自然灾害非参数统计方法

裴铁璠 梁文举 于系民 著

科学出版社

序

作者简介

裴铁璠，男，57岁，中国科学院沈阳应用生态研究所研究员，博士生导师。多年来，一直从事生态水文与生态气候研究，发表论文50余篇。

作者简介

梁文举，男，35岁，中国科学院沈阳应用生态研究所研究员，理学博士。多年来，从事农业生态学和土壤生态学研究，发表论文30余篇。

于系民，男，59岁，沈阳区域气象中心研究所研究员。多年来，一直从事农业气象研究，发表论文50余篇。

目　　录

附录：非参数统计常用表

引　论

自然灾害危害人民生命，破坏经济建设，阻碍社会进步，历来为社会各界所关注。各国政府部门和科学技术界，一直在努力探索自然灾害发生、发展的规律，以及防灾减灾技术与对策，尤其是近二三十年来随着广泛应用卫星遥感、雷达探测和计算机技术，使得自然灾害研究以及防灾减灾技术与对策取得长足的进展。尽管如此，肆虐的自然灾害依然不能被有效地遏制。分析原因大体有三个方面。一是自然灾害具有广泛性、复杂性和偶然性，因而人们认识自然灾害需要一个相当漫长的过程。目前，人们对许多自然灾害成因的认识仍然存在着盲目性，比如，地震、厄尔尼诺现象、全球气候变暖等等；人们对自然灾害的认识，仍然未走出必然王国的境地。二是人们认识自然灾害和实施防灾减灾措施，紧紧依赖于技术进步，而目前的防灾减灾技术水平，满足不了国民经济对防灾减灾的要求。三是研究方法上的原因。自然灾害多为随机事件，以往研究自然灾害多应用概率论与数理统计中的参数统计方法，较少应用数学分析、数理方程方法，更少应用非参数统计方法。然而，自然灾害中除有数量数据外，还有许多诸如分类或等级的非数量数据，比如能见度等级、地震震级与烈度、卫星遥感图像、洪水描述记录等。上述数据用参数统计法难于统计计算，而应用非参数方法则易于处理，而且可以得出较为确切的规律。此外，应用非参数统计方法不受总体分布的限制，它是一种定量与定性相结合的方法。我们撰写此书，旨在说明在研究自然灾害中，既要应用参数统计法、数学分析方法、数理方程方法，也要应用非参数统计方法；要依据自然灾害属性与数据状况，采

用不同方法乃至两种方法的有机结合，服务于防灾减灾。本书主要论述非参数统计方法在自然灾害研究中应用，供读者讨论与借鉴，当然更希望感兴趣者一齐来应用非参数法，使它成为防灾减灾的卓有成效的工具，为防灾减灾，加快经济建设，促进社会进步做出应有的贡献！

本引论主要包括：概率论与数理统计在自然灾害研究中的应用；非参数统计方法应用于自然灾害研究的意义和问题；非参数统计方法在防灾减灾中应用。

第一节　概率论与数理统计在自然灾害研究中的应用

人们生活在自然界中，为了充分合理地利用它的资源，一直在不断地对它探索与研究。自然界同其他事物一样，也具有两重性：它在给人类以资源的同时，也会给人类带来损失乃至灾难，后者也就是我们常说的自然灾害。人们熟知的自然灾害有水灾、旱灾、泥石流、台风、冰雹、地震、森林火灾以及植物病虫害等等。这些灾害有些频繁发生，比如植物病虫害，年年都会发生；有些是时有发生，诸如水灾、旱灾；有些为稀有事件，像烈度较高的地震，一旦发生危害极大，造成损失惨重。由于自然灾害的广泛性、普遍性和严重性，人们自古以来想方设法分析、诊断和预测自然灾害，并尽量采取措施防灾减灾。我国古代有许多科学家总结劳动人民同自然灾害进行斗争的措施和经验。这些经验和措施在自然科学的一些古典书籍中有所记载和论述，如《夏小正》《吕氏春秋》《四民月令》《氾胜之书》《齐民要术》《梦溪笔谈》《农书》《徐霞客游记》《农政全书》以及《沈氏农书》等。比如，《氾胜之书》中有“稗既堪水旱，种无不收之时”，“宜种之以备凶

