单变量水文序列频率计算原理与应用

宋松柏　康 艳　宋小燕　王小军　金菊良/著

Frequency Calculation Principle and Their Application of Single Variable Hydrological Sequence

科 学 出 版 社

单变量水文序列频率计算原理与应用

宋松柏　康　艳　宋小燕　王小军　金菊良　著

科学出版社

北　京

内 容 简 介

本书是继《Copulas 函数及其在水文中的应用》出版后，又一部力求反映国内外关于单变量水文序列频率计算理论前沿研究进展的著作．全书结合统计水文学、微积分、概率论与数理统计和数值计算等原理，系统地推导了单变量水文序列频率计算的有关计算公式．其中，许多计算原理和方法尚未见于中文文献；同时，更正了目前文献中一些印刷或其他方面的错误，给出了许多较为详细的推导过程，以帮助青年学生系统学习水文频率计算原理．本书主要内容包括：水文序列经验频率计算方法、P-Ⅲ型分布水文序列频率分布参数的常用计算方法、基于熵原理的 P-Ⅲ型分布参数估计、非参数核密度估计原理与应用、高阶概率权重矩原理与应用、基于贝叶斯理论的水文频率分布参数估计、部分概率权重矩原理与应用、洪峰流量的理论概率分布、重现期计算、截取分布在水文中的应用、非一致水文序列频率计算原理等．除叙述上述计算原理和方法外，书中附有大量的计算实例，供读者阅读和理解．

本书可作为学习水文统计学原理的工具书或参考书，也可供水文与水资源工程、农业水土工程、水利水电工程、环境科学、气象科学、土木工程和统计等专业高年级本科生、研究生以及相关领域教学、科研与工程技术人员使用．

图书在版编目(CIP)数据

单变量水文序列频率计算原理与应用/宋松柏等著．—北京：科学出版社，2018.2

ISBN 978-7-03-056600-3

Ⅰ. ①单…　Ⅱ. ①宋…　Ⅲ. ①水文计算　Ⅳ. ①P333

中国版本图书馆 CIP 数据核字(2018) 第 036246 号

责任编辑：李　欣　赵彦超／责任校对：彭　涛
责任印制：张　伟／封面设计：陈　敬

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

北京教图印刷有限公司 印刷
科学出版社发行　各地新华书店经销
*

2018 年 2 月第　一　版　开本：720 × 1000　1/16
2018 年 2 月第一次印刷　印张：46 1/4
字数：912 000

定价：198.00 元

(如有印装质量问题，我社负责调换)

前　言

水文频率分析是运用水文学、概率论与数理统计和其他数学原理，利用水文资料分析水文现象的统计规律，定量描述水文变量设计值与发生频率 (或重现期) 之间的关系，也是各类水利、土木工程规划、设计确定工程规模和管理决策的主要依据. 工程规划的目的不是完全消除所有自然灾害，而是最大程度地减少导致灾害事件的发生次数. 因此，水文事件的发生频率必须正确计算. 但是，水文事件的 "输入" 受控于自然条件，导致这一问题发生机理十分复杂. 降水和径流是水文频率计算最为经常采用的分析数据. 1880—1890 年，Herschel 和 Freeman 第一次进行了径流序列频率分析. 1914 年，Fuller 被认为是首次进行综合性水文频率计算的学者. 1941—1942 年，Gumbel 第一次研究洪水极值事件概率分布. 早期研究水文频率计算的还有: 1936 年，Yarnell 分析了美国 5—24min 历时的降水序列频率. 1953 年，Chow 研究了 Illinois 和 Chicago 州的降水频率分析. 1954 年，Chow 又进一步扩充了极值概率分布研究. 1964—1992 年，美国 Hershfield(1964)、Weather Bureau(1964)、Miller 等 (1973)、Frederrick 等 (1977)、Huff 和 Angel (1989, 1992) 等开展了降水频率分析. 1954 年和 1963 年，Gumbel 应用极值分布研究枯水径流和干旱发生频率. 1975 年英国颁布了洪水研究报告. 1947 年，我国学者陈椿庭先生开展了长江等 5 条河流的洪水流量频率计算研究. 20 世纪 80 年代，丁晶、宋德敦、马秀峰、刘光文等先后提出了 P-III型分布参数估计的概率权重法、单权函数法和双权函数法，他们的研究成果至今被广泛地应用于水文频率分布的参数估计或被许多计算手册、教科书引用. 水文频率分析距今已有一百多年的研究和实践. 目前，这一方法已经广泛地应用于水质、海洋等许多随机事件的概率特征分析.

