统计天气预报

朱伯承

上海科学技术出版社

统 计 天 气 预 报

朱 伯 承

上 海 科 学 技 术 出 版 社

内 容 简 介

本书介绍怎样利用数理统计方法作天气预报．共分十三章，包括概率和统计的基本知识、数据分型、预报因子的筛选、预报量和预报因子处于各种情况下的预报方法、预报质量的评定等等．文字通俗，并有大量实例．可供气象台站预报员、大专院校有关专业师生参考，也可供水文、地震预报工作者参考。

统 计 天 气 预 报

朱 伯 承

上海科学技术出版社出版

（上海瑞金二路450号）

新华书店上海发行所发行　上海商务印刷厂印刷

开本787×1092 1/16　印张27.25　字数656,000

1981年4月第1版　1981年4月第1次印刷

印数 1—4,000

书号：13119·875　　定价：(科四) 2.50元

前　言

近几年来，我国广大气象台站逐步采用概率统计方法来预报天气，并与其他预报方法相结合，对提高天气预报准确率和提高预报服务质量起到了一定作用.

天气现象（或天气过程）的发生，包含有必然性和偶然性. 统计方法即从天气现象（或天气过程）具有偶然性的一面出发，通过大量历史资料，去探索它内部隐藏着的必然性，从而作出天气预报来. 换言之，用概率统计方法分析天气演变的统计规律性以及分析预报因子和预报量之间的数量关系，建立数学模式，预报未来天气，这就是统计天气预报.

为适应我国基层气象台站的需要，本书介绍常用的统计预报方法. 从概率统计的基本知识开始，由浅入深地介绍统计预报中的各个课题，即数据分型、预报因子的筛选方法、各种情况下的预报方法、预报质量的评定等等. 为使读者易于掌握，书中有大量的实例、较详细的计算步骤、预报体会等.

在本书的编著过程中，征求过很多同志的意见. 由于给予编者帮助的同志太多，无法一一列出. 在此，一并表示衷心感谢. 由于编者水平有限，书中一定存在不少缺点与错误，请读者批评指正.

编　者

目　录

第一章 基础知识

在县站预报中，经常要接触许多数据．这些数据提供了有关未来天气的有用情报，它能帮助人们认识天气演变的内在规律．如何从大量的数据中提取对未来天气有指示作用的情报呢？概率统计方法是解决这个问题的有力工具．作为全书的开端，本章介绍有关概率统计与县站统计预报的基础知识．

§1 事件与概率

一、事　件

在大气中存在各种各样的天气现象：梅雨季节我国华东、华南地区可能出现连阴雨；冬季强寒潮入侵时引起的剧烈降温；盛夏受副热带高压控制会有酷热等等．这些现象在概率统计中称为事件，用字母 A, B, C, … 表示．事件由它出现的可能性被分为三种：随机事件、必然事件与不可能事件．

1．随机事件

在一定的条件下，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．

2．必然事件与不可能事件

在一定条件下进行大量重复试验时，每次必然会发生的事件称为必然事件，用字母 U 表示；反之，在一定条件下进行大量重复试验时，每次必然不发生的事件称为不可能事件，用字母 V 表示．它们是随机事件的两种极端情况．必然事件与不可能事件之间存在着密切联系，如果在一定条件下某个事件是必然事件，则在同样条件下那一事件的反面就必然是不可能事件．反过来也一样．

3．事件的和

事件 A 与事件 B 之和 $A+B$ 仍是事件，它表示“A 与 B 中至少出现一个”时的事件．如果 A, B 两事件分别用两个圆表示，则两个圆相交外围线所围成的部分表示事件 $A+B$ 之和，如图 1-1 所示．

