

基于高阶概率权重矩的广义极值分布参数估计

肖 玲, 宋松柏

(西北农林科技大学水利与建筑工程学院, 陕西 杨凌 712100)

摘 要:研究基于高阶概率权重矩的广义极值分布参数估计。根据高阶概率权重矩法原理,建立了广义极值分布高阶概率权重矩估算参数模型。以陕北地区4个水文测站的年最大洪峰流量序列为例,结果表明:高阶概率权重矩法能赋予大洪水值更多的权重。蒙特卡洛试验表明:适当提高阶数可以减小误差,但阶数过高反而会增大误差。

关键词:高阶概率权重矩;广义极值分布;参数估计;蒙特卡洛试验

中图分类号:P333

文献标识码:A

文章编号:1000-0852(2013)06-0001-05

1 前言

水文频率分析是水利工程规划设计的重要依据。一个测站的年最大洪峰流量序列设计值的估算往往需要选用一个合理的分布模型来进行拟合计算,然后,应用该分布模型进行重现期的洪水设计值估算。目前,国内外使用的单变量频率分布模型主要有P-分布类、 Γ 分布类、极值分布类、Wakeby分布类和Logistic分布类^[1]。近年来,许多学者采用广义极值分布(Generalized extreme value distribution,简记为GEV)理论,广泛地应用于水文气象极端事件研究领域^[2-3]。

频率分析统计推断的精度除了取决于样本和选用的分布模型外,模型参数估计方法对计算结果也具有较大的影响。参数估计的常用方法有极大似然(ML)、矩法(MOM)、线性矩(LM)、概率权重矩(PWM)、最小二乘(LS)和间隔最大积(MPS)等^[3]估计方法。20世纪90年代,Wang Q.J.博士提出了采用高阶概率权重矩洪水频率分布参数估计方法,取得了较好的应用效果。目前,我国还没有引入这一理论方法进行洪水频率分布参数估计。本文根据Wang Q.J.^[4]高阶概率权重矩方法,以陕北地区4个水文测站的年最大洪水序列为例,研究了高阶概率权重矩在推求GEV分布参数中的应用,并对GEV分布参数的高阶与低阶概率权重矩法进

行比较与评价,探索研究高阶概率权重矩在我国洪水频率分布参数估计的普适性,以期推求洪水频率分布参数估计提供合理的计算方法。

2 广义极值(GEV)分布

20世纪30年代,Fisher和Tippett^[5]证明极值具有渐进分布特性,概括了3类极值分布模型。

极值型(Gumbel)分布:

$$F_X(x) = P\{X < x\} = \exp[-\exp(\frac{x-\mu}{\sigma})], -\infty < x < +\infty \quad (1)$$

极值型(Fréchet)分布:

$$F_X(x) = P\{X < x\} = \begin{cases} \exp[-(\frac{x-\mu}{\sigma})^\xi], & x > \mu \\ 0, & x \leq \mu \end{cases}; \quad (2)$$

极值型(Weibull)分布:

$$F_X(x) = P\{X < x\} = \begin{cases} \exp[-(\frac{x-\mu}{\sigma})^\xi], & x < \mu \\ 1, & x \geq \mu \end{cases}; \quad (3)$$

式中: ξ, α, k 分别为分布的形状参数、位置参数和尺度参数。

Jenkinson^[6]、Coles^[7]根据极值分布理论,证明当极值的渐进分布存在且为非退化时,可以将以上三种类型的极值分布合并为一个统一的三参数极值分布函数,即广义极值分布(GEV)。

收稿日期:2012-07-30

基金项目:国家自然科学基金项目(51179160,50879070,50579065);高等学校博士学科点专项科研基金(20110204110017)

