

Estimation of GEV Distribution Parameters Using Partial Probability Weighted Moments*

Xiuhong Yuan, Songbai Song[#]

College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling
Email: yuanxiuhong911@126.com, [#]ssb6533@nwsuaf.edu.cn

Received: Jun. 18th, 2012; revised: Jul. 3rd, 2012; accepted: Jul. 12th, 2012

Abstract: In order to reduce the uncertainty of the design flood estimation with larger return period, the principles and methods of Partial probability weighted moments (PPWM) were used to estimate parameters of Generalized extreme distribution (GEV). MATLAB programs were used to achieve the parameters' numerical solution of the GEV from PPWM on the basis of previous studies. Monte Carlo experiments were performed to assess the statistical properties of parameters estimation of the GEV distribution by the method of PPWM from different lower bound censored samples. The research results indicate that the method of PPWM from lower bound censored samples still has good effectiveness in high quantile estimation.

Keywords: Partial Probability Weighted Moments; Generalized Extreme Value Distribution; Monte Carlo Experiments; Parameter Estimating

基于部分概率权重矩的洪水频率参数估计方法*

原秀红, 宋松柏[#]

西北农林科技大学, 水利与建筑工程学院, 杨凌
Email: yuanxiuhong911@126.com, [#]ssb6533@nwsuaf.edu.cn

收稿日期: 2012年6月18日; 修回日期: 2012年7月3日; 录用日期: 2012年7月12日

摘要: 在洪水频率分析中, 为了避免小洪水估算大重现期设计洪水值呈现出的滋扰行为, 本文应用部分概率权重矩进行洪水频率分布参数估计的原理和方法, 探索部分概率权重矩(PPWM)在广义极值分布(GEV)参数估计中的应用。采用 MATLAB 编程进行基于部分概率权重矩的广义极值分布参数数值求解。通过 Monte Carlo 试验, 研究了不同低删失样本的部分概率权重矩法估计量的统计特性, 并获得了相应的统计试验结果。结果表明, 低删失样本的部分概率权重矩法在高分位数估计方面呈现出良好的有效性, 可以应用于以推求大重现期设计洪水为目的的洪水频率分析。

关键词: 部分概率权重矩; 广义极值分布; Monte Carlo 试验; 参数估计

1. 引言

洪水频率分析是水利水电工程规划设计与管理

的重要依据。Q. J. Wang(1990)博士指出, 由于小洪水与大洪水的发生几乎没有关系, 用包含小洪水值的系列进行大重现期设计洪水值估算时, 这类小洪水值往往呈现出滋扰行为。Q. J. Wang^[1]根据部分概率权重矩法(PPWM)的原理, 利用低删失样本进行洪水频率分布参数估计, 取得了较好的应用效果。

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

[#]通讯作者。

作者简介: 原秀红(1989-), 女, 山西河津人, 在读硕士生, 主要从事流域水文模拟及水文预报研究。

概率权重矩法(PWM)是 Greenwood 等人(1979年)提出的一种新型参数估计方法。国外大量的研究表明, 该法在水文概率分布中都具有较好的参数估计特性。丁晶、宋德敦等也先后将概率权重矩法应用于 P-III型分布、对数正态分布、指数 Γ 分布、正态分布等线型^[2-4], 采用数值计算的方法, 将概率权重矩法的应用范围从分布函数的反函数能解析表达推广到分布函数的反函数不能解析表达的 P-III型分布^[2]。李松仕、林洪孝等提出了求解概率权重矩的解析方法, 使概率权重矩法在参数估计中的应用得到了进一步的扩展^[5]。目前, 我国还没有引入 PPWM 方法进行洪水频率分布参数估计。本文应用积分原理和特殊函数, 研究广义极值分布(GEV)的部分概率权重矩(PPWM)以及低删失样本部分概率权重矩的无偏估计计算应用。运用 MATLAB 平台, 进行低删失样本下基于部分概率权重矩法的广义极值分布参数的数值计算, 编写基于部分概率权重矩的广义极值分布模拟计算程序, 并通过 Monte Carlo^[1,6,7]试验研究部分概率权重矩估计量的统计特性, 进一步评价在不同低删失样本下该方法的优劣和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. 部分概率权重矩(PPWM)及其无偏估计

Wang(1990)扩充了概率权重矩(PWM)^[1]的概念, 提出了用部分概率权重矩法来估计删失样本下广义极值分布参数估计的方法。对于低删失, 部分概率权重矩(PPWM)有下式

$$M'_{p,r,s} = \int_{F_0}^1 [x(F)]^p F^r (1-F)^s dF \quad (1)$$

($F_0 = F(x_0)$, x_0 为低删失门限值)

当 $p=1, s=0$ 时, 有

$$\beta'_r = M'_{1,r,0} = \int_{F_0}^1 x(F) F^r dF \quad (2)$$

给定一个长度为 n 的有序样本

$x_1 \leq x_2 \leq x_3 \cdots \leq x_n$, β'_r 的无偏估计值 b'_r 为

$$b'_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)\cdots(i-r)}{(n-1)(n-2)\cdots(n-r)} x_i^* \quad (3)$$