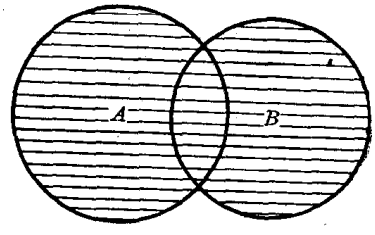

图 1-1　事件 $A+B$ 的图示

事件之和的一般定义为：如果 A 发生等价于 A_1, A_2, …, A_n 中至少有一个发生，则称事件 A 是事件 A_1, A_2, …, A_n 的和，记为

$$A=\sum_{i=1}^{n}A_i=A_1+A_2+\cdots+A_n \tag{1-1}$$

例如，某站进行三次大风预报，设事件 A_0 为“三次预报中，一次也没报对”；事件 A_1 为“三次预报中，报对一次”；事件 A_2 为“三次预报中，报对二次”；事件 A_3 为“三次预报中，报对三次”．则

$$A_1+A_2+A_3=\text{“三次预报中，报对不少于一次”},$$

$$A_0+A_1+A_2=\text{“三次预报中，报对不多于二次”}.$$

由“事件和”的定义，不难得出下列三个等式：

$$A+A=A \tag{1-2}$$

$$A+U=U \tag{1-3}$$

$$A+V=A \tag{1-4}$$

4．事件的积

若事件 C 是由事件 A 与 B 同时发生所构成的，则称 C 为事件 A 与 B 之积，记为

$$C=AB$$

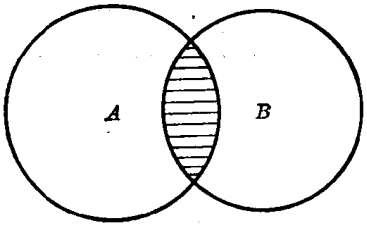

图 1-2　事件 AB 的图示

如果 A，B 两事件分别用两个圆表示，则两个圆相交部分即是 AB．由图 1-2 所示．

事件积的一般定义为：如果 A 发生等价于 A_1，A_2，…，A_n 同时发生，则称事件 A 是事件 A_1，A_2，…，A_n 的积，记为

$$A=\prod_{i=1}^{n}A_i=A_1A_2\cdots A_n \tag{1-5}$$

例如，A 为“降雨”事件，B 为“降雪”事件，则 AB 为“雨夹雪”事件，即

$$\text{“降雨”}\cdot\text{“降雪”}=\text{“雨夹雪”}$$

由事件积的定义可以得出下列等式：

$$AA=A \tag{1-6}$$

$$AU=A \tag{1-7}$$

$$AV=V \tag{1-8}$$

5．互不相容事件与相容事件

如果事件 A 与事件 B 不可能在试验结果中同时出现，即

$$AB=V \tag{1-9}$$

则事件 A 与事件 B 是两个互不相容的事件（或称互斥事件）．

如果事件 A 与事件 B 能在试验结果中同时出现，则称 A 与 B 为相容事件．

6．互逆事件

不能同时发生而又必然发生其一的两事件称为互逆事件．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．对于互逆事件有

$$A+\overline{A}=U \tag{1-10}$$

$$A\overline{A}=V \tag{1-11}$$

7．相互独立事件

如果事件 A 的出现与否，对事件 B 的出现与否毫无影响，则称事件 B 对事件 A 是独立的．

8．互不相容的完备事件组

若有 n 个事件 A_1，A_2，…，A_n，它们两两为互不相容事件，而每次试验中必然出现其中之一个事件，则称这 n 个事件构成互不相容的完备事件组．

二、概　率

1．概率的定义

在大气中随机事件是大量存在的，它们出现的可能性大小是不同的．这些随机事件(如雷暴、降雨)出现可能性的大小用一个数 $P(A)$ 来表示，这个数 $P(A)$ 叫做事件 A 的概率．在县站统计预报中，也常用符号 P_A 表示事件 A 的概率．

在实际工作中如何求得概率 P_A 呢？假如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 A 在 n 次试验中出现的频率．如果试验的次数 n 逐渐增大时，事件 A 的频率越来越稳定地接近某一个常数 P，我们便说事件 A 发生的概率是 P，即

$$P_A=P \tag{1-12}$$

在一般情况下，常数 P 是不可能精确测到的．因此通常以 n 充分大时的频率作为概率的近似值，即

$$P_A=P\approx\frac{m}{n} \tag{1-13}$$

由此可见，频率是概率的近似估计值．

例如，我们想要知道上海地区雨日出现的概率，则可根据定义应用上海雨日的历史资料(见表 1-1)，通过计算频率而近似求得．

表 1-1　上海 1881～1970 年逐年雨日

年份	年份									
	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 1-1 给出的资料中发现，在九十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 或 366)，就得到每年雨日的频率．频率最大的是 45.8%，最小的是 24.3%．假如从 1881 年开始，五年、十年、十五年、…、九十年各求一个频率，其结果见表 1-2．

