数理统计在天气预报中的某些应用

"数理统计学习班"讲义

新疆维吾尔自治区革委会气象局气象台

一九七三年元月

“读书是学习，使用也是学习，而且是更重要的学习。从战争学习战争——这是我们的主要方法。没有进学校机会的人，仍然可以学习战争，就是从战争中学习。革命战争是民众的事，常常不是先学好了再干，而是干起来再学习，干就是学习。”

毛泽东：《中国革命战争的战略问题》

“要理智地、自觉地、有效地投身于革命，就必须学习。”

列　宁：《论彼得格勒公共图书馆的任务》

“要建设，就必须有知识，必须掌握科学。而要有知识，就必须学习。顽强地、耐心地学习。”

斯大林：《在苏联列宁共产主义青年团第八次代表大会上的演说》

前　　言

统计研究方法的广泛应用，早以为我们大家所熟知。在社会科学中，应用统计的研究方法，形成了社会经济统计学的一套体系。而在自然科学中，探讨随机变动的数量（即受一些偶然因素的影响而变动的数量）的特征和规律性的统计研究方法，则称作数理统计学。它主要是应用数学的原理和方法，根据所研究对象的全部数字资料中的一部份资料，去推断全部对象的数量特征和规律性来。它无论在物理学、化学、天文、水文、地质、生物、工业技术（如产品检验）、农业试验和气象等各门学科中，都得到了广泛的应用。

在气象学中，对大气运动规律的认识和怎样做好天气预报等问题，长期存在着两种基本观点：一种认为，大气运动规律是确定的，因此，只要台站网的分布是足够广泛和稠密，观测资料足够多，那么就可按照流体力学和热力学的原理，对大气运动进行数学描写，借助于大型的数字电子计算机，使天气预报逐步准确。这就是流体力学数值预报。另一种观点则认为：天气现象不是确定的，而是随机的，即使观测资料足够多，也不能把大气过程完全描写出来，因而必须应用数理统计的方法，预报未来天气出现的可能性，这就是概率统计预报。由于目前流体力学数值预报的局限性，尚不能满足国防和国民经济建设对天气预报的需求，许多复杂的天气现象的预报、局部天气预报和中长期天气预报，还只能依靠统计方法来制作；同时，随着概率统计预报的发展，人们还有可能通过大气运动的统计规律的了解，进一步增进，加深对大气动力学规律的认识，促进数值预报和天气学预报方法的发展，更重要的还有，概率统计预报能利用少数台站的资料，作出较大

范围或较远地区的天气预报，对战时的天气预报，具有其独特的价值。因此，研究和发展概率统计预报，是具有很大的现实意义的。尽管它还存在着一些严重的弱点，例如，缺乏物理基础，不能报出历史资料中未出现过的天气现象，在气候演变发生特殊变化时，预报常常失败；甚至大气运动中是否确实存在概率统计的规律，也还不能取得一致的看法等等。但由于它具有上述的实际应用价值，以致近年来，仍然越来越引起人们的重视，并获得了较大的发展。

概率统计天气预报的发展，最初主要是根据预报人员的经验和天气学的知识，利用本站或外站前期的地面资料，挑选数量较少的预报参数，采用多元迴归等方法，建立统计预报关系式，预报未来的天气。当然，远在天气图出现之前，劳动人民当中就流传着大量的天气谚语和测天经验，实际也是天气现象之间统计关系的生动反映。后来，由于现代计算技术和流体力学数值预报的发展，概率统计预报也发展到不仅可以利用本站和外站的高空观测资料，而还可以利用流体力学数值预报的环流形势预报等结果；统计分析方法，也发展到可以借助于电子计算机，应用筛选法，从大量的（如数百个）预报参数中，挑选一定数量（如几个到几十个）的预报参数进行迴归，作出效果较好的预报关系式；应用判别分析等方法作气象要素的等级预报；应用经验正交函数的分析方法，时间序列的外推方法以及其它一些数理统计方法，如马尔可夫链，二阶矩过程，线性规划，统计决策等等，进行天气分析和制作预报。近年来，则开始了将概率统计预报和流体力学数值预报由双方平行发展转向两者相互结合起来的试验研究。

