本越小，频率与概率的偏离可能越大。

概率有两层意义，既表示多次观测中某一事件发生的频繁程度，又表示一次观测中该事件发生的可能性。比如，7 月雨日概率是 0.64，它表示，就平均而言 7 月份的 31 天里要出

现雨日 19.8 天（31 天×0.64=19.8 天）。如果雨日彼此独立，某一天是否下雨与其他日子是否下雨无关，则概率 0.64 又意味着 7 月每一天是雨日的可能性为 64%。

人们有时说某一年的大旱是百年一遇的；水利工程设计有时采用千年一遇或万年一遇的暴雨洪水为标准。这类“百年”“千年”“万年”称为重现期。重现期 T 与概率 p 的关系怎样呢？若某一次大旱是百年一遇的，重现期 $T=100$。百年中有一年发生此种现象，它的概率 $p=0.01$，显然

$$T=100=\frac{1}{0.01}=\frac{1}{p}$$

可知，重现期是概率的倒数。重现期习惯以年为单位，相应的概率也应以年的资料来计算。重现期不是周期，在 1954 年出现了百年一遇的洪水并不意味着下一次将出现在 2054 年。百年一遇只表示平均每一百年出现一次，至于具体出现在哪一年，重现期不能反映出来。

前面我们说明了概率的意义。下面将进一步研究几种天气概率的关系。

第二节　几种天气概率的关系

有些天气是可能同时发生的，有些天气是不可能同时发生的，需要区别这两种情况来讨论几种天气的概率的关系。

1.互不相容事件发生的概率

按照观测规范，每一天只能属于无雨日、小雨日、中雨日、大雨日、暴雨日中的某一种，不能既算这一种，又算那一种。那么，无雨日、小雨日、……暴雨日是互不相容事件。几种事件中任意两事件不可能同时出现，这几种事件是互不相容事

件。

某地 6 月份大雨日概率是 2/30，暴雨日概率是 1/30，则大—暴雨日(即大雨或暴雨)的概率显然是$\frac{2}{30}+\frac{1}{30}=0.1$。事件 A 和事件 B 若互不相容，则出现 A 或 B 两者之一，称为“A,B 两互不相容事件之和”，其概率 $P(A+B)$ 等于 A, B 两事件概率之和。

$$P(A+B)=P(A)+P(B) \tag{1.1}$$

推广之，若有 m 种互不相容事件 A_1, A_2, ……A_m, 它们各自的概率是 $P(A_1)$, $P(A_2)$, ……$P(A_m)$, 出现 A_1 或 A_2 或……或 A_m 其中一种事件的概率，即 m 种互不相容事件之和的概率

$$\begin{aligned}&P(A_1+A_2+\cdots\cdots+A_m)\\&\quad=P(A_1)+P(A_2)+\cdots\cdots+P(A_m)\end{aligned} \tag{1.2}$$

这就是互不相容事件的概率加法。

某一事件 A 的对立面 $\tilde{A}$，称为 A 的对立事件（或逆事件）。例如，降雹与无雹是对立事件；降水量≥1200 毫米和降水量<1200 毫米是对立事件。一次观测中若出现事件 A，必不出现其对立事件 $\tilde{A}$，这两种事件中必出现一种而且只出现一种，两者构成概率为 1 的必然事件，

$$P(A)+P(\tilde{A})=1 \tag{1.3}$$

若预报明天“有雨”的概率是 0.2，它的对立事件“无雨”的概率是 1−0.2=0.8。

2. 相容事件发生的概率

相容事件是可以同时发生的事件。雷暴和下雨是相容事件。雨日和中雨日也是相容事件，因为中雨日同时又是雨日。

假定一段时期的总日数是 n；雷暴日 (A) 的频数是 m_A，$P(A)=\frac{m_A}{n}$；雨日 (B) 的频数是 m_B，$P(B)=\frac{m_B}{n}$；雷暴日同

时又是雨日的频数是m_{AB}，$P(AB)=\frac{m_{AB}}{n}$；无雷暴又是不下雨的频数是m_0。这几种频数的关系如图 1.1 所示。

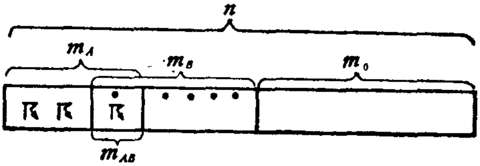

图 1.1

ꓘ 为雷暴　　　　• 为下雨

可见，雷暴日和雨日两者至少出现一种（包括两者同时出现）的频数

$$m_{A+B}=m_A+m_B-m_{AB}$$

$$P(A+B)=\frac{m_{A+B}}{n}=\frac{m_A+m_B-m_{AB}}{n}$$

$$=\frac{m_A}{n}+\frac{m_B}{n}-\frac{m_{AB}}{n}$$

$$P(A+B)=P(A)+P(B)-P(AB) \qquad (1.4)$$

如两事件互不相容，不可能同时发生，$P(AB)=0$，(1.4)式就成为(1.1)式。

要弄清几种天气之间较为复杂的关系，还要研究条件概率。

第三节　直接应用条件概率作天气预报

1. 条件概率

某站一个月 30 天内出现东风 9 天，东风同日有雨 8 天，出现东风条件下有雨的概率是 8/9，这个概率就是在东风条件下有雨的条件概率。用符号表示，在事件 B 出现的 m_B 次中，事件 A 出现 m_A 次，则在事件 B 的条件下出现事件 A 的