表 1-2　频数表

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

由表 1-2 的结果可画出频率图(见图 1-3)：

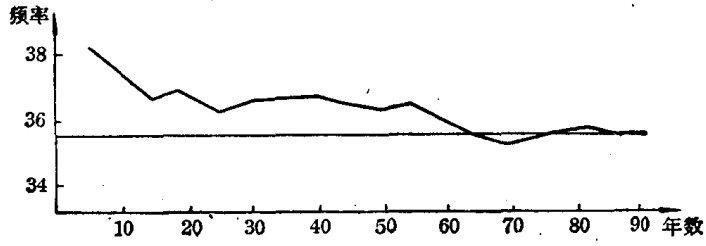

图 1-3 频率图

从表 1-2 与图 1-3 中可以看出，随着年数不断增加，雨日频率有稳定于某一常数的趋势，我们可将九十年的频率 35.5% 作为概率 P 的近似估计值. 应当注意的是，频率与概率并不是完全相等的概念. 频率值随试验次数不同而异，概率值是与试验次数无关的常数. 只有当试验次数不断增加，频率逐渐稳定时，才能用它来代替概率值. 否则，资料年代短，频率不稳定，则事件的频率就不能代替概率. 在气象统计预报中，往往把有限次观测得到的频率来代替"概率"，但必须注意它们之间的区别.

2. 概率的基本性质

根据概率的定义，可以得到如下三个基本性质：

① 必然事件的概率等于 1，即

$$P_U=1 \tag{1-14}$$

② 不可能事件的概率等于零，即

$$P_V=0 \tag{1-15}$$

③ 随机事件 A 的概率永远不会小于 0，也不会大于 1. 即

$$0\leqslant P_A\leqslant 1 \tag{1-16}$$

若在 n 次观测中，事件 A 出现了 m 次，因为 m 不可能 <0，也不可能 $>n$，所以随机事件 A 的频率 $\frac{m}{n}$ 不可能小于零，也不可能大于 1. 当 n 足够大时，频率 $\frac{m}{n}$ 趋近常数 P_A，概率 P_A 当然也不可能小于零，也不可能大于 1.

在实际观测中，某事件 A 的概率接近于 0，我们称为实际不可能事件，在观测预报中可断定事件 A 不可能出现. 若概率接近 1，则称为实际必然事件，在预报中可断定事件 A 出现. 但小到什么程度才能当作实际不可能事件呢？这就要随预报对象的实际情况而定.

3. 概率的加法定理

如果 A 与 B 是两个互不相容的事件，则

$$P_{A+B}=P_A+P_B \tag{1-17}$$

这是由于在 n 次试验中，若事件 A 出现了 m_1 次，事件 B 出现了 m_2 次. 由于 A、B 互不相容，在事件 A 出现 m_1 次中事件 B 皆未出现，在事件 B 出现的 m_2 次中事件 A 皆未出现，因此在 n 次试验中 A 出现或 B 出现的次数为 m_1+m_2，由 $\frac{m_1+m_2}{n}=\frac{m_1}{n}+\frac{m_2}{n}$ 则得(1-17)式.

例如，某地从多年资料得出某月晴天(A)出现的概率 $P_A=\frac{12}{30}$，多云(B)出现的概率 $P_B=\frac{5}{30}$，那么出现"晴或多云"的概率为

$$P_{A+B}=P_A+P_B=\frac{12}{30}+\frac{5}{30}=\frac{17}{30}$$

加法定理还可作如下的推广：如果 A_1, A_2, …, A_k 等 k 个事件是两两互不相容的，则有：

$$P_{A_1+A_2+\cdots+A_k}=P_{A_1}+P_{A_2}+\cdots+P_{A_k} \tag{1-18}$$

例如，某站7月分吹南风的概率为23%，南南东风的概率为14%，东南风的概率为12%，南南西风的概率为6%，西南风的概率为2%，因此该站7月分偏南风的概率为

$$23\%+14\%+12\%+6\%+2\%=57\%$$

若 $\{A_1, A_2, \cdots, A_k\}$ 构成完备事件组，则

$$P_{A_1+A_2+\cdots+A_k}=1$$

4．条件概率

在运动变化着的大气中，各种天气现象是相互影响、相互制约的，一些天气现象会引起另外一种天气现象的出现，由此引入条件概率的概念．条件概率用 $P_{A|B}$ 或 $P_{B|A}$ 表示．$P_{A|B}$ 为事件 B 出现的条件下事件 A 出现的概率，或称事件 A 关于事件 B 的条件概率．$P_{B|A}$ 为事件 A 出现的条件下事件 B 出现的概率，或称事件 B 关于事件 A 的条件概率．由条件概率的定义可知

