游程概率
统计原理及其应用

马秀峰　夏　军　著

科学出版社

内 容 简 介

本书是著者对游程长度、数量的概率统计问题历经17年的研究成果。着重向读者介绍如何应用“游程分析”的数学工具，揭示我国历史文献灾情记录中蕴含的重现规律。书中详细地介绍了著者独特的研究思路和方法：首先根据游程长度与数量的基本定义，依据概率论的基本原理，生成可能发生互不重复的全部样本，用“概念、数字、解析技术”寻求样本游程长度、数量的概率密度、分布及数字特征等一整套解析公式；然后用随机模拟的方式，生成大量独立或非独立样本，对解析公式的可靠性、适用性进行检验。对于长期困扰统计学界的非独立样本游程概率问题，也创造性地给出了简练、有效的解决方法。书中还介绍了估算黄河流域连旱重现期的典型算例，以及应对环境变化、多年连旱的应用实例。本书以一般理工科专业师生能够顺利阅读为宗旨进行撰写，不但在学术上，而且在方法论上都希望给读者有益的启迪。

本书可供希望利用历史文献，构建符号序列，研究灾情演化规律的专业人士参考；也可供防灾、减灾、水文、气象、水资源、海洋、保险、农业等部门有关科技人员及大专院校的师生参考。

图书在版编目(CIP)数据

游程概率统计原理及其应用/马秀峰，夏军著. —北京：科学出版社，2011

ISBN 978-7-03-031602-8

Ⅰ. 游… Ⅱ. ①马…②夏… Ⅲ. ①概率论-研究 Ⅳ. ①O211

中国版本图书馆CIP数据核字(2011)第115059号

责任编辑：杨帅英 杨 然/责任校对：邹慧卿

责任印制：钱玉芬/封面设计：耕者设计工作室

科学出版社 出版

北京东黄城根北街16号

邮政编码：100717

http://www.sciencep.com

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2011年7月第 一 版 开本：787×1092 1/16

2011年7月第一次印刷 印张：23 1/2

印数：1—1 500 字数：534 000

定价：89.00元

（如有印装质量问题，我社负责调换）

前　言

我国是一个人口、资源、环境矛盾十分尖锐、自然灾害严重而频发的国家，其中尤以大范围多年连旱的危害最为触目惊心。《国语·周语》有“伊洛竭而夏亡，河竭而商亡，三川竭而周亡”的记述，明崇祯五～十五年(公元 1632～1642 年)，连续发生 11 年的大旱，旱情最重的 1639～1641 年，无雨期长达 17～19 个月，“黄沁微微，伊洛断流，飞蝗遍地，野绝青草”，耗尽了政府与民间所有粮食储备，“大饥、人相食，饿殍盈野，死者相续，十室九空，流民塞道”的记述，不绝于书；又因政治腐败，税赋酷重，农民起义，天下大乱，遂导致明王朝快速灭亡。

据各地史志记载，从 1922～1925 年，黄河上中游地区每年都有局部性旱灾发生。1926～1927 年连片干旱扩大到整个上中游地区，1928～1931 年灾情最重，泾渭河“断流，车马可由河道通行”，甘肃“大饥，积尸载道”，宁夏“死者枕藉，十室九空”，诸如此类的史志记载不胜枚举。我国历代善为政者，无不把持续性的自然灾害，特别是多年持续的严重旱灾，上升到关乎国家兴衰治乱的高度，慎重对待。

改革开放以来，我国经济快速稳定增长，人民生活大有改善，基础设施快速发展，抗灾的能力显著增强。然而，如果像明代和民国年间那样我国东半部“精华”地区大范围的多年连旱事件一旦重演，再加上我国今日的人口负担，将会对国家的水安全、粮食安全和社会经济造成重大的影响，极端灾害的风险仍然在我们的身旁徘徊。

