



水灾害防治中的 多变量概率问题

谢华 著



中国水利水电出版社

www.waterpub.com.cn



水灾害防治中的 多变量概率问题

谢华 著



中国水利水电出版社
www.waterpub.com.cn

内 容 提 要

针对多变量概率问题在水科学领域广泛存在的事实和该问题的解决对水利防灾减灾风险概率评估的重要作用,本书围绕多变量联合概率分布问题,以水科学领域中几个典型的多变量概率分布问题为研究对象,从理论方法探索和实际问题解决两个角度,研究了平原河网地区排涝标准、跨流域调水的丰枯遭遇、海洋工程中的波高和潮位联合风险、洪潮遭遇以及洪水多特征频率分布等,并以这些具体问题为实例,研究了各种多变量联合概率分布模型的性质、特点、适用性及应用方法。

本书可供从事水灾害风险分析、水文频率分析、水利工程设计标准、水资源调度、工程结构可靠度等领域的科研人员和从事水利规划、设计、管理的工程技术人员参考。

图书在版编目(CIP)数据

水灾害防治中的多变量概率问题 / 谢华著. — 北京:
中国水利水电出版社, 2014.12
ISBN 978-7-5170-2881-9

I. ①水… II. ①谢… III. ①水灾—灾害防治—多变量—概率—评估 IV. ①P426.616

中国版本图书馆CIP数据核字(2015)第014176号

书 名	水灾害防治中的多变量概率问题
作 者	谢华 著
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路1号D座 100038) 网址: www.waterpub.com.cn E-mail: sales@waterpub.com.cn 电话: (010) 68367658 (发行部)
经 售	北京科水图书销售中心(零售) 电话: (010) 88383994、63202643、68545874 全国各地新华书店和相关出版物销售网点
排 版	中国水利水电出版社微机排版中心
印 刷	北京瑞斯通印务发展有限公司
规 格	184mm×260mm 16开本 8.75印张 207千字
版 次	2014年12月第1版 2014年12月第1次印刷
印 数	0001—1000册
定 价	28.00元

凡购买我社图书,如有缺页、倒页、脱页的,本社发行部负责调换

版权所有·侵权必究

前言

水灾害是世界上发生频率最高、损失最大的自然灾害之一。水灾害通常包括洪灾、涝灾、干旱、水污染等，其中洪涝、干旱等灾害主要由自然因素引起，水污染则由人类活动引起。我国是遭受各种水灾害危害最严重的国家之一。水灾害问题已经严重威胁我国经济社会的健康发展，对工农业生产、国民生活水平的提高构成严重危害。

新中国成立以来，国家投入了巨大的人力物力兴建了大量的水利工程，不断提高工程设防标准，希望能够避免各种水灾害。但是投入了巨资修建的各种工程，却难以遏制水灾害频率增加、损失持续上升的趋势。随着社会经济的发展和工程建设的进展，人们已经意识到紧紧依靠各种水利工程措施来控制各种水旱灾害是远远不够的，需要转变传统的单纯依靠工程抗灾的观念，通过各种工程性和非工程性措施的综合运用，对各种水旱灾害实施综合灾害风险管理。

为了减灾防灾、兴水利避水害，人们采用了各种工程措施（包括堤防、水库、大坝、沟渠、泵站、水闸、水电站等）来满足防洪、排涝、抗旱、发电、给水、灌溉的需求，并出台了各种规范、法律、法规等对工程进行规划和管理。而对各种不同频率的致灾随机水文事件进行科学的评估和预测，是保证各种工程和非工程措施科学性的基础。

对绝大多数的洪灾、涝灾、旱灾来说，灾害的发生往往是多种致灾因子共同作用的结果。这些致灾因子相互作用、相互伴生，构成了一个结构复杂、层次众多的系统，每个因子在不同的水灾害中表现方式、发挥的作用都有所不同。如何客观地描述多种致灾因子共同作用下的各种水灾害的概率，从而制定科学的防灾减灾工程标准一直是一个难题。迄今为止，各种水文灾害风险预测评估方法，以及各种水利工程的规划设计等，都只考虑了灾害事件的某一单一特征变量，采用单变量概率分析方法评估水文极值事件的风险概率。显然，传统的单变量分析方法不能综合全面地描述各种致灾因子对灾害的贡献，难以描述各特征属性统计规律和相互之间的相关关系。由于水灾害和致灾水文因子的多样性、随机性、关联性的特点，对灾害和随机水文事件风险

