

气候统计

么 枕 生 著

科学出版社

气候统计

么枕生 著

科学出版社

1963

內 容 簡 介

本书主要阐述了数理统计学在气候统计中的应用和气候资料的统计整理方法两个方面。全书共分为五章,在第一章緒論中,首先阐述了气候统计的概念与必要性。在第二章中,则以气候统计为例,概要地介绍了数理统计学的基本原理。第三章至第五章主要是应用第二章中的数理统计学原理,详细地叙述了如何统计各种气候要素(包括气温、降水、风向与风速、湿度、云量、日照以及大气现象等)的方法,并根据国内资料作了统计分析。

气候统计资料是分析研究天气和气候变化规律的基础,也是农林、水利、建筑等部门生产、规划、设计所必须参考的依据。因此,本书不但可供气象、气候业务工作者和研究工作者以及有关高等院校师生阅读参考,同时还可供农林、水利、建筑等部门有关工作者参考。

气 候 统 计

么 枕 生 著

*

科学出版社出版 (北京朝阳门大街 117 号)

北京市书刊出版业营业许可登记证出字第 061 号

中国科学院印刷厂印刷 新华书店总经售

*

1963 年 12 月 第 一 版	书号: 2929 字数: 365,000
1963 年 12 月 第一次印刷	开本: 787×1092 1/16
(京) 0001—2,650	印张: 16

定价: 2.20 元

序

本书是著者自 1952 年以来,在南京大学气象系天气气候专业气候专门组多次讲述气候统计的讲稿,经逐年修改而成的,其中资料和一些数理统计应用方法与推导方法,都是著者多年来研究工作中已发表或未发表成果。本书的主要内容包包括数理统计学在气候中的应用与气候资料的整理方法两个方面,所以本书命名为“气候统计”。

本书的主要目的,就是叙述如何把数理统计学应用到气候统计方法中去,所以在内容的编排方面也以此为主。在第一章绪论中,首先阐述了用数理统计学处理气候统计的概念与必要性。在第二章中,以气候统计为示例,概要地介绍了数理统计学。从第三章到第五章,主要是应用第二章中的理论与方法,叙述了如何统计各种气候要素。数理统计学的内容是很为广泛的,仅仅在第二章以一章的篇幅叙述全部数理统计学,当然是有困难的,因此在后面各章中所提到的某些数理统计学理论或方法,有的在前面第二章中并未提到。凡遇有这种情况,著者大都附注文献,请读者另行参考,或直接简略加以叙述。

气候统计是研究气候学的重要工具,所以本书编写的另一个目的,就是为了推动气候学研究更进一步的发展。目前气候学的研究,固然在气候形成的理论研究上已具有很大的成就,但在气候资料的利用方面,却仍偏于概要的描述,因而使这门科学还不能很好地为国民经济建设服务。我们可以把同一气候资料用各种各样的方法来进行统计分析,然后得出各种各样经过加工的气候资料。这样的资料,不但可以直接供应国民经济各部门的需要,并且还能反映出更能密切结合国民经济各方面需要的,尤其是结合农业部门需要的气候学规律。当然,本书内容尚嫌贫乏,还不能很好地满足这样高的要求,这仅仅是著者的愿望而已。

因为本书主要是企图阐明如何把数理统计学应用到气候统计中去,所以对有些学者的气候统计理论,阐述得并不详尽。为了弥补这一方面的缺点,著者正在总结各学者在气候统计理论上的成就,希望将来能以书的形式提供大家参考。