$$P_{A|B}\neq P_{B|A}$$

如果 A 与 B 是相互独立事件，事件 B 的存在与否并不影响 A 的出现，因此有

$$P_{A|B}=P_A, \qquad P_{B|A}=P_B \tag{1-19}$$

5．概率的乘法定理

概率的乘法定理主要是解决相容事件同时出现情况下概率的计算问题．先看一个例子，由某地多年资料统计得出：某月(30天)出现东风11天，其中有雨9天，无雨2天；其它风向19天，其中有雨4天，无雨15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．已知一个月内出现东风的概率 $P_A=\frac{11}{30}$，东风与有雨共同出现的概率 $P_{AB}=\frac{9}{30}$，在出现东风的条件下有雨的条件概率 $P_{B|A}=\frac{9}{11}$，由这三个概率可以看出

$$P_{AB}=\frac{9}{30}=\frac{11}{30}\times\frac{9}{11}=P_AP_{B|A}$$

同样，我们知道一个月内有雨的概率 $P_B=\frac{13}{30}$，在有雨的条件下出现东风的概率

$$P_{A|B}=\frac{9}{13},$$

则

$$P_{AB}=\frac{9}{30}=\frac{13}{30}\times\frac{9}{13}=P_BP_{A|B}$$

因此

$$P_{AB}=P_AP_{B|A}=P_BP_{A|B} \tag{1-20}$$

这就是概率的乘法定理，它说明了两事件同时出现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条件下的条件概率．k 个事件 A_1, A_2, …, A_k 同时出现的概率为

$$P_{A_1A_2\cdots A_k}=P_{A_1}P_{A_2|A_1}\cdots P_{A_k|A_1A_2\cdots A_{k-1}} \tag{1-21}$$

如果两事件 A 与 B 是互不相容的，显然有

$$P_{AB}=0$$

如果相容的两事件 A 与 B 是独立的，则由(1-19)式可得

$$P_{AB}=P_AP_B \tag{1-22}$$

当 k 个事件 A_1, A_2, …, A_k 互相独立，即其中任一事件是否出现对其它事件的出现与否均无影响，则有

$$P_{A_1A_2\cdots A_k}=P_{A_1}P_{A_2}\cdots P_{A_k} \tag{1-23}$$

6．全概率公式

若事件 B 只能与互不相容的完备事件组 A_1, A_2, …, A_k 中的一个同时发生，那么事件 B 的概率称为全概率，它的公式为

$$P_B=\sum_{i=1}^{k}P_{A_i}P_{B|A_i} \tag{1-24}$$

例如，某站研究6月下旬的暴雨过程(记为事件 B)，发现它们只与"锋面、切变线、低涡经过本站"这三个事件(分别用 A_1, A_2, A_3 表示)有关，且知 $P_{A_1}=0.2$, $P_{A_2}=0.3$, $P_{A_3}=0.5$. 设在锋面过境时，6月下旬出现暴雨的概率 $P_{B|A_1}=0.1$；切变线过境时，6月下旬出现暴雨的概率 $P_{B|A_2}=0.3$；低涡过境时，6月下旬出现暴雨的概率 $P_{B|A_3}=0.4$.

试求利用这三个系统影响时，该站预报6月下旬出现暴雨的概率.

从题意可知，A_1, A_2, A_3 构成互不相容完备事件组，B 只能与 A_1, A_2, A_3 中之一同时发生，故用全概率公式

$$\begin{aligned}P_B&=\sum_{i=1}^{3}P_{A_i}P_{B|A_i}=P_{A_1}P_{B|A_1}+P_{A_2}P_{B|A_2}+P_{A_3}P_{B|A_3}\\&=0.2\times0.1+0.3\times0.2+0.5\times0.4=0.28\end{aligned}$$

即6月下旬出现暴雨的概率为0.28.