历史上各种持续性自然灾害轮番出现的客观事实向人们展现出两个发人深省的问题：

第一，类似这种长期持续性的旱情，历史上曾经是否更多次发生？重现期有多长？

第二，一旦再度发生类似这种长期持续性的旱情，该如何应对？

这两个问题，仅仅依靠现有不多的实测水文气象记录很难回答。于是我们开始思考利用史志中的灾情记录，回答持续干旱的重现规律问题。

在这个强烈愿望的驱使下，我们查阅、统计了历代治河典籍和各地灾情史志资料，逐步发现这些丰富的史料具有以下三个突出特点：

第一，我国古代以农立国，对旱涝灾情特别关注，自有文字可考的历史开始，就一直保留着如实记录重大灾情事件的优良传统，因而我国史志资料的可信度高。今天我们能够检索出的旱涝史料，尽管都是古人根据当时的直观感受，定性记述的灾情事件，难以用来定量计算降雨的丰枯量值，但旱涝灾情的定性分类则真实可靠，而自然界的客观规律，恰好就寓于真实的历史记录之中。

第二，我国史志资料分布广、数量大、时序长。位于甘肃省以东和长城以南的省、地、县三级政府，几乎都存有史志典籍，尽管由于各种原因曾经造成过部分记录的中断，但根据“旱灾呈片，涝灾呈线”的一般规律，有可能借鉴邻近地、县的同步史料，结合图形显示、逻辑旁证与科学判断，对缺失的部分进行插补，梳理成时间和空间上的系统资料。

第三，提供统计分析的成果形式，不是基于连续观测的水文气象数据，而是用“旱、正常、涝”等文字符号，定性地表述的离散的灾情事件，即用文字符号表示的时序系列。

上述第一个特点显示出我们国家独有的悠久而丰厚的历史文化底蕴，使我们认识和坚定了利用史料揭示我国旱涝规律的信心。第二个特点使我们预感到这一研究成果必将具有广泛的应用前景。第三个特点使我们认识到这个研究课题与规划设计行业常用的频率计算存在着重大差别。因而必须从现代科学体系中选择或研制一套适用于这种离散符号序列的理论与方法，借以揭示历史灾情演化的基本规律。这三个特点汇集在一起，再加上人们的努力，应当能够产生当代科学技术与我国丰厚文化底蕴密切结合的研究成果。

我们在遴选技术方法的过程中发现，数理统计学科中的“游程分析”恰好具有这样的功能，它突出的特点是，只讨论事件的有无，不过问有无的程度，略去过程的细节，将复杂的随机过程简化为只有两个文字符号交替产生的过程，借以从时间域上描述随机变量的统计规律。和常规的数理统计方法相比，该方法确实显得非常粗糙，可以称为“粗糙分析法”，然而恰恰是这种粗糙的分析方法，往往能揭示出隐藏在事物内部，或“潜伏”在表面上截然不同的事物之间更为本质、更为重要的公共属性。也恰恰是这种粗糙的分析方法，可以从文字符号表示的离散的时序系列中，揭示出历史灾情演化的基本规律。

我们从大量文献中了解到，科学家们对游程问题的研究已经有了相当长的历史。

然而时至今日游程一词尚未得到学者们的一致认同，在不同的数理统计专著中，学者们往往使用其他术语表述这个含义相同的内容。例如，把游程称为“轮次”、“连”、“连贯”、“交互”、“交叉”等，从而在不同学科中出现了内涵接近但词汇不同的术语。

尽管在不同的著作中采用的术语不同，但在进行统计检验时，几乎都使用 A. M. Mood 导出的样本游程个数的期望值与方差公式：或进行趋势性检验；或用来检验两个样本是否来自同一总体，检验样本内部各个元素之间是否相互独立。这说明不同学者尽管采用的术语不同，但所用术语的内涵是相同的。

本书把时间序列样本中连续出现相同元素的历程称作“游程”，英文术语为 run。有时为了文字叙述、列表或图中注字的简练方便，往往使用“连”或“轮”作为“游程”的同义语，使用“连长”或“轮长”作为“游程长度”的同义语，使用“连数”或“轮数”作为“游程个数”的同义语。

与此同时，我国学者将有关游程理论的论述大多作为数理统计中的一个条目，从国外文献转译，而且内容大致雷同。对连续随机事件的重现期，还没有公认的定义；对非独立随机过程的游程特征，尚不能用解析的方法描述；更缺乏全面系统论述游程理论及其应用的专著。总之，这还不足以用来揭示我国史志文献中持续性灾害记录的统计规律。