著者才疏力薄,知识有限,书中缺点甚或错误,在所难免,希望读者不吝赐教,概予指正,以便再版时加以修改。

本书中的计算工作,主要由朱培如、马宜珍等同志多年来代为计算的;原稿完成后又承蒙施永年同志阅读,提出了宝贵的意见,著者一并在 此 谨 致 谢 意。

1963 年 1 月 15 日 么枕生志于南京大学

目 录

序	ix
第一章 緒 論	1
§ 1. 气候統計的目的与含义	1
§ 2. 气候統計的应用范围	1
§ 3. 气候統計的基础——概率論	2
§ 4. 气候統計的发展	2
第二章 数理統計学概要	5
§ 1. 概率論基础	5
1.1 概率的定义	5
1.2 基本定理	5
(1) 概率的加法定理(总概率定理) (2) 概率的乘法定理(复合概率定理)	
1.3 二項分布	6
1.4 多項定理	8
§ 2. 一个变数的实验頻数分配	8
2.1 頻数表	8
2.2 頻数分配的图示法	9
(1) 直方图与頻数多边形 (2) 累积頻数多边形 (3) 頻数曲綫 (4) 累积頻数曲綫	
2.3 頻数分配的量数	11
(1) 矩 (2) 集中量数 [平均数 众数 中位数] (3) 离散量数 [极差(变程) 平均差 标准差 四分位差]	
2.4 偏度与峯度	16
§ 3. 一个变数的理論頻数分布	16
3.1 分布函数	17
(1) 連續分布 (2) 不連續分布 (3) 平均数与标准差 [随机变数的平均数 随机变数的函数平均数 标准差]	
3.2 正态分布	19
(1) 正态分布函数 (2) 正态分布的計算	
3.3 偏斜分布	22
(1) 皮尔孙頻数曲綫 (2) 对数正态分布 (3) 爱治沃斯級数	
3.4 二項分布	25
3.5 泊松分布	26
§ 4. 多变数的頻数分配	27
4.1 两个变数分布的特性	27
(1) 連續分布 (2) 不連續分布	
4.2 矩、边缘平均数与边缘方差	29
4.3 边缘平均数与边缘方差的計算	29

4.4 条件平均数与条件方差	31
(1) 条件平均数 (2) 条件方差	
4.5 条件平均数与条件方差的计算	32
4.6 随机变数线性函数的平均值与方差	32
§ 5. 迴 归	33
5.1 直綫迴归	33
(1) 直綫迴归方程 (2) 直綫迴归方程的计算 (3) 函数直綫式的计算	
5.2 曲綫迴归	35
§ 6. 相 关	36
6.1 直綫相关	36
(1) 协方差 (2) 相关系数 (3) 相关系数的计算	
6.2 非直綫相关	38
§ 7. 多变数的迴归与相关	39
7.1 多变数的迴归	39
7.2 多变数的相关	41
(1) 偏相关 (2) 多重相关	
§ 8. 抽样分布与显著性检验	43
8.1 随机抽样与抽样分布	43
8.2 平均数的抽样分布	43
8.3 分位数	44
8.4 平均数的显著性检验	44
8.5 m 的置信限界	45
8.6 二平均数差值显著性的检验	45
8.7 百分数差值显著性的检验	46
8.8 相关系数显著性的检验	46
8.9 相关系数差异显著性的检验	48
§ 9. χ^2 分布	48
9.1 χ^2 的分布函数	48
9.2 χ^2 分布的应用	49
(1) 方差的分布 (s^2 分布) (2) s^2 分布的分位数 (3) σ^2 的置信限界	
(4) 适度配合的检验	
§ 10. t 分布	52
§ 11. F 分布	53
第三章 气温的统计方法	56
§ 1. 温度观测数列的订正	56
1.1 温度观测数列的检查与订正	56
1.2 差值订正法与一般订正法	58
(1) 相关图与联系方程 (2) 不均匀性的订正 (3) 订正到长年时期	
(4) 订正适当性的标准 [差值订正法的适当性标准 一般订正法的适当性标准]	
§ 2. 平均温度	69
2.1 平均温度的计算与检验	69

2.2 日平均温度	70
2.3 月、年平均温度	74
2.4 等温线的绘制	74
§ 3. 温度的谐波分析(温度周期分析 I)	75
§ 4. 温度的概率与累积概率	81
4.1 概率的意义与计算方法	81
4.2 分位数的计算	84
§ 5. 温度的变率	86
5.1 平均绝对变率(平均差)与平均相对变率(相对平均差)	86
5.2 标准差与相对标准差	88
§ 6. 温度的序列变率	89
§ 7. 极端温度	91
7.1 极端温度的基本统计方法	91
7.2 极端温度的订正	92
7.3 极值的频数分配	93
(1) 龔貝理論 (2) 秦肯遜理論	
§ 8. 非季节性暖期与冷期	100
§ 9. 指标温度与定限温度出现日期、持续时期及其概率	102
9.1 指标温度平均出现日期与持续时期	102
9.2 指标温度出现日期与持续时期的概率	103
9.3 由定时观测记录计算超过定限温度的概率与持续时间	105
§ 10. 霜 冻	108
§ 11. 积温与度日	110
11.1 积 温	110
11.2 度 日	114
§ 12. 温度的时间序列分析	117
12.1 温度趋势的确定	117
(1) 最小二乘法 (2) 正交多项式 (3) 移动平均 [简单移动平均 移动多项式]	
12.2 温度周期振动的确定(温度周期分析 II)	129
(1) 周期图分析 (2) 相关图分析 [序列相关 自回归序列 相关图振动的周期]	
第四章 降水量的统计方法	135
§ 1. 降水观测数列的检查与订正(比值订正法)	135
1.1 相关图与联系方程	135
1.2 不均匀性的检查、订正与插补	137
1.3 订正到长年时期	140
1.4 订正适当性的标准	141
§ 2. 降水量	142
2.1 年、月降水量	142
2.2 5 日与 10 日的降水量	143