我们从广大劳动人民的丰富的测天经验以及自己长时期的天气预报实践中，积累和发现了多种多样的前兆未来天气变化的预报经验和指标，有些在动力气象和天气学中一时还得不到解释，而用数理统

计的方法，却能收到实际的预报效果，也使我们越来越感到学习和应用一些数理统计的方法来提高气象工作服务质量的必要性和迫切性。它正是我们举办这期学习班的主要目的和任务。

但是由于我们的水平很低，条件有限，在这期学习班里，还只能和大家一起共同学习一些数理统计中比较基本的知识及其在天气预报某些方面的初步应用。同时，为遵照毛主席关于理论联系实际的教导，我们将着重介绍一些基本的概念，以及在某些应用过程中的基本操作方法，而尽量省略掉那些比较复杂的数学推导。难免会有这样或那样的错误和缺点，欢迎在学习中共同讨论，共同提高。

目　次

第一章　　概率论的基础知识

自然界的各种现象，有些是具有明显的、确定的因果关系的，而有些现象具有一定的偶然性，我们通常称这类现象为随机事件。例如，掷一个硬币，那一面向上？任选一天是否下雨？下多大的雨，等等。它们都受到很多因素的影响，而这些因素我们都无法控制，但我们却不能说这种类型的随机现象就毫无规律的了。其实，它们也是受着一定的规律所约束、所支配的。也就是说，还是有一定的规律可寻的。概率论就是研究这类随机现象在某些情况下出现可能性的大小和它的某些规律性的一门数学科学分支。

在这一章里，我们将结合一些气象现象，介绍一些概率论最基本的知识。

§ 1　概率的概念·经验概率

1.事件，事件的种类

在概率论中，一般把所研究的现象称之为事件。

在进行一项试验之前，如果我们可以断定，某一现象必定会在试验结果中出现，我们就把这一现象叫做必然事件。

例如，我们做掷骰子的游戏（骰子为六面正方体，各面上分别刻有1.2.3.4.5.6的数字）每掷一次（一项试验），"骰子上向上一面的数字小于7"，就是一个必然事件。

又如，某地遭到寒潮侵袭时（成立），则"气温下降"也是一个

必然事件。

在进行一项试验之前，我们可以断定某一现象决不会在试验结果中出现，则称这种现象为不可能事件。

例如，掷一个骰子，“骰子向上一面的数字大于7”就是一个不可能事件，某地区一天里，“同时出现当地年的极端最高、最低气温”也是一个不可能事件。

还存在着另一种现象，它在一项试验结果中，可能出现，也可能不出现，我们就把这类现象叫作随机事机（或称概然事件、偶然事件）。

例如，一粒骰子投掷一次，骰子的向上一面，，或出现为1、或为2、或为3、或……或为6，共有六种可能性，此时，出现1—6当中的那一个数字，就是随机事件。同样，某地受到寒潮侵袭时，气温下降10℃（或者9℃、11℃等等），也是随机事件。

2 概率的定义

上述随机事件，在自然现象中是大量存在的，只不过有些随机事件，或许有较大的出现可能性， 另一些随机事件，则具有较小的出现可能性。而它们出现的可能性大小，又是由它们本身的客观规律所确定的，并且是可以用数量来表示的。

例如，掷骰子的次数多了，就会发现每掷5000次中，骰子的某一面（如刻有数字1的一面）经常出现在1000次左右，显然，用 $\frac{1000}{6000}=1/6=0。17$ 来表示 在每次投掷中，刻有数字1出现的可能性，是适当的。这种描写概然事件在试验结果中出现可