Vijay P. Singh 和 W. G. Strupczewski 认为水文频率分析方法大致可以分为 4 类 (2002): ①经验法; ②现象法; ③动力法; ④随机模型结合蒙特卡罗模拟法. 按照采用测站的多少，分为单站频率分析和区域频率分析. 经验法是上述 4 类方法中应用较广泛的方法，采用经验法进行单站频率分析是工程规划设计使用最多的方法. 水文频率分析主要有参数统计和非参数统计两种途径. 参数统计方法是国内外研究和应用较多的方法，需事先假定水文序列的分布模型，利用参数估计方法估算样本参数，根据样本统计特征值与分布参数的关系，求出分布参数. 这种方法涉及 6 个步骤 (Singh, Strupczewski, 2002; Meylan et al., 2012): ①水文样本数据选择和数据检验; ②选择经验公式 (绘点位置公式) 计算样本经验概率; ③选择概率分布函数，采用合适的参数估算技术拟合水文样本; ④水文分布模型检验; ⑤水文设计值

不确定性分析; ⑥给定设计频率, 进行水文设计值计算.

水文样本数据一般根据需要选取, 并形成某类特征值数据序列, 如一定时间和空间尺度的极值、月值、枯水值和年值数据等. 对于暴雨和洪水来说, 也涉及暴雨和洪水的场次划分. 按照选样方法, 可形成年值序列或年极值序列、超定量序列和年多值序列. 洪水计算除选择实测洪水外, 有条件的情况下, 还需要通过水文调查, 选取较为可靠的历史洪水. 上述选取水文数据序列必须满足下述计算前提条件 (Meylan et al., 2012; Favre et al., 2012; Musy et al., 2012): ①数据正确地揭示水文变化规律; ②形成数据序列的物理机制没有发生变化 (一致性, consistent), 满足平稳性 (stationary) 和同质性 (homogeneous); ③数据序列满足随机简单样本特性. 随机性 (random) 是指样本数据服从同一概率分布, 而简单样本则指一个样本数据不影响后续值的发生, 即数据间满足独立性; ④数据序列应具有足够的长度. 数据检验分为参数检验 (parametric test) 和非参数检验 (nonparametric test) 两大类, 包括样本特征参数与分布参数的一致性检验 (conformity test); 两个样本分布的同一性检验 (homogeneous test); 样本服从某一概率分布的检验 (goodness of fit test); 样本数据间的相依性检验 (autocorrelation test). 水文分布模型检验方法有图形法、卡方检验、Kolmogorov-Smirnov 检验、GPD 检验、Anderson-Darling 检验、矩图法 (diagrams of moments) 、线性矩图法 (diagrams of L-moments), 以及分布函数模型选优比较法 (AIC 法和 BIC 法). 上述水文频率计算中的①和④步骤所涉及的原理和方法一般在《水文统计学》或《概率论与数理统计》教科书均有所介绍, 读者可参阅有关文献学习, 本书不再重复叙述.

