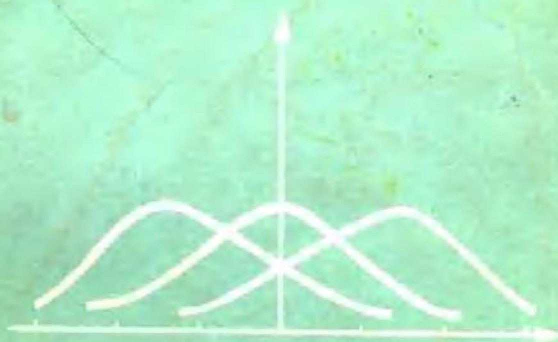
气象资料的整理和统计方法

王树廷　王伯民等编著



气象出版社

气象资料的整理和统计方法

王树廷　王伯民等编著

气象出版社

内 容 简 介

本书比较系统地介绍了我国地面、高空、船舶洋面气象资料的整理、统计方法和有关的技术规定，并从数理统计理论的角度，简明扼要地阐述了概率统计的基本知识和资料序列的审查、订正方法。书末还附有查算表，可供整理、统计资料时用。

本书是一本较好的业务工作用书，可供气象台站和气候资料部门及有关部门参考使用，亦可供中级气象科技人员自学和中等气象学校教学参考之用。

气象资料的整理和统计方法

王树廷　王伯民等编著

责任编辑：史秀菊

*　*　*

气 象 出 版 社 出 版

（北京西郊白石桥路46号）

房山张坊三合印刷厂　　新华书店北京发行所发行

*　*　*

开本：787×1092 1/16　印张：19.5　字数：481 千字

1984年7月第一版　　1984年7月第一次印刷

印数：1—12,000　　统一书号：13194·0165

定价：3.90元

前　言

本书是由国家气象局北京气象中心气候资料室的有关同志，根据气候资料整编工作的需要，结合有关的技术规定而编著的。比较系统地总结了我国地面、高空、船舶洋面气象资料的整理、统计方法和技术规定。着重介绍了主要气象要素、基本整编项目的整理、统计方法，并简明扼要地阐述了概率统计的基本知识和资料序列的审查、订正方法的数理统计理论。另外，在有关章节中还介绍了国外部分整编项目的统计方法。

本书第一章至第六章分别由孙安健、王效瑞、宋超辉、王树廷、王伯民和花灿华等同志编写。全书由张家诚同志审查定稿。

气象资料的整理和统计，就工作过程看，应包括从观测记录的订正、查算和日、候、旬、月、年值的计算，直到历年值、累年值的统计等全部内容。但考虑到《地面气象观测规范》、《高空气象观测手册》和《高空气象观测规范》等，对观测记录的整理和记录报表的编制方法已作了详尽的规定，为了减少重复，有关这方面的内容，本书第三、四章中只简要地作了叙述。同时，为了能在实际工作中得以应用，编写时我们参考了国家气象局气候资料室制定的有关统计方法和技术规定，并尽量地与其保持一致。

目前，我国气象资料的统计工作多数还是以手工的方式进行的，因此本书的主要内容也是在此基础上编写的。今后，随着电子计算机技术在气象资料整理、统计中的逐步应用和推广，不仅统计效率将会显著提高，而且在统计方法上也将会有新的发展；与此同时，有些仅适用于手工统计方式的技术规定和表格形式，也将作必要的调整和修改。

作者

一九八二年十月

主要字母、符号说明

A：较差，极差(变幅)

C：常数

C_s：偏度系数

C_v：相对标准差(离差系数、变异系数)

C_{ov}：协方差

D：风向，间隔差

d：偏差(距平)

d_v：相对偏差(距平百分率)