7．贝叶斯公式

设事件 B 能而且只能与互不相容的完备事件组 A_1, A_2, …, A_k 之一同时发生，则事件 B 出现的条件下，事件 A_i 出现的概率

$$P_{A_i|B}=\frac{P_{A_i}P_{B|A_i}}{\sum_{j=1}^{k}P_{A_j}P_{B|A_j}}\quad i=1,\ 2,\ \cdots,\ k. \tag{1-25}$$

(1-25)式称为贝叶斯公式.

证明：

由乘法定理 $$P_{A_iB}=P_BP_{A_i|B}=P_{A_i}P_{B|A_i}$$

可得 $$P_{A_i|B}=\frac{P_{A_i}P_{B|A_i}}{P_B}$$

将全概率公式(1-24)代入上式中的 P_B，即得(1-25)式. 在刚才举的例子中，求出在暴雨出现的条件下，影响系统过境的条件概率各为：

$$P_{A_1|B}=\frac{P_{A_1}P_{B|A_1}}{\sum_{j=1}^{3}P_{A_j}P_{B|A_j}}=\frac{0.2\times0.1}{0.5\times0.4+0.3\times0.2+0.2\times0.1}=0.071$$

$$P_{A_2|B}=\frac{P_{A_2}P_{B|A_2}}{\sum_{j=1}^{3}P_{A_j}P_{B|A_j}}=\frac{0.3\times0.2}{0.5\times0.4+0.3\times0.2+0.2\times0.1}=0.214$$

$$P_{A_3|B}=\frac{P_{A_3}P_{B|A_3}}{\sum_{j=1}^{3}P_{A_j}P_{B|A_j}}=\frac{0.5\times0.4}{0.5\times0.4+0.3\times0.2+0.2\times0.1}=0.714$$

§2 随机变量及其概率分布

一、随机变量

随机变量的概念，是概率论中最重要的基本概念之一. 随机变量概念比事件概念在解决随机现象的问题上是更为普遍的工具. 为引出随机变量的概念，我们先来看看单站资料具有什么特性.

1. 单站资料的决定性与随机性

单站资料具有决定性与随机性的特征. 如果把几年的逐日 14 时气温点成一条曲线，就会看到由冬到夏气温逐渐升高，由夏到冬气温逐渐降低，有一个显著的年变化周期，这是由于地球绕太阳公转而造成四季变化，这种变化可以说是决定性变化. 在这条曲线上迭加许多扰动，比如说四月份的气温是逐渐上升的，但由于一次冷空气侵袭，气温可能降低. 而"四月五日气温为该月的最低点"这一变化相对于气温的年变化来说是一种随机型的变化. 因为气温最低点可以出现在四月五日，也可能不出现在四月五日，事先不能确定.

决定性和随机性在理论上虽有明确的含义，但就具体的一串观测数据来说，区分决定性和随机性是相对的.

2. 随机变量的概念

上面我们已谈到，单站资料是一些既受决定性因素制约又受随机性因素影响而变动的数量. 我们将受随机因素的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随机形式取各种不同的值，并以确定的概率取这些值. 随机变量一般以 X、Y 表示.

例如，长江中下游梅雨的入梅日期就是一个随机变量. 我们事先并不能精确地知道某年入梅日期，因为它是受许多随机因素影响的. 其中副热带高压脊线位置的变动就是随机因素之一.

引出随机变量概念之后就可以对大量的随机事件进行定量的研究，对一些非数量的随机事件也可以进行数量化处理，变成特征随机变量. 例如，"无雨"与"有雨"分别用 $X=0$ 和 $X=1$ 表示，即

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

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，都对应着确定的概率.

3. 随机变量的种类

按照随机变量的性质，通常可以把它分为离散型与连续型两类.

① 离散型随机变量

随机变量 X 所取的可能值，能全部列出来，如 $(x_1, x_2, \cdots, x_n)$，并且以确定的概率 $(P_1, P_2, \cdots, P_n)$ 取这些值，即

$$P(X=x_i)=p_i \tag{1-26}$$

则称 X 为离散型随机变量. 因为 x_i 是在所有可能值 $x_1, x_2, \cdots, x_n$ 中取某一个值, 则 $x_1, x_2, \cdots, x_n$ 构成互不相容的完备事件组, 所以有

$$\sum_{i=1}^{n} p_i = 1 \tag{1-27}$$

当我们知道了随机变量 X 的所有可能值以及 X 的每个取值 x_i 所对应的概率 p_i 时, 我们就可以说已完全认识了这个随机变量.