概率的统计和预测，在数学上往往属于多变量联合概率分布问题。

多变量联合概率问题在水科学领域广泛存在。比如，洪水的危害不仅仅在于洪峰流量的大小，洪水总量、峰现时间、洪水历时等多种因素也起着关键作用。而在大多数的洪水频率分析中，往往只是采用单变量极值分布（如：皮尔逊Ⅲ型）对洪峰流量进行频率分析，这种单变量的洪水频率分析忽略了洪水总量、峰现时间、洪水历时等因素与洪峰流量的相关关系，其对洪水特征的描述是不全面的，对多特征属性洪水的频率分析是一个典型的多变量概率问题。此外，各种水文事件的遭遇概率问题以及各种环境极值事件的联合分布问题都属于多变量概率问题，比如跨流域水资源调配规划时，常需要分析不同水文区降雨和径流的丰枯遭遇情况；拟定设计洪水的地区组成时，往往需要考虑不同区间的来水遭遇；沿海地区城市防洪规划需要考虑风暴潮、上游洪水、区间暴雨等水文事件的遭遇概率；旱灾的发生是区域降雨和地表径流同时减少的结果；海洋平台工程失效破坏是风、浪、流等多种环境荷载共同作用的结果，需要采用多变量概率分析方法描述多种荷载的联合分布。上述这些问题从本质上讲都属于多变量联合概率分布问题。随着社会经济的发展和人们生活质量的提高，对水利工作也提出了更高的要求，要求确切而不笼统、具体而不空洞地分析研究各种水旱灾害。在生产实际中迫切希望能有易于理解和掌握的多变量概率分析方法，用于解决水科学领域的各种多变量概率问题。

针对多变量概率问题的复杂性和该问题在水科学领域的广泛性，本书以水科学领域中几个典型的多变量概率分布问题为研究对象，从理论方法探索 and 实际问题解决两个角度，研究水科学领域中各种形式的多变量概率分布问题的解决途径和解决方法。期望通过本书的研究，既能从理论上找到易于理解和掌握的多变量概率分析方法，又能通过本书中各种实际多变量概率问题的解决，为工程技术人员提供参考。

本书是国家自然科学基金项目“不同尺度条件下考虑多种致灾因子的洪涝灾害防治标准研究”（编号 50909074）、“旱涝急转发生机理与减灾方法研究”（编号 51339004）、“十二五”国家科技支撑计划项目“大型灌区节水及设备技术研究示范”（编号 2012BAD08B03）等科研项目成果的总结和提升。

本书能够顺利撰写完成和正式出版，离不开作者的博士生导师——武汉大学水利水电学院黄介生教授的指导与关心，受到了武汉大学水利水电学院谈广鸣教授的热忱鼓励与支持，得到了水利科技专著译著出版项目和武汉大学水资源与水电工程科学国家重点实验室的资助。项目研究期间，冯卫民教授和王修贵教授给予了极大的关心和帮助。在此向他们表示衷心的感谢！

本书写作过程中参阅了大量国内外文献资料，作者已尽量将所有引用、参考的文献一一标示出来，但难免出现遗漏的问题。在此，谨向这些文章的作者们表示衷心的感谢！由于水科学领域多变量概率分布问题的广泛性和复杂性，作者对该问题的认识只是管中窥豹，仅仅把自己的一点点体会和成果写成此书，以便于学术界同仁相互交流。限于作者的学识水平，疏漏和不足之处在所难免，敬请广大读者批评指正。

作者

2014 年 1 月



目 录

前言

第 1 章 绪论	1
1.1 研究背景	1
1.2 水灾害的基本状况及特点	2
1.3 本书的研究内容	5
第 2 章 多变量概率问题研究述评	7
2.1 单变量水文频率分析方法	7
2.2 多变量概率分布问题解决方法	8
2.3 本章小结	11
第 3 章 多变量联合概率分布	13
3.1 两变量联合概率分布	13
3.2 两变量联合概率分布模型	18
3.3 三变量联合概率分布	24
3.4 本章小结	26
第 4 章 Copula 函数理论	27
4.1 概述	27
4.2 Copula 函数及特点	31
4.3 Copula 函数族	35
4.4 阿基米德族 Copula 函数及其性质	37
4.5 最优 Copula 函数的确定及拟合优度检验	41
4.6 Copula 函数表示的联合分布及条件分布	47
4.7 本章小结	49
第 5 章 平原河网地区排涝标准研究	50
5.1 概述	50
5.2 平原河网地区涝灾风险分析	51
5.3 基于风险分析的排涝标准研究	66
5.4 本章小结	70
第 6 章 跨流域调水中不同水文区丰枯遭遇分析	72
6.1 概述	72
6.2 边际分布函数确定	73