e：水汽压，自然对数底

f：频率

$f(x)$：概率密度函数

$F(x)$：分布函数

g：重力加速度

H：高度

K：模比系数

M：数学期望

m, n, N：次数，年数等

P：气压，概率

$P(A|B)$：条件概率

q：比湿，四分位数

r：相关系数

R：降水量，干空气比气体常数，复相关系数

s：混合比，标准差

T：绝对温度，重现期

t：摄氏温度

t_d：露点温度

v：风速

$\bar{x}$：平均值

$\hat{x}$：众数

α：信度

β：峰度

$\Gamma(\alpha)$：伽马函数

γ：峰度系数

η：相关比

μ：总体平均值

μ_r：r 阶中心距

ρ：空气密度，总体相关系数

σ^2：总体方差

σ：均方差

ν_r：r 阶矩

Φp：离均系数

目 录

绪论 …………………………………………………………………………………………（1）
第一章 概率统计基本知识 ……………………………………………………………（2）
§ 1.1 事件与概率 ………………………………………………………………（2）
§ 1.2 随机变量的数字特征 ……………………………………………………（6）
§ 1.3 随机变量的概率分布 ……………………………………………………（20）
§ 1.4 数字特征的估计 …………………………………………………………（31）
§ 1.5 假设检验 …………………………………………………………………（36）
§ 1.6 线性统计推断 ……………………………………………………………（42）
第二章 气象资料序列的审查、订正方法 ……………………………………………（46）
§ 2.1 序列的审查 ………………………………………………………………（46）
§ 2.2 序列订正方法 ……………………………………………………………（52）
§ 2.3 序列订正的误差 …………………………………………………………（60）
§ 2.4 超短资料序列订正 ………………………………………………………（62）
§ 2.5 月平均值允许缺测天数的统计推断 ……………………………………（66）
第三章 地面气象资料的整理和统计方法 ……………………………………………（69）
§ 3.1 概述 ………………………………………………………………………（69）
§ 3.2 地面气象资料统计总则 …………………………………………………（72）
§ 3.3 气压 ………………………………………………………………………（83）
§ 3.4 气温 ………………………………………………………………………（89）
§ 3.5 空气湿度 …………………………………………………………………（106）
§ 3.6 降水 ………………………………………………………………………（109）
§ 3.7 风 …………………………………………………………………………（117）
§ 3.8 其它要素 …………………………………………………………………（130）
第四章 高空气象资料的整理和统计方法 ……………………………………………（147）
§ 4.1 概述 ………………………………………………………………………（147）
§ 4.2 各等压面资料 ……………………………………………………………（150）
§ 4.3 各温度层资料 ……………………………………………………………（160）
§ 4.4 逆温层资料 ………………………………………………………………（163）
§ 4.5 对流层顶资料 ……………………………………………………………（168）
§ 4.6 各高度风资料 ……………………………………………………………（171）
§ 4.7 各厚度合成风资料 ………………………………………………………（178）
§ 4.8 大风层和最大风速层资料 ………………………………………………（186）
§ 4.9 东西风转换层资料 ………………………………………………………（190）
§ 4.10 高空天气报资料 …………………………………………………………（193）
第五章 船舶洋面气象资料的整理和统计方法 ………………………………………（203）

§ 5.1 船舶气象观测的特殊性和资料时空分布的特点 ……………………(203)
§ 5.2 船舶气象资料的审查和处理 ……………………………………(207)
§ 5.3 海区的划分 ……………………………………………………(216)
§ 5.4 船舶气象资料统计中的数字特征 ………………………………(219)
§ 5.5 船舶气象资料统计值的代表性 …………………………………(229)
§ 5.6 船舶洋面气象资料的基本统计项目 ……………………………(234)
§ 5.7 统计值的分析 …………………………………………………(245)
第六章 电子计算机在气象资料整理统计中的应用 ……………………(249)
§ 6.1 程序设计基本概念 ……………………………………………(249)
§ 6.2 气象资料信息化 ………………………………………………(253)
§ 6.3 电子计算机在气象资料整理统计中的应用 ……………………(255)
附表
1. 标准化正态密度函数 $\phi(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$ ……………………(261)
2. 标准化正态分布函数 $\Phi(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}e^{-\frac{\mu^2}{2}}d\mu$ ……………………(262)
3. χ^2 分布 $P(\chi^2\geq\chi^2_\alpha)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_{\chi^2_\alpha}^{\infty}e^{-\frac{u}{2}}u^{\frac{n}{2}-1}du=\alpha$ 的数值表 ……………(263)
4. t 分布 $P(|t|\geq t_\alpha)=\frac{2}{\sqrt{\pi}B\left(\frac{1}{2},\frac{n}{2}\right)}\int_{t_\alpha}^{\infty}\frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}dt=\alpha$ 的数值表… (264)
5. F-分布 ……………………………………………………………(265)
6. 普哇松分布 $P(X=r)=\frac{\lambda^r}{r!}e^{-\lambda}$ 的数值表 ……………………(269)
7. 皮尔逊第 III 型曲线的离均系数 Φ 表 ……………………………(270)
8. 当 $C_s=2C_v$ 时第 III 型频率曲线的 K_p 值表 ……………………(272)
9. 初终日累计日数查算表（按年度统计）……………………………(274)
10. 初终日累计日数查算表（按年份统计）……………………………(275)
11. 风速的空气密度订正系数表 ………………………………………(276)
12. 东风、北风分量查算表 ……………………………………………(277)
13. 厚度合成风查算表(测风球用) ……………………………………(280)
14. 厚度合成风查算表(探空球测风用) ………………………………(281)
15. 概率为 K、资料数量为 n 时，频率值在 $K\pm\Delta K$ 范围内的保证率 ……(292)

