么枕生 编著

气候统计学基础

科学出版社

-173.

气候统计学基础

(统计气候学理论 I)

么枕生 编著

科学出版社

1984

内 容 简 介

气候统计学是一门气候学结合数理统计学的边缘学科，目前正在发展，尚未定型．编写本书的目的就是把数理统计学概念引进到气候学中，即应用数理统计学中的分布函数、抽样理论、统计推断与随机过程，并进一步予以发展和创新，以解决现有气候记录样本所给信息的最优模拟记录总体，估计与推断记录总体的特征．本书就是有关这些问题的基础理论部分．

本书内容不仅是统计气候学的基本理论，也是长期统计预报不可缺少的基础；不仅提供研究统计气候学基本理论的重要手段，也为一切与气候有关的其他科学研究、资源区划、工程设计、经济规划以及生产管理提供统计理论与方法．

气候统计学基础

（统计气候学理论 I）

么枕生 编著

责任编辑 郑秀灵

科学出版社出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

*

1984年8月第一版　开本：850×1168 1/32

1984年8月第一次印刷　印张：19

印数：0001—4,000　字数：499,000

统一书号：13031·2631

本社书号：3625·13—15

定价：3.50元

前　　言

概率论起源于十七世纪，始于研究赌博上的概率问题．在十八世纪，概率论开始应用于科学，这样就使概率论得到很大的发展，到二十世纪，概率论的发展更为迅速．

概率论是纯数学，数理统计学是应用数学．概率论是数理统计学的基础，所以数理统计学是在概率论出现后才逐步发展起来的（也包括了一些概率论的内容），到本世纪40年代以后才开始有完整的系统．现在数理统计学的理论发展愈来愈快，其应用范围也愈来愈广，除应用于气象科学外，已经应用到所有其他学科，如生物、物理，化学、地质、地理、工农业技术、经济、商业、医疗、体育以及环境保护等等各个方面．

气候记录是实验性质的观测记录，也是观测的时间序列（随机过程），现有记录在时间和空间上形同瀚海，要在这样庞大而复杂的气候资料中分析归纳气候变化规律，并且在气候规律的基础上去预报未来的气候变化，就必须引进数理统计学与概率论中已有的理论与方法，并结合气候学的特殊情况加以发展与创新．因此，气候统计学就是在数理统计学与概率论基础上，结合气候学而发展起来的一门边缘学科．

统计学方法可以分为两大类，即叙述法与解析法．叙述法是把大量的观测记录化为少数统计量，例如平均数、方差、偏度、峰度与相关系数等．这样的统计学内容最早被引进气候学中．气候学家 V. Conrad 在本世纪10—50年代就一直努力把统计学的叙述法引进到气候学中，且于1950年在 Pollak 的协作下，写出《气候学方法》一书．解析法是用分布模式与回归模式等（总称为统计模式）去拟合观测记录（样本），估计、推断总体参数和表征总体中的分布与变化特点．在本世纪的20—50年代，气候学家 C. E. P.

Brooks 在解析法如何应用于气候学方面做了许多工作，且于 1953 年在 Carruthers 的协助下，出版了《气象学中统计方法手册》著作.

早在本世纪40年代初，作者在农业气象学与气候学的教学中，就开始认识到数理统计学方法在气候学中的重要性，并自 1958 年以来专门从事气候统计的教学与研究工作. 近年来，作者更认识到气候统计学与统计气候学是不同的两门科学. 所谓统计气候学是用气候统计学方法为工具去分析与预报气候的时空问题；所谓气候统计学，就是统计气候学的理论基础. Conrad 的工作多属于统计气候学的范围，而 Brooks 的工作就已由统计气候学范围初步进入气候统计学领域. 作者的研究工作也是以 1958 年为分界线的，在这以前的研究工作属于统计气候学范畴，以后的研究工作属于气候统计学范畴.

近年来，国际上对于统计气候学与气候统计学的研究更是蓬勃发展，且于 1979 年底在日本东京举行过第一次国际统计气候学会议. 会议论文集——《统计气候学》(Statistical climatology) 于 1980 年出版，其中有由日本人铃木写了一篇统计气候学发展方面的总结论文，由另一日本人吉野写了一篇会议总结. 吉野也认为气候统计学与统计气候学并不相同，不过由于两者很难严格分开，所以把包括气候统计学和统计气候学的论文集命名为《统计气候学》. 作者是把气候统计学和统计气候学严格分开的，所以本书的副标题为统计气候学理论.