6.3	三维 Copula 函数参数的确定及拟合优度检验	79
6.4	不同水文区的丰枯遭遇分析——多变量联合分布及条件分布	80
6.5	多维 Copula 方法与一维转换方法比较	83
6.6	本章小结	86
第 7 章	相依结构和边际分布对联合分布概率的影响	88
7.1	相关性的准确度量对联合概率的影响	88
7.2	边际分布的选择对联合概率的影响	92
7.3	本章小结	94
第 8 章	多变量概率问题解决流程及应用实例	96
8.1	多变量概率分布问题解决流程	96
8.2	多特征变量洪水频率分析问题研究	97
8.3	感潮河网地区洪潮遭遇概率研究	109
8.4	本章小结	120
结束语		121
参考文献		122

第 1 章 绪 论

1.1 研 究 背 景

灾害是自然系统与人类物质文化系统相互作用的产物,是对所有造成人类生命财产损失或资源破坏的自然和人为现象的总称。自然灾害是指发生在生态系统中的自然过程,它可以导致社会经济系统失去稳定与平衡,使社会财产产生损失或导致社会在各种原生的和有机的资源方面出现严重的供需不平衡^[1]。

全球每年有 20%~50% 的人口遭受洪涝、干旱、地震、台风等各种自然灾害的威胁。近 100 年内,发生死亡人数在 1000 人以上的灾害事件达数千起。随着世界人口的剧增,人类社会活动的增强,一方面加速了自然环境的退化,另一方面增加了社会财富的易损性,导致区域灾害损失潜力增大^[2-3]。20 世纪 90 年代以来,各种灾害有增无减,频率增加了 3.2 倍,年死亡和受伤人数分别增加 5.2 倍和 6.9 倍,这意味着全球平均每 31 人中有 1 人遭灾,每 3.1 万人中有 1 人死亡^[4]。2002 年全球发生的各种灾害共造成 400 亿美元的损失,并导致 1.9 万人死亡。其中有 80 多个国家的 1700 多万人遭受水灾,受灾面积超过 800 万 km²。2003 年全球有超过 5 万人被地震、洪水、暴风雪等自然灾害夺去生命;根据联合国减灾署发布的数据^[5],2011 年全球发生洪水、地震、热带风暴等重大自然灾害共 302 起,2.06 亿人受灾,其中一半受到洪涝灾害影响,另有 6000 万人遭遇旱灾,3400 万人受风暴影响。据统计,洪涝、暴风、地震、干旱、台风、泥石流、雪灾、虫害是当今世界面临的最频繁的自然灾害,其中前四种灾害的损失占总损失的 90%,而洪涝灾害就占约 40%^[6]。

水灾害是世界上发生频率最高、损失最大的自然灾害之一。水灾害通常包括洪灾、涝灾、干旱、水污染等。我国是遭受各种水灾害危害最严重的国家之一。根据德国慕尼黑再保险公司 1950—1999 年全球重大自然灾害的统计结果^[7],在 20 世纪全球 100 个损失最大的自然灾害中,中国占了 17 个,而在中国的 17 个灾害事件中,仅洪水就占了 12 个,损失额 678.7 亿美元,中国是世界上洪水灾害最频繁的国家之一,且同时与印度并列为亚洲干旱最频繁的国家。在中国几千年的发展历程中,与水有关的各种灾害对中华民族构成了最严重的威胁。据不完全统计,公元前 206—1949 年共 2155 年间,我国发生过的较大水灾有 1092 次^[8];1950—2000 年,全国水灾共造成 26.3 万人死亡,平均每年造成的受灾农作物和成灾农业面积分别占耕地面积的 10% 和 5%^[9]。此外,干旱缺水对我国经济社会的发展也构成了重大危害。1949—1990 年,全国受旱耕地面积累计共 8.2 亿 hm²,其中成灾面积 3.2 亿 hm²,成灾比(成灾面积与受旱面积之比)为 39%^[10]。随着经济的快速增长,水资源需求持续增加,水环境破坏日益严重,各种干旱、缺水等问题呈蔓延趋



势。在沿海三角洲河口地区,由于台风风暴潮等海洋灾害引发的风暴增水、暴雨等常常导致严重的洪涝,水污染、盐水入侵等灾害,对堤围、公路、海塘、农田、生态等构成严重危害。

水灾害问题已经严重威胁我国经济社会的健康发展,对工农业生产、国民生活水平的提高构成严重危害。新中国成立以来,国家投入了巨大的人力物力兴建了大量的水利工程,不断提高工程设防标准,希望能够避免各种水灾害问题。但是投入了巨资修建的各种工程,却难以遏制水灾害频率增加、损失持续上升的事实。水利工程标准不可能无限制地提高,水灾害的风险也不可能完全消除,如何通过各种工程性和非工程性措施的综合运用,提高综合防灾抗灾能力,获得最大的安全保障,建立人与自然共存、人水和谐的可持续发展的社会,是各国政府普遍关注的问题。