能性的大小的数量，就是概率。

概率的定义可写作：如果事件A在同样的n次试验中出现了m次，则我们把$\frac{m}{n}$称为事件A在n次试验中出现的频率。如果试验的次数n逐渐加大时，事件A的频率愈来愈稳定地在一个常数p的附近做微的摆动，我们便说事件A的概率是p，并用符号p（A）表示事件A的概率。亦即：

$$p(A)=p$$

在一般情况下，常数p是不可能精确得到的。因此，通常便以n充分大时，事件A的频率作为事件A的概率p的近似值。

$$p(A)=p\approx\frac{m}{n}$$

例如，根据阿拉山口气象站61年至69年来每天观测的大风（≥17。2米／秒）日数的积累资料（如表1），我们可以近似用0。46来近似地表示阿拉山口大风日数的概率，也就是说，该地每一天里出现大风的可能性近似地为0。46。

表1　　阿拉山口气象站的大风日数的积累资料

年　代	61	62	63	64	65	66	67	68	69
积累观测日数	365	730	1095	1460	1825	2190	2555	2920	3285
积累大风日数	163	330	512	678	844	1015	1192	1355	1500
大风频率	0.45	0.45	0.47	0.46	0.46	0.46	0.47	0.46	0.46

3.经验概率：

在数理统计中，把用多次试验的频率估计出的概率称为经验概率。

当试验次数n相当大时，经验概率可以根据已有资料，按照频率公式来进行计算。但我们气象台站一般资料年代较短，尤其是对于长期天气预报，所用资料不足50年，即$n\leq 50$，对此，有时采用经验概率的修正公式：

$$p^{*}=\frac{m}{n+1}$$

进行计算。或称之为经验频率更为合适一些。

§2　概率的基本性质

根据概率的定义，可以得到概率的几个基本性质：

1. $0\leq P(A)\leq 1$。即事件A的概率不小于0，不大于1。

$\because 0\leq m\leq n$

2. P（必然事件）＝1，即必然事件的概率等于1。

3. P（不可能事件）＝0，即不可能事件的概率等于0。

4. 概率的加法定理：

如果事件A与事件B不可能在试验结果中同时出现，则称事件A与事件B是两个互不相容事件。

概率的加法定理可写作：

$$P(A\text{或}B)=P(A)+P(B)$$

其中事件A与事件B互不相容。

由此类推，概率的加法定理可写作：

$P(A_1\text{或}A_2\text{或}A_3\text{或}\cdots\cdots\text{或}A_n)=P(A_1)+P(A_2)+$

$\cdots\cdots + P(A_k)$ 其中事件 A_1、A_2、$A_3\cdots\cdots A_k$ 等 k 个事件是两两互不相容的。

例如，某站定时观测百叶箱内的气温，每次观测只能有一个读数，这种情况下，"气温为18.1℃"与"气温为18.2℃"、"气温为18.3℃"、"气温为18.4℃"和"气温为18.5℃"这五个事件，都是两两 互不相容的。假定它们在该站出现的概率分别为0.03、0.04、0.05、0.06、0.07，则该站定时观测的百叶箱气温在18.1到18.5℃之间的概率应是：

P(18.1℃或18.2℃或18.3℃或18.4℃或18.5℃)＝0.03＋0.04＋0.05＋0.06＋0.07＝0.25

5.独立事件的概率乘法定理：

如果在进行两项试验时，无论第一项出现什么结果，都对于第二项试验出现何种结果毫无影响，则称这两项试验是相互独立的或这两项试验是独立试验。

在两相互独立的试验中，第一项试验结果中的任何事件都与第二项试验结果中的任何事件相互独立，即相互之间毫无影响关系。或称不相关事件。

独立事件的乘法定理就是：两独立事件同时出现的概率等于该两事件的概率的乘积。即：

$$P(A\text{及}B) = P(A)\cdot P(B)$$

由此类推，如果事件 A_1、A_2、$A_3\cdots\cdots A_k$ 都相互独立，则

$p(A_1及A_2及A_3及\cdots\cdots及A_k)=p(A_1)\cdot p(A_2)\cdot p(A_3)\cdots\cdots\cdot p(A_k)$。