我们可以用下面的分布列

x_1	x_2	$\cdots$	x_n
p_1	p_2	$\cdots$	p_n

来描述离散型的随机变量 X.

为了更直观起见, 可用图形来表示. 横坐标为随机变量的可能取值 x_i, 纵坐标为随机变量取这些值的概率 p_i, 把各随机变量可能值与相应的概率值点在坐标图中, 并把这些点连接起来, 所得的图形是一个分布多边形, 由图 1-4 所示.

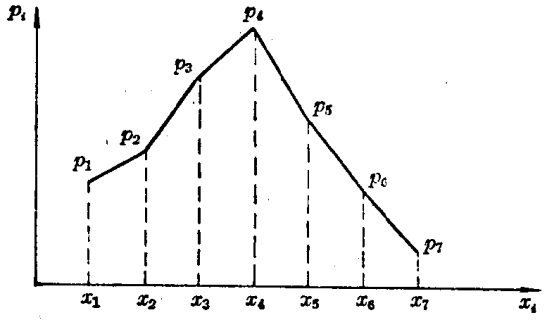

图 1-4 分布多边形

分布列与分布多边形是离散型随机变量分布律的两个表现形式.

在单站统计预报中, 常把风向(方位)、云量、能见度、降水、雾、雷暴等各种天气的日数, 以及霜冻初终日期、入(出)梅日期等等都作为离散型随机变量来处理.

② 连续型随机变量

随机变量 X 所取的可能值不能一一列出来, 而是连续充满数轴的某一区间 $[a, b)$ 时, 则称为连续型随机变量. 这里 $[a, b)$ 表示随机变量的所有可能值出现在大于或等于 a 以及小于 b 的范围内.

气温就是一个连续型的随机变量. 我们可以求出某地夏季温度高于 35℃ 的概率或冬季温度低于 −10℃ 的概率, 而不能求出气温为某一特定值的概率. 由于连续型随机变量取值充满整个区间, 即可取无穷多个值, 因此分配到任何一个特定值上的概率都是零. 这样, 对于连续型随机变量来说, 研究它取某特定值的概率是没有意义的.

对于连续型随机变量, 只能研究 $c \leqslant X < d$ 的概率, 即研究 X 落在某区间 $[c, d)$ 上的概率. 其中 $[c, d)$ 为 $[a, b)$ 的子区间, 即 $a \leqslant c < d < b$.

令: 事件 A 由 $X < d$ 组成, 则 $P(A) = P(X < d)$; 事件 B 由 $X < c$ 组成, 则 $P(B) = P(X < c)$; 事件 c 由 $c \leqslant X < d$ 组成, 则 $P(c) = P(c \leqslant X < d)$. 由事件的定义不难得出: $A = B + C$. 由于 B 与 C 是互不相容的, 因此

$$P(A) = P(B) + P(C)$$

即

$$P(c \leqslant X < d) = P(X < d) - P(X < c) \tag{1-28}$$

因此, 只要求出 $P(X < d)$ 与 $P(X < c)$ 也就求出 X 在区间 $[c, d)$ 上的概率. 所以, 研究连续型随机变量, 实质上就是研究 $X < x$ 的概率.

4．随机变量的准确定义

由上面的讨论，我们可得到随机变量的准确定义：

定义：如果每一个试验的结果可以用一个变数 X 来表示，而且对于任何实数 x，“$X<x$”有着确定的概率，则称 X 是随机变量．

二、随机变量的概率分布

随机变量出现不同数值的可能性有多大的问题，这就不能用一个概率来表示它，而要了解它出现不同数值的概率，即概率分布．下面介绍概率分布的两种主要形式：分布函数与分布密度．