当前,各国政府和学术界已经普遍认识到紧紧依靠工程措施控制水灾害是远远不够的,需要统筹考虑各种工程、法律、行政、经济、技术等手段,实施综合灾害风险管理。所谓综合自然灾害风险管理是指人们对可能遇到的各种自然灾害风险进行识别、估计和评价,并在此基础上综合利用法律、行政、经济、技术、教育与工程手段,通过整合的组织和社会协作,通过全过程的灾害管理,提升政府和社会灾害管理和防灾减灾的能力,以有效地预防、回应、减轻各种自然灾害,从而保障公共利益以及人民的生命、财产安全,实现社会的正常运转和可持续发展^[1]。对自然灾害进行综合风险管理首先要解决的问题就是要对灾害风险进行科学的预测和评估。

水利防灾减灾的各种工程和非工程措施建立的基础,就是要科学地预测分析水灾害发生的可能性及其危害。自然灾害往往不是孤立的。水灾害的发生往往是多种致灾因子共同作用的结果,各致灾因子之间相互联系,相互伴生,具有较强的相关关系。由于水灾害和致灾因子的复杂性,在水灾害防治领域,往往忽略了多种致灾因子这一事实,各种水文灾害风险预测评估方法以及各种水利工程的规划设计等都只考虑了灾害事件的单一致灾因子或单一特征变量,采用单变量概率分析方法评估水文极值事件的风险概率。显然,对于具有多种特征属性的水文极值事件和多致灾因子共同作用下的水灾害事件,传统的单变量分析方法不可避免忽略了许多重要的信息,不能综合全面地描述各种致灾因子对灾害的贡献,其对水灾害事件的预测评估是不全面的。

对各种量级灾害事件的发生概率进行科学预测和评估,是确定各种水利工程设计标准的依据。对水灾害风险概率的预测评估要综合考虑灾害之间的相互联系和致灾因子之间的相关性,进行多变量综合风险分析。然而,当前各种水灾害防治标准往往将灾害作为孤立的单因子事件处理,以某一量级的单一致灾因子的重现期作为工程标准,这不可避免地忽略了自然灾害多致灾因子的特点,不符合灾害的本质特征。本书的选题和研究就是建立在上述的背景之上。

1.2 水灾害的基本状况及特点

1.2.1 水灾害的基本概念

水灾害是因为水的诱因而导致对人类生命财产造成危害的灾害。水灾有两层含义,一



层是“水”，另一层是“灾”^[12]。“水”的含义就是指自然界的物质部分，包括以降雨、融雪、洪水等方式运动的自然事件；“灾”的含义则指水对其影响范围内的人类社会造成的损害，即为“灾害”。显然，水灾害与人类社会的发展密切相关。事实上，引发灾害的各种水文现象，如暴雨、洪水、水土流失等是人类出现以前早已存在的自然现象。人类为了保护自己，初始时是顺应自然，避开洪、涝、旱的威胁而迁居，继而采取各种措施防御水灾害对人们的侵袭。在长期与水旱灾害斗争的历史过程中，社会经济不断发展，同时也发展了各种防御、减轻和适应水旱现象的对策和措施。

水灾害的具体表现形式主要包括洪灾、涝灾、干旱等。中外文献对于“洪”和“涝”很少有严格的定义。早期对于洪涝是不分的，及至后来修筑了堤防、圩垸等防洪工程，天然的来水受到人为的分割，河道泄流条件发生改变，逐步有了洪涝之分。一般认为河流漫溢或堤防溃决造成的灾害为洪灾；当降雨过多，不能排除的积水灾害为涝灾。由于“内水”和“外水”是相对的，因此很难严格区分，在世界其他国家，不论“内水”和“外水”都称为洪水。洪水和内涝在水文特性上有区别，洪水来势凶猛；雨涝一般来势较缓，强度较弱。二者造成的灾害特点也不相同，洪水破坏性强，可以破坏各种基础设施，会对工农业生产造成毁灭性破坏，而涝灾的破坏性相对较弱。此外，防洪治涝的对策措施也有很大不同，防洪工程措施主要是水库、堤防和蓄滞洪区等，而除涝则主要通过沟渠、泵站、水闸等排除积水^[10]。

洪水、干旱、内涝等对人类社会造成危害的程度与区域人口、工农业布局和经济水平密切相关。如果洪泛区是没有人类生产生活的荒地，那么即使发生较大的洪水，也不至于对人造成很大危害。相反，如果洪水影响范围内会对人类生产生活造成危害，就对防洪提出了要求。干旱也是如此，气象条件是形成干旱的重要自然因素，但干旱的气象条件不一定形成旱灾。同理，涝灾害能否形成，除了降雨地形等自然因素外，还与当地的除涝能力和外部洪水情势有关。