基于以上思想认识，著者 1984 年起动手探索随机现象中的游程问题。先分别用黄河流域的年降水量和年径流量系列，以年为时间单位，研究连续干旱长度的重现期问题；然后再全面收集我国各地史志中的干旱记录，争取构建自汉代以来我国农耕区干旱游程的符号系列，借以揭示连旱特征在时间和空间上的变化规律。

在探索游程重现期的过程中，我们最初曾模仿工程水文频率计算的概念，把游程长度的重现期定义为游程长度概率分布的倒数。后来发现把大于或等于某长度的游程作为一

个随机事件，是连续多年的统计量，因而游程长度的概率分布与工程水文频率计算中的概率分布存在重大差别，不可以机械地模仿工程水文频率计算中关于重现期的定义。查阅国内外大量参考文献后，没有发现关于游程长度的重现期的论述。因此，著者在本书第3章试探性地给出了游程长度重现期的定义。

著者以回答实践问题为初衷，探索游程问题。为了尽快找到答案，必须求助于赌盘模型，但从人们的直观感受观察赌盘模型与时间序列中的游程现象似无共同之处，因而在整个探索过程中，著者不能不担心根据赌盘模型推出的结论不切合实际。为澄清这种疑虑，我们在撰写与探索过程中，遵循着两个技术途径并行探讨：

第一个途径是按照赌盘模型的基本性质和游程的判别准则生成完备的样本空间，统计样本游程数目或长度的概率密度与数字特征，这些统计量的突出特点是没有误差的。因此，可以用“概念数字的解析技术”寻求其统计法则。这些统计法则，集中表现为著者给出的样本连长与样本连数的概率密度、概率分布及其数字特征的一整套解析公式。

第二个途径是用多样性的随机过程发生器，随机生成大量的样本，按照游程的判别准则，分析样本游程的数量、长度与频次的统计特征，这些统计量的突出特点是，计算误差随着抽样个数的增大而减小，并且当抽样个数逐步增大时，能够依概率逼近赌盘模型的相应统计量。进行这种统计试验的目的是，模拟各种各样的随机过程的游程变化，观察它们与赌盘模型游程法则的联系，对赌盘模型游程法则的适用性作出评价。

按照上述技术途径，通过大量和多样性的统计试验，我们获得了始料未及的惊奇发现：

对于一切平稳独立的随机过程，不论它们原来服从何种概率分布，也不论其变化机制如何复杂，只要“来到”登录游程的窗口，就都会自动忘掉自己的“出身”，服从赌盘模型确立的游程法则。

对于平稳非独立的随机过程，数理统计学至今尚未给出简练的解析方法，我们的惊奇发现是，不论平稳非独立的随机过程原来服从何种概率分布，也不论其变化机制如何复杂，只要“来到”登录游程的窗口，也会自动忘掉自己的“出身”，服从“迁移参数法”确立的游程法则。

这一发现告诉我们，可以放心大胆地使用赌盘模型导出的游程法则，处理回答生产实践中各种各样的游程问题。本书第10章集中介绍了根据黄河实测径流和历史文献，估计持续性干旱的概率与多年连旱的重现期，以及样本独立性检验的具体算例，可供同行们参考。

我们阅读过不少关于游程分析的文章或书籍，看到许多学者在进行逻辑推理之前，几乎都是先假定随机变量的分布（通常是假定服从正态分布），然后再进行逻辑推理。根据我们的研究心得，认为在探索游程的统计问题时，不论是否独立，只要过程具有平稳性，都没有必要先假定研究对象服从何种概率分布。这是因为，游程分析是粗糙的研究方法，其精髓在于只关心事件的有无，不关心有无的程度。

此外，在应用理论方法探索实际的游程问题时，不可能得到大量的属于同一总体的随机样本，更不可能得到完备的样本空间。通常只能得到容量较小的单个或若干个样本，这相当于从总体中随机地抽取一个或若干个子样，因而这些子样又不可避免地存在着代表性不足的问题。人们可能做到的是尽量收集和增补资料，扩大样本容量，以及通过考察、