作者简介:肖玲(1987-),女,白族,贵州大方人,硕士研究生,从事水量转化与调控研究。E-mail: lindasun1988@163.com

通讯作者:宋松柏(1965-),男,陕西永寿人,博士,教授,博士生导师,从事水文水资源研究。E-mail: ssb6533@nwsuaf.edu.cn

设 x_1, \dots, x_m 为服从 GEV 分布的独立随机变量, 其分布函数表达式为:

$$F(x) = \begin{cases} \exp\left\{-\left[1-\frac{k}{\alpha}(x-\xi)\right]^{\frac{1}{k}}\right\}; & k \neq 0 \\ \exp\left\{-\exp\left[-\frac{1}{\alpha}(x-\xi)\right]\right\}; & k=0 \end{cases} \quad (4)$$

式中: ξ, α, k 分别为分布的形状参数、位置参数和尺度参数。当 $k \rightarrow 0$ 时为极值型, 即 Gumbel 分布; 当 $k < 0$ 时为极值型, 即 Fréchet 分布; 当 $k > 0$ 时为极值型, 即 Weibull 分布。

式(4)的逆函数形式为:

$$x(F) = \begin{cases} \xi + \frac{k}{\alpha} [1 - (-\ln F)^k]; & k \neq 0 \\ \xi - \alpha \ln(-\ln F); & k=0 \end{cases} \quad (5)$$

式中: ξ, α 和 k 意义同上。

3 高阶概率权重矩法

对于常用的参数估计方法(极大似然法、线性矩法^[8]和部分概率权重矩法^[9]), 或多或少存在一定的局限性。在使用低删失样本进行计算时, 小洪水段对分布函数的拟合影响不大, 即使考虑其影响, 最终拟合结果在很大程度上还是会依赖于大洪水段。

3.1 高阶概率权重矩定义

Greenwood(1979)等人^[10]首先提出概率权重矩法, 假定随机变量 x 的分布函数为 $F(x)$, 则概率权重矩的定义为:

$$M_{p,r,s} = \int_0^1 [x(F)]^p F^r (1-F)^s dF \quad (6)$$

式中: p, r, s 为实数; $x(F)$ 为分布函数的逆函数。当 $p=1, s=0$ 时, 式(6)变为:

$$\beta_r = M_{1,r,0} = \int_0^1 x(F) F^r dF \quad (7)$$

式中: $r=0, 1, 2, \dots$ 为概率权重矩的阶数。式(7)表明, 当概率权重矩的阶数增大时, 会赋予大值 x 较多的权重。

理论上讲, 只要分布函数已知, 就能求出其概率权重矩分布参数的解析表达式。给定一长度为 n 的升序排列样本 $x_1 \leq x_2 \leq \dots \leq x_n$, 高阶概率权重矩的无偏估计量可通过式(8)来计算:

$$b_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} x_i \quad (8)$$

3.2 广义极值分布的概率权重矩及参数表达式

根据式(5), Wang Q.J.^[4]推求了概率权重矩表达式。

(1) 当 $k \neq 0$ 时, GEV 分布的概率权重矩为:

$$(r+1)\beta_r = \xi + \frac{\alpha}{k} \left[1 - \frac{\Gamma(k+1)}{(r+1)^k} \right]; \quad k \neq 0 \quad (9)$$

式中: $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx$ 为 gamma 函数。

由式(4)可知, 要求解 GEV 分布函数中 ξ, α, k 三个参数, 需要列三个方程式进行求解。令概率权重矩的阶数 $r=\eta, \eta+1, \eta+2$, 并分别代入式(9), 有

$$(\eta+1)\beta_\eta = \xi + \frac{\alpha}{k} \left[1 - \frac{\Gamma(k+1)}{(\eta+1)^k} \right] \quad (10)$$

$$(\eta+2)\beta_{\eta+1} = \xi + \frac{\alpha}{k} \left[1 - \frac{\Gamma(k+1)}{(\eta+2)^k} \right] \quad (11)$$

$$(\eta+3)\beta_{\eta+2} = \xi + \frac{\alpha}{k} \left[1 - \frac{\Gamma(k+1)}{(\eta+3)^k} \right] \quad (12)$$