总体而言，各种水旱灾害的形成非常复杂，有自然的、也有社会的。各种致灾因子具有明显的随机性和不确定性。因此，各种防洪、除涝、抗旱等工程和非工程措施的出台，首先要做的就是对各种灾害及其危害的统计资料进行分析，研究灾害发生的强度和频度。根据概率论和统计学的基本原理和方法，评估各种量级的灾害发生的风险，为制定科学有效的防灾减灾、兴利避害的措施提供依据。

1.2.2 水灾害的基本特征

由于形成水灾害的因素复杂多样，水灾害的表现形式也多种多样。各种形式的灾害相互联系、相互伴生、互相演变，构成了一个结构复杂、层次众多而且规模庞大的系统，具有致灾因子多样性、随机不确定性、复杂性和交叉性等特征。科学地分析水灾害这些基本特征是采取有效防治措施的基础。

1. 致灾因子多样性

水灾害是由各种性质不同的自然现象引起，如暴雨、洪水、干旱、台风、暴雨、融雪、地震等各种自然现象，都可能引发各种规模不等、形式多样的水灾害。灾害的发生往往是多种致灾因子共同作用的结果。

致灾因子多样性的特点要求在评估和预测灾害风险和危害时，必须区分矛盾的主要方



面,同时要考虑致灾因子的对灾害的综合影响。

2. 随机性和可预测性

水灾害的随机性源自致灾因子和灾害的多样性、差异性。由于致灾因子、水文气象状况、地形地貌的随机性,导致水灾害的随机性。其随机性体现在灾害区域、灾害规模、灾害时程的随机性。可预测性是指灾害发生、发展、危害等,是具有一定规律性的,是可以预测的。

随机性和可预测性的意义在于可以利用描述变量随机性特点的数学工具预测水灾害。

3. 复杂性和交叉性

水灾害是开放的组织系统^[13]。系统内部之间以及系统与外部之间不断进行着物质、能量和信息的交换,系统状态可以由一种无序结构突变到另一种有序结构,水灾害的群发就是证明^[13]。由于自然生态环境的复杂性,各种自然灾害往往相伴而生,形成灾害链。比如暴雨作为一种常见的水灾害,可能会触发滑坡、泥石流等灾害,形成暴雨→洪水→滑坡→泥石流等多灾并发的灾害链。而地震可能会诱发海啸和洪涝灾害。

灾害的复杂性和交叉性特点要求灾害的预测和评估必须考虑灾害之间以及致灾变量之间的相依关系和相关性。

4. 区域性和阶段性^[13]

水灾害的区域性,一是表现在经济发达地区,水灾害发生的频率增多,如超强度开发水资源加重了干旱和水污染等;二是社会经济发达,人类活动对自然环境的改造及对水灾害的干预程度较高;三是经济发达地区人口集中、产业密集,故水灾害承受面广,受灾机会多,经济损失大。水灾害的阶段性主要体现在随着社会经济发展的不同阶段,水灾害的破坏作用和成灾方式也呈现不同的变化阶段。

区域性和阶段性的特点要求各种水利工程必须符合区域的实际情况,针对不同时期的社会发展需求满足不同的需要。

1.2.3 水灾害防治中的多变量概率问题

为了减灾防灾、兴水利避水害,人们采用了各种工程措施,包括堤防、水库、大坝、沟渠、泵站、水闸、水电站等,以满足防洪、排涝、抗旱、发电、给水、灌溉的需求,并出台了各种法律法规对工程进行规划和管理。而对各种不同频率的致灾随机水文事件进行科学的评估和预测,是保证各种工程和非工程措施科学性的基础。由于水灾害和致灾水文因子的多样性、随机性、关联性的特点,对灾害和随机水文事件风险概率的统计和预测,从数学上讲属于多变量联合概率分布问题。

水文频率分析是水利工程规划、设计、管理的重要基础。随机水文事件通常具有多方面的特征属性,为了全面了解其统计规律,需要从多个角度进行定义和描述。由于资料 and 有效数学工具的缺乏,实际中往往只挑选某个最重要的特征属性来代表水文事件,进行单变量的频率分析。比如,洪水作为研究最广泛的水文极值事件,其特征属性包括洪峰流量、洪水总量、峰现时间、洪水历时等,而在大多数的洪水频率分析中,往往只是采用单变量极值分布(如皮尔逊Ⅲ型)对洪峰流量进行频率分析,这种单变量的洪水频率分析忽略了洪水总量、峰现时间、洪水历时等诸多因素对洪灾的作用,以及各种特征属性之间的相关关系,其对洪水特征的描述是不全面的;同样的,对于设计暴雨而言,单变量的频率