绪　论

气象是自然环境的重要组成部分，气象条件对于人类的生活和生产活动都有不同程度的影响。因此，气象资料已成为经济建设、科学研究和环境保护等不可缺少的重要依据之一，并且越来越被人们所重视。

我国幅员辽阔，地形复杂，气象条件与国民经济的关系尤为密切，随着各方面工作的开展，对气象资料的需要十分迫切。在农业上，作物布局、引种改制、推广双季稻等，都广泛地使用了各地的气象资料；在农田水利建设上，我国长江三峡的水利工程、黄土高原的综合治理和三江平原的开发利用等，也都把气象资料作为可行性研究的依据之一；在工业上，根据气象资料进行合理的城市规划、工业布局、工程设计和能源利用等，也都取得了一定的经济效果；在科学研究、文化教育和环境保护等方面，气象资料也得到了广泛的应用，对发展科学文化技术，为改造人类的生活环境，起了一定的作用。

分布在全国各地的广大气象台站的气象工作者，日以继夜地监视着天气的变化，记录了各种气象数据，为气象资料服务提供了可靠的依据。随着气象探测手段的不断发展和资料年代的逐步增加，大量的原始记录与日俱增，而使用部门对气象资料的要求越来越高，从而资料整理、统计的工作量也就越来越大。因此，气象资料的整理、统计方法应不断地研究和改进，使之方便、合理。这是提高气象资料服务质量，更加适合使用部门的要求的关键。

选择正确的气候统计特征量和合理的统计方法，这在气象资料统计工作中是很重要的。所谓气候统计特征量，就是指不同时间、不同地点的观测记录，经过加工整理能反映气候特征的数值。在气象资料统计中，基本的气候统计量有平均值、极端值、距平值、变率和频率值等。其中就平均值而言，又可分为算术平均数、加权平均数、中位数和众数等。这些统计量虽然都表示气候的平均情况，但在代表的意义上又各有不同。例如，三次观测站日平均气温的统计方法，若用08，14和20时三次观测记录统计，其平均值只能代表白天的温度情况，与02，08，14和20时四次观测记录统计的平均值相差很大；如用08时记录加权平均，或用当天最低气温加前一天20时气温的平均值进行加权平均，其值与四次观测记录统计的平均值就比较接近。

资料统计量大小的确定，也是气象资料统计工作中经常需要考虑的问题。在充分了解资料的使用情况，尽量满足用户要求的前提下，适当地简化统计方法，或减少参加统计的资料量，对提高统计效率是十分重要的。例如，求算累年月平均海平面气压时，根据用户的不同要求，可用三种不同的统计方法：精度要求较高的，应用各定时的本站气压和气温进行统计，但工作量很大；精度要求中等的，可用历年月平均本站气压和月平均气温进行统计，工作量大为减小；精度要求不高的，可用累年月平均本站气压和月平均气温直接反查而得，工作量最小。又如日平均值的统计方法，用分布均匀的02，08，14和20时四次记录统计的日平均值，与用01，02…23和24时共二十四次记录统计的日平均值非常接近。因此，在一般情况下，都用四次记录进行统计。

第一章　概率统计基本知识

自然界中的大气现象——风雨冷暖，变幻无常，当前尚难严格控制它的变化，准确预测它的未来。为了探索大气现象中存在的客观规律，一般都是根据大量观测到的气象资料，进行综合分析，从中归纳出它的基本特征。这种“综合”、“归纳”可以借助于概率统计方法。事实上，在气象资料的整理和统计中，普遍使用的统计量是平均值、标准差和频率，而它们则分别是数学期望、均方差和概率的估计值。在处理由于迁站、仪器更换等引起的前后两段气象资料序列能否合并统计的问题时，就要运用有关假设检验与统计推断的概念和方法。为了获得较长的资料序列而将短资料序列订正延长时，或探求两个(或多个)气象要素之间的统计关系时，将会遇到相关系数和线性迴归方程之类的概念。本章的任务是，在讨论地面、高空和洋面气象资料的整理和统计方法之前，首先介绍一下气象资料整理统计过程中经常用到的概率统计方法。

§1.1　事件与概率

一、事件的概念

在自然界里有许多现象，我们可以预言它们在一定条件下是否会出现。例如，“温度降到0°C以下时，结冰必定发生”，“寒潮来临时，气温就得下降”等等是一定会出现的；而上述现象的反面，即“温度降到0°C以下时，没有结冰”，“寒潮来临，气温升高”等等是必然不会出现的。这种在一定条件下必然出现的现象叫必然事件，而必然不出现的现象叫不可能事件。

