气象统计预报讲义

中国人民解放军空字六二三部队

一九七五年一月

马 克 思 语 录

在科学上面是没有平坦的大路可走的，只有那在崎岖小路的攀登上不畏劳苦的人，有希望到达光辉的顶点。

恩 格 斯 语 录

……；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，如此等等。

列 宁 语 录

我们不需要死读硬记，但是我们需要用基本事实的知识来发展和增进每个学习者的思考力，因为不把学到的全部知识融会贯通，共产主义就会变成空中楼阁，就会成为一块空招牌，~~共产主义者~~也只会是一些吹牛家。你们不仅应当领会~~你们学到的知识~~，并且要用批判的态度来领会这些知识，~~使自己的头脑~~不被一堆无用的垃圾塞满，而能具备现代~~有学识的人~~所必备的一切实际知识。

毛主席语录

你们学自然科学的，要学会用辩证法。

在生产斗争和科学实验范围内，人类总是不断发展的，自然界也总是不断发展的，永远不会停止在一个水平上。因此，人类总得不断地总结经验，有所发现，有所发明，有所创造，有所前进。

胸中有“数”。这是说，对情况和问题一定要注意到它们的数量方面，要有基本的数量的分析。任何质量都表现为一定的数量，没有数量也就没有质量。

要自学，靠自己学。

目　录

第一章 概率统计基础

“人的认识物质，就是认识物质的运动形式”。人们对大气运动规律的认识，长期以来存在着两种基本观点。一种认为：大气运动规律是确定的，只要观测资料足够多，根据流体力学和热力学的原理，对大气运动进行数学描述，借助于大型数字电子计算机的计算，做出天气预报，这就是流体力学数值预报。另一种观点则认为：大气运动规律不是完全确定的，而是随机的，即使观测资料足够多，也不能把大气运动规律完全描写出来，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，这就是数理统计天气预报或称概率统计天气预报，简称统计预报。

统计预报是应用概率论与数理统计的方法，分析大量的天气演变的历史资料，概括得出预报量（对象）和预报因子（指标）之间的统计性的关系。

概率论与数理统计是在三大革命实践的基础上产生的。它在气象科学上越来越得到广泛的应用，它是统计预报的理论基础。因此，我们在学习具体的统计预报方法之前首先应对概率论与数理统计的基本知识要有所了解。

概率论与数理统计方法是研究大量随机现象的数量特征和规律性的统计研究方法。所谓随机现象就是偶然出现的现象。恩格斯说：**“偶然的东西是必然的，而必然的东西又是偶然的”**。因此，随机现象也是具有规律性的，这种规律性是应用概率论与数理统计的研究方法揭露出来的。例如在天气预报中，寒潮侵袭一定要降温这是必然的，但气温可能降多少则具有偶然性的一面，可以把它作为“随机现象”来研究。我们只要累积大量有关的资料，应用概率论与数理统计的方法，就可以找到在不同情况下的寒潮入侵气温可能降多少的规律性，并运用这种规律预报未来寒潮降温的可能数值。

概率论与数理统计方法主要是根据研究对象的全部资料（称总体）中的一部份资料（称样本）去推断总体的规律性。同样以寒潮降温为例，我们所掌握的有关寒潮天气的资料（即样本）总是有限的，而自从有寒潮天气以来直到将来还要发生寒潮的年代（即总体）是无限的。当我们从样本资料（比如五年历史资料）应用概率论与数理统计方法找到了寒潮降温的某些“规律”，而这种“规律”是否能代表（或接近）总体的规律（即真正的规律）呢？如果与真正的规律相差很远那就没有预报意义了。因此这一问题也要应用概率论与数理统计的方法去解决。

应用统计预报方法，不管大气现象多么复杂，通过对大量资料的统计分析，就能比较客观地概括出它们的统计规律性，进而作出比较客观的天气预报。但是，统计预报不能保证每一次预报都是正确的，尤其对那些与历史规律相差较大的特殊情况一般报不出来。因此，要求我们要十分重视与其他预报方法（如天气图方法等）配合使用，取长补短，以利提高预报准确程度，搞好军事气象保障工作。