分析亦不能描述暴雨强度、雨量、降雨历时等因素之间的联合分布关系。显然,对于具有多种特征属性的水文极值事件,传统的单变量分析方法难以描述各特征属性统计规律和相互之间的相关关系,需要采用多变量联合概率分析方法。

多变量联合概率分布问题在水科学领域广泛存在。比如,我国的许多沿江平原地区和三角洲河网地区大多地势低平,暴雨涝水能否顺利排除不仅取决于设计暴雨标准,还受到外江水位的极大制约,设计暴雨和外江水位的遭遇问题,本质上就是一个两变量联合分布问题;跨流域水资源调配规划时,常需要分析不同水文区降雨和径流的丰枯遭遇情况^[14];拟定设计洪水的地区组成时,往往需要考虑不同区间的来水遭遇^[15-16];沿海地区城市防洪规划需要考虑风暴潮、上游洪水、区间暴雨等水文事件的遭遇概率^[17];旱灾的发生是区域降雨和地表径流同时减少的结果;海洋平台工程失效破坏是风、浪、流等多种环境荷载共同作用的结果^[18],需要采用多变量概率分析方法描述多种荷载的联合分布。上述这些问题从本质上讲都属于多变量联合概率分布问题。

多变量联合分布函数是研究两种或两种以上随机变量联合影响下的某种自然现象概率分布的有效数学工具。多变量联合分布概率可采用式(1.1)表示^[18-19]:

$$F(x_1, x_2, \dots, x_n) = \iiint \dots \int f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \quad (1.1)$$

式中: $F(x_1, x_2, \dots, x_n)$ 为多因素共同作用的联合概率; $x_1, x_2, \dots, x_n$ 为随机变量; $f(x_1, x_2, \dots, x_n)$ 为多变量概率密度函数。

对于式(1.1)所示的多变量联合分布函数,只有在各变量均属正态分布时,其联合分布函数才会有解析表达式。对于非正态、相关的多维随机变量,因其复杂性而成为理论研究和生产应用中的难点。

1.3 本书的研究内容

水利防灾减灾和兴利避害是构建可持续发展社会的基础性工作。由于水灾害及其致灾因子的多样性、随机性和相关性的特点,多变量联合概率分布问题在水科学领域广泛存在。由于多变量概率问题的复杂性,迄今为止,各种水文灾害风险预测评估方法以及各种水利工程的规划设计等都只考虑了灾害事件的某一单一特征变量,并采用单变量概率分析方法评估水文极值事件的风险概率。单变量概率分析方法不可避免忽略了许多重要的信息,其对水文极值事件的预测评估是不全面的。随着社会经济的发展和生活水平的提高,人们对工程技术及工程管理水平提出了更高的要求。生产实际中期望能有易于理解和掌握的多变量概率分布方法,用于解决水科学领域的各种多变量联合分布问题。

本书围绕水科学领域中的多变量联合概率分布问题展开研究,以水科学领域中几个典型的多变量概率分布问题为研究对象,从理论方法探索和实际问题解决两个角度,研究了平原河网地区排涝标准问题、跨流域调水的丰枯遭遇问题、海洋工程中的波高和潮位联合风险问题、洪潮遭遇问题,以及洪水多特征属性频率分布问题等;并以这些具体问题为实例,研究了各种多变量联合概率分布模型的性质、特点、适用性及应用方法。主要包括以下几个方面:



(1) 系统总结了现状各种多变量概率问题求解方法和模型, 分析了多变量概率分布问题以及各种分布模型的特点, 比较了各种模型的优缺点, 探讨了各模型在水科学领域中的适用性和局限性。

(2) Copula 函数方法作为一种新的构建多变量联合分布模型的方法, 能够克服现有模型的不足, 在水科学领域具有广阔的应用前景。本书介绍了 Copula 函数的性质特点、建模方法。重点介绍了阿基米德族 Copula 函数的特性、参数估计方法、拟合优度检验方法, 以及基于 Copula 函数的各种形式的联合概率、条件概率、重现期的确定方法。

(3) 平原河网地区的涝灾是暴雨和外江水位等多个致灾因子共同作用的结果。本书以暴雨和外江水位的联合分布概率为实例, 将 Copula 函数方法应用于解决涝灾风险问题, 研究了二维 Copula 函数用于解决实际问题的方法和步骤, 并在此基础上建立了基于涝灾风险率的排涝标准。

(4) 当变量维数 $n \geq 3$, 多变量联合概率分布问题因其复杂性难以有明确的解析表达式。本书以长江流域、淮河流域及黄河流域不同水文区丰枯遭遇概率研究为实例, 将三维 Copula 函数引入水科学领域中多变量概率问题的求解, 构建了基于 Copula 函数的三维联合概率分布模型, 分析了长江流域、淮河流域及黄河流域的径流量的联合概率和条件概率问题, 并与费永法方法 (多维分布转换为一维分布) 进行比较, 进一步分析了在三维条件下 Copula 函数的应用方法和特点。