年”;《齐民要术》引《汉书·食货志》语，称“种谷必杂五种，以备灾荒”，意即混种黍、稷、麻、麦、豆，以期达到总体减灾稳产之目的。古代农书中的某些论述，今天仍有价值，并可用现代一些科学知识加以解释。比如贾思勰所说的“天雨初晴，北风寒切，是夜必霜”，指霜恰是空气湿度大夜间辐射冷却所致，这是该条件下霜发生规律的概括语言。当然，由于古代科学技术水平的限制，对灾害的认识只能是粗浅的、定性的。比如明代地理学家徐霞客只能凭手的感觉，记录温度的状况。真正定量地认识自然灾害，是在17—18世纪物理学、数学、生物学和技术迅速发展时代。在西方，数学知识，比如欧几里德的《几何原本》被明代科学家徐光启引进中国的时候，许多人没有认识其作用。徐光启当时说，但愿百年后，中国人能认识到几何学在农业中的作用，这包括在防洪减灾筑堤中的应用。然而，令人遗憾的是，约百年之后还很少有人问津，直到大约250年后，清代李善兰才继续完成了徐光启的工作，并有所发展。

马克思曾说，一种科学只有在成功地运用数学时，才算达到了真正完善的地步。数学和其他自然科学都源于人们生产和社会活动实践，然而，数学的抽象与概括程度远远高于其他自然科学，因而使它成为最普遍的科学工具，又被人们应用到生产活动实践中去。因为自然灾害涉及生物科学和地球科学方面的众多学科及其交叉学科，所以数学在成为地球物理学、气象学、水文学、生态学、地震学、海洋学、遗传学、心理学等各学科不可少的工具的同时，理所当然地成为研究自然灾害的有力工具。现代数学方法应用于自然灾害研究，在我国始于本世纪初。著名地理气象学家竺可桢院士是我国现代地理学和气象学的奠基人，也是数量地理学的倡导者和最早的实践者。竺可桢早在哈佛大学攻读博士学位时，就开始用数学方法研究中国降水与旱涝规律，撰写论文发表于美国的 *Monthly Weather Review* 杂志上。著名地质学家李四

光在研究地质学同时，着手研究地震预报。到本世纪三十、四十年代，高等数学在我国受到重视，当时大学理工科都已开这方面课程，随之而来数学在地学、生物学中应用越来越多。将数学合理地应用于自然灾害研究，因而这方面研究成果水平有了提高，用于灾害预报、警报业务，效益渐增。涂长望、赵九章等气象学家，谢家泽、刘昌明等水文学家，顾功叙、傅承义等地球物理学家，罗宗洛、吴仲贤、马世骏等生物学家，崔启武、李博等数学生态学家等都曾比较成功地将高等数学方法应用于自然灾害研究中，取得了一批有学术价值和实际意义的成果，用于暴雨、洪水、台风、地震、病虫害等分析决策之中，为社会作出贡献，同时也培养了这方面的科学人才。

新中国建立后，特别是中国科学院有关所成立以来，从国外归来的许多科学家和技术专家，积极投入自然灾害规律及防灾减灾研究之中。这不仅包括直接研究灾害的地学、生物学、农学方面的专家，而且有一些数学家直接或间接地投入研究工作之中，使数学在自然灾害研究中的应用有了长足进展。40 年来，所运用的数学范围涉及数学分析、解析几何、微分方程、概率论与数理统计、组合数学、微分拓扑学、信息论和控制论、泛函分析等多种数学分支。其中利用微分方程数值解法的研究，我国当代著名气象学家叶笃正、曾庆存和黄荣辉等在研究大气环境、干旱、大气红外遥感等方面取得举世瞩目的成就，他们为提高自然灾害预报能力作了许多基础工作。但在一些以实用为目标的研究中，由于灾害涉及生物的复杂性、广泛性，以及许多农业自然灾害的偶然性，因而应用数学分析和微分方程建模无能为力，即使列出微分方程，也求不出解析解，即使求出数值解也由于是在许多脱离实际的方程简化前提条件之下，故得到的数值解也往往与实际状况差别较大。与微分方程相比，概率论与数理统计在自然灾害研究中应用较多，之所以如此，灾害一般是随机事件。概率论与数理