第一节　事件与概率

一、事　件

在大气中有各种各样的天气现象，如夏半年冷锋过境可能出现雷暴；冬季强寒潮入侵时可能引起霜冻；盛夏副热带高压脊控制时会有酷热等等。这些现象在概率统计中常被称为事件。在大气中各种事件的发生是各式各样的，为了研究它们的规律性，常常把事件分为三种类型。

（一）随机事件

随机事件的定义是：在一定条件下，若某种事件可能发生也可能不发生，这种事件就称为随机事件。例如，夏季冷锋影响某测站，该站可能有雷暴发生，但也可能无雷暴发生，故"冷锋影响某站出现雷暴"就是随机事件。又如某站在偏东风情况下有利于出现低云，但不一定一吹东风就出现低云，因此这种"某站在偏东风情况下出现低云"就是随机事件。还有"南京夏季日最高气温≥35°C的日数"也是随机事件。总之天气预报的对象都是随机事件。

随机事件在大量重复试验（观测）中，就其个别来说出现与否是具有偶然性的，但是就其总体来说是有统计规律性的。因此我们的任务就是要从大量的气象要素变化中去寻找、总结预报对象与预报因子之间的统计规律性，利用这种规律性作出天气预报。

随机事件在一次观测中可能出现也可能不出现，即具有偶然性，这样是否就认为它不受客观规律支配了呢？不是的。从辩证法的观点来看，自然界所发生的现象，总是同时具有必然性和偶然性。恩格斯指出：**"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"**，因此二者是对立的统一，如普遍性和特殊性是对立的统一一样。例如在南京一年中夏季气温最高，这是必然性的一面，但夏季日最高气温≥35°C日数一年有多少就是偶然性的一面。事物的必然性是由一种或几种主要的，相对稳定的因素决定的，有明确的因果关系（如南京一年中夏季气温最高，是由太阳高度角较高，日照时间较长所决定的）。事物的偶然性则是众多的、相对来说是次要的、变化不定的因素所造成（如南京夏季日最高气温≥35°C日数一年有多少，则还与太平洋副热带高压强度、是否控制南京地区，天空状况，风和降水的情况等很多因素有关）。因此我们说，随机事件就其个别来说也是有因果性的，受客观规律支配的。

（二）必然事件和不可能事件

那种在一定条件下进行大量重复试验时每次都肯定发生的事件称为必然事件（用U表示）；反之，那种在一定的条件下进行大量重复试验时每次都肯定不发生的事件，就称为不可能事件（用V表示）。显然它们是随机事件的两种极端情形。例如，在标准大气压力下，"水加热到100°C时沸腾"是必然事件，而"水加热到50°C沸腾"是不可能事件。又如，"某平原测站在寒潮侵袭时气温下降"是必然事件，"某地在同一天出现当年的极端最高、

极端最低气温值”是不可能事件。

从上面所举的例子中可以看出，必然事件和不可能事件之间有着密切的联系。如果在一定的条件下某个事件是必然事件，那么在同样条件下那一事件的反面就必然是不可能事件。反过来也是一样。

二、概　　率

（一）概率的概念和性质

1．概率的概念：

前面曾介绍过的随机事件，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，而另一些随机事件则具有较小的出现可能性。如在我国北方冬季出现下雨的可能性比南方要小一些，而北方冬季出现下雪的可能性要比南方大一些。这些随机事件（如下雨、下雪等）出现可能性的大小，都是由它们本身的客观规律所确定的。也就是说，随机事件发生的可能性的大小是事件本身的客观规律所确定的。即随机事件发生的可能性的大小，是事件本身所固有的不随人们主观意愿而改变的一种属性。随机事件的这种属性，正是可以对它发生的可能性大小进行度量的客观基础。因此，我们可以用一个数$P(A)$来作为事件A发生的可能性大小的数值表征。称这个数值$P(A)$叫事件A发生的概率。

例如，在乒乓球比赛开始前，裁判员掷一枚均匀硬币，看它出现正面还是反面，来决定谁优先发球和选取场地，这是基于出现正、反面具有等同可能性的认识。在历史上，有人曾在相同条件下，把一枚均匀硬币重复掷多次，所得试验结果见表1—1及图1—1。