(5) 无论是传统的多变量联合概率模型还是 Copula 函数, 其构建的多 (两) 变量联合概率模型都取决于变量的边际分布和描述变量相互关系的相依结构。本书以海洋工程中极值风速和有效波高的联合风险概率为实例, 研究了变量相关性及其对联合概率的影响, 为多变量概率分布模型的选优提供了决策依据。

(6) 以河口地区的洪潮遭遇问题、洪水多特征频率分布问题为实例, 总结了当面对一个实际工程中的多变量联合概率分布问题时, 应该采用的解决方法和流程, 模型选择, 参数的确定等问题。

第 2 章 多变量概率问题研究述评

2.1 单变量水文频率分析方法

长期以来,生产实际中评估和预测洪水、暴雨等水文随机变量的发生概率时,都只是应用一维的概率分布模型对其分析;对于水科学领域经常遇到的诸如水文事件的遭遇概率问题以及各种环境极值事件的联合分布等问题,要么采用实测资料进行定性分析,要么将各水文变量分别进行概率分析,然后进行简单的叠加。当前,只考虑单一因素的一维概率分布模型日趋成熟,已经广泛应用于水利行业的各个方面,同时也为多变量概率分布问题的研究奠定了基础。

对于单变量水文极值数据系列的分析,世界各国制定有关水文计算手册或规范时,大多采用皮尔逊Ⅲ型(简称 P-Ⅲ型)、极值Ⅰ型、对数皮尔逊Ⅲ型等分布模型。我国常采用皮尔逊Ⅲ型分布;英国、法国及其原殖民地国家选用广义极值分布(GEV),但英国的最新设计洪水估算手册提出采用广义逻辑分布(GL);美国、加拿大和澳大利亚、印度等国采用 Gumbel 分布;日本采用对数正态分布;泰国等东南亚国家一般选用 P-Ⅲ分布作洪水和暴雨频率分析^[20]。世界各国采用的频率分布线型见表 2.1^[20-21]。

表 2.1 各国采用的水文频率分布线型^[20]

分 布 线 型	国 家
皮尔逊Ⅲ型分布	中国、奥地利、保加利亚、匈牙利、波兰、罗马尼亚、瑞士、泰国
对数皮尔逊Ⅲ型分布	美国、澳大利亚、加拿大、新西兰、墨西哥及一些南美国家
广义极值分布	英国、法国、爱尔兰和非洲一些国家
极值Ⅱ型,极值Ⅲ型分布	英国、法国和非洲一些国家
二/三参数对数正态分布	日本
极值Ⅰ型分布	比利时、德国、瑞典、土耳其
K-M 分布	俄罗斯及一些东欧国家

1. 皮尔逊Ⅲ型分布

皮尔逊Ⅲ型分布是我国主要应用的概率分布模型,是一种三参数 Gamma 分布函数。其概率密度函数如下:

$$f(x)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}(x-\gamma)^{(\alpha-1)}e^{-\beta(x-\gamma)} \quad (2.1)$$

式中: $\Gamma(\alpha)$ 为 α 的 Gamma 函数; α 为形状参数; β 为尺度参数; γ 为位置参数。



2. 极值 I 型分布 (Gumbel 分布)

极值 I 型分布亦称 Gumbel 分布, 指的是 n 个观测值中极大值的概率分布, 其不超过概率分布函数, 如式 (1.3) 所示:

$$P(X \leq x) = \exp[-e^{-(x+a)/c}] \quad (2.2a)$$

$$\text{其中} \quad a = rc - \bar{x} \quad (2.2b)$$

$$c = \frac{\sqrt{6}}{\pi} \sigma \quad (2.2c)$$

式中: P 为不超过概率; x 为水文统计量; a 、 c 为统计参数; r 为欧拉常数, $r = 0.57721$; σ 为水文统计量的标准差。

3. 对数正态分布

对数正态分布是从正态分布的对数变换推演而来的, 其概率密度函数为:

$$f(x) = \frac{1}{x\sigma_y \sqrt{2\pi}} \exp\left[-\frac{\ln x - \alpha_y}{2\sigma_y^2}\right], x > 0 \quad (2.3)$$

$$Y = \ln X \quad (2.4)$$

式中: α_y 、 σ_y 分别是随机变量 Y 的数学期望和均方差。

4. 两参数 Gamma 分布

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{(\alpha-1)} e^{-\beta x} \quad (2.5)$$