1．分布函数

定义：设有随机变量 X，事件 $X<x$ 的概率 $P(X<x)$ 是 x 的函数，则称 $P(X<x)$ 是随机变量 X 的分布函数，记作 $F(x)$，即

$$F(x)=P(X<x) \tag{1-29}$$

分布函数 $F(x)$ 有时又称为分布的累积函数，它完整地描述了随机变量的统计特征．无论是连续型还是离散型的随机变量都存在分布函数．

对于连续型随机变量，若 $a<b$，有

$$P(a\leqslant X<b)=P(X<b)-P(X<a)=F(b)-F(a) \tag{1-30}$$

即连续型随机变量落在给定区间上的概率等于分布函数在这个区间上的增量．

对于离散型随机变量 X，有

$$F(x_k)=P(X<x_k)=P(X=x_1)+P(X=x_2)+\cdots+P(X=x_{k-1}) \tag{1-31}$$

2．分布函数的性质

分布函数的主要性质有以下几个：

① $$0\leqslant F(x)\leqslant 1$$

因为 $F(x)=P(X<x)$，任何事件的概率总有 $0\leqslant P(X<x)\leqslant 1$，所以 $0\leqslant F(x)\leqslant 1$．

② $F(x)$ 是所有自变量 x 的非降函数．即

当 $x_2>x_1$ 时，有 $F(x_2)\geqslant F(x_1)$．

因为事件的概率总不能小于零的，即 $P(x_1\leqslant X<x_2)\geqslant 0$，同时 $P(x_1\leqslant X<x_2)=F(x_2)-F(x_1)$，所以 $F(x_2)-F(x_1)\geqslant 0$，亦即 $F(x_2)\geqslant F(x_1)$．可见 $F(x)$ 是非降函数．

③ $$F(-\infty)=0,\qquad F(+\infty)=1.$$

将 x 沿坐标轴无限地向左移动，随机变量 X 落在 x 左侧变成不可能事件，因此，

$$P(X<-\infty)=P(V)=0.$$

同样将 x 无限地向右移动，此时事件 $X<x$ 变为必然事件．因此

$$P(X<+\infty)=P(U)=1.$$

④ $F(x)$ 是左连续的．即

$$F(x)=F(x-0),$$

根据分布函数的性质可知，任何一个非降的、左连续的函数 $F(x)$，如果满足条件

$$F(-\infty)=0,\qquad F(+\infty)=1,$$

则该函数一定是某个随机变量的分布函数．

3．分布密度

定义：一个随机变量 X 有可微的分布函数 $F(x)$，若极限

$$\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=F'(x)$$

存在，则称分布函数的导数 $F'(x)$ 为 X 的分布密度，或称概率密度．记为

$$f(x)=\frac{dF(x)}{dx}=F'(x) \tag{1-32}$$

概率密度 $f(x)$ 是单位长度区间上的概率，它决不是随机变量 X 取某值 x 的概率．由分布密度的定义可知，随机变量 X 落入区间 $[a, b)$ 的概率为