统计是研究随机事件规律的科学。下面作一简单说明。如果当条件组 S 实现时，事件 A 有时发生有时不发生，则称 A 是对于条件组 S 而言的随机事件。概率论是研究随机事件规律的。由于概率规律呈现在大量实验过程中，因之要发现这种规律就要进行大量试验。在各个充分长的试验（即实现条件组 S）序列中所得到事件 A 实现的频率 $v=\frac{u}{n}$ 大致相同并且接近于概率 p。俄罗斯数学家 Чебышев 认为，频率 $\frac{u}{n}$ 与概率 p 之间的偏差的级为 $\frac{1}{\sqrt{n}}$。$\left|\frac{u}{n}-p\right|<\varepsilon$ 当 $n\to\infty$ 时，ε 趋于 0。数理统计学是通过观察某些现象的频率而研究其现象规律的。在自然灾害研究中运用数理统计方法，也是以概率论原理为基础的，通过观察灾害发生频率，去探索自然灾害的统计规律。如我国气候灾害统计研究奠基人么枕生教授指出，相对频数就是概率的实验值，灾害概率的理论值就是某种灾害出现的真实相对频数或相对频数的实验值。数理统计这门科学关系到自然灾害有关要素本身及与其有关的自然现象、工农业生产、航空航海活动、人们对灾害的心理活动、思维和经济损失规律的认识。在自然灾害研究中的应用，主要有两大分支：描述统计（descriptive statistics）和推断统计（inferential statistics）。说到描述统计，不妨让我们分析一下“统计”这个名词的词汇原意：英文是 statistics，法文是 statistique，德文是 statistik，俄文是 статистика，其他一些欧洲语言词汇与之相似，它们都源于拉丁文中 status 一词，其词意为“情况”。因为在 18 世纪，当统计成为一门科学时，统计这个名词乃是和表示国家状况的种种事实记录联系在一起的。自然灾害的事实记录正是国家记录的一部分。我国历史悠久，灾害记录相当丰富，在欧洲也有一定的历史灾害记录，在日本有专项记录，如坪井八十二利用收集到的史料写出日本几百年的冷害、干旱大事表。在没有数理统计学的年代，虽有很多灾情记录，但只录于地方或全国性志书之中，并没有一套定

量的研究方法。描述统计学正是在概率论理论指引下，为适应状况分析之需应运而生的数理统计分支，它可以定量地总结灾害事件的发生、发展及后果等规律。比如在给定区域内，干旱或洪涝灾害发生的季节、危害程度与类型构成，其平均值、均方差、偏度、峰度等属于参数统计；中位数等属于非参数统计。统计推断即前面提到的推断统计学，主要是依据有限的观测事实来推断出大量事件本质规律的手段。统计推断的任务，就是研究者依据概率论的基本原理作出二者择一的决策。观察到的差异是出于偶然，还是样本抽自总体，进而说明差别是否为本质的。

概率论和近代数理统计学近百年来发展很快，新的概念、定理和方法不断涌现。从数学上被证明为可靠者，随时为其他科学领域的人们应用。与自然灾害有关的研究者，往往紧跟其发展，运用这些知识解决自然灾害研究中遇到的一些难题。比如，数理统计学中的逐步回归（step-wise regression）方法是 1946 年由美国麻省理工学院 Bryan 提出，到 1960 年比较成熟，于 1961 年正式发表。当时在中央气象局研究所工作的年轻技术员史久恩从国外数学期刊读到，很快将该法用于自然灾害的长期预报方面并写出论文，于 1964 年发表在《气象学报》上。此后，为国内外许多自然灾害研究、教学、业务工作者多次引用。几十年来这种方法用得越来越广泛，越来越深入。开始在中央气象台应用，继而到大区中心、省、市气象台，现已广泛用于县气象站业务之中。从应用范围，不仅有长期，也用于中、短期乃至超短期（短时）灾害预报。所用因子已从原有的常规气候数据拓广到天气图、数值预报的模式输出（model output）。就其领域而言，除气象外，还有水文、农业、林业、牧业等诸多领域。譬如，某河段出现超过某一流量 Q_e 的概率为 10%，试问最近 10 年内均不出现和仅出现一次的概率各为多少？通过计算不出现的概率，$B(0)=34.7\%$，仅出现一次的概率为 $B(1)=38.7\%$。在水利工程设计中，像此问