2.3 日降水量	143
(1) 雨日平均降水量 (2) 最大日降水量 [平均最大日降水量与绝对最大日降水量 5年移动平均最大日降水量 绝对最大日降水量的频数分配]	
(3) 具有各种日降水量的日数 [降水日数的订正 降水日数的曲线配合与日降水量的频数分配]	
2.4 雪 被	151
2.5 等雨量綫图的繪制	152
§ 3. 降水的概率	152
3.1 雨日概率	152
3.2 各地同时具有雨日的概率	152
3.3 年、月降水量的概率与保证概率	156
(1) 图解算法 (2) 理論計算法	
§ 4. 湿期与干期的概率	162
4.1 一定湿期与一定干期长度的概率	163
4.2 湿期与干期各种序列的概率	164
(1) 柏努利模式 (2) 馬尔科夫模式 (3) 波亚模式 (4) 天气形势与地形对于湿期与干期的影响	
4.3 对数定律	173
§ 5. 湿月、干月与干湿月气候循环的概率	174
5.1 湿月与干月概率	174
5.2 湿月序列与干月序列	180
(1) 湿月序列与干月序列的概率 (2) 湿月序列与干月序列的平均长度与方差	
5.3 干湿月气候循环	185
(1) 干湿月气候循环的概率 (2) 干湿月气候循环长度的平均值与方差	
§ 6. 降水的持續性	187
6.1 持續性的量数	188
(1) 貝松持續系数 (2) 持續比	
6.2 雨日持續概率与雨日持續序列的长短	189
§ 7. 降水的变率	191
7.1 平均绝对变率与平均相对变率	191
7.2 降水变率的其他量数	194
7.3 降水变率的精度檢驗	196
7.4 降水变率所要求的測年长短	198
§ 8. 暴 雨	201
8.1 暴雨的定义与实验式	201
8.2 暴雨概率	203
8.3 暴雨频数分配	203
§ 9. 降水的周期振动与干湿年气候循环	208
第五章 其他气候要素的統計方法	211
§ 1. 气压的統計方法	211
§ 2. 风的統計方法	212
2.1 风 向	212

2.2 风 速·····	214
2.3 极端风速·····	215
2.4 盛行风向·····	217
2.5 合成风向与合成风速·····	219
2.6 风的常定度·····	221
2.7 风速的谐波分析·····	221
2.8 风的频数分配·····	222
§ 3. 湿度的统计方法·····	223
§ 4. 云的统计方法·····	225
§ 5. 日照的统计方法·····	227
§ 6. 大气现象的统计方法·····	229
6.1 雾、雷暴与雪暴资料的检查与基本统计方法·····	229
6.2 大气现象的概率·····	230
6.3 概率论基本定理的应用·····	231
(1) 各种风向时霁的概率 (2) 各种风向时的降水概率 (3) 一般概率的计算	
6.4 大气现象的频数分配·····	235
参考文献 ·····	243

第一章 緒 論

§ 1. 气候統計的目的与含义

自然科学研究的目的,就是寻求自然規律,以便掌握和控制自然。为了达到这个总的目的,气候統計的目的有三:(1)推求量数,以便分析气候的客观規律;(2)运用实验分布函数或推导分布函数分析气候,以便节省統計劳力;(3)訂正記錄,以便求得均匀的气候記錄。

根据数理統計学的方法,我們可以把杂乱无章的記錄分析清楚,找出記錄中的气候客观規律。我們知道,現有气候記錄只是总体(可能有的所有記錄)中的一个随机样本。气候統計的目的,就是根据測得的样本,利用分布函数去分析未知的总体,以免被样本的偶然性所矇蔽。不过,这只能根据概率論与数理統計学,在某种概率标准上对有关总体方面的假設加以測驗而已。在拟定这些假設时,一方面必須应用数理統計学的理論,另一方面还必须参考气候学中有关气候形成的規律。所謂气候規律,就是各种气候要素与現象,因为受到气候因素(緯度、位置、水陆分布、地貌与地面特性等)的影响,所形成的在時間与空間上的变化特性。在气候統計工作中,固然应当广泛地应用数理統計学方法,但是也不能仅仅形式上采用一些統計公式,还应当考虑到統計假設是否和基本气候規律相符合。因此,气候統計与一般的数理統計学并不尽同,它应当在数理統計学的基础上,寻求符合于气候規律的另一种統計理論。其次,在气候統計中更为特殊的,就是在統計气候記錄以前,首先需要訂正記錄。我們知道,各地气候記錄长短不一,因此在統計以前,首先必須把記錄訂正到同样长短的年代,否則統計結果就有不同程度的誤差或錯誤。