条件概率是

$$P(A|B)=\frac{m_A}{m_B}$$

这里我们把 B 条件下出现 A 的概率记为 P(A|B)。

该月东风(B)的概率$P(B)=\frac{9}{30}$，该月内东风同时有雨(AB)的概率 $P(AB)=\frac{8}{30}$。上面已说明东风条件下有雨的条件概率 $P(A|B)=\frac{8}{9}$。这几种事件的频数的关系参看图 1.2。

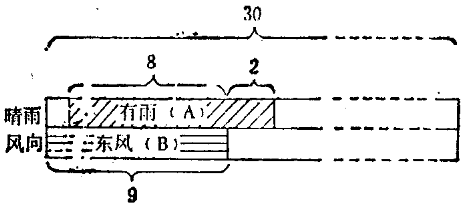

图 1.2

显然

$$\frac{8}{30}=\frac{9}{30}\times\frac{8}{9}$$

即

$$P(AB)=P(B)\times P(A|B) \tag{1.5}$$

或

$$P(A|B)=\frac{P(AB)}{P(B)} \tag{1.6}$$

假定该月雨日共 10 天，雨日概率 $P(A)=\frac{10}{30}$。雨日同时有东风 8 天，则雨日条件下出现东风的条件概率 $P(B|A)=\frac{8}{10}$，显然

$$\frac{10}{30}\times\frac{8}{10}=\frac{8}{30}$$

即 $$P(A)\times P(B|A)=P(AB) \tag{1.7}$$

或

$$P(B|A)=\frac{P(AB)}{P(A)} \tag{1.8}$$

由(1.5)和(1.7)式得出

$$\begin{aligned} P(AB)&=P(B)\times P(A|B) \\ &=P(A)\times P(B|A) \end{aligned} \tag{1.9}$$

(1.9)式表达了两种事件的几种概率的关系。

直接应用条件概率可以构成许多简易的统计预报方法，下面介绍综合概率法，又称多因子综合相关。

2.综合概率法

这种方法的实质，是将各个因子出现一定数值的条件下

表 1.3

序号 \ 列数	1	2	3	4	5
序号 \ 变量	x_1	x_2	x_3	y	$\hat{p}$
1	0	0	0	0	0.28
2	1	1	0	1	0.62
3	0	0	1	0	0.35
4	0	1	1	1	0.55
5	0	1	0	0	0.48
6	1	0	0	0	0.42
7	1	0	1	0	0.49
8	1	1	0	1	0.62
9	1	0	1	1	0.49
10	1	1	1	1	0.69

预报对象出现的条件概率，加以平均，作为预报的依据。

(1) 两级预报举例。

将预报对象 y 和因子 x_1, x_2, x_3 分别根据一定的标准编成 0 或 1，称为(0，1)变量，其资料见表 1.3 的第 1—4 列。第 5 列 $\hat{p}$ 用来衡量当 x_1, x_2, x_3 出现某种组合①的条件下，$y=1$ 的可能性，是有待于计算的。

先作出 x_1 和 y 的联合频数表(表1.4 a 左半部分)

表 1.4a

		y		合计	$P(y\|x_1)$	
		0	1			
x_1	0	3	1	4	$P(y=0\|x_1=0)=\frac{3}{4}=0.75$	$P(y=1\|x_1=0)=\frac{1}{4}=0.25$
	1	2	4	6	$P(y=0\|x_1=1)=\frac{2}{6}=0.33$	$P(y=1\|x_1=1)=\frac{4}{6}=0.67$
合计		5	5	10		

再统计出条件概率 $P(y|x_1)$。例如，$x_1=0$ 共 4 次，其中对应 $y=0$ 有 3 次，$y=1$ 有 1 次，则

$$P(y=0|x_1=0)=\frac{3}{4}=0.75$$

$$P(y=1|x_1=0)=\frac{1}{4}=0.25$$

记在表 1.4 a 的右半部分。

同样分别作出 x_2, x_3 和 y 的联合频数表(表 1.4 b,c 左半部分)，并统计出条件概率(右半部分)。

① 这里所指因子的某种组合，是指各因子分别取一定数值的状况，如"$x_1=0, x_2=1, x_3=0$"算是一种组合。两个(0，1)化的因子有四种组合，三个(0，1)化的因子有八种组合。可表达为

$$Q=2^n$$

Q 为组合数，n 为因子数。