表1—1

掷硬币次数 n	出现正面次数 m	出现正面的频率m/n
4048	2048	0.5080
12000	6019	0.5016
24000	12012	0.5005

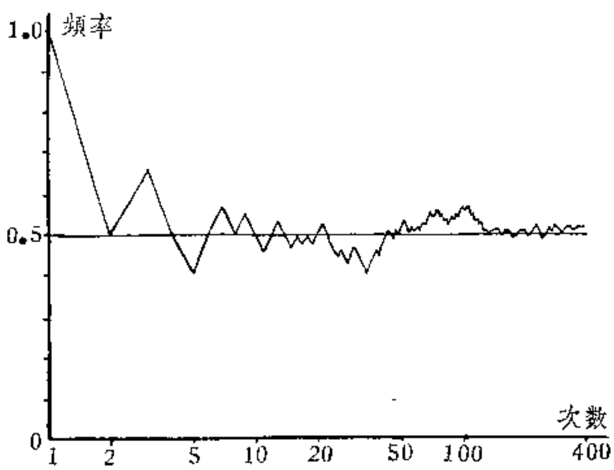

图1—1

我们从图和表中看到，出现正面的频率在0.5附近摆动，当n越大时，在0.5附近摆动的幅度越小。于是可以认为掷一枚硬币出现正面与反面具有等可能性，我们便取0.5作为出现正面的概率。

2．概率的定义：

上面所述是概率的概念，现在我们给出概率的定义如下：假设我们重复地进行同一试验，如果随机事件A在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于某一个常数p的附近作微小的摆动，我们便说事件A发生的概率是p，亦即：

$$P(A)=p$$

在一般情况下，常数p是不可能精确地得到的。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。亦即：

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

因此，频率是概率的近似估计值，它通过大量的重复试验将概率显示出来了。

这一点从上面的例子已经明显的看出来了，下面我们再举一个气象上的例子。例如，我们要想知道上海年雨日出现的概率，则可以根据概率的定义，应用上海雨日的历史资料（见表1—2），通过计算频率去近似的估计。

表1—2　上海1881—1970年逐年雨日资料

	1	2	3	4	5	6	7	8	9	10
1880	135	149	142	144	127	117	119	132	167	135
1890	141	116	132	122	122	139	146	134	136	133
1900	107	126	119	124	147	142	132	141	148	140
1910	152	139	120	133	121	144	118	137	138	147
1920	132	133	126	128	123	130	123	112	113	147
1930	154	134	143	108	153	134	150	120	96	89
1940	115	92	96	109	93	113	96	101	121	147
1950	148	147	143	161	120	134	147	127	133	135
1960	139	129	116	132	126	128	106	113	140	133

从表1—2所给出的资料中发现，在90年中雨日最多的年份有167天(1889年)，最少的年份只有89天（1940年），如果把每年雨日分别除以该年的总日数（365天或366天），就得到了每年雨日的频率。频率最大的是45.8%，最小的是24.3%，假如把五年、十年、十五年、二十年、……、九十年各求一个频率，则可得到如表1—3所给出的结果（以%为单

位）根据表1—3的结果绘出图1—2。

表1—3

年数	5	10	15	20	25	30	35	40	45
频率	38.2	37.4	36.5	36.8	36.2	36.6	36.6	36.7	36.5
年数	50	55	60	65	70	75	80	85	90
频率	36.3	36.5	36.1	35.5	35.2	35.5	35.6	35.6	35.5