§ 2. 气候統計的应用范围

用統計理論去整理气候資料的目的,不只是为了研究气候形成的規律,以及研究不同地区气候体系的特征,更主要的是为了广泛地为国民經济建設提供必要的气候資料。例如,在城市建設方面,工厂、机关与住宅的布置,首先需要了解当地的盛行风。在建造高大建筑物时,如鉄塔、水塔与烟囱等,就要考虑到风压。在建筑水力发电、灌溉工程、防洪防汛工程、鉄路、公路、矿山以及电力輸送綫时,必須考虑降水強度、暴雨特性、連續降水日数、降水年变化、降水变率等資料。又如在冻土地带鋪設地下水管、开凿下水道或各种工程的基础时,必須参考土温与冻土的資料。土壤温度的資料,不但对于地下工程建造的設計有关,对于其他的許多实际問題(如估計农作物成长的条件)也是必需的資料。有些工程,如水泥工程、暖气工程与冷气工程也必须考虑温度資料。在設計粮食倉庫、制葯厂、印刷工业、造纸工业、冷气工程等还要参考湿度的資料。

随着社会主义建設事业的迅速发展,气候資料的基本統計与分析結果,已經不能滿足經济建設方面日益增长的需要。为了使气候資料能密切配合国民經济部門的需要,特别是农业生产方面的迫切需要,必須进一步应用統計理論改进現有的統計方法。例如,对我

对农业生产影响最大的灾害性现象——干旱与干风、霜冻与冰雹,以及暴风与长期雷雨等,气象部门首先必须根据数理统计理论提出其发生的规律,计算其出现的概率与累积概率,以便推测哪些气象灾害在哪些地区和在哪些季节内是机会最多的,或是最少的,以供生产部门参考。一般的温度概率,在国民经济建设方面应用也是很广泛的。例如,建筑设计机构可以根据温度的统计资料,推断哪种温度情况是机会最多的,或者哪种温度情况是不易发生的,以便考虑决定房屋的取暖设计方案等。

总之,气候统计资料应用的范围是十分广阔的,因而气候统计的方法应当大力改进,以满足许多科学部门、国民经济、国防建设部门,尤其是支援农业上的迫切需要。

§ 3. 气候统计的基础——概率论

一个随机变量或随机变数就是一个量,其值和机会有关。随机变数可以具有某些值,而这些值又具有其相应的概率。因为这些变数是和概率相联系的,所以称为随机变数。所谓随机变数就是规定于样本空间的函数。在样本空间中,每一个样本点都有一个相应变数的值。例如,在 n 个柏努利 (Bernoulli) 测验中,成功数目 r 就是一个函数。此函数规定于包括 2^n 个样本点的样本空间,所以 r 就是随机变数。在柏努利测验中,每个样本点的出现都具有概率 $1/2^n$ 。

例如,我们可以把干湿年的演变看作是一种试验。在此试验中,我们并不能确切预知干年的出现(用 0 表示),也不能确切预知湿年的出现(用 1 表示)。但是,在气候演变过程中,所有可能出现结果的总体或集 $\{0, 1\}$ 是完全可以描述的。因此,我们虽不能确切预知千年或湿年的出现,但千年或湿年出现的概率是可以计算的。如果历年气候干湿的变化彼此之间并无影响,那末这种试验就称为随机试验。试验的结果就是典型的随机变数。这种所有可能出现的总体(集),就称为试验空间或样本空间。这些试验结果(千年或湿年)就是样本点。

各种气候要素都具有不同的变化性,这些变化性就是气候变化的客观规律。这种客观规律就是由于自然世界的内在因素与外在因素随机(偶然)配合的结果。因此,气候现象,如温度、湿度、雨量、风向与风速以及其他气候现象等,都是随机变数。我们知道,气候变化的外界条件与初始条件(如太阳辐射的规律、地面不同的性质、目前的大气环流形势)固然是已知的,但是空气流动以后,大气和太阳辐射及其与地面性质间的错综复杂的情况是难以掌握的。某一时期的气候情况,只是这种错综复杂的大气过程中一种不可预测的机会事件。这种机会事件也可以看作随机变数。概率论的主要目的,就是寻求随机变数(机会事件)的出现概率,其大小可用数字表示。气候要素与现象既然是随机变数,所以气候统计的基础就是概率论。