例如，假定某地3月出现异常冷暖的概率为0。3，6月份出现的异常冷暖的概率为0。2，10月份出现的异常冷暖的概率为0.1，而这三个月是否出现异常冷暖，彼此毫无关系，则在同一年中，这三个月份都为异常冷暖的概率是：

$$0.3\times0.2\times0.1=0.006$$

6.事件与其逆事件的概率互补：

我们把A不出现的事件称为事件A的逆事件记作$\bar{A}$，因为事件A或者出现，或者不出现，二者必居其一，也就是说，事件A与事件$\bar{A}$即不可能同时出现，也不可能都不出现。因此可以得到：

$$p(A)+p(\bar{A})=1$$

反言之，如果两事件的概率之和等于1，我们就该称两事件互补。

由此类推，如$A_1A_2A_3\cdots\cdots A_n$是n个两两互不相容的事件，且试验结果中，必定有其中之一出现，则：

$$p(A_1)+p(A_2)+p(A_3)+\cdots\cdots+p(A_n)=1$$

例如，对总云量的观测规定按10个分量来记载，0，1，2，……10分量的11个数字，在每次观测中，必须记载其中的一个数字，而且只能记载其中的一个数字（两两互不相容），这11个记载云量的数字出现的概率总合等于1。

还有一种情况，若$A_1A_2A_3\cdots\cdots A_n$等n个事件是某一组试

验的全部可能结果，它们两两互不相容，且都有同样的可能性出现，此时，因

$$p(A_1\text{或}A_2\text{或}A_3\text{或}\cdots\cdots\text{或}A_n)=p(A_1)+p(A_2)+$$
$$+p(A_3)+\cdots\cdots+p(A_n)=1$$

以及 $p(A_1)=p(A_2)=p(A_3)=\cdots\cdots=p(A_n)=p$

而得到：
$$p(A_1)+p(A_2)+p(A_3)+\cdots\cdots+p(A_n)=$$
$$=p+p+p+\cdots\cdots+p=np=1$$

亦即：$p=\frac{1}{n}$

例如，掷骰子的游戏中，因骰子是结构匀称的六面正方体，每一面的出现（指在正上面）都有同等的可能性，因此每一面（从所刻数字1到6）出现的概率都是1／6。

§3　随机变量及其概率分布

1.随机变量的概念

我们逐日所观测到的温、压、湿、天空中的云量，阴、雨、降水量以及阴雨的日数等各种气象要素，都是受许多因素的影响而变动的数量。这些因素，我们无法控制它们（至少目前是这样），使它们不变，因而受这些因素影响的数量也就不可能不变。

我们把这些无法控制以使它们不变的这些因素称为随机因素。把受随机因素的影响而变动的数量称为随机变量。

2.随机变量的两个种类。

㈠、离散的随机变量：通常是指它所可能取得的数值，是可以例举出来的那种随机变量。

㈡、连续的随机变量：则是指它可以取得一个区间内的任意实数值的那种随机变量。

在气象统计中，常常把气压、温度、降水量、湿度、云量、高空等压面的高度、风速等当作连续的随机变量来处理，而把诸如降水日数、冰雹、雾日数以及其它气象现象所出现的日数、初终期等作为离散的随机变量来处理。

3.随机变量的概率（或频率）分布。

㈠、随机变量概率分布的概念。

我们先举个例子，假定某射击手根据他多次射击记录得到，在他10000次的总射击次数中，没有射中环靶（记作0环）的为100次，射中1环（图1·1）的有400次，射中2环的500次，射中3环的1000次，射中4环的3000次，射中红心（5环）的5000次，则射中各环的概率可例如表1.2：