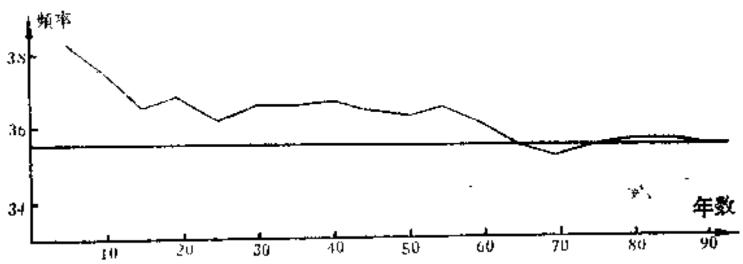

图1—2

从上面的图和表中可以看出，随着计算频率时间的增加(即年数不断增加)，雨日频率有稳定于某一个常数值（即为p）的趋势，这里九十年的频率35.5%就是概率p的近似估计值。这里要注意，频率与概率之间不是完全相等的概念，频率是个实测值而概率却是个理论值。频率值随试验次数不同而异，概率值却是一个不变的常数。只有当试验次数不断增加，频率渐稳定时，才能用它来代替概率值。否则，资料年代短，频率不稳定，则事件频率不等于其概率。所以在实际工作中有较长的记录，才可以把计算的频率作为概率的近似值。

3．概率的基本性质：

根据概率的定义，可以得到它的几个基本性质：

（1）必然事件的概率等于1。即

$$P(U)=1$$

这是很明显的，因为必然事件在n次试验中必定出现n次，所以必然事件的频率恒为1，则频率稳定于某一个常数值p当然就是1。

在实际应用中，若一事件的概率接近1，则我们认为该事件叫实际必然事件。例如用某种统计预报方法计算出本站未来24小时内出现雷暴的概率为0.92，则我们把未来24小时内出雷暴看成是实际必然事件，因此预报未来24小时内将会有雷暴出现。

（2）不可能事件的概率等于零。即

$$P(V)=0$$

这个性质也同样是很明显的，因为不可能事件在n次试验中必定出现零次，所以不可能事件的频率恒为零，则实际频率稳定所趋的那个数值当然是零，即概率p为零。

在实际应用中，若一事件发生的概率很小，接近0，则称该事件叫实际不可能事件。例如用某种统计预报方法计算出本站未来出低云的概率是0.16，则我们认为未来出低云的可能性很小，看成实际不可能事件，报未来不出低云。

（3）随机事件A的概率永远不会小于0，也不会大于1。即

$$0 \leqslant P(A) \leqslant 1$$

因为$0 \leqslant m \leqslant n$，所以随机事件$A$的频率$\frac{m}{n}$决不能小于零（$\frac{0}{n}=0$），也决不可能大于$\frac{n}{n}=1$，则频率趋于稳定的数值$p$当然不可能小于零或大于1。

（二）概率的加法定理与乘法定理

上面我们讨论了事件和它的概率这两个概率论中最基本的概念，今后我们将用A、B、C等表示事件，而用符号$P(A)$、$P(B)$、$P(C)$等表示A、B、C等所对应的概率。

大家知道，**"在自然界中没有孤立发生的东西"**。如果我们只是一个个地、孤立地来研究事件及其概率是很不够的，实际生活中往往要求我们在考虑任何一个随机事件时，还需同时考虑与之联系的种种事件以及这些事件之间的关系，它们的概率之间的联系等等。

1．概率的加法定理：

在未介绍加法定理之前，我们先说一下互不相容事件和相容事件的含义。如果事件A与事件B不可能在试验结果中同时出现，则事件A与事件B是两个互不相容的事件。例如某站同一时刻"晴"与"阴"构成互不相容事件，因两者不可能同时发生。如果两事件能够同时发生则称为相容事件。例如A表示"云量在4成以上"的事件，B表示"云量在8成以下"的事件，则A、B可以同时出现，故为相容事件。

假如在n次（n相当大时）试验中，事件A出现m_1次，事件B出现m_2次。由于A、B互不相容，所以在A事件出现m_1次中事件B皆未出现，同样在B事件出现m_2次中事件A皆未出现。因此在n次试验中A出现或B出现的次数共m_1+m_2次，故知"A或B出现其一"的概率为：

$$P(A+B) \approx \frac{m_1+m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} \approx P(A)+P(B)$$

即

$$P(A+B)=P(A)+P(B) \qquad (1—1—1)$$

上式说明，如果事件A与B互不相容，则"事件A或事件B出现其一"的概率$P(A+B)$。等于事件A的概率$P(A)$与事件B的概率$P(B)$之和（即事件和的概率等于各事件概率之和）。这就是概率的加法定理。