我们现在的气候记录仅仅是无穷气候记录总体中的一个样本. 这个样本不是按人为规定或控制出现的，而是有某种概率(机会)随机出现的随机样本. 我们过去的记录是随机样本，未来的记录也是未来的随机样本. 我们的记录样本既然是随机样本，那末未来记录就可以是这样的记录样本，也可以是那样的记录样本. 我们由现有记录直观求得的气候规律，往往只是真实气候特征一时的实现，从这样片断的记录不会直观认识到既全面而又真实的气候特征，只有等到无穷年代后，才能从记录总体中找出完全正确的

气候变化规律，但是无穷年代是永远不会达到的. 数理统计学中的分布函数、抽样理论、统计推断与随机过程，以及气候统计学本身的特殊发展与创新，就替我们解决了这个难以解决的问题，使我们能够根据现有记录样本所给的信息最优模拟记录总体，估计与推断记录总体的特征. 本书内容就是有关这些问题的基础部分. 作者编写本书的目的之一，就是希望读者能初步理解数理统计学如何必须引进到气候学科中，又如何还要在引进的基础上去发展与创新，使气候统计学发展成为一门边缘科学. 所以本书内容只阐述气候统计学的基础. 至于作者多年来根据气候统计学基础进一步在气候变化、气候循环与气候预报方面所进行的研究，将在下一部书《气候统计学(统计气候学理论 II)》中探讨，该书拟包括下列内容:(1)气候统计基本问题，(2)气候变化性与极值统计，(3)循环分析与旱涝游程，(4)周期分析，(5)统计模式的建立与预报，(6)回归模式，(7)趋势模式，(8)逐步回归与简化逐步回归，(9)正交回归与相关筛选回归，(10)修匀模式，(11)随机模式，(12)概率模式与概率预报，(13)客观统计分析与预报，(14)二分变量统计分析与预报.

气候统计学不但是气候学领域内正在发展的理论学科，其本身是研究统计气候学基本理论的手段，而且也是一切与气候有关的其他科学研究、资源区划、工程设计、经济规划以及生产管理等所应参考的统计理论与方法. 因此我们可以说，气候统计学在实现祖国四化宏图和推动国民经济发展上确实起着重要作用. 正因为气候统计学的研究有这样重要的作用，所以在 1979 年的第一次国际统计气候学会议之后， 1983 年在里斯本和 1986 年将在维也纳分别召开第二次、第三次国际统计气候学会议. 我们可以说，这门学科在国际上近年来发展之迅速，已达到日新月异的程度. 使作者感到遗憾的是，正当这门科学日益发展，水平日益提高之际，作者已进入高龄，能力本来不强，精力现又衰退. 但是，作者认为，我们不能见高山而仰止，应以老马识途的精神，为广大青年指出艰苦攀登高峰的捷径;更应以老当益壮的气概，与广大青年共同携手

奋斗，在攀登科学高峰的道路上克服一切艰难险阻．这就是作者编写本书的另一目的．

最后还必须指出，由于作者高年和其他原因，本书错误难免，殷切希望广大读者随时惠予指正．

么 枕 生

1982 年 6 月 1 日于南京大学气象系

目　　录

第一章 概率与概率基本运算

§1 概率论的基础

概率论的基本出发点就是事件．气象学中的事件就是指在一定的条件下所发生的各种大气现象．不同的事件可用各种方式结合在一起而成为另一种事件．例如，连阴雨的大气现象就是由许多雨日结合成的．这种过程可以很好地用集合论去叙述，所以我们先介绍简单的集合论，做为气候统计理论的基础，这也是模糊数学在气候统计学中应用的基础．

我们认为大气现象是随机的，气象观测是一种随机试验，气象观测序列就是随机试验的不同结果．每一个气象观测结果称为样本点，所有可能样本点的总体称为样本空间． 设气候变化的随机试验只是观测湿年与干年，这样就具有两种结果或样本点，这种样本空间就仅有 2 个样本点．如设 M 代表湿年，D 代表干年，则连续 3 年气候变化可有下列型式：DDD，DDM，DMD，MDD，DMM，MDM，MMD，MMM．这里共有 $2^3=8$ 个可能结果，每一可能结果就是一个样本点．因此，这种随机试验的样本空间又共有 8 个样本点．如把天气现象划分为晴、昙、阴、雨四种，在一个月内就有 4^{30} 个可能结果或样本点，因而这种随机试验的样本空间在一个月内包括 4^{30} 个样本点．设在气候观测序列中划分为 k 种不同的记录 x_1，x_2，$\cdots$，x_k，这 k 种记录的数目分别为 n_1，n_2，$\cdots$，n_k，而 $n_1+n_2+\cdots+n_k=n$． 这就是说，我们的记录序列为 $\{x_{ij}\}$，$i=1,2,\cdots,k$，$j=1,2,\cdots,n$，这里第二下角标 j 必须结合第一个下角标 i．序列 $\{x_{ij}\}$ 就代表样本空间，这 k 种记录就是此样本空间所包括的 k 个样本点．