经过数学推导, 有

$$\frac{(\eta+2)\beta_{\eta+1} - (\eta+1)\beta_\eta}{(\eta+3)\beta_{\eta+2} - (\eta+1)\beta_\eta} = \frac{(\eta+1)^{-k} - (\eta+2)^{-k}}{(\eta+1)^{-k} - (\eta+3)^{-k}}; \quad k \neq 0 \quad (13)$$

(2) 当 $k=0$ 时, GEV 分布的概率权重矩为:

$$(r+1)\beta_r = \xi + \alpha[\varepsilon + \ln(r+1)]; \quad k=0 \quad (14)$$

式中, $\varepsilon=0.5772156649\dots$ 为欧拉常数。同理有

$$\frac{(\eta+2)\beta_{\eta+1} - (\eta+1)\beta_\eta}{(\eta+3)\beta_{\eta+2} - (\eta+1)\beta_\eta} = \frac{\ln(\eta+2) - \ln(\eta+1)}{\ln(\eta+3) - \ln(\eta+1)}; \quad k=0 \quad (15)$$

上述式中 GEV 分布的样本概率权重矩 $\beta_\eta, \beta_{\eta+1}, \beta_{\eta+2}$ 可用式(8)计算出的无偏估计量 $b_\eta, b_{\eta+1}, b_{\eta+2}$ 代替, 采用适当的公式可求出参数 k, ξ, α 。由于没有给出 k 的直接表达式, 采用常规的迭代法计算量很大, 因此, 可采用计算机数值算法来求解。

重写式(13)有:

$$\frac{(\eta+2)\beta_{\eta+1} - (\eta+1)\beta_\eta}{(\eta+3)\beta_{\eta+2} - (\eta+1)\beta_\eta} = \frac{\ln(\eta+2) - \ln(\eta+1)}{\ln(\eta+3) - \ln(\eta+1)} = \frac{(\eta+1)^{-k} - (\eta+2)^{-k}}{(\eta+1)^{-k} - (\eta+3)^{-k}} = \frac{\ln(\eta+2) - \ln(\eta+1)}{\ln(\eta+3) - \ln(\eta+1)}$$

令上式右边为

$$z = \frac{(\eta+1)^{-k} - (\eta+2)^{-k}}{(\eta+1)^{-k} - (\eta+3)^{-k}} = \frac{\ln(\eta+2) - \ln(\eta+1)}{\ln(\eta+3) - \ln(\eta+1)} \quad (16)$$

给定 η 值, 可采用二次函数拟合式估算 k 值:

$$k = a_0 + a_1 z + a_2 z^2 \quad (17)$$

设 $-0.5 \leq k \leq 0.5 (k \neq 0)$, 分别给定 $r=1, 2, 3, r=2, 3, 4, r=3, 4, 5, r=4, 5, 6$, 利用式(16)求出不同阶数对应的 z 值, 并对 z 和 k 的关系进行拟合, 得出不同阶数下式(17)中的系数 a_0, a_1, a_2 , 结果见表 1。

通过式(16)计算出不同阶数对应的 z 值后, 结合

表 1 不同高阶概率权重矩下式(17)中系数 a_0, a_1 和 a_2 的取值及绝对值偏差统计表

Table 1 Coefficients a_0, a_1 and a_2 for Formula (17) and the absolute error based on different higher probability weighted moments

阶数	a_0	a_1	a_2	绝对值偏差
$r=0,1,2$	0.0000	7.8514	2.9388	0.0001
$r=1,2,3$	0.0000	11.9038	2.7765	0.0001
$r=2,3,4$	0.0000	15.9285	2.7266	0.0001
$r=3,4,5$	0.0000	19.9431	2.7045	0.0001
$r=4,5,6$	0.0000	23.9527	2.6927	0.0001

表1中相应阶数下各系数值,代入式(17)中求出GEV分布的形状参数 k 。将形状参数 k 值代入式(10)和式(11)中,计算GEV分布的尺度参数 α 和位置参数 ξ 。