$$P(a\leqslant X<b)=\int_a^b f(x)dx \tag{1-33}$$

它的几何意义是：这个概率等于以这个区间为界的分布密度曲线 $f(x)$ 下面的曲边梯形的面积，由图 1-5 所示．

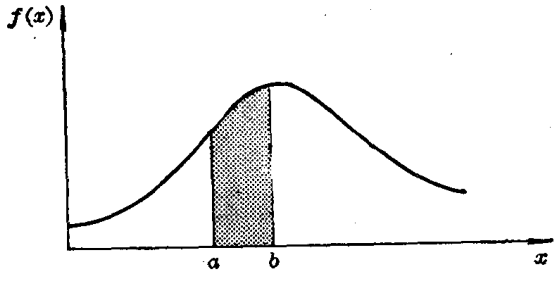

图 1-5

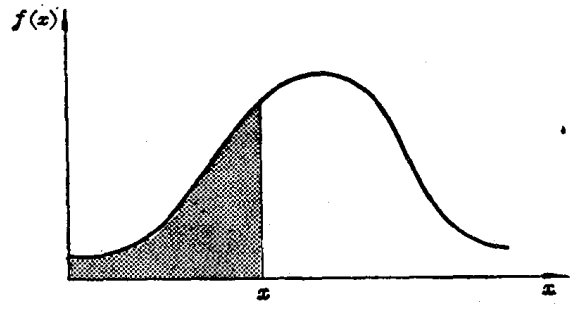

图 1-6

由分布密度 $f(x)$ 的定义可得出它与分布函数 $F(x)$ 的关系为：

$$F(x)=\int_{-\infty}^{x} f(x)dx \tag{1-34}$$

如用图来表示的话，$F(x)$ 是位于点 x 左边的分布密度曲线下的面积，由图 1-6 所示．

4．分布密度的性质

① $f(x)\geqslant 0$，即分布密度是非负函数．

因为 $f(x)=\lim\limits_{\Delta x\to 0}\dfrac{\Delta F(x)}{\Delta x}$，而 $F(x)$ 是非降的，即 $\Delta x>0$ 时，$\Delta F(x)\geqslant 0$，所以 $f(x)\geqslant 0$．从几何图形上看（见图 1-5 与图 1-6），密度曲线位于直角坐标系的第一、二象限．

② $$\int_{-\infty}^{+\infty} f(x)=1$$

因为 $$F(x)=\int_{-\infty}^{x} f(x)dx,$$

所以 $$F(+\infty)=\int_{-\infty}^{+\infty} f(x)dx=1.$$

上式的几何意义是随机变量落在整个数轴上的事件是必然事件，即横轴上方与分布密度曲线之间围成后面积等于 1．

分布函数与分布密度是息息相关的，它们都表示随机变量的概率分布形式．在单站统计预报中，分布密度的概念用得较多．

三、分布密度的求得

分布密度 $f(x)$ 是全面描述随机变量 X 的一个函数，我们如何根据单站资料来求得分布密度呢？下面介绍常用的方法——制作频数表与直方图.

1．频数表

气象站记录的数据是很多的，很难直接看出这些数据的特征，必须将它们进行分组，制成频数表. 这样就能从庞杂的资料中抽出要点，了解这些数据的特征. 例如某站记录了如下一百个相对湿度(%)数据：

76	89	83	81	77	80	72	82	87	79	81	76	80	74	82	82	85	75	82	79	84
82	79	82	82	70	74	82	77	76	77	74	77	77	84	85	72	88	80	85	79	86
79	93	76	88	80	79	78	80	76	85	90	83	78	83	81	88	79	85	77	77	79
85	71	81	84	84	82	87	75	76	79	80	78	75	82	83	82	82	82	80	81	77
86	76	77	67	77	78	82	74	83	82	81	81	84	88	95	77					

频数表的制作步骤如下：

第一步找出最大值与最小值

在上述相对湿度的数据中，最大值为95%，最小值为67%. 最大值与最小值之差在统计学中常称为极差，记为 R.

第二步决定组距与组数

在样本容量较大时，通常分为10～20组；样本容量小于50时，分成5～6组. 根据经验，样本容量 N 与组数有如下关系

$$\text{组数}\sim 5\lg N$$

本例中 $N=100$，因此组数为10较妥. 组距的决定与极差 R 大小有关，本例中 $R=95-67=28$，看来组距定为3较好. 需要说明的是，并不是所有情况下都一定要采用等距分组，有时还可采用不等距分组法.

第三步决定分点

分点应比原测定值的精度高一位，如果按原来的精度分组，若分为66～69，69～72，…那么当相对湿度正好等于69的数据，是分到66～69这一组呢还是分到69～72这一组？这就要作出具体规定，若刚好等于组限的数值应都分在上一组，或都分在下一组. 为了避免这种麻烦，我们也可以分成如下十组：

66.5～69.5，69.5～72.5，72.5～75.5，75.5～78.5，78.5～81.5，
81.5～84.5，84.5～87.5，87.5～90.5，90.5～93.5，93.5～96.5.

第四步频数

数据落在每组内的数目，称为频数. 在计算频数时，将数据列成表1-3的形式，这种表称为频数分布表. 频数与资料总数之比为频率. 有了频数分布表，我们可以看出这一百个数据的概貌，也可以求得相对湿度(随机变量 X)取值在某区间($a\leqslant X<b$)的频率.

2．直方图

求出频数分布表(表1-3)后，我们可以看出数据变化的规律. 为了更形象地看出这些变化，可以画直方图. 在横坐标上标出分组的点，纵坐标表示频数. 以组距为底长，频数为高度画出一个个小矩形，如图1-7所示，称为直方图.