例如，某地从多年资料得出某月晴天（A）出现的概率$P(A)=12/30$，阴天（B）出现的概率$P(B)=8/30$，那末出现"晴天或阴天"的概率$P(A+B)$为

$$P(A+B)=P(A)+P(B)=12/30+8/30=20/30$$

概率的加法定理可以应用于许多互不相容事件的概率相加。即

$$P(A_1+A_2+\cdots\cdots+A_k)=P(A_1)+P(A_2)+\cdots\cdots+P(A_k) \qquad (1-1-2)$$

例如，某地一月份吹北风的概率为23%，东北风的概率为26%，西北风的概率为8%。因此某地一月份偏北风的概率为23% + 26% + 8% = 57%。

利用概率的加法定理，可以得到互逆事件的概率之和为1。

所谓互逆事件，是指不能同时发生而又必然发生其一的两事件。例如A为“有雨”事件，则A的逆事件为“无雨”事件，记为$\overline{A}$，显然互逆事件必然是互不相容事件。对某地某一时刻来说，A与$\overline{A}$只能出现其中一种而必然出现其中一种。因此“A或$\overline{A}$出现”的概率之和为1。即

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

上式可得：

$$P(A)=1-P(\overline{A}) \qquad (1-1-3)$$

利用此式可由$P(\overline{A})$来间接计算$P(A)$往往是很简便的。例如作某站降水概率预报时，若报出无雨的概率为$P(\overline{A})=0.3$，则$P(A)=1-0.3=0.7$就是出现降水的概率。

如果有A_1，A_2，……，A_k个事件，它们两两互不相容，而且在试验(观测)中必然出现其中一件，则这些事件构成一个互不相容的完备事件组$\{A_1, A_2, \cdots A_k\}$。一个互不相容的完备事件组中各概率之和为1。即

$$P(A_1+A_2+\cdots+A_k)=P(A_1)+P(A_2)+\cdots+P(A_k)=1$$

例如，某地一月份的天空状况，晴天（A_1）出现的概率为30%，少云（A_2）出现的概率为40%，多云（A_3）出现的概率为20%，阴天（A_4）出现的概率则为10%。而晴、少云、多云和阴包括了天空状况的全部可能出现情况，而且是互不相容的，又必然出现其中之一，故这四种事件构成了一个互不相容的完备事件组$\{A_1, A_2, A_3, A_4\}$。它的各概率之和为：

$$P(A_1+A_2+A_3+A_4)=P(A_1)+P(A_2)+P(A_3)+P(A_4)$$
$$=30\%+40\%+20\%+10\%=100\%=1\text{。}$$

2. 概率的乘法定理：

概率的乘法主要是讨论相容事件共同出现情况下的概率运算问题。

例如，由某地多年资料统计得出，某月（30天）出现东风9天，其中有雨8天，无雨1天。其他风向出现21天，其中有雨只有5天（全月有雨则共有13天）。可以看出该地东风（事件A）与有雨（事件B）是两个相容事件，而且有密切关系。从这些资料中，我们导出概率的乘法定理。

已知一个月内出现东风的概率记作$P(A)=9/30$，东风与有雨共同出现的概率记作$P(AB)=8/30$，此8/30也就是A与B乘积的概率，在出现东风的条件下有雨的概率（这种有前提条件的概率称为条件概率）记作$P(B|A)=8/9$。

由上面所列举的三种概率可以看出，有如下的关系：

$$P(AB)=8/30=9/30\cdot 8/9=P(A)\cdot P(B|A)$$

同样，我们还知一个月内有雨的概率记作 $P(B)=13/30$，在有雨的条件出现东风的条件概率记作 $P(A|B)=8/13$。则

$$P(AB)=8/30=13/30\cdot 8/13=P(B)\cdot P(A|B)$$

因此 $$P(AB)=P(A)P(B|A)=P(B)P(A|B) \qquad (1-1-4)$$

上式的意义，即两事件乘积的概率等于其中一个事件的概率乘上另一事件在该事件发生的条件下的条件概率。这就是概率的乘法定理。

如果两事件是互不相容的，则显然有 $P(AB)=0$。

但是，如果相容的两事件是独立的，概率的乘法应该是怎样的呢？所谓独立，就是事件 A 的出现与否，对事件 B 的出现与否毫无影响，则称事件 A 和事件 B 是独立的。