在自然界还有另一类现象，它们在一定条件下的重复试验或观测中，有时出现有时不出现，这种现象通常称为随机事件，或简称为事件。例如，“明年上海年降水日数为 130 天”，“本世纪末全球的气候要变冷”等等都是随机事件。

如果对于一次观测预报，两个事件不能同时出现(比如本站晴天和雨天事件)，则称这两事件为互斥事件或互不相容事件。如果观测预报的结果必然要在某事件组(例如无雨、小雨、中雨、大雨、暴雨天气事件组)中出现一件，则称这样的事件组为完备事件组(群)。若由两个不相容事件构成的完备事件组[例如本站出现无雨(A)与本站出现雨天(B)组成的事件组]，称此两事件为对立事件，称事件A的对立事件B为A的逆事件，记为$\bar{A}$(即A不出现的事件)。

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A和事件B是相互独立的。例如，若东风的出现并不影响降雹的出现与否，则东风与降雹两种事件是相互独立的。

二、事件的运算

在事件之间存在着一些重要关系，并能对此进行如下的运算：

(一) 如果事件A的出现必然导致事件B出现，就说A是B的特款，或者说B包含A。例如，A：“降雨”是B：“降水”的特款；B：“日最高气温≥35°C”包括A：“日最高气温≥40°C”。

(二) “二事件A，B中至少有一个出现”也是一事件，称此事件为A，B的和。类似地，事件“$A_1, A_2, \cdots, A_k$ 中至少有一个出现”称为事件 $A_1, A_2, \cdots, A_k$ 的和。例如，“降水”是“降雨”、“降雪”、“降雹”、“降冰粒”、“降霰”诸事件之和，只要其中之一发生，降水这一事件

就发生。

（三）“二事件 A，B 都出现”也是一事件，称为事件 A,B 的积。类似地，“k 个事件 A_1，$A_2,\cdots,A_k$ 都出现”称为事件 $A_1,A_2\cdots,A_k$ 的积。例如，“雷雨”是既“打雷”又“下雨”两个事件都发生的复合事件，即“雷雨”是事件“打雷”，“下雨”的积。

三、概率的概念

概率论是数学的一个分支，它研究的对象是随机事件的数量规律性。观察随机事件的方法常常是在相同条件下对事件进行重复试验，即所谓随机试验。当我们对事件多次做某一随机试验时，经常会察觉到某些事件出现的可能性大些，而另一些事件出现的可能性要小些。各事件出现的可能性既然有大有小，就使人想到该用一个数字 $P(A)$ 来表征事件 A 出现的可能性。这数字 $P(A)$ 就称为事件 A 的概率。

在深入讨论概率之前，先引进频率的概念。频率的意义是：在 n 次观测中，若观测到事件 A 出现 m 次，则事件 A 的频率 $f_A=\frac{m}{n}$，f_A 亦叫相对频数，以百分数表示。这里 m 称为事件 A 在 n 次观测中的出现频数。显然，当 A 出现的可能性愈大时，频率 f_A 也愈大，反之亦然。例如，某一风向的频率就是观测到该风向的次数占观测总次数的百分率。

频率与概率之间有着紧密的关系。事实表明，随着观测次数的增加，事件 A 的频率就会愈来愈稳定地在一个常数 p 的附近波动。当观测次数趋于无穷时，f_A 之极限即为 p，这个 p 就是 A 的概率 $P(A)$。虽然气象观测的总体是无限的，但实际观测是有限的，为此应尽可能用较长年代的记录计算频率来代表概率。这是因为当 n 充分大时，其频率可以作为概率的近似值。例如，用上海 1874—1972 年共 99 年 7 月份的暴雨日数资料，计算每 11 年平均暴雨日及其频率，即平均暴雨日数除以 31（见表 1.1）。由表可见，每 11 年的暴雨日频率的随机波动较大，变化在 0.58—3.81%之间。但随着年数的增加，暴雨日频率则稳定在 2.50%上下。因此，可以近似地认为上海 7 月份暴雨日概率是 0.025。

表 1.1 上海 1874—1972 年 7 月暴雨日数统计值

记录年代	每11年 平均暴雨日数	每11年 暴雨日频率(%)	记录年代	平均暴雨日数	暴雨日频率(%)	实有年数
1874—1884	0.18	0.58	1874—1884	0.18	0.58	11
1885—1895	0.55	1.77	1874—1895	0.36	1.16	22
1896—1906	1.09	3.52	1874—1906	0.61	1.97	33
1907—1917	1.18	3.81	1874—1917	0.75	2.42	44
1918—1928	0.91	2.94	1874—1928	0.78	2.52	55
1929—1939	1.00	3.23	1874—1939	0.82	2.65	66
1940—1950	0.82	2.65	1874—1950	0.82	2.65	77
1951—1961	0.73	2.35	1874—1961	0.81	2.61	88
1962—1972	0.45	1.45	1874—1972	0.77	2.48	99