4 实例应用

为了验证高阶概率权重矩法在洪水频率分析中的应用,本文以陕北地区4个水文测站(志丹、绥德、交口河和刘家河)年最大洪峰序列为例,应用高阶概率权重矩法,进行GEV分布的参数估算研究。

选用GEV分布函数作为洪水频率分布模型,利用

前文中列出的公式,用高阶概率权重矩法对不同阶数下GEV分布参数进行估计。结果见表2。点绘出不同测站的年最大洪峰流量经验频率点据,由表2中计算出来的参数,代入式(1)中,得到各测站在不同阶数下GEV分布的分布值,并与经验点据进行对比,结果见图1~4。

表 2 基于高阶概率权重矩的4测站GEV分布参数估计结果统计表

Table 2 The estimated GEV distribution parameters of the four stations based on higher probability weighted moments

阶数		$r=0,1,2$	$r=1,2,3$	$r=2,3,4$	$r=3,4,5$	$r=4,5,6$
志丹	ξ	365.2542	366.0777	343.6914	289.6026	204.9424
	α	291.9050	289.8246	324.1867	389.1858	480.6571
	k	-0.3212	-0.3243	-0.2789	-0.2078	-0.1264
绥德	ξ	482.7455	463.0907	421.6762	365.9796	304.1374
	α	388.4998	474.0813	561.6622	652.7659	741.3789
	k	-0.1992	-0.0844	0.0054	0.0803	0.1406
交口河	ξ	540.5353	525.1451	482.2043	412.1700	329.5017
	α	434.1296	472.3046	532.0218	607.8082	685.6047
	k	-0.4300	-0.3962	-0.3520	-0.3045	-0.2626
刘家河	ξ	874.2642	853.4229	768.0253	605.5168	372.2801
	α	714.4633	769.6683	894.3189	1082.5823	1325.3544
	k	-0.3992	-0.3680	-0.3098	-0.2377	-0.1618

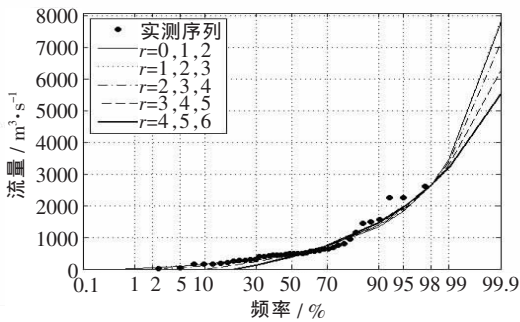


图 1 志丹站不同高阶概率权重矩下 GEV 分布曲线与年最大洪峰流量频率曲线拟合

Fig.1 Fitting of the GEV distribution of different higher PWMs to annual maximum flow in the Zhidan station

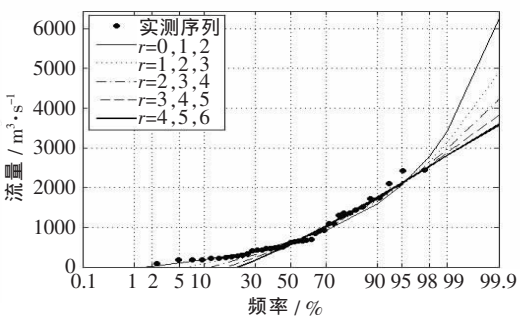


图 2 绥德站不同高阶概率权重矩下 GEV 分布曲线与年最大洪峰流量频率曲线拟合

Fig.2 Fitting of the GEV distribution of different higher PWMs to annual maximum flow in the Suide station

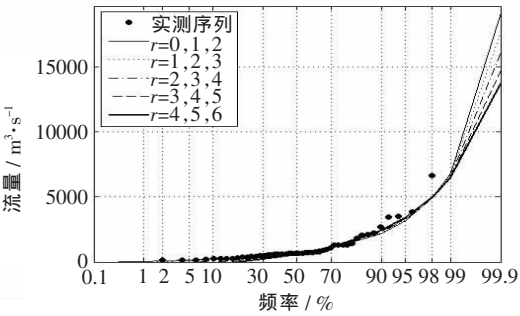


图 3 交口河站不同高阶概率权重矩下 GEV 分布曲线与年最大洪峰流量频率曲线拟合

Fig.3 Fitting of the GEV distribution of different higher PWMs to annual maximum flow in the Jiaokouhe station