常用的水文频率线型有 20 多种, 主要分为以下四大类: ①Γ 分布类, 包括指数分布、两参数 Γ 分布、P-III 型分布和对数 P-III 型分布等. ②极值分布类, 包括极值 I 型、II 型、III 型分布及广义极值分布. ③正态分布类, 包括正态、对数正态和三参数对数正态分布. ④Wakeby 分布类, 主要包括五参数 Wakeby 分布、四参数 Wakeby 分布和广义 Pareto 分布. 参数估计方法主要有: ①矩法; ②极大似然法; ③概率权重和线性矩法; ④最小二乘法; ⑤最大熵原理; ⑥混合矩法; ⑦广义矩法; ⑧不完整均值法; ⑨单位脉冲响应函数法. 其中, 矩法、极大似然法和概率权重法是最广泛的参数估计方法. 美国地质勘查局也使用指数洪水分析 (index flood method) 和回归分析 (regression analysis) 进行区域频率分析. 除经验法外, 现象法 (phenomenological method)、动力法 (dynamic method) 和随机模型结合蒙特卡罗模拟法目前还处于学术研究层面, 工程实际应用较少. 非参数统计方法主要指在所处理对象总体分布族的数学形式未知情况下, 对其进行统计研究的方法. 而非参数估计就是在没有参数形式的密度函数可以表达时, 直接使用独立同分布的观测值对总体的密度函数进行估计的方法. 主要包括概率密度核估计和非参数回归估计模型. Rao A. Ramachandra 和 Hamed Khaled H. 在他们的《洪水频率分析》专著中,

系统地总结了目前水文频率的常用计算方法, 给出了大量详细的应用实例.

上述水文频率计算的前提条件在实际中难以满足. 第一, 单站和区域的水文概率分布函数是未知的, 甚至有多种物理机制形成径流, 例如, 降雨径流、融雪径流等. 显然, 在一些大流域中, 选用一个概率分布函数描述径流的统计规律是不合理的. 第二, 人们通过各种修建水库、引水灌溉等水利工程进行水资源调控. 这种大规模的水事活动虽然对河川径流形成了一定程度的调控能力, 缓解或解决了流域水资源的供需矛盾, 但是, 也改变了河川径流的天然状态. 同时, 土地开发、水土保持等一系列大规模生产活动, 不同程度地改变了流域的下垫面条件, 气候变化也引起了河川径流情势的改变 (谢平等, 2009). 因此, 在受气候变化影响较大和人类水事活动频繁发生的流域, 径流序列在人类水事活动影响前后将不能认为是来自同一总体. 第三, 观测仪器分辨率的限制, 使得低于仪器最小测定限的数据不能被观测, 形成不完整数据序列 (删失数据序列), 应用完整数据序列分析方法推断这种不完整数据序列的统计规律是不合理的. 第四, 许多分布函数的取值范围为 $(-\infty,\infty)$ 或 $(0,\infty)$, (a,∞), 实际中, 水文取值可能是最小值与最大值之间, 不可能是无穷大值, 这种取值范围不符合水文值的实际取值范围, 其计算结果必然使计算值增大. 在工程实际中可近似地认为分布取值均在可能最小与最大之间, 其分布可能是由常规分布转化为截取分布.

水文频率计算不是一个崭新的研究领域, 涉及高等数学、概率论与数理统计、水文学、数值计算、优化计算等学科的交叉和渗透. 由于数据序列非一致性和删失数据序列存在, 其计算复杂度大大增加, 面临一系列亟待解决的科学问题. 2008 年 3 月至 2009 年 3 月, 作者在美国 Texas A & M 大学访问合作研究期间, 同国际著名学者 Vijay P. Sigh 教授开展了一些合作研究, 有机会阅读了一些国外文献. 回国后, 在参加几次国际学术会议期间, 有幸同一些国内外著名的水文专家交流学习, 先后在国家自然科学基金 (51479171、51579059、51409222、51179160、71273081) 相关课题项目和西北农林科技大学“农业高效用水与区域水安全”学科群支持下, 吸收了大量国外同行的理论和方法, 开展了水文频率计算步骤中第②—⑥步骤所涉及的原理和方法研究. 另外, 作者结合统计水文学、微积分、概率论与数理统计和数值计算等原理, 花费了较大的时间和精力, 系统地推导了目前单变量水文序列频率的有关计算公式, 其中一些计算原理和方法尚未见于中文文献, 同时更正了文献的一些印刷或其他方面的错误, 提出和建立了一些计算模型, 并给出了相应的计算机实现方法, 力求使高年级本科生和研究生掌握国外这些先进的原理和方法, 为他们从事水文频率分析和其他统计特征值计算提供参考. 由于篇幅限制, 本书仅列举了各类方法的典型应用实例, 使用较大篇幅阐述各类计算公式的来龙去脉.