例如，某地从多年资料统计得出，某月（30天）出现东风（A）10天，其中有雹（B）1天；出现其他各种风向20天，其中有雹2天。则 $P(B|A)=1/10$，$P(B)=3/30=1/10$，即 $P(B|A)=P(B)$。这表示在东风条件下有雹的条件概率和降雹的概率相等，说明东风与降雹是没有关系的，是相互独立的。

在 A、B 两事件相互独立的情况下，因 $P(B|A)=P(B)$ 或 $P(A|B)=P(A)$，故

$$P(AB)=P(A)P(B) \qquad (1-1-5)$$

上式说明两独立事件乘积的概率等于该两事件概率的乘积。这就是独立事件的概率乘法定理。

乘法定理可以推广到 k 个事件（A_1，A_2，……，A_k）相互独立的情形，即其中任一事件是否出现对其他事件是否出现的概率互不影响的情形，则有：

$$P(A_1A_2\cdots\cdots A_k)=P(A_1)P(A_2)\cdots\cdots P(A_k) \qquad (1-1-6)$$

式中 $P(A_1A_2\cdots\cdots A_k)$ 表示这 k 个事件同时出现的概率。

例如，已知上海6月份降水量大于 300mm 的概率为 10%，又设前一年6月降水多少与该年6月降水多少无关，求连续两年、连续三年6月降水量大于 300mm 的概率各为多少？

设 A 为“第一年降水量大于 300 mm”这一事件，

B 为“第二年降水量大于 300 mm”这一事件，

C 为“第三年降水量大于 300 mm”这一事件。

根据假设 A,B,C 互相独立，因此连续两年6月降水量都大于 300 mm 的概率为：

$$P(AB)=P(A)P(B)=10\%\cdot 10\%=1\%$$

连续三年6月降水量都大于 300 mm 的概率为：

$$P(ABC)=P(A)P(B)P(C)=0.1\%$$

也就是说平均1000年出现一次连续三年6月降水量都大于 300 mm 的事件。

独立事件的乘法定理在统计预报中很有用处，因为它可以使计算大大简化。

（三）条件概率

概率的乘法定理(1—1—5)式是指A事件与B事件为两个相互独立的事件，但是这种情况在气象预报的实践中常常是不能全部满足的，**“因为一切客观事物本来是互相联系的和具有内部规律的”**。运动着的大气中，各现象之间也是相互影响、相互制约的，一些现象的发生是依一定条件而存在的，因此就要计算条件概率。

条件概率以$P(A|B)$或$P(B|A)$表示之。$P(A|B)$表示在事件B出现的条件下事件A出现的概率，或称A对B的条件概率。$P(B|A)$表示在事件A出现的条件下事件B出现的概率，或称B对A的条件概率。条件概率的计算公式可由（1—1—4）式得出，即：

$$\left.\begin{aligned} P(A|B)&=\frac{P(AB)}{P(B)}\\ P(B|A)&=\frac{P(AB)}{P(A)}\end{aligned}\right\} \qquad (1—1—7)$$

下面举例说明条件概率的计算。

现有x、y两组资料(见表1—4)，根据这些资料计算$A=\{x=1\}$条件下$B=\{y=1\}$的条件概率[即计算$P(y=1|x=1)=P(B|A)$]。

表1—4

序　号	1	2	3	4	5	6	7	8	9	10	11	12
x	1	1	1	1	0	0	1	0	1	0	1	1
y	1	1	0	1	1	0	1	0	1	1	1	0

由上表可知，$P(AB)=6/12$，$P(A)=8/12$。由（1—1—7）式计算，$P(B|A)=P(AB)/P(A)=0.75$。因此当$x=1$条件下$y=1$的条件概率$P(y=1|x=1)$为75%。