题颇多。从这个例子可知，如果 $P\{Q>Q_a\}=10\%$，并用 Q_a 去设计该河段的断面，以排泄内涝，那末在未来10年中，该河段附近不受内涝的概率为34.7%，而要受内涝灾害的概率为65.3%，即受灾的可能性比不受灾的可能性几乎大1倍。我们制定水利计划中，10年中允许出现1次内涝，其农业歉收由丰收年补足，可以不算受灾。那末能实现我们计划的概率为 $B(0)+B(1)=73.4\%$，而不能实现我们计划的概率为26.6%，也就是说，实现我们计划的概率是比较大的，因它比不能实现计划的可能性几乎大2倍。本书强调概率统计方法在自然灾害研究中的应用，并不意味着不愿应用其他方法，事实上，作者们也多次应用微分方程、计算数学、运筹学等其他方法。用什么方法合适，应视问题性质、数据状况而定。如有可能，应提倡不同方法的有机结合。比如，中期数值天气预报的成功，说明数值方法在灾害预报中有重要贡献。但是，这种方法的出现，不宜削弱统计方法的应用；恰恰相反，正是由于有了中期数值预报中心发布的120h之内的预报图，使统计预报拓宽了选因子的范围。这促进了统计方法在中期天气预报推广中的应用。

第二节　非参数统计应用于自然灾害研究的意义和问题

把概率论与数理统计的一些经典方法用于自然灾害研究若干年后，有经验的研究者可能发现，用经典的概率统计方法找到的一些自然灾害指标，开始几年用起来还可以，然而时间久了，就不合适了。究其原因有两方面，一是生物学问题，一个是样本容量太小。比如某年干旱地区的水分条件只适合某种谷子，不能种小麦。但是随着科技进步，育出了抗旱小麦品种后，必然以种

小麦为抗旱增产的好措施。研究者应该根据生产实际否定以前的统计推断和预测结论，这是生物学问题。至于第二个问题，可以举个简单例子，在 60 年代，研究旱涝灾害时，由于受资料限制，许多人只用 7—8a 的资料寻求相关关系，一旦资料序列长了，这一关系也就不适用了。就拿投掷硬币试验来说，Buffon 试验了 4 040 次，Pearson 试验了 2.4 万次，才观察到正面出现频率近于 0.5。也就是说，即使一个地方，旱、涝年出现的概率实际上都接近 1/2，但在某些较短时段上，旱年完全可以是涝年的几倍，反之亦然。由于发现上述问题，一些曾经使用数理统计方法研究自然灾害者，相当一部分人走向两个极端：一部分人认为，概率统计乃至其他数学方法用于自然灾害研究，无非如此而已。因此，想放弃上述方法，还是去定性描述。另一部分研究者则致力于深奥的数学知识，至于在自然灾害研究中能否用上，不予重视，将活用数学而解决灾害问题，视为“太简单”。

作为灾害研究人员，学习数学的目的全在应用。其实，数学在自然灾害研究中用得好不好，不在于所用数学定理有多深，方法有多新，公式有多长，而在于是否符合客观规律，是否能够解决自然灾害研究或减灾运筹中的实际问题。对于那些费力但又不切合实际的数学方法，不宜忙于运用；而对那些既省力又合理的数学方法，应提倡早用、快用。将概率论与数理统计用于自然灾害研究的正确态度，应当是从前者以及作者们本人以往实践的基础上，指出正反两个方面的经验教训。带着这些经验教训，回过头来重新阅读理论著作及其有关新的文献。比如，“旱涝指标何以失灵”，Brooks 的 *Handbook of Statistical Methods in Meteorology*，Дроздов 的*Методы Климатической Обработки Метеорологических Наблюдений*（1960 年俄文版）、Fisz 的《概率论及数理统计》（1962 年中文版），Cramer 的《统计学的数学方法》（1964 年中文版）以及 Frend 的 *Mathematical Statistics*（1963 年英文版）等书，值得