旁证和密切联系实际的分析，来提高样本元素和分析结果的可靠性。接下来人们仍需要面对有限的少量样本，“因陋就简”地进行统计分析。

由于面对实际问题有较多现实困难，因此在提供结论性成果时，应遵循“多种途径，综合分析，联系实际，合理选用”的基本原则。

毛泽东同志的两个哲学观念，给本书著者以巨大的启发：①矛盾的普遍性即寓于矛盾的特殊性之中；②就人类认识运动的秩序说来，总是由认识个别的和特殊的事物，逐步地扩大到认识一般的事物。人们总是首先认识了许多不同事物的特殊的本质，然后才有可能更进一步地进行概括工作，认识诸种事物的共同的本质……这是两个认识的过程：一个是由特殊到一般，一个是由一般到特殊。

这本专著，可以说是著者自觉运用这两个哲学观念、探讨科学问题的一个具体样例。著者在本书中推导的几乎所有通用的计算公式，都是在这一哲学理念的指导下，先探索最简单最特殊条件下的表达式，再探索其他条件的特殊表达式；然后将若干个不同条件下的特殊表达式放在一起，发掘它们的共性，用文字符号取代表达式中的具体数字，试探性地写出比较一般的解析表达式；最后把比较一般的解析表达式放在更多的各种各样特殊条件下，进行检验，发现问题，修正、完善、提高已有的认识。

由于著者水平有限，不当之处，在所难免，恳请指正。

本书的出版得到了国家重点基础研究 973 项目(2010CB428406)“气候变化对我国东部季风区陆地水循环与水资源安全的影响及适应对策”的资助。华北水利电力学院的彭高辉同志为本专著做了大量文献统计工作，中国科学院地理科学与资源研究所博士生佘墩先、严子奇、邱冰为本书图文编辑付出了心力，还有不少同志对书稿的梳理校对给予了帮助，著者在此特地向他们一并表示衷心的谢意。

最后，我们真诚地感谢与本书第一著者马秀峰相濡以沫的伴侣龚庆胜女士，她是马秀峰先生的学妹，曾在同一所大学攻读同一个专业。她不但悉心照料和关心老伴的健康，而且是本书各章节初稿的第一读者，更是最及时指正书稿中某些缺陷、提出改进意见的第一人，可以说，这部专著也凝结着她的心血与智慧。在这里我们向她表示真挚的问候与谢意！

马秀峰　夏　军

2011 年 1 月 8 日

目　录

第1章 绪 论

持续发生相同属性的随机事件称为游程现象，相同随机事件持续的历程称为游程，游程似周期而非周期。游程理论是描述持续性随机事件统计规律的数学工具，是数理统计学科的重要分支，可用来揭示时间序列中游程现象的发生概率，回答持续性灾害事件的重现规律，并对样本的独立性进行统计检验。

1.1 屡见不鲜的游程现象

在人们的日常生活、生产与社会实践中，常常遇到各种各样的游程现象：

几个人一起打扑克，其中某人曾一连多局抽到大小王牌。

赌徒掷骰子往往一连数次掷得完全相同的点数。

某个地区或某一流域，往往会一连多年发生干旱、洪涝或某种持续性的异常天气现象；某地区或某城市的年度气象特征值（气温、气压、降水量等）往往会一连多年高于或低于正常值。

河道的水位往往在一段时间内持续低于正常通航水位。

某车间在正常条件下，由于许多复杂因素的综合影响，有时会一连数次出现次品。

高速飞行的飞机，在气流涡动的作用下，机翼或某一部件的振动往往在一段时间内持续超过某一风险应力。

泥沙颗粒受自重和水流上举力的作用，在河床上做起伏跳跃的运动，往往在某一段时间内，持续跳跃在某一门槛高度之上。

湍流内部某点的瞬时流速，在一段时间内持续低于（或高于）该点的时均流速。

观测仪表的探头在探测某一物理量时，仪表读数往往在一段时间内持续出现正、负号相同的误差。

服务台前往往在一段时间内持续出现排队现象。

……

从以上列举的许多关于游程的例子可以看出，不论是独立的随机事件（如赌博、掷骰子等），还是非独立事件（如湍流内部某点的瞬时流速、观测仪表的误差符号等），都存在游程现象，可以说游程是一切随机事件的普遍现象。