另外，在实际计算条件概率时，有时可以直接从条件概率的概念得出。如根据上表求$x=1$条件下$y=1$的条件概率时，首先求得$x=1$出现的次数(由表中可知共出现8次)，和$x=1$及$y=1$共同出现的次数（由表中可知共出现6次），两个的比值（$x=1$及$y=1$共同出现的次数/$x=1$出现的次数）$6/8=0.75$，这就是$x=1$条件下$y=1$的条件概率$P(B|A)$的值。

第二节　随机变量及其概率分布

一、随机变量

（一）随机变量的概念

大气中的各种天气现象，如大风、降水、雷暴、雾、风沙等都是随机事件，它们的出现与否都受着许多随机因素的影响。又如长江中下游梅雨的入梅日期每年不同，它与副高脊线的随机移动有关。南京地区某月雷暴出现的日数或月降水量的多少，在未进行观测之前是无法确切地知道其数值是多少，如此等等。我们把这些受随机因素影响而变动的数量称为随机变量。随机变量随着试验或观测结果的不同以偶然形式取各种不同的值，并且以确定的概率取这些值。

引出随机变量的概念之后，就可以从数量上来研究随机事件和它的概率了。例如，用X表示某地2月份回流低云云高，那么$P(200<X<500)=0.80$就表示随机事件"2月份回流低云云高"为200至500米之间的概率等于80%；又如，用"$X=0$"表示"某地冬季12月份上午10时的能见度小于4公里"这一事件，则$P(X=0)=0.25$就清楚地表明"某地冬季12月份上午10时能见度小于4公里"这一事件发生的概率为25%。

为了研究的方便，我们把随机变量分为两类：凡随机变量X所取的可能值能够一一地全部列举出来者，就称为离散型随机变量。如入梅日期是6月5日或6日、7日、……、30日，这就是离散型随机变量的一例。如果随机变量X所取的可能值不能够一一列出来，而是连续地充满某个数值区间，即它可以取得该区间内的任意实数值，这样的随机变量我们就叫做连续型随机变量。例如，某站的地面气压可取数值在1000mb到1030mb之间变化，则此气压值就是连续型随机变量。在统计预报中也常将连续型随机变量进行分组统计，称此为"离散化"处理。

（二）特征随机变量

如前所述，我们引出随机变量的目的是为了对随机事件进行定量化的描述。但在我们所研究的预报对象中，如"有雨"与"无雨"并无数量的概念。"出低云"或"不出低云"也不便于直接进行定量计算。为此，如果用随机变量X表事件，"无雨"与"有雨"分别用$X=0$和$X=1$表示，则：

$$X=\begin{cases}0 & (\text{无雨})\\ 1 & (\text{有雨})\end{cases}$$

设"有雨"的概率为0.3，即$P(X=1)=0.3$，那么，"无雨"的概率则为0.7，即$P(X=0)=0.7$。由此可见随机变量X每取一定值时，它都对应着确定的概率。

又如，"不出低云"与"出低云"分别用$Y=0$和$Y=1$表示，则有

$$Y=\begin{cases}0 & (\text{不出低云})\\ 1 & (\text{出低云})\end{cases}$$

气温的升、降也可用随机变量来表示，如果$X=0$表示24小时降温（$\Delta T_{24}<0$），$X=1$表示24小时升温（$\Delta T_{24}>0$），则有

$$X=\begin{cases}0 & (\Delta T_{24}<0)\\ 1 & (\Delta T_{24}>0)\end{cases}$$

由上可见，将非数量的随机事件经过“数量化”处理后就构成“特征随机变量”，或简称“特征变量”。无论是预报因子或预报对象都可以将它们的实际观测资料转换成特征资料“0、1”，这样可使计算大大简化。特征随机变量的概念在简易统计预报中有极其重要的作用。

二、随机变量的概率分布

尽管随机变量取什么值事先并不知道，它决定于试验或观测的结果，但这决不是说随机变量的取值是杂乱无章、没有规律可寻。正如恩格斯所指出：**“而所谓偶然的东西，是一种有必然性隐藏在里面的形式”**。今天问明天最高气温取值应是多少？虽然我们今天还不知道是多少，但是我们可以从大量的统计资料中知道，在相同的天气形势下气温出现在某个数值区间的概率有多大。因此，随机变量每取一个值，都对应着一定的概率。我们把这种随机变量取值及其对应的概率关系叫做它的概率分布。随机变量与概率分布之间的关系，体现了偶然性和必然性之间的对立统一关系。在偶然性和必然性的辩证关系中，**“是那种以偶然性为其补充和表现形式的必然性占统治地位”**。因此，我们不能离开概率分布去研究随机变量。随机变量的概率分布形式有分布列、分布函数、分布密度三种，下面将一一介绍。