全书由宋松柏、康艳、宋小燕和王小军统稿, 引用了研究生李宏伟、袁超、李扬、成静清、谢萍萍、张雨、李雪月、于艺、曾智、王剑峰、肖可以、赵丽娜、刘丹

丹、侯芸芸、刘斌、陈子全、马明卫、王红兰、郭成、原秀红、肖玲、王俊珍、殷建、牛林森、梁骏、魏婷、史黎翔、马晓晓、王誉杰、杨惠、赵明哲和王炳轩等计算实例. 在此向他们表示衷心的感谢!

本书感谢 Vijay P. Sigh 教授的悉心指导和鼓励, 也感谢西北农林科技大学水利与建筑工程学院、南京水利科学研究院水文水资源研究所、合肥工业大学土木与水利工程学院的大力支持. 书中参考了大量国内外学者的研究成果和文献, 大部分在书中和参考文献中列出, 在此一并致谢.

书中叙述了许多尚未见于中文文献的计算原理和方法, 其相关理论方法还在进一步研究和发展. 由于作者水平有限, 书中计算公式较多, 推导过程复杂, 虽经多次核对和修改, 难免存在一定的不足之处, 敬请有关专家学者和读者批评指正, 以利本书今后进一步修改和完善.

作　者
西北农林科技大学水利与建筑工程学院水文水资源研究所
南京水利科学研究院水文水资源研究所水资源配置与管理研究室
合肥工业大学土木与水利工程学院水资源与环境系统工程研究所
2017 年 2 月

目　录

第 1 章　水文序列经验频率计算方法

经验频率, 也称绘点位置 (plotting positions) 或秩次概率 (rank-order probability), 是根据样本按递减 (或递增) 顺序排列, 采用一定的计算方法估计样本每项值的频率, 这个估计频率值称为经验频率. 目前, 经验频率公式种类很多, 代表性的计算公式有 California (1923) 公式、Hazen (1930) 公式、Weibull (1939) 公式、Leivikov (1955) 公式、Blom (1958) 公式、Tukey (1962) 公式、Gringorten (1963) 公式、Cunnane (1978) 和 Hosking (1985) 公式. 我国《水利水电工程设计洪水规范》推荐采用 P-III 型曲线和图解适线法推求设计频率对应的设计洪水值. 在选定线型和适线准则的前提下, 经验频率是评价参数估计优劣的依据. 经验频率与理论频率偏差越小, 分布参数估计越好. 因此, 经验频率的计算尤为重要. 常用的适线法确定分布参数, 其计算精度主要取决于经验频率. 20 世纪 50 年代后, 我国学者在吸收国外研究的基础上, 在实际水文分析中, 广泛采用数学期望公式 (Weibull, 1939). 朱元甡、郭生练、金光炎等认为数学期望公式具有一定的理论基础, 与其他公式相比, 上端偏大, 下端偏小, 对工程设计偏于安全. 另外, 以不合适的绘点位置为基准, 评估各种适线准则的优劣, 或抽取 “理想” 样本, 进行各种参数估计方法的比较, 这些做法都是不太合理的. 因此, 本章首先详细叙述并推导目前国内外主要采用的经验频率公式.

1.1　次序统计量分布

目前, 已有大部分经验频率公式是基于次序统计量分布原理推导而来的 (郭生练和叶守泽, 1992). 本节参考一些次序统计量分布的文献 (金光炎, 1994, 2002, 2012; 胡宏达, 1991; 王善序, 1979, 1990; 李裕奇等, 2010), 推导有关计算公式, 叙述次序统计量分布的基本原理.