上面关于游程的例子，从时序变化上可以归并为连续和离散两大类：

一类是连续随机变量在时间变化过程中的游程问题，如水文气象记录连续偏离正常值、仪表读数的正负偏差、湍流的瞬时速度等问题；

另一类是离散随机事件流的游程问题，如一连多局抽到大小王牌，一连数次掷得完全相同的点数，一连多年发生干旱、洪涝或某种异常天气现象。

诸如此类现象，其持续时间的长短、发生次数的多少和发生概率的高低，成了人们关心的主题。研究和掌握这类现象变化的统计规律，对指导生产建设，进行科学试验，具有重要的实用价值。

例如，机械制造的设计师，希望了解超限振动的持续历时与发生概率之间的函数关系，以便采取措施，将破坏性风险限制在允许的范围以内。制造观测仪表的设计师，希望了解仪表读数同符号误差的持续历时与发生概率之间的函数关系，以便采取措施，将这种误差的发生概率限制在允许的范围以内。

又如，1922～1932 年黄河的天然河川径流量持续低于黄河天然年径流的多年平均值，专家们根据历史文献考证，在明代崇祯年间(1632～1642 年)，黄河流域也曾经出现过连续 11 年的大旱和特大旱灾。进而用当代实测黄河天然年径流序列进行统计，发现黄河天然年径流还有连续 2 年、3 年、4 年……多种连续偏枯的现象，并且还发现连续偏枯的概率随连续偏枯时段的增长而衰减。治理黄河的工作者很希望知道黄河天然年径流连续偏枯的概率或重现期与连续偏枯时段长度之间的依变关系，用来指导制定防旱减灾与水资源利用的对策措施。

1.2 游程理论的发展简史

科学家们对游程问题的研究已经有了相当长的历史，但是不同领域的专家学者往往用不同的术语描述此类现象。

1897 年皮尔逊(Karl. Pearson)最早使用过“轮次”的术语，他在文献[1]中讨论获自蒙特卡罗轮盘的试验资料时，指出游程分布是多项式分布的一种特殊形式。1899 年 Karl. Marbe 根据二项分布理论推导出了在给定样本容量条件下，某一随机事件连续发生的轮次平均值的计算方法。A. M. Mood 于 1940 年在《统计数学年鉴》上发表了用轮次理论检验两个样本是否来自同一总体的论文，他指出：轮次分布的理论研究，曾经有过轰轰烈烈的历史，该理论大约开始于 19 世纪末。

在 20 世纪 30 年代，研究概率论的数学家们在探索 n 次二项式展开时，把连续出现相同的字母称为轮次，把一个轮次中含有连续相同字母的个数称为轮长，并给出了样本轮次的期望值与方差的计算公式。有些数理统计专著借鉴布朗粒子随机游动的术语，把连续发生同类的试验结果，或者把时间序列中连续出现相同类型的元素称作“游程”[5]。然而游程一词尚未得到学者们的一致认同，在不同的数理统计专著中，学者们往往使用其他术语表述这个含义相同的内容。例如，把游程称为“轮次”、“连”、“连贯”、“交互”、“交叉”等。

1944～1945 年赖斯(Rice)最早把游程分析的概念应用于时间序列分析，提出了“水平交叉”的概念，并计算了独立正态随机变量在单位时间内低于或超过某一门限值的平均次数以及每交叉一次相隔的时间间隔[2]。此后，游程分析技术日益广泛地应用于生产实践与各种科学研究，从而在不同学科中出现了内涵接近但词汇不同的术语。例如，美籍水文统计学家南斯拉夫人 V. 叶非耶维奇(Yevjevich)，1972 年最早用“轮次分析”的术语描述年径流序列的丰枯变化，并把轮次分析理论纳入他的名著《水文随机过程》一书(成都工