这里必须指出，频率和概率之间虽然有着密切的关系，但其概念并不相同。概率是在一定条件下随机事件出现可能性的数量标志，它不随观测次数的多少而改变，频率却往往随着观测次数的不同而不同。随着观测次数的增加，频率将摆动地逼近概率。

与概率有关的另一概念是重现期 T。其关系式为

$$T=\frac{1}{p} \tag{1.1}$$

即重现期是概率的倒数。

例如，上海 10 月份出现大暴雨年份的概率为 0.02，则据(1.1)式，上海 10 月份大暴雨的重现期为$T=\frac{1}{0.02}=50$(年)。也就是说，上海 10 月份出现大暴雨是 50 年一遇。

四、概率的性质

（一）必然事件的概率等于 1；不可能事件的概率是 0；可能发生但非必然发生事件之概率是小于 1 的正数。因此，概率值必在 0 与 1 之间，即 $0\leqslant P(A)\leqslant 1$。

在实际问题中，经常会遇到概率接近于 1 但不等于 1 的"实际必然"事件和非常接近于 0 但不等于 0 的"实际不可能"事件。这种概率很接近于 1 的事件称为大概率事件，而概率很接近于 0 的事件称为小概率事件。

（二）对任意事件 A，有

$$P(A)+P(\bar{A})=1 \tag{1.2}$$

亦即对立事件的概率之和为 1。

例如，预报"明天有雨"事件是"明天无雨"的对立事件，若预报"明天有雨"的概率是0.1，它的对立事件"明天无雨"的概率就是 $1-0.1=0.9$。

（三）如果事件 A,B 是相容事件，在 n 次试验中，事件 A 和事件 B 出现的频数各为 $m(A)$ 和 $m(B)$，A,B 事件同时发生的频数为 $m(AB)$，则 A 事件和 B 事件二者至少出现一种的概率为

$$\begin{aligned}P(A+B)&=\frac{m(A)+m(B)-m(AB)}{n}\\&=\frac{m(A)}{n}+\frac{m(B)}{n}-\frac{m(AB)}{n}\\&=P(A)+P(B)-P(AB)\end{aligned} \tag{1.3}$$

例如，上海 1971—1975 年 7 月雨日(A)和雷暴日(B)各出现 59 天与 46 天，其中，有 8 天下了雷雨，则雷暴与雨日二者至少出现一种的概率为：

$$P(A+B)=\frac{59}{155}+\frac{46}{155}-\frac{28}{155}=0.50。$$

（四）如果事件 A,B 互不相容，则有

$$P(A+B)=P(A)+P(B) \tag{1.4}$$

例如，上海 7 月大雨日的概率为 1.0/31，暴雨日的概率为 0.8/31，于是上海 7 月大—暴雨日的概率为 $1.0/31+0.8/31=1.8/31=0.06$

（五）若 $A_1,A_2,\cdots,A_k$ 构成了互不相容的完备事件群，则有

$$\sum_{i=1}^{k}P(A_i)=1 \tag{1.5}$$

例如，上海 7 月无雨日、小雨日、中雨日、大雨日、暴雨日的概率分别为 19.8/31，7.1/31，2.3/31，1/31，0.8/31，而这 5 个事件恰好构成了互不相容的完备事件群。因此，这些事件的概率之和必定等于 1，即 $19.8/31+7.1/31+2.3/31+1/31+0.8/31=1$。

五、条件概率

在气象资料统计中，有时不仅要知道事件 A 的概率 $P(A)$，而且还需要知道在"事件 B 已出现"的条件下，事件 A 出现的条件概率 $P(A|B)$。由于增加了"事件 B 已出现"这个条件，

所以 $P(A|B)$ 一般与 $P(A)$ 不同。

设在事件 B 出现的 $m(B)$ 次中，事件 A 出现 $m(AB)$ 次，则在事件 B 出现的条件下，事件 A 的条件概率定义为

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

同理，

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

例如，某站在一个月的 30 天中，有 15 天下雨，雨日概率 $P(A)=\frac{15}{30}$；雨日同时吹东北风的有 10 天，则雨日条件下出现东北风的条件概率为 $P(B|A)=\frac{10}{15}$，即 $P(B|A)=\frac{10}{30}\div\frac{15}{30}=\frac{10}{15}$。

由条件概率的概念可以引伸出下面三条重要概率定理。

1．概率的乘法定理

据(1.6)式和(1.7)式得

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

类似地，设 $A_1,A_2,\cdots,A_k$ 为 k 个事件。其中 $k\geq2$，并满足 $P(A_1A_2\cdots A_{k-1})>0$，则有

$$P(A_1A_2\cdots A_k)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_k|A_1A_2\cdots A_{k-1}) \tag{1.9}$$