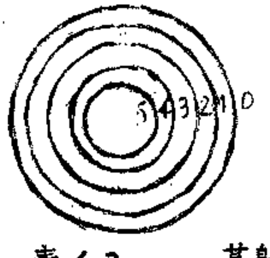

图1 1射击环靶

表1.2　　某射击手射中各环的概率表

射中的环数	0	1	2	3	4	5
相应的概率 P	0.01	0.08	0.05	0.10	0.30	0.50

上例中，射手射中的环数即为随机变量所可能取得的值，它们都各有自己不同的出现概率。与这种随机变量取得各种可能值相对应的出现概率，就构成随机变数的概率分布。通常，随机变量在各项试验中出现那一个值是偶然的，但它的各个可能值所出现的概率则是有客观规律的。对于这种概率分布规律的研究，在概率论当中是占据着很重要的地位的。

(二)、随机变量概率分布的表示法。

它一般有三种表示法，一是表列法，二是图示法，三是解析函数表示法。

在气象业务中，由于观测年代的限制，一般都直接用频率来代替概率进行统计。

例如，新源站曾根据1956年至1963年的资料，统计过夏季天气图上各类影响系统入侵时，本站各等级降水量的出现频率，以强锋区（指塔什干——阿拉木图一带500mb西风风速≥16米／秒）短波小槽的入侵为例，其降水等级的频率分布可列如下表：

表1.3. 强锋区小槽入侵时，新源站各降水等级的频率分布

各降水等级	大雨	中雨	小雨	无雨	合计
出现次数	9	4	34	3	50
相对频率	0.18	0.08	0.68	0.06	1.0

一般气象要素和天气现象的频率，通常采用图示法来表示，如图1.2表示了乌鲁木齐冬季（12月——2月）各种月降水日出现频率的分布图。图中表明8天到10天的降水日数出现的频率最高，其次是5天到6天。全月无降水日以及降水日数多于17天的可能性是极小的，（在这部份资料里没有出现过）。

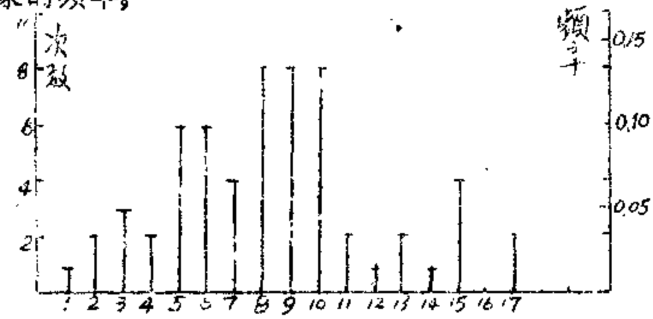

图1.2 乌鲁木齐冬季(12-2月)月降水日数频率分布图 (1951-1970)

图1.3.表示了乌鲁木齐七月份每天4次（02、08、14、20时）定时观测的各级总云量的频率分布图，其中无云的频率最高，阴天（8——10分量的云）的频率次之，多云的频率则最低表现为凹形的图形。

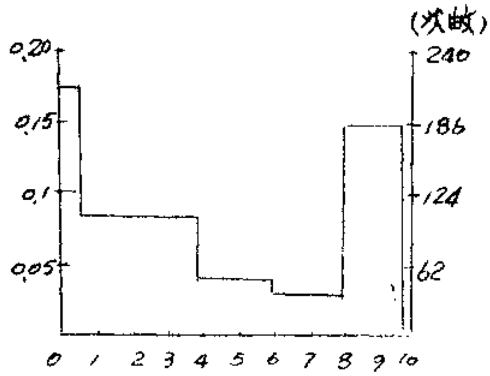

图1.3. 乌鲁木齐七月份定时观测各级总云量的频率分布图（1961～1970年共1237次记录）

图1.4是乌鲁木齐七月份各种日降水量的频率分布，从微量到5毫米之间的频率急剧下降，此后，随着降水量的增大，它们出现的频率都比较小，并且逐渐地减小，表现为乙形的图形，图中的平滑曲线