概率乃是一个统计量。如果我们能够确定某种气候要素或现象出现的频数,那末如记录相当长,其概率就可以掌握。设 p 代表其概率,则 p 的意义就是:在一系列 n 次可能出现的会(即 n 次观测)中,某种气候要素或现象平均出现 pn 次。 pn 仅仅是平均数,个别出现的频数是环绕着平均数 pn 而振动的。概率论的目的就是要估计这些个别频数。

§ 4. 气候统计的发展

气候统计的发展是由于概率论应用于气候统计的结果。概率论乃是在 17 世纪起源

于解决賭博問題，至于建立概率論的理論基础，則應該归功于柏努利（Bernoulli, 1654—1705）与拉普拉斯（de Laplace, 1749—1827）等人的努力。自拉普拉斯时代以后，对于概率論貢獻最大的是苏联学者，而近代的概率論也主要是由苏联学者如 А. Н. 科尔莫格罗夫（Колмогоров）等所建立起来的。

气候数理統計学是近 40 年来逐渐发展起来的，尤其在苏联这门科学发展得最快。在苏联由于计划經濟的实施，各方面都需要各种精确的气候資料，所以这门科学在苏联发展得最为迅速。例如，苏联学者 А. А. 卡明斯基（Каминский）、Е. С. 魯宾什晉（Рубинштейн）、О. А. 德洛夫多夫（Дроздов）与 Е. С. 庫茨涅佐夫（Кузнецов）等在奠定气候統計的理論基础方面都具有很大的成就。他們在把数理統計应用到气候統計的过程中，設計了許多气候統計方法，尤其在資料的訂正理論方面更有貢獻。德洛夫多夫（1957）在“气象观测的气候学整理方法”一书中，就总结了苏联在气候統計方面的貢獻。

奥地利气候学家 V. 康拉德（Conrad）自本世紀 20 年代以来，在气候統計方面也作了很多工作，并且在 1944 年写出了“气候学中的方法”一书，总结了以前各学者的工作以及他自己的工作。該书在 1950 年又由 L. W. 波拉克（Pollak）加以增补。L. W. 波拉克在本世紀 20 年代以来，特別在周期分析方面著述很多。英国气候学者 C. E. P. 布罗克斯（Brooks）从本世紀 20 年代到 50 年代，在气候統計方面也具有貢獻，他和 N. 卡陆塞斯（Carruthers）在 1953 年合著了“气象学統計方法手册”一书，該书在数理統計学如何应用于气候統計的問題上，作了詳細的总结。高桥浩一郎也在 1944 年与 1956 年著有“气象統計”一书，在气候統計方面也提出了一些理論与方法。H. A. 潘諾夫斯基（Panofsky）与 G. W. 布瑞尔（Brier）1958 年曾著有“統計学在气象学中一些应用”一书。該书虽嫌簡略，但尚有独到之处。

近 30 年来，国内学者在气候統計方面也作了很多工作。尤其在祖国解放以后，由于党和毛主席的正确领导，在三面紅旗的光輝照耀下，我国气候工作者在气候統計方面也取得了很大的成績，在气候数理統計的建設工作中，也貢獻出应有的力量。

在这些工作中，张宝堃（1929—1934）^[1]曾統計分析了南京风向与天气的关系，并且在 1934 年統計过候温，又根据物候現象提出了中国的四季划分与分配^[2]。吕炯（1938）^[3]曾統計过逐候平均温度的年变化，提出了春季迴寒的問題。卢鋈（1939）^[4]統計过东亚各地的月平均气压，并根据其分布分析了地面气流。朱炳海（1949）^[5]曾統計过我国的降水相对系数，并用以說明我国的气候特征。楊鑑初（1949）^[6]曾根据 1947 年我国夏季降水量的离差，統計分析了水旱災害。朱炳海（1955）^[7]根据雨日降水量的計算，分析过我国夏季的降水強度，并且計算了某一年代內可能出現的絕對最大日降水量。徐淑英（1956）^[8]曾計算过黄河流域各地区同时出現各种旱涝的概率（相对頻数），并且还計算了干旱的持續概率。朱炳海（1957）^[9]根据他在降水強度以及降水变率与旱涝方面的研究，作出了中国的降水区划。著者曾分別在 1957 年^[10]与 1959 年^[11]計算过农业指标温度与积温，以及我国东部境内各地降水量的保証概率。楊鑑初（1958）^[12]还計算过我国 1951—1955 年各月平均日际气压变率、平均日际变高与平均日际变温，以便分析气候变化的季节性。他并且于 1959 年^[13]根据北半球海平面气压場的月偏距中心，討論了某些地区的气候特性。