上式称为乘法公式。它的直观意义是：$A_1A_2\cdots A_k$ 同时出现的概率，等于先出现 A_1、在 A_1 出现的条件下出现 A_2、在 A_1A_2 出现的条件下出现 A_3、$\cdots$在 $A_1A_2\cdots A_{k-1}$ 出现的条件下出现 A_k 的概率的乘积。

2．全概率公式

如果 $A_1,A_2,\cdots,A_k$ 构成了互不相容的完备事件群，且事件 B 只能与事件 A_i 同时出现，即 $B=BA_1+BA_2+\cdots+BA_k$，由于 BA_i 互不相容，据加法和乘法定理求得

$$P(B)=\sum_{i=1}^{k}P(B|A_i)P(A_i) \tag{1.10}$$

上式称为全概率公式。

3．巴叶斯公式

由条件概率的定义及全概率公式得

$$P(A_i|B)=\frac{P(A_iB)}{P(B)}=\frac{P(A_i)\ P(B|A_i)}{\sum_{i=1}^{k}P(A_i)P(B|A_i)} \tag{1.11}$$

上式称为巴叶斯公式。

巴叶斯公式通常用在下列实际问题中：即只可能出现 A_1，A_2，$\cdots$有穷或可列多种不同的情况，而事件 B 只能伴随这些情况之一发生，今在 B 已出现的条件下，试求发生了情况 A_i 的条件概率

例如，已知某地各风向频率和各风向中大风的频率，欲求大风的各风向频率。现设出现大风为事件 B，出现某一方向的风为事件 A_i，A_i 构成了互不相容的完备事件群，而事件 B 能只能与 A_i 中的一种同时发生，于是某地大风各风向的频率及大风出现频率可以分别利用

表 1.2 某地大风频率统计值

各种概率 \ 风向	NW	N	NE	E	SE	S	SW	W	风向不定
$P(A_i)$	0.10	0.25	0.10	0.10	0.10	0.10	0.10	0.10	0.05
$P(B\|A_i)$	0.20	0.15	0.10	0.10	0.20	0.10	0.10	0.10	0.00
$P(A_i)$、$P(B\|A_i)$	0.02	0.04	0.01	0.01	0.02	0.01	0.01	0.01	0.00
$P(A_i\|B)$	0.15	0.31	0.08	0.08	0.15	0.08	0.08	0.08	0.00

(1.11)式和(1.10)式求得，其结果列于表1.2。由表得知，大风出现概率 $P(B)=0.02+0.04+0.01+0.01+0.02+0.01+0.01+0.01+0.00=0.13$。$P(B|A_i)$表示在某一风向条件下大风的出现概率，其中 NW,SE 风向的大风概率大。$P(A_i|B)$表示在大风发生条件下该大风属于某一风向的概率，可以看出，大风主要来自北风。

六、独立性

如果事件A的出现与否对事件B的出现与否毫无影响，即 $P(B|A)=P(B)$，

或

$$P(AB)=P(A)P(B)$$

则称事件 A,B 是相互独立的。

类似地，若有 $A_1,A_2,\cdots,A_k$ 为 k 个事件，满足

$$P(A_1A_2\cdots A_k)=P(A_1)\times P(A_2)\times\cdots\times P(A_k) \tag{1.12}$$

则称 $A_1,A_2,\cdots,A_k$ 为两两相互独立的事件。

§ 1.2 随机变量的数字特征

一、随机变量

在确定的相同条件下进行的一组观测，每次观测结果可以用一个数量x表示，这是一个变量，随着观测结果取各种数值，但不能事先知道是什么数，不过取各种数值的概率是确定的，这样的变量称为随机变量。

随机变量具有离散和集中两方面的特性。首先，随机变量取值不确定，即可以取各种不同的数值，这就是随机变量的离散特性。然而，随机性并非无规律可循，常可见到随机变量有取值的稳定性，即偶然性中存在着某种基本规律，亦即所谓集中特性。例如，从表1.3中可以看出，上海年降水日数在100年中不是一个固定不变的数值，而是逐年都在变化。即在地点（上海）、时段（1年）固定不变的条件下，降水日数的观测结果是不确定的，事先也无法知道，因此，降水日数是一个随机变量。但是，年降水日数的变化也有一定的范围，100年中最多的年降水日数为167天(1889年)，最少为94天(1873年)。而且在年降水日数为94—167天区间内之不同小区间取值的机会各不相同。例如，在91—100天和161—170天这两个小区间内，100年中各只出现了3次和2次，可是在131—140天和141—150天的小区间内则分别出现31次和22次。这说明它有集中性的特点。因此，在概率论中，我们不仅关心随机变量取什么为值，而且更主要的是关心它取某些数值的可能性的大小，也就是说，关心它取某些数值的概率有多大。