（一）分布列

将武汉历史上六十二年的年降水量资料进行整理、分组统计(即离散化处理)得出下列的雨量频率表：

表 1—5

组界（区间）	组中值(mm)	频数	频率	累积频率
[500, 700)	600	1	0.016	0.016
[700, 900)	800	2	0.032	0.048
[900, 1100)	1000	15	0.242	0.290
[1100, 1300)	1200	18	0.290	0.580
[1300, 1500)	1400	13	0.210	0.790
[1500, 1700)	1600	7	0.113	0.903
[1700, 1900)	1800	3	0.048	0.951
[1900, 2100)	2000	2	0.032	0.983
[2100, 2300)	2200	1	0.016	0.999

分组统计后，年雨量X的可取值（组中值）为600，800，1000，……，2200。这些可取值对应的概率(用频率代替)分别为0.016，0.032，0.242，……，0.016。将X之可取值及其对应的概率列成下表：

表1—6

x_i	600	800	1000	1200	1400	1600	1800	2000	2200
p_i	0.016	0.032	0.242	0.290	0.210	0.113	0.048	0.032	0.016

上面的这种表格叫做离散型随机变量的分布列。一般，设x_i ($i=1,2,\cdots\cdots n$) 是离散型随机变量X的所有可能值，而p_i ($i=1, 2, \cdots\cdots n$) 是X取值x_i时的概率，即

$$P(X=x_i)=p_i \qquad (i=1, 2, \cdots\cdots n)$$

同时有 $$\sum_{i=1}^{n} p_i=1$$

那么，我们可以用下述分布列来表示离散型随机变量取值的概率分布规律：

$$\begin{pmatrix} x_1 & x_2\cdots\cdots x_n \\ p_1 & p_2\cdots\cdots p_n \end{pmatrix}$$

分布列清楚地描述了随机变量的一切可能取值，以及以怎样的概率取这些值。

（二）分布函数

从武汉年雨量频率表(表1—5)中可以看出：年雨量落在区间〔1100，1300）内的概率最大，其值为29.0%，落在区间〔500，700）和区间〔2100，2300）内的概率为最小都是1.6%。可见区间不同，随机变量(年雨量)出现的概率可以不同。如果我们能从大量历史资料中，按各种不同的组界分组统计出随机变量（年雨量）落在〔400，500），……〔800，850），……〔1000，1250），……〔1500，1510），……〔2000，2400）等各种区间内出现的概率，则我们就可以用 $P(x_1\leqslant X<x_2)$表示随机变量X落在任意区间〔x_1，x_2）上的概率。显然，$P(x_1\leqslant X<x_2)$是区间〔x_1,x_2）的函数，为了把这个区间的函数化为点x的函数，以便于数学上的处理。由互不相容事件的加法定理：$P(x_1\leqslant X<x_2)=P(X<x_2)-P(X<x_1)$可知，研究$X$落在区间〔$x_1,x_2$）上的概率实质上是研究$X$的值小于任意实数$x$的概率$P(X<x)$，当$x$取不同值时，$P(X<x)$也对应有确定的数值。我们令

$$F(x)=P(X<x) \qquad (1—2—1)$$

则函数$F(x)$就叫做随机变量X的概率分布函数，简称X的分布函数。它是用来描写随机变量可取值与其对应的概率关系的函数。注意，$F(x)$决不是$X=x$点处的概率，而是小于x的区间上所有概率之总和，它是某区间上的累积概率，所以，有人又称$F(x)$叫分布累积函数。

对离散型随机变量而言有：

$$F(x)=P(X<x)=\sum_{x_i<x} P(X=x_i) \qquad (1—2—2)$$