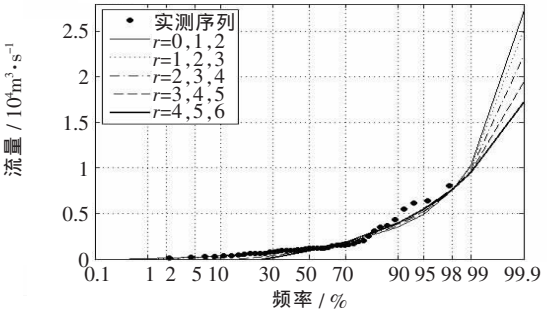


图 4 刘家河站不同高阶概率权重矩下 GEV 分布曲线与年最大洪峰流量频率曲线拟合

Fig.4 Fitting of the GEV distribution of different higher PWMs to annual maximum flow in the Liujiahe station

从图 1~4 可以看出,所选的水文测站均体现了运用高阶概率矩的优势,随着概率权重矩阶数的提高,其 GEV 分布曲线与年最大洪峰流量中的大洪水段点据拟合越好,而与小洪水段点据拟合较差。在水利工程规划设计中,较多的考虑大中洪水段对水利工程的影响,说明高阶概率权重矩法具有很大的优势。

5 蒙特卡洛试验(Monte Carlo experiments)

蒙特卡洛试验是对使用不同高阶概率权重矩法估计 GEV 分布参数时对不同重现期设计洪水值的影响评价。在实际工作中,对于已知的洪水资料,其分布函数是未知的,且常常异于所给出的已知分布。当选用一个已知分布模型进行计算时,需要对其适用性进行分析。同样,使用高阶概率权重矩法计算广义极值分布参数,需要对高阶与低阶进行分析比较,最终为分布模型

参数计算选出合适的方法^[4,10-11]。

本文利用具有更高灵活度的 Wakeby (5 参数)分布模型^[12],模拟 $N=100\ 000$ 组序列,每组序列的长度 $n=30$ 。利用模拟出来的序列计算不同高阶概率权重矩下 GEV 分布参数值,并分别计算分位数 $x(F=0.98)$ 和 $x(F=0.99)$ 的偏差(Bias)、标准误差(SE)和均方根误差(RMSE),结果如图 5~7 所示。

从图 5~7 可以看出,基于 4 类 Wakeby 分布模拟的数据,利用 GEV 分布函数估计出来的不同高阶概率权重矩下分位数 $x(F=0.98)$ 和 $x(F=0.99)$ 的 Bias、SE 和 RMSE 随着阶数的增大总体呈先减小后增大趋势。其中,Bias 的这种趋势特别显著。另外,基于 WA-4 和 WA-5 分布模拟样本计算的 SE 和 RMSE 较小。由此表明,使用高阶概率权重矩法估算广义极值分布的参数时,可以适当的提高阶数,但不能盲目的提高阶数,

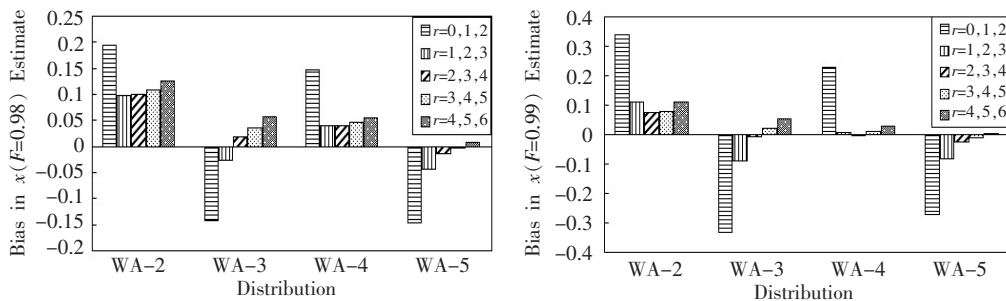


图 5 基于 Wakeby 分布样本的广义极值分布在 $x(F=0.98)$ 和 $x(F=0.99)$ 时估计值的偏差

Fig.5 Bias in $x(F=0.98)$ and $x(F=0.99)$ estimates from fitting the GEV distribution to the Wakeby samples