气象要素大都可看作是随机变量，且可以按照取值的特点分成两类：离散型和连续型。离散型随机变量的特征是可以在事先列举出的数值中间取值。例如，一年中各种天气现象的日

表 1.3　上海 1873—1972 年历年年降水日数

年份 年代	0	1	2	3	4	5	6	7	8	9
1870				94	132	113	103	147	153	123
1880	133	135	149	142	144	127	117	119	132	167
1890	135	141	116	132	122	122	139	146	131	136
1900	133	107	126	119	124	147	142	132	141	143
1910	140	152	139	120	133	121	144	118	137	139
1920	148	133	133	126	128	123	130	123	112	113
1930	147	154	134	145	107	153	134	150	133	140
1940	117	141	120	122	133	100	134	111	120	143
1950	153	148	147	143	161	120	134	147	127	133
1960	135	132	129	116	132	126	128	106	113	140
1970	134	98	148							

数就只能是在 0,1,2,3,•••,364,365,366 这些数字之中。连续型随机变量则可以取一定区间内的任何数。例如，某站地面气压在 1000—1030 毫巴之间变化。

二、总体与样本

气象要素的时间数列是一个无限数列。例如,上海年降水日数,其数列的全部应该是自古迄今再展延至未来极其长远岁月中的所有上海年降水日数。这种包括整个情况的数列，称之为总体或母体。但是,我们所掌握的实测资料只能是总体的一个很小部分,亦即是一个有限数列。这有限的数列就叫做此总体的一个样本或子样,这数列的长度 n 称为样本容量、样本大小或样本数。 像上海 1873—1972 年年降水日数就是上海年降水日数总体中的一个样本容量为 100 的子样。

三、数学期望

如前所述，上海年降水日数是一随机变量，而且历年年降水日数有一半集中在 131—150 天的范围之内。但是,上海年降水日数的集中位置在那儿,则是通常最关心的一个数字特征。在气象资料统计中,一般采用平均值来表征随机变量集中位置的量数。此外,集中量数还有众数，中位数等。

(一) 平均值。设有记录数列 $x_1,x_2,\cdots,x_n$。则其平均值为

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{1.13}$$

若记录数列较长,平均值可应用下列分组公式计算：

$$\bar{x}\cong\frac{1}{n}\sum_{j=1}^{k}y_j m_j \tag{1.14}$$

式中,y_j 为第 j 个组值,m_j 为第 j 个组中的频数，k为组数,n 为总次数,$n=\sum_{j=1}^{k}m_j$。

如果随机变量 x能取各种可能值 $x_1,x_2,\cdots,x_k$,x_i 各具有不同的比重(即有不同的出现次数),就可以采用加权方法,求得加权的算术平均值。

若 x_1 出现 m_1 次,x_2 出现 m_2 次,$\cdots$,x_k 出现 m_k 次，其中 $m_1+m_2+\cdots+m_k=n$，是总次数。则加权的算术平均值为

$$\bar{x}=\frac{1}{n}(x_1m_1+x_2m_2+\cdots+x_km_k)=\frac{1}{n}\sum_{i=1}^{k}x_im_i \tag{1.15}$$

如果用 x_i 的频率 f_i 来代替 $\frac{m_i}{n}$，则上式可改写为

$$\bar{x}=\sum_{i=1}^{k}x_if_i \qquad 显然\sum_{i=1}^{k}f_i=1 \tag{1.16}$$

对离散型随机变量来说，若 x 取 k 个可数的值 x_i，其相应的概率为 p_i，则称 $\sum\limits_{i=1}^{k}x_ip_i$ 为 x 的数学期望，记作 Mx，以 μ 表示。即

$$Mx=\mu=\sum_{i=1}^{k}x_ip_i \tag{1.17}$$

从(1.17)式可知，离散型随机变量的数学期望就是以概率为权的加权平均。将(1.16)式与(1.17)式进行比较，不难看出其表现形式极为相似。事实上，当观测记录足够多时，观测值的算术平均值就逼近于(概率意义下的)它的数学期望(或称总体平均值)。