学院1978年翻译但未正式出版)[3]。英国的水文学家科特戈达(Kottegoda)1980年在《随机水资源技术》一书中,把研究水文时间序列起伏变化过程的理论称为"交互理论"[4]。

尽管在不同的著作中采用的术语不同,但在进行统计检验时,几乎都使用Mood导出的样本游程个数的期望值与方差公式:或进行趋势性检验;或用来检验两个样本是否来自同一总体[5],检验样本内部各个元素之间是否相互独立[6]。这说明不同学者尽管采用的术语不同,但所用术语的内涵是相同的。

本书把时间序列样本中连续出现相同元素的历程称作"游程",有时为了文字叙述、列表或图文的简练方便,往往使用"连"或"轮"作为"游程"的同义语,使用"连长"或"轮长"作为"游程长度"的同义语,使用"连数"或"轮数"作为"游程个数"的同义语。

1.3 游程分析的主要特点

1. 时域、频域相结合的统计属性

众所周知,对于某一指定的概率分布函数,一个容量为 n,均值、方差、偏态系数分别为 $\overline{X}$、D、C_s 的随机时序系列样本(也称为一个现实)将唯一地决定该样本的一个相应的概率分布。如果维持样本容量及其中的每一个元素的数值不变,仅任意改变样本元素的时序位置,则可以得到 $n!$ 个排序不同的样本,按照现行方法作频率计算,它们将仍然服从完全相同的概率分布,仍然给出完全相同的概率预测[7]。

可见,工程规划设计中的频率计算,仅仅是从频域一个侧面描述随机时间序列的方法。这种方法无法辨知容量、均值、方差、偏态系数完全相同,而元素排序不同的样本之间的任何差别。

然而从人类社会实践和生态环境对河川径流的密切依存关系看,容量、均值、方差、偏态系数完全相同,而元素排序不同的样本,将有不同的社会与生态效果,在某些极端的情况下,如持续递增或持续递减的时间序列,可能对应着灾难性的社会与生态后果。大量的实践经验证明,随机时间序列在时域上和频域上都具有丰富的统计属性。因此,仅从频域一个侧面描述随机时间序列是不够的,需要从频域和时域两个方面对随机时间序列作出全面的描述。

实践中,水文工作者常用选典型年或典型洪水过程线的办法弥补水文频率计算的局限性。然而,这种选典型的方法不但具有很大的任意性,而且很难体现随机时间序列在时域上丰富的统计属性。

总之,游程是一切随机变量共有的重要属性。而游程分析则是将时域和频域密切结合,揭示随机事物变化属性的重要工具。

2. 检验样本独立性的有效工具

大量统计试验表明,自相关系数很小的随机过程,其样本游程个数的概率密度函数与数字特征,与相同条件的独立过程相比,将产生明显的变化。这个特点恰好可以用来作为检验样本独立性的工具[4]。

3. 粗糙的分析技术

游程分析的突出特点是：只关心事件的有无，不计较有无的程度，忽略变化的细节，专门考察持续出现在门槛以上或门槛以下的随机事件的统计性质，因而其被称为粗糙的统计分析技术。

表 1-3-1 中的观测值 $x(t)$是由 20 个数据组成的样本，其均值为 10。图 1-3-1 是观测值 $x(t)$随时间变化的折线图。图 1-3-2 是观测值 $x(t)$随时间变化的柱状图。通常人们把观测值 $x(t)$随时间变化的过程称为随机序列或随机过程。

表 1-3-1 观测值 $x(t)$与随机符号序列 $y(t)$随时间变化过程表

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$x(t)$	11	15	12	6	11	7	6	6	11	15	17	15	13	8	4	14	4	3	8	14
$y(t)$	0	0	0	1	0	1	1	1	0	0	0	0	0	1	1	0	1	1	1	0