涂长望在 1937 年^[14, 15]曾配合过迴归方程，作为我国低温和水旱預告的根据。

刘恩兰（1950）^[16]、朱炳海（1957）^[17]与著者（1958）^[18]都曾先后統計过降水的变率，朱炳

海并且根据降水离差统计了我国旱涝的规律。著者(1950)^[19]在“論杭州气候”一文中,更努力于把统计学方法应用于区域气候的分析中。

程純枢(1942)^[20]与朱崗昆(1947)^[21]都曾分別作过气压年、日变化的分析与谐波分析,著者(1951)^[22]更用谐波分析方法分析了中国大量台站的年温变化。謝义炳(1943)^[23]、涂长望与张汉松(1944)^[24]曾用历史文献資料作过周期分析。章基嘉等(1959)^[25]曾根据自相关函数,計算过候降水量的自相关性,以便寻求其周期性。王紹武(1961—1962)^[26,27]与史久思、徐羣(1962)^[28]还根据气象观测的資料,利用周期图法,分別分析了我国大气活动中心的周期,以及长江中下游5个台站5—9月降水量的周期。

在数理统计理論如何应用于气候统计的問題上也有許多工作,例如,朱崗昆(1942)^[29]提出变异分析法在气象记录中的应用問題。徐尔灝(1950)^[30]曾論述过年降水量是常态分配的。中央气象局气象研究所气候研究室的同志們(1959)^[31]曾提出計算平均差的簡化公式。最近,著者(1962)^[32]提出所謂干湿气候循环的概率分布函数,并且推导出我国气候变迁具有4年左右的基本周期。

气候資料的统计方法虽已有相当巩固的理論基础和各种各样的统计方法,但是随着科学的迅速发展与国民經济部門要求的不断提高,气候统计理論还远远落后于客观的需要,因此,气候工作者必須滿足社会主义建設的迫切要求,繼續高举三面紅旗,为攀登更能結合实际的气候数理统计科学新高峯而共同努力。

第二章 数理統計学概要

§ 1. 概 率 論 基 础

1.1 概 率 的 定 义

某种事件出現的次数对于可能出現次数之比,就称为該事件的概率。这就是概率的古典定义。根据此定义,概率用 0 与 1 間的数字代表之。在极端情况下,如概率为零,就說明某事件出現是不可能的。反之,在另一极端情况下,如概率为 1,則某事件是必然出現的。

如在 n 次同等观测中,某事件出現的次数为 ν , 則 ν 称为在 n 次观测中此事件的绝对頻数(見表 1 第 3 行),而比值

$$f = \frac{\nu}{n}$$

称为相对頻数(見表 1 第 5 行)。在这里相对頻数也称之为概率。实际上,相对頻数 f 就是概率的实驗值(样本值),概率的理論值就是某事件出現的真实相对頻数或相对頻数的极限值。当观测次数愈多,則相对頻数愈能代表概率。当观测次数大到无限大时,則相对頻数就是真正的概率。记录愈长,則計算概率的精度就愈大,不过并不是观测记录无限增长时,概率的精度就无限增大,而是渐近于一常数。因此,即使具有最长的观测数列,但仍有些疑問。我們計算的概率仅仅是概率的估計值,真正概率我們是无法知道的。因此,統計概率时,必須計算其誤差,給出其置信界限 (Confidence limits) (參見第三章 § 4.1)。

根据习惯,我們用大写字母 $A, B, \dots$, 代表各事件。复合事件 AB 代表 A 与 B 同时出現的事件。 $\bar{A}$ 代表 A 不出現的事件。 $A \cup B$ 代表 A 或 B 出現的事件。 E 代表永远出現的必然事件。事件 A 的概率用 $P(A)$ 代表之。于是,

$$P(E) = 1.$$

1.2 基 本 定 理

(1) 概率的加法定理(总概率定理)

定理 1 如兩事件 A 与 B 是彼此相斥的, 那末或此或彼的出現概率就等于每个事件出現概率的和:

$$P(A \cup B) = P(A) + P(B). \quad (1.1)$$

設有二个以上的事件 $A_1, A_2, \dots, A_n$, 其中每两个事件是彼此相斥的, 則一般化或此或彼的概率为

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n). \quad (1.2)$$

設 $A = A_1 \cup A_2 \cup \dots \cup A_n$, 則

$$P(A) = P(A_1) + P(A_2) + \dots + P(A_n). \quad (1.3)$$

概率 $P(A_1), P(A_2), \dots, P(A_n)$ 为 A 的各个相斥形式的部分概率, $P(A)$ 为 A 的总概率, (1.3)式就是下面的定理.

定理 2 事件 A 的概率为其各个相斥形式 $(A_1, A_2, \dots, A_n)$ 的概率之和.