例如，从表 1.1 中的资料可知，各观测时段的上海 7 月暴雨日平均值逐步逼近 0.8 天，于是 0.8 天可近似地代表上海 7 月暴雨日的总体平均值。

1．数学期望的性质：

(1) 常数的数学期望等于它本身，即 $M(C)=C$。

(2) 随机变量乘以任一常数后的数学期望等于随机变量的数学期望乘以同一常数。即 $M(Cx)=CM(x)$。

(3) 随机变量之和的数学期望等于各个随机变量数学期望之和。即 $M(x+y)=M(x)+M(y)$。

(4) 如果随机变量 x，y 相互独立，则 $M(xy)=Mx\cdot My$。

2．平均值的性质：

(1) 平均值是数学期望的无偏估计量。

设 μ 是随机变量 x 的数学期望，则平均值 $\bar{x}$ 的数学期望值为 $M(\bar{x})=M\left(\frac{1}{n}\sum\limits_{i=1}^{n}x_i\right)=\frac{1}{n}\sum\limits_{i=1}^{n}M(x_i)=\frac{1}{n}\sum\limits_{i=1}^{n}\mu=\mu$。$\bar{x}$ 是 μ 的无偏估计量，表明它对总体平均值无系统性的偏差，仅有随机性偏差。

(2) 随机变量 x 对平均值 $\bar{x}$ 的离差(即距平)的代数和等于 0。

实际上，
$$\sum_{i=1}^{n}(x_i-\bar{x})=\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\bar{x}=\sum_{i=1}^{n}x_i-n\bar{x}$$
$$=\sum_{i=1}^{n}x_i-n\times\frac{1}{n}\sum_{i=1}^{n}x_i=0$$

(3) 随机变量 x 对平均值 $\bar{x}$ 的离差平方和，比变量对其他任何常数的离差平方和要小。

设某常数C代表数列$x_1,x_2,\cdots,x_n$的平均状况，则有$\sum_{i=1}^{n}(x_i-C)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-C)^2$，于是 $\sum_{i=1}^{n}(x_i-C)^2\geqslant\sum_{i=1}^{n}(x_i-\bar{x})^2$永远成立。欲使$\sum_{i=1}^{n}(x_i-C)^2$达最小值，则上式$n(\bar{x}-C)^2$应为0，即$C=\bar{x}$。这说明在全部可选的常数中，只有平均值$\bar{x}$与数列$x_1,x_2,\cdots,x_n$的离差最小。

(4) 平均值具有稳定性。即将原数列等分若干组取平均值后，总的平均值不变。

设数列$x_1,x_2,\cdots,x_n$，将其分成k组，每组都有g个数，于是各组的平均值为$\bar{x}_j=\frac{1}{g}\sum_{i=1}^{g}x_{ij}$，$j=1,2,\cdots,k$。各组平均值的平均值为

$$\frac{1}{k}\sum_{j=1}^{k}\bar{x}_j=\frac{1}{k}\sum_{j=1}^{k}\frac{1}{g}\sum_{i=1}^{g}x_{ij}=\frac{1}{kg}\sum_{j=1}^{k}\sum_{i=1}^{g}x_{ij}=\frac{1}{n}(x_1+x_2+\cdots+x_n)=\bar{x}。$$

表 1.4 北京 1930—1969 年历年 7 月平均气温(℃)

年代 \ 年份	0	1	2	3	4	5	6	7	8	9
1930	27.4	26.3	26.3	26.2	26.7	26.5	27.0	26.6	—	—
1940	26.6	26.3	27.1	27.7	27.5	27.1	26.8	26.0	25.7	24.1
1950	25.5	25.9	26.3	26.3	23.4	26.6	25.8	25.5	26.2	25.7
1960	26.6	27.6	26.2	27.2	25.7	26.9	25.6	25.3	26.9	24.6

表 1.5 北京 1930—1969 年 7 月平均气温频数表

组距	组值	频数
23.1—24.0	23.6	1
24.1—25.0	24.6	2
25.1—26.0	25.6	10
26.1—27.0	26.6	18
27.1—28.0	27.6	7

3．算术平均值的简便计算方法：在某些气象要素中，例如气压、降水量等，有效数字位数很多，若将所有数字参加运算，不仅计算很繁，而且容易出错。现介绍平均值的简便计算方法。

设一数列$x_1,x_2,\cdots,x_n$。任选一常数a，将原数列变换为$y_1,y_2,\cdots,y_n$。这里$y_i=x_i-a$。于是平均值为

$$\bar{x}=a+\bar{y} \tag{1.18}$$

4．平均值计算举例：北京 1930—1969 年 7 月平均气温的逐年资料及其频数分配，分别列于表 1.4 和表 1.5。利用表 1.4 资料，根据(1.13)式求得北京累年 7 月平均温度为

$$\bar{x}=\frac{1}{38}(27.4+26.3+\cdots+24.6)=\frac{1}{38}\times 997.7$$
$$=26.3(℃)$$

简化计算，可令$x_i=20.0+x_i'$。

则
$$\bar{x}=20.0+\frac{1}{38}(7.4+6.3+\cdots+4.6)=20.0+6.3$$
$$=26.3(℃)$$