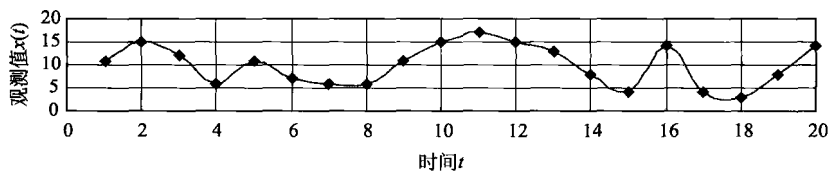

图 1-3-1 观测值 $x(t)$随时间 t 变化过程

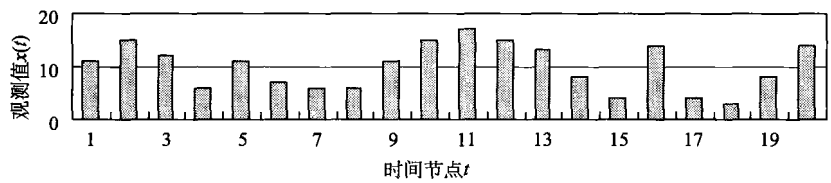

图 1-3-2 观测值 $x(t)$随时间 t 变化过程

表 1-3-1 中 $y(t)$代表用均值为 10 作为门槛，将观测值 $x(t)$进行区分的符号变化过程：规定凡是小于或等于均值的随机变量记为 1，凡是大于均值的随机变量记为 0(作出相反的规定并不影响最后的结论)。

这样就把复杂的随机过程简化为仅由 0 和 1 两个数字符号组成的“随机符号序列 y”，然后只对随机符号序列 y 的交互变化进行研究。

在人们的日常生活、生产、科学试验与各种社会实践中，类似于图 1-3-1 和图 1-3-2 的观测过程不胜枚举。因此，可以进行游程分析的对象极其广泛。可以看出，游程分析的突出特点是：略去随机过程的细节，将随机变量简化为两个文字符号，借以从时间域上描述随机变量的统计规律。和常规的数理统计方法相比，确实显得非常粗糙。然而这种粗糙的研究方法，往往能揭示出隐藏在事物内部，或潜伏在表面上截然不同的事物之间的更为

本质的重要属性。

游程分析的这种“只问有无,不拘细节”的粗粒性质,正好用来揭示我国历史资料中蕴含的旱、涝或各种自然灾害的统计规律。

我国古代以农立国,旱涝灾情是影响国家兴衰治乱的重大问题,历朝各代都对旱涝灾情给予极度的关注。自有文字可考的历史开始,就有如实记录重大灾情事件的优良传统,为今天人们分析认识旱涝规律留下了极其珍贵的史志资料,这些史志资料具有以下三个突出特点。

一是定性可靠。

我国史志资料可信度高。“秉笔直书”自古就是我国史官们行为自律的基本准则,历史上涌现过许多不畏权势、据情实录的人物和事例。而一切客观的规律都潜伏于真实的记录之中,因而,我国史志资料的真实性质构成了认识我国历史旱涝规律的最重要的资料基础。

二是分布广、数量多、年代久。

我国史志资料不但分布广、数量多,而且时间跨度长。位于甘肃省以东和长城以南的省、地、县三级政府,几乎都存有悠久的史志典籍,尽管由于各种原因曾经引起过部分记录的中断,但根据“旱灾呈片,涝灾呈线”的一般规律,有可能借鉴邻近地、县的同步史料,结合逻辑旁证,科学判断,对缺失的部分进行插补,梳理成空间上连片和时间上持续的系统资料。

三是符号序列。

根据我国史料可以提供统计分析的成果形式,不是基于连续观测的水文气象数据,而是用“旱、正常、涝”等文字符号定性表述的离散的灾情事件,即用文字符号表示的离散的时序系列。这个特点正好符合游程分析“只问有无,不拘细节”的粗粒性质。

基于我国史志资料的上述三大特点,游程分析技术可以作为分析我国史志资料,揭示历史旱涝规律的工具。而且这一研究成果,必将具有广泛的应用前景。

当代美国著名物理学家福德(Ford)专门在物理学杂志上发表文章,把这种仅用少量文字符号取代复杂过程的分析方法称为“符号动力学方法”,指出符号动力学方法对物理学基础将产生深远的影响。著者发现,使用符号动力学的基本概念可以导出游程分析中的全部基础公式。