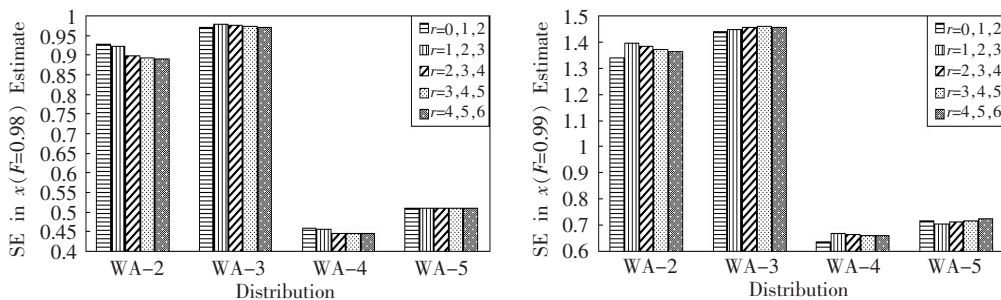


图 6 基于 Wakeby 分布样本的广义极值分布在 $x(F=0.98)$ 和 $x(F=0.99)$ 时估计值的标准误差

Fig.6 SE in $x(F=0.98)$ and $x(F=0.99)$ estimates from fitting the GEV distribution to the Wakeby samples

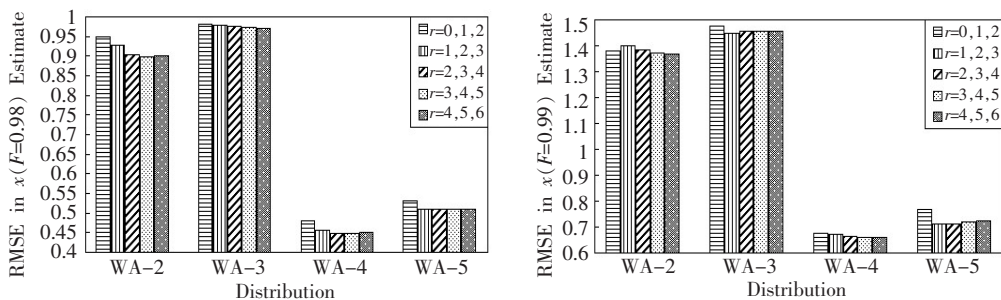


图 7 基于 Wakeby 分布样本的广义极值分布在 $x(F=0.98)$ 和 $x(F=0.99)$ 时估计值的均方根误差

Fig.7 RMSE in $x(F=0.98)$ and $x(F=0.99)$ estimates from fitting the GEV distribution to the Wakeby samples

当 $\eta > 3$ (即阶数 $r > 3, 4, 5$) 时, 误差会增大。

6 结论与展望

在水文频率分析计算中, 理论频率曲线是一条光滑的曲线, 但是实测点据常常出现两段或多段的分散区, 如果强行使用一条理论频率曲线来拟合所有的点据, 会产生复杂的组合频率曲线, 在估计重现期较大的洪水时会出现较大的偏差。因此可以通过赋予大中洪水段更多的权重, 从而得到更具有参考价值的大洪水设计值。

本文根据 Wang Q.J.^[4] 高阶概率权重矩法的 GEV 分布参数的计算公式, 通过对陕北地区 4 个水文测站的年最大洪峰流量序列进行分析, 结果表明, 随着概率权重矩阶数的提高, GEV 分布曲线与年最大洪峰流量中的大中洪水段点据拟合越好, 而与小洪水段点据拟合较差。另外, 本文进行了蒙特卡洛试验, 结果显示, 使用高阶概率权重矩法估算广义极值分布的参数时, 可以适当的提高阶数, 但当阶数越高时, 其相应的误差也增大。在实际应用中, 建议取值不超过 3。