把样本 $(x_1, x_2, \cdots, x_n)$ 按由小到大的次序排列成递增顺序 $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(i)} \leqslant \cdots \leqslant x_{(n)}$ 时, $X_{(i)}$ 取值 $x_{(i)}$, 则有

$$X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(i)} \leqslant \cdots \leqslant X_{(n)} \tag{1}$$

统计量 $X_{(i)}$, $i = 1, 2, \cdots, n$, 称为次序统计量. 次序统计量 $X_{(i)}$ 的边际密度函数有以下三种推导方法.

1.1.1 由次序统计量的联合密度函数推导 $X_{(i)}$ 的密度函数

定理 1 设 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 是连续分布总体 $F(x)$、密度函数为 $f(x)$ 的样本次序统计量, 则 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 的联合密度函数为

$$g\left(x_{(1)}, x_{(2)}, \cdots, x_{(n)}\right) = n! \prod_{i=1}^{n} f\left(x_{(i)}\right) \tag{2}$$

式中, $-\infty < x_{(1)} < x_{(2)} < \cdots < x_{(n)} < \infty$.

定理 2 设 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 是连续分布总体 $F(x)$、密度函数为 $f(x)$ 的样本次序统计量, 则当 $x_{(i)} < x_{(j)}, 1 \leqslant i < j \leqslant n$ 时, $\left(X_{(i)}, X_{(j)}\right)$ 的二维边际密度函数为

$$\begin{aligned} g_{ij}\left(x_{(i)}, x_{(j)}\right) =& \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}\left[F\left(x_{(i)}\right)\right]^{i-1}\left[F\left(x_{(j)}\right) - F\left(x_{(i)}\right)\right]^{j-i-1} \\ & \cdot \left[1 - F\left(x_{(j)}\right)\right]^{n-j} f\left(x_{(i)}\right) f\left(x_{(j)}\right) \end{aligned} \tag{3}$$

以下以 3 维积分区域 (图 1) 为例, 证明定理 2.

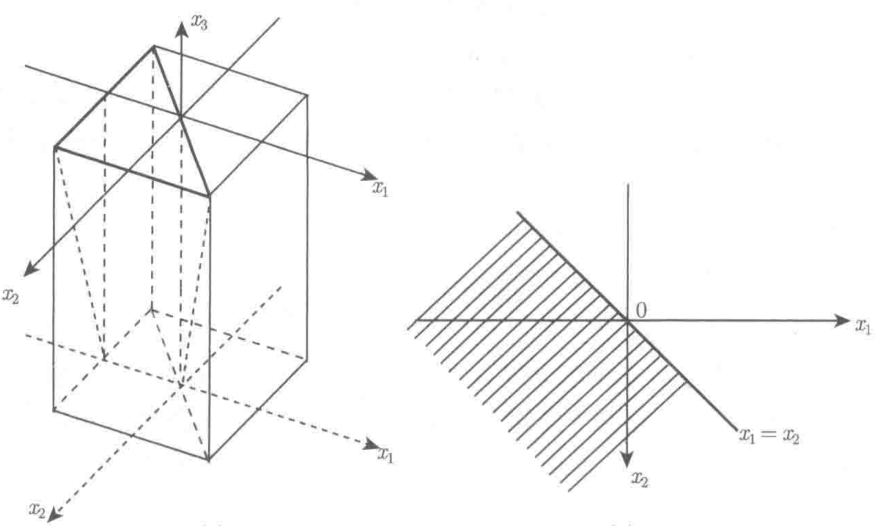

图 1 积分区域示意图

根据边际密度函数与多元变量联合密度函数的关系, 我们对多维空间 $D = \left(x_{(1)}, x_{(2)}, \cdots, x_{(i-1)}\right) \cup \left(x_{(i+1)}, x_{(i+2)}, \cdots, x_{(j-1)}\right) \cup \left(x_{(j+1)}, x_{(j+2)}, \cdots, x_{(n)}\right)$ 进行积分, 即可得到 $\left(X_{(i)}, X_{(j)}\right)$ 的联合密度函数 $g_{ij}\left(x_{(i)}, x_{(j)}\right)$. 当 $x_{(i)} < x_{(j)}$ 时, 有