1.4 本书的技术途径

按照确立的概念进行统计试验,构成完备的样本空间,再据之以寻求游程变化的函数关系,然后在不同条件下进行检验,发现问题,补充完善,循序渐进,逐步提高,是本书的基本思路。针对旱涝灾害发生的成因,包括气候变化以及人类活动影响的相关研究的问题,进一步开展实例研究与应用对策研究,借鉴到我国生产实际中[16~20]。

探讨游程统计规律,采用了两条并行的技术途径。

第一条技术途径是:制定出判别游程的基本定义,并以赌盘模型为基本工具,指定某个样本容量,根据概率统计的基本原理,详尽地做出隶属于某一指定状态的完备的样本空

间，统计样本游程的长度、数量及其发生频次，分析样本游程的各种统计特征的变化法则。这个技术途径的突出特点是，导出的各种计算公式和统计数表精确无误。

基于这个重要的特点，在执行第一条技术途径时，我们摸索出“概念数据的解析技术”。

这套技术可以用语言表述如下：

根据概念，先生成若干特殊简单条件下的数据，寻求这些特殊简单条件下描述数据变化规律的解析公式；在此基础上，使用可以涵盖个别属性的抽象的文字符号，取代个别表达式中相应的具体数字，进而将个别表达式改变为共同的形式，完成从具体到抽象，从特殊到一般，寻求普遍适用的结论和法则。

第二条技术途径是：构造出具有不同属性的“二分随机模型”，生成 m 个容量为 n 的样本，根据游程的基本定义和概率统计的基本原理，统计样本中隶属于某一指定状态的游程的长度、数量及其发生频次，分析游程各种统计特征的变化法则。按照这个技术途径导出的各种计算数据图表是有误差的。

如果按照第二条技术途径得到的各种结论，随着样本容量和样本数量的增大，能够逐渐逼近第一条技术途径的结论，则认为第一条技术途径的结论是第二条技术途径结论的吸引子，因而可以用来处理那些与第二条技术途径具有共同属性的统计问题。

由于在处理各种实际的统计问题时，不可能像统计试验那样，获得众多真实的样本，人们面对的是各种各样由实际观测资料构成的容量有限的单个样本，因此人们只能根据实际观测的资料，尽量展延单个样本的容量；尽量通过对资料的审查，去伪存真，提高样本的代表性。同时，还需要假定样本的统计量能够代表总体的统计特征。在这样的前提下，借鉴上述第二条技术途径的结论，回答具体样本的游程问题。

第二条技术途径的目标是，通过统计检验，回答在何种条件下可以利用第一条技术途径的各种结论，回答类似的游程问题。因此，第二条技术途径的关键是，密切联系实际，提出或选择有实用意义的随机模型，设计统计试验的计算程序，充分发挥电子计算机快速运算的优势，进行大量的统计试验，取得稳定可靠的结论。

不少学者在探索独立样本游程问题时，往往先假定随机变量的原始分布，而且多假定为正态分布。我们发现，不论是样本游程长度的概率密度与数字特征，还是样本游程数目的概率密度与数字特征，对于独立样本，都与随机变量的原始分布无关。可见，这种假定是多余的。

本书把游程分析大致划分为两大部分：

第 2 章～第 5 章，集中探讨样本游程长度的概率密度与数字特征。第 6 章和第 7 章，集中探讨样本游程数目的概率密度与数字特征。在使用游程理论回答实际问题时，往往要牵涉时序系列的独立性问题。例如，要回答年径流系列中丰水段或枯水段的重现期时，首先要分析相邻年份之间的径流是否存在着相关关系。因为不独立的时序系列与相互独立的时序系列相比，游程长度或游程数目的统计特征将发生明显差异。所以，在讨论了独立样本的问题之后，需要进一步讨论非独立样本的游程问题（参见第 7 章）。第 8 章使用初级的符号动力学方法，探索游程的统计法则；讨论了样本游程长度与个数期望之间可以互换的恒等关系，还讨论了我们意外发现的求和恒等式与积分恒等式。第 9 章集中讨论了平行试验的交集的游程问题。第 10 章集中讨论了游程分析理论在水文水资源中应用