如事件 $A_1, A_2, \dots, A_n$ 不仅是相互排斥的,而且是穷举的(即在 A 的各个形式中必有一个形式出现),于是

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1. \quad (1.4)$$

例如, $P(A) + P(\bar{A}) = 1$, 所以

$$P(\bar{A}) = 1 - P(A), \quad (1.5)$$

并且我們知道 $P(A)$ 至多为 1, 即

$$0 \leq P(A) \leq 1. \quad (1.6)$$

(2) 概率的乘法定理(复合概率定理)

定理 3 两事件 A 与 B 都出现的概率等于一个事件的绝对概率和另一个事件的条件概率的乘积:

$$\begin{aligned} P(AB) &= P(A)P(B|A) \\ &= P(B)P(A|B), \end{aligned} \quad (1.7)$$

此处 $P(B|A)$ 是在 A 出现的条件下, B 的条件概率. 为了和 $P(B|A)$ 分别起见, $P(B)$ 常称为 B 的绝对概率. $P(A|B)$ 与 $P(A)$ 的意义同此.

如一般化, 则

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2|A_1) \dots P(A_n|A_1 A_2 \dots A_{n-1}). \quad (1.8)$$

定理 4 如二事件是随机无关的, 那末二事件都出现的概率就等于每个事件出现概率的乘积:

$$P(AB) = P(A)P(B). \quad (1.9)$$

如为上述一般化的情况, 且每一个事件对于其余事件都是随机无关的, 则

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2) \dots P(A_n). \quad (1.10)$$

設 $A_1, A_2, \dots, A_n$ 为 n 个相斥的事件, 其中一个事件必然出现, 即

$$E = A_1 \cup A_2 \cup \dots \cup A_n.$$

其次, 設任何事件 B 仅能联同事件 A_i 出现, 或写作如下符号:

$$B = B A_1 \cup B A_2 \cup \dots \cup B A_n. \quad (1.11)$$

因为 $B A_i$ 是相互排斥的, 所以根据加法定理与乘法定理, 求得

$$P(B) = \sum P(B|A_i)P(A_i). \quad (1.12)$$

这就是总概率定理的另一形式.

1.3 二项分布

設一事件 A 在单个測驗时具有成功(出现)的概率 p , 失败(不出现)的概率 $q=1-p$. 我們作 n 次測驗, 不計其次序, 在每次測驗中其結果和以前的結果是随机不相联系的, 且 p 不变. 这种測驗就是所謂柏努利測驗. 設在 n 次柏努利測驗中, r 次出现的概率为 $p_n(r)$, 則

$$p_n(r) = \binom{n}{r} p^r q^{n-r}. \quad (1.13)$$

換句話說， n 个柏努利測驗的样本空間包括有 $\binom{n}{r}$ 个样本点，每个点具有概率 $p^r q^{n-r}$ 。

例如，設湿月(月降水量 > 月标准值)的概率为 p 。如果沒有月降水量恰好等于标准值的月份，則干月(月降水量 < 月标准值)的概率为 $1 - p = q$ 。現在設 $p = q = \frac{1}{2}$ ，問在一年之內湿月与干月出現的概率如何？

在我們的問題中， $n = 12$ ， $p = q = \frac{1}{2}$ ， $r = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$ 。