与常用的低阶概率权重矩法相比, 高阶概率权重矩法并不复杂, 只要能写出分布函数的概率权重矩公式, 就能使用高阶概率权重矩法估计模型参数。另外, 高阶概率权重矩法进行我国普遍选用 P-III 型分布参数估计的应用仍需进一步研究。

参考文献:

- [1] 李宏伟. 水文频率参数计算方法与应用研究[D]. 杨凌: 西北农林科技大学, 2009. (LI Hongwei. Methods and Application of Estimating Parameters in Hydrologic Frequency[D]. Yangling: Northwest A & F University, 2009. (in Chinese))
- [2] 金光炎. 广义极值分布及其在水文中的应用[J]. 水文, 1998, (2): 9 - 15. (JIN Guangyan. General extreme value distribution and its applica-

tion to hydrology [J]. Journal of China Hydrology, 1998, (2): 9 - 15. (in Chinese))

- [3] 陈子燊, 刘曾美, 陆剑飞. 广义极值分布参数估计方法的对比研究[J]. 中山大学学报(自然科学版), 2010, 49(6): 105 - 109. (CHEN Zishen, LIU Zengmei, LU Jianfei. Comparative analysis of parameter estimation methods of general extreme value distribution [J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2010, 49(6): 105 - 109. (in Chinese))
- [4] Wang, Q.J.. Using higher probability weighted moments for flood frequency analysis [J]. Journal of Hydrology, 1997, 194(1): 95 - 106.
- [5] Fisher R. A., Tippett L.H.C.. Limiting forms of the frequency distribution of the largest or smallest member of a sample [J]. Mathematical Proceedings of the Cambridge Philosophical Society, 1928, 24(2): 180 - 190.
- [6] Jenkinson A. F.. The frequency distribution of the annual maximum (or minimum) values of meteorological elements [J]. The Quarterly Journal of the Royal Meteorological Society, 1955, 81(348): 158 - 171.
- [7] Coles S.. An Introduction to Statistical Modeling of Extreme Values [M]. New York: Springer Verlag, 2001.
- [8] Hosking, J.R.M.. L-moments: Analysis and estimation of distributions using linear combinations of order statistics [J]. Journal of Royal Statistical Society (Series B - Methodological), 1990, 52(1): 105 - 124.
- [9] Wang, Q.J.. Using partial probability weighted moments to fit the extreme value distributions to censored samples [J]. Water Resources Research, 1996, 32(6): 1767 - 1771.
- [10] Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R.. Probability weighted moments: definition and relation to parameters of distribution expressible in inverse form [J]. Water Resources Research, 1979, 15(5): 1049 - 1054.
- [11] Wang, Q.J.. Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution [J]. Journal of Hydrology, 1990, 120(1-4): 115 - 124.
- [12] Landwehr, J.M., Matalas, N.C.. Estimation of parameters and quantiles of Wakeby distributions 2. Unknown lower bounds [J]. Water Resources Research, 1979, 15(6): 1373 - 1379.

Parameter Estimation of Generalized Extreme Value Distribution Based on Higher Probability Weighted Moments

XIAO Ling, SONG Songbai

(College of Water Resources and Architectural Engineering, Northwest A & F University, Yangling 712100, China)

Abstract: This paper made research on parameter estimation of generalized extreme value distribution based on higher probability weighted moments. The model of higher probability weighted moments for estimating parameters of generalized extreme value distribution was established by the principle of higher probability weighted moments. The results of examples of annual maximum flow series in four stations indicate that the higher PWMs gives more weight to large flood values. The Monte Carlo experiments indicate that it is useful to raise the order appropriately.

Key words: higher probability weighted moments; generalized extreme value distribution; parameter estimation; Monte Carlo experiment