某种样本点的集合体称为集，样本点就是集的元素． 样本空

间就是元素的可能最大集，其中各个元素称为样本点或简称为点．样本空间可以包括有限的样本点或无限的样本点． 例如，上述各种样本空间都是包括有限的样本点．如果集 E_1 的每个元素(样本点)也是集 E_2 的一个元素，则 E_1 称为 E_2 的子集，写如 $E_1 \subset E_2$. 因此，每一个集就是自己的子集．无元素的集是为空集 Φ，空集就是各集的子集．全集就是样本空间．如果一个集包括样本空间除 E 以外的所有元素，则此集称为集 E 的余集，用 $\bar{E}$ 或 $R-E$ 代表，这里 R 代表样本空间．

一个事件就用样本空间的子集代表．所谓样本空间内的一个事件，就是用样本空间内各样本点组成的子集规定．当然一个事件也可以仅仅包括一个点．设 0 代表干年，1 代表湿年，3 年内出现 3 个干年的事件就用子集 $\{(0, 0, 0)\}$ 代表，此子集只包括一个样本点． 3 年内一个湿年的事件就用子集$\{(1,0,0),(0,1,0),(0,0,1)\}$代表，此子集包括 3 个样本点． 3 年内 2 个湿年的事件就用子集 $\{(0,1,1),(1,0,1),(1,1,0)\}$ 代表，此子集包括 3 个样本点． 3 年内出现 3 个湿年的事件就用子集 $\{(1, 1, 1)\}$ 代表，此子集包括 1 个样本点． 样本空间就包括 3 年内 0，1，2，3 年成为湿年的事件，其中样本点共有 8 个． 只包括一个样本点的事件称为初等事件，也常称为简单事件，3 年内出现 3 个干年的事件$\{(0, 0, 0)\}$或 3 年内出现 3 个湿年的事件 $\{(1, 1, 1,)\}$就是这样的事件．如果一个事件可以分解成为许多简单事件，则称为复合事件，3 年内出现两个以上湿年的事件 $\{(0,1,1), (1,0,1), (1, 1, 0),(1, 1, 1)\}$ 就是这样的事件．因此，所谓事件就是观测的结果和其集合体．

我们所讨论的事件是和概率相联系的子集．上述的样本空间就可以说成为概率空间，样本点就是概率空间的元素． 事件 E 出现的概率就是指 E 的任何样本点出现的概率． 例如，我们问 3 年内有 1 年成为湿年的概率如何？ 这个概率就是在三个样本点 $(1, 0, 0)$，$(0, 1, 0)$ 与 $(0, 0, 1)$ 中任何样本点出现的概率． 为了进一步了解概率运算法则，我们先讲述集的结合与分解法则． 在下

面的讨论中，各种集如应用闭合平面图即所谓文图(Venn diagram)表示，较为方便．在这个图中用一个长方形代表一个样本空间，其元素位于长方形之内．长方形内某一区域就代表子集．各种有关集的关系都可以用适当的图形简单证明．

设每一个代表事件的子集 E_1，E_2，… 包括样本空间 R 中的一些样本点．E_1 与 E_2 的并集就是一个集，其中每个元素或属于 E_1 或属于 E_2，或属于二者(即至少属于其一)．E_1 与 E_2 的并集(用 $E_1 \cup E_2$ 或 E_1+E_2 表示)代表一个事件，这个事件用图 1 中画有斜线部分代表．我们可以看出，若 $E_2 \subset E_1$，则 $E_1 \cup E_2 = E_1$.

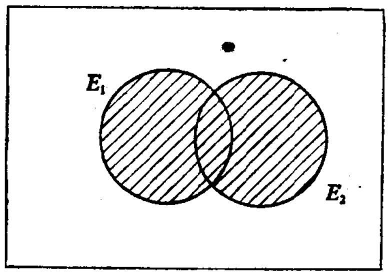

图 1

E_1 与 E_2 的交集就是一个集，其元素既属于 E_1 也属于 E_2. E_1 与 E_2 的交集(用 $E_1 \cap E_2$ 或 E_1E_2 表示)也代表一个事件，这个事件用图 2 中一闭合平面图形所范围的部分(斜线部分)代表．我们可以看出，若 $E_2 \subset E_1$，则 $E_1 \cap E_2 = E_2$.

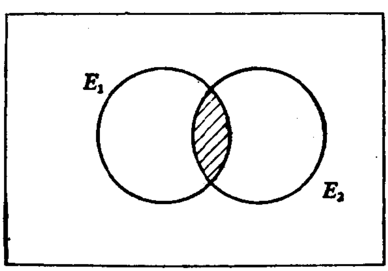

图 2

所谓 E_1 与 E_2 不相交，即二者无一共同元素（图 3)，或二者