$$\begin{aligned}
 p_{12}(0) &= \binom{12}{0} \cdot \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^{12} = 1 \cdot \frac{1}{4096} = \frac{1}{4096} = 0.0002441, \\
 p_{12}(1) &= \binom{12}{1} \cdot \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^{11} = 12 \cdot \frac{1}{4096} = \frac{12}{4096} = 0.0029297, \\
 p_{12}(2) &= \binom{12}{2} \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^{10} = 66 \cdot \frac{1}{4096} = \frac{66}{4096} = 0.0161133, \\
 p_{12}(3) &= \binom{12}{3} \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^9 = 220 \cdot \frac{1}{4096} = \frac{220}{4096} = 0.0537109, \\
 p_{12}(4) &= \binom{12}{4} \cdot \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^8 = 495 \cdot \frac{1}{4096} = \frac{495}{4096} = 0.1208497, \\
 p_{12}(5) &= \binom{12}{5} \cdot \left(\frac{1}{2}\right)^5 \cdot \left(\frac{1}{2}\right)^7 = 792 \cdot \frac{1}{4096} = \frac{792}{4096} = 0.1933593, \\
 p_{12}(6) &= \binom{12}{6} \cdot \left(\frac{1}{2}\right)^6 \cdot \left(\frac{1}{2}\right)^6 = 924 \cdot \frac{1}{4096} = \frac{924}{4096} = 0.2255860, \\
 p_{12}(7) &= \binom{12}{7} \cdot \left(\frac{1}{2}\right)^7 \cdot \left(\frac{1}{2}\right)^5 = 792 \cdot \frac{1}{4096} = \frac{792}{4096} = 0.1933593, \\
 p_{12}(8) &= \binom{12}{8} \cdot \left(\frac{1}{2}\right)^8 \cdot \left(\frac{1}{2}\right)^4 = 495 \cdot \frac{1}{4096} = \frac{495}{4096} = 0.1208497, \\
 p_{12}(9) &= \binom{12}{9} \cdot \left(\frac{1}{2}\right)^9 \cdot \left(\frac{1}{2}\right)^3 = 220 \cdot \frac{1}{4096} = \frac{220}{4096} = 0.0537109, \\
 p_{12}(10) &= \binom{12}{10} \cdot \left(\frac{1}{2}\right)^{10} \cdot \left(\frac{1}{2}\right)^2 = 66 \cdot \frac{1}{4096} = \frac{66}{4096} = 0.0161133, \\
 p_{12}(11) &= \binom{12}{11} \cdot \left(\frac{1}{2}\right)^{11} \cdot \left(\frac{1}{2}\right)^1 = 12 \cdot \frac{1}{4096} = \frac{12}{4096} = 0.0029297, \\
 p_{12}(12) &= \binom{12}{12} \cdot \left(\frac{1}{2}\right)^{12} \cdot \left(\frac{1}{2}\right)^0 = 1 \cdot \frac{1}{4096} = \frac{1}{4096} = 0.0002441.
 \end{aligned}$$

上面各数分別代表一年內湿月随机出現 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 与 12 次的概率。因为湿月概率 p 等于干月概率 q ，所以干月随机出現 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 与 12 次的概率也有如上列一样的各数，并且干月和湿月組合的方式如下：

0 湿月与 12 干月	4 湿月与 8 干月
1 湿月与 11 干月	5 湿月与 7 干月
2 湿月与 10 干月	6 湿月与 6 干月
3 湿月与 9 干月	7 湿月与 5 干月

8 湿月与 4 干月	11 湿月与 1 干月
9 湿月与 3 干月	12 湿月与 0 干月
10 湿月与 2 干月	

現在假設有 100 年降水資料,則湿月与干月的各种組合应有下列頻数:

一年內各月都为湿月	0.02
一年內 1 个月为干月, 11 个月为湿月	0.29
一年內 2 个月为干月, 10 个月为湿月	1.61
一年內 3 个月为干月, 9 个月为湿月	5.37
一年內 4 个月为干月, 8 个月为湿月	12.08
一年內 5 个月为干月, 7 个月为湿月	19.34
一年內 6 个月为干月, 6 个月为湿月	22.56
一年內 7 个月为干月, 5 个月为湿月	19.34
一年內 8 个月为干月, 4 个月为湿月	12.08
一年內 9 个月为干月, 3 个月为湿月	5.37
一年內 10 个月为干月, 2 个月为湿月	1.61
一年內 11 个月为干月, 1 个月为湿月	0.29
一年內各月都为干月	0.02

一年之內 12 个月都为湿月或都为干月的概率最小,一万年內只有 2 年。一年中 6 个月为湿月, 6 个月为干月的概率最大,一百年內有 23 年如此。干湿月实际是不相联系的,所以如有长年資料,干湿月的分配就接近于此。

1.4 多項定理

設有事件 $A_1, A_2, \dots, A_k$ 。我們現在要求在 n 次測驗中, A_1 出現 r_1 次, A_2 出現 r_2 次, $\dots, A_k$ 出現 r_k 次的概率为 $P_{r_1, r_2, \dots, r_k}$, 而 $r_1 + r_2 + \dots + r_k = 1$, 則

$$P_{r_1, r_2, \dots, r_k} = \frac{n!}{r_1! r_2! \dots r_k!} p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}. \quad (1.14)$$

(1.14)式就称为多項定理。二項定理就是 $k = 2$ 的特殊情形。

§ 2. 一个变数的实验頻数分配

2.1 頻 数 表

为了节省時間,記錄要分組,即記錄相差不远者分为一組。要分多少組(組距的大小)是和記錄的长短与精度有关的。

根据經驗,一般以分 10—20 組为宜。組数决定以后,就要确定組界。最后計算每組应有的記錄次数,这就是把記錄分組为頻数表(如表 1)。

在表 1 中,把頻数从上向下累加时,就求得累积頻数,此值說明上海 10 月絕對最低温度在某值以下的总頻数。