式中: $\Gamma(\alpha)$ 为 α 的 Gamma 函数; α 为形状参数; β 为尺度参数。

2.2 多变量概率分布问题解决方法

对于式 (1.1) 所示的多变量联合分布函数, 只有在各变量均属正态分布或相互独立时, 其联合分布函数才会有解析表达式。对于非高斯、相关的多维随机变量, 许多学者对其做了研究, 提出了许多具有代表性的求解方法。

2.2.1 多维分布转换为一维分布

费永法^[22-23]利用事件积原理推导出二维及多维随机变量转化为一维随机变量的计算方法, 并用来求解不同流域丰枯遭遇问题。多维联合概率的计算公式为:

$$P(X_1 \geq x_1, X_2 \geq x_2, \dots, X_n \geq x_n) = P(Z \geq z) \quad (2.6a)$$

式 (2.6a) 中随机变量 Z 的系列为:

$$Z_k = \min(x_{1k}, x_{2k} + a_2, \dots, x_{nk} + a_n), k = 1, 2, 3, \dots, n \quad (2.6b)$$

式 (2.6b) 中, $a_j = x_1 - x_j$, $j = 2, 3, \dots, n$; $z = x_1$; x_j 为随机变量 X_j 相应频率的设计值。

同理可得联合不超过概率的计算公式:

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = P(Z \leq z) \quad (2.6c)$$

式 (2.6c) 中随机变量 Z 的系列为:

$$Z_k = \max(x_{1k}, x_{2k} + a_2, \dots, x_{nk} + a_n) \quad (2.6d)$$

各符号含义同式 (2.6b)。



2.2.2 正态化变换方法

由于正态随机变量具有明确的概率分布函数表达式, 如果将原始数据通过某种变换方法转换为正态随机变量, 则可以利用多维正态分布函数描述随机变量的联合分布概率。利用正态分布进行概率分析的关键是原始数据的正态化问题。常用的正态化变换方法主要有幂变换、多项式正态变换 (PNT)、NQT 变换等。

1. 幂变换^[24]

设 $x_i > 0, i=1, 2, \dots, n$, 令

$$Y_{\lambda i} = \begin{cases} \frac{(x_i^\lambda - 1)}{\lambda g^{\lambda-1}}, & \lambda \neq 0 \\ g \ln(x_i), & \lambda = 0 \end{cases} \quad (2.7)$$

其中

$$g = \left(\prod_{i=1}^n x_i \right)^{1/n}$$

式中: x_i 为原始样本; $Y_{\lambda i}$ 为变换后的样本; λ 为变换参数; g 为几何平均值。

Box 和 Cox^[25] 提出了式 (2.8) 所示的略有修正的正态变换:

$$Y_{\lambda i} = \begin{cases} \frac{x_i^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(x_i), & \lambda = 0 \end{cases} \quad (2.8)$$

幂变换的优点在于可以根据不同数据系列的特点确定不同的变换, 对于数据有较好的适应性。对于不同的统计参数 λ , 变换具有不同的类型, 包括了对数变换 ($\lambda=0$), 平方根变换 ($\lambda=0.5$), 倒数变换等各种变换。Box-Cox 变换在水文科学领域有较广泛的应用。

2. 多项式正态变换 (PNT 变换)

一个原始偏态分布的变量 X 可以通过三阶的多项式变换为标准正态分布^[26]:

$$X = a_0 + a_1 Z + a_2 Z^2 + a_3 Z^3 \quad (2.9)$$

式中: X 为原始变量; Z 为标准正态分布的变量; a_0, a_1, a_2, a_3 为多项式系数。

对于给定的一组 a_0, a_1, a_2, a_3 , 要求 X 与 Z 的关系为单调递增的一一对应关系, X 与 Z 的单调性关系满足式 (2.10):

$$a_2^2 - 3a_1 a_3 > 0, a_3 > 0 \quad (2.10)$$

梁忠民和戴昌军^[27-29] 分别对 Box-Cox 变换和 PNT 变换的特点、变换效果以及应用特点做了详细研究。

3. 分位数正态变换 (NQT 变换)

设变量 X 的原始分布为 $F(x)$, 则采用以下的分位数变换:

$$\Phi(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w e^{-\frac{1}{2}t^2} dt \quad (2.11a)$$

$$x' = \Phi^{-1}[F(x)] \quad (2.11b)$$

式中: Φ 为标准正态分布函数; Φ^{-1} 为标准正态分布反函数; x' 为由 X 经分位数正态化 (NQT) 以后服从标准正态分布的随机变量。

一般而言, 正态化变换可以改进正态的近似程度, 但并不能保证变换后的数据充分接近于正态分布, 因此对正态化变换产生的结果, 应该检验其是否违背正态性的假设^[30]。