$$
\begin{aligned}
&g_{ij}\left(x_{(i)},x_{(j)}\right)\\
=&n!f\left(x_{(i)}\right)f\left(x_{(j)}\right)\int\limits_{\left(x_{(1)},x_{(2)},\cdots,x_{(i-1)}\right)}f\left(x_{(1)}\right)f\left(x_{(2)}\right)\cdots f\left(x_{(i-1)}\right)dx_{(1)}dx_{(2)}\cdots dx_{(i-1)}\\
&\cdot\int\limits_{\left(x_{(i+1)},x_{(i+2)},\cdots,x_{(j-1)}\right)}f\left(x_{(i+1)}\right)f\left(x_{(i+2)}\right)\cdots f\left(x_{(j-1)}\right)dx_{(i+1)}dx_{(i+2)}\cdots dx_{(j-1)}\\
&\cdot\int\limits_{\left(x_{(j+1)},x_{(j+2)},\cdots,x_{(n)}\right)}f\left(x_{(j+1)}\right)f\left(x_{(j+2)}\right)f\left(x_{(n)}\right)dx_{(j+1)}dx_{(j+2)}\cdots dx_{(n)}
\end{aligned}\tag{4}
$$

对于式 (4) 中的积分 $I_1=\int\limits_{\left(x_{(1)},x_{(2)},\cdots,x_{(i-1)}\right)}f\left(x_{(1)}\right)f\left(x_{(2)}\right)\cdots f\left(x_{(i-1)}\right)dx_{(1)}\cdot dx_{(2)}\cdots dx_{(i-1)}$ 来说，$\left(x_{(1)},x_{(2)},\cdots,x_{(i-1)}\right)$ 满足条件 $-\infty<x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(i-2)}\leqslant x_{(i-1)}<x_{(i)}$，则 $i-1$ 重积分为

$$
\begin{aligned}
I_1=&\int\limits_{\left(x_{(1)},x_{(2)},\cdots,x_{(i-1)}\right)}f\left(x_{(1)}\right)f\left(x_{(2)}\right)\cdots f\left(x_{(i-1)}\right)dx_{(1)}dx_{(2)}\cdots dx_{(i-1)}\\
=&\int_{-\infty}^{x_{(i)}}\int_{-\infty}^{x_{(i-1)}}\cdots\int_{-\infty}^{x_{(3)}}\int_{-\infty}^{x_{(2)}}f\left(x_{(1)}\right)f\left(x_{(2)}\right)\\
&\cdots f\left(x_{(i-1)}\right)dx_{(1)}dx_{(2)}\cdots dx_{(i-2)}dx_{(i-1)}\\
=&\int_{-\infty}^{x_{(i)}}f\left(x_{(i-1)}\right)dx_{(i-1)}\int_{-\infty}^{x_{(i-1)}}f\left(x_{(i-2)}\right)dx_{(i-2)}\\
&\cdots\int_{-\infty}^{x_{(3)}}f\left(x_{(2)}\right)dx_{(2)}\int_{-\infty}^{x_{(2)}}f\left(x_{(1)}\right)dx_{(1)}\\
=&\int_{-\infty}^{x_{(i)}}f\left(x_{(i-1)}\right)dx_{(i-1)}\int_{-\infty}^{x_{(i-1)}}f\left(x_{(i-2)}\right)dx_{(i-2)}\\
&\cdots\int_{-\infty}^{x_{(3)}}f\left(x_{(2)}\right)dx_{(2)}\ F\left(x_{(1)}\right)\Big|_{-\infty}^{x_{(2)}}\\
=&\int_{-\infty}^{x_{(i)}}f\left(x_{(i-1)}\right)dx_{(i-1)}\int_{-\infty}^{x_{(i-1)}}f\left(x_{(i-2)}\right)dx_{(i-2)}\\
&\cdots\int_{-\infty}^{x_{(3)}}\left[F\left(x_{(2)}\right)\right]f\left(x_{(2)}\right)dx_{(2)}
\end{aligned}
$$