其中 $x_i^* = \begin{cases} 0; & x_i \leq x_0 \\ x_i; & x_i > x_0 \end{cases}$

3. 基于部分概率权重矩(PPWM)的广义极值分布(GEV)参数数值求解

3.1. 广义极值分布(GEV)的分布函数为

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

式中, ξ 为分布函数的位置参数, k 为形状参数, α 为尺度参数。

其分布函数的逆函数为

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

应用积分原理和特殊函数, Q. J. Wang(1990)推导出广义极值分布(GEV)的 PPWM 公式

$$\beta'_r = \left(\xi + \frac{\alpha}{k}\right) \frac{1}{r+1} (1-F_0^{r+1}) - \frac{\alpha}{k} \frac{\Gamma(1+k)}{(r+1)^{1+k}} P(1+k, -(r+1)\log F_0) \quad (6)$$

式中, $P(\cdot)$ 为不完全 gamma 函数, 即

$$P(1+k, -(r+1)\log F_0) = \int_0^{-(r+1)\log F_0} \frac{x^k e^{-x}}{\Gamma(1+k)} dx \quad (7)$$

3.2. 基于 PPWM 的 GEV 参数数值求解

1) 当 $F_0 = 0.0$ 时, Hosking(1985)^[8]针对部分概率权重矩进行广义极值分布参数估计进行了推导, 有公式

$$\hat{k} = 7.8590C^2 + 2.9554C^2 \quad (8)$$

式中: $C = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\ln 2}{\ln 3}$, b_0, b_1, b_2 分别为样本概率权重矩的无偏估计量。

参数 α, ξ 的估计公式为

$$\hat{\alpha} = \frac{(2b_1 - b_0)\hat{k}}{\Gamma(1+\hat{k})(1-2^{-\hat{k}})} = \frac{l_2 \hat{k}}{\Gamma(1+\hat{k})(1-2^{-\hat{k}})} \quad (9)$$

$$\hat{\xi} = b_0 + \frac{\hat{\alpha}}{\hat{k}} \left[\Gamma(1+\hat{k}) - 1\right] = l_1 + \frac{\hat{\alpha}}{\hat{k}} \left[\Gamma(1+\hat{k}) - 1\right] \quad (10)$$

2) 当 $F_0 \neq 0.0$ 时, 先确定低删失门限值 x_0 , 按顺序将有序样本中各个 $P_i = i/n+1$ 与 F_0 做比较, 确定第一个大于 F_0 的 P_i 对应的数值 x_j , 如果 $j=1$, 则有序样本中 x_1 即为删失门限值 x_0 ; 若 $j>1$, 则有序样本中 x_{j-1} 为删失门限值 x_0 。然后由 x_0 和给定的样本 x_i ,

确定删失样本 x_i^* 。

把 $r=0$ 、 $r=1$ 、 $r=2$ 代入式(6), 有

$$\beta'_0 = \left(\xi + \frac{\alpha}{k} \right) (1 - F_0) - \frac{\alpha}{k} \Gamma(1+k) P(1+k, -\log F_0) \quad (11)$$

$$\frac{2\beta'_1}{1-F_0^2} - \frac{\beta'_0}{1-F_0} = -\frac{\alpha\Gamma(1+k)}{k} \left[\frac{P(1+k, -2\log F_0)}{2^k(1-F_0^2)} - \frac{P(1+k, -\log F_0)}{1-F_0} \right] \quad (12)$$

$$\frac{2\beta'_1}{1-F_0^2} - \frac{\beta'_0}{1-F_0} = \frac{P(1+k, -2\log F_0)}{2^k(1-F_0^2)} - \frac{P(1+k, -\log F_0)}{1-F_0} = \frac{1}{z} \quad (13)$$

$$\frac{3\beta'_2}{1-F_0^3} - \frac{\beta'_0}{1-F_0} = \frac{P(1+k, -3\log F_0)}{3^k(1-F_0^3)} - \frac{P(1+k, -\log F_0)}{1-F_0} = \frac{1}{z}$$

结合式(13), 给定一组 F_0 值($F_0 = 0.0 - 0.5$)和一组 k 值($k = -0.5 - 0.5$), 应用 MATLAB 编程, $z-k$ 关系见图 1。

由图 1 可知, 当删门限值变大时, 曲线的斜率迅速降低了, 当 $F_0 = 0.5$, 曲线有非常小的斜坡。由于 k 值与对应的 z 值函数曲线具有光滑性, 所以, 对于给定的 F_0 , 采用近似回归曲线 $k = a_0 + a_1z + a_2z^2$, 在 MATLAB 环境下运行程序, 其参数计算见表 1。

在实际计算中, 由 $\hat{z} = \frac{3b'_2}{1-F_0^3} - \frac{b'_0}{1-F_0} - \frac{2b'_1}{1-F_0^2} + \frac{b'_0}{1-F_0}$ 计算出 \hat{z} 值

后, 再由 $\hat{k} = a_0 + a_1\hat{z} + a_2\hat{z}^2$ 得到参数 k 的估计量 \hat{k} 。

参数 ξ 、 α 可由式(14)、(15)得到其估计量 $\hat{\xi}$ 、 $\hat{\alpha}$

$$\hat{\xi} = \frac{b'_0}{1-F_0} + \frac{\hat{\alpha}}{\hat{k}} \left[\Gamma(1+\hat{k}) \frac{P(1+\hat{k}, -\log F_0)}{1-F_0} - 1 \right] \quad (14)$$

$$\hat{\alpha} = -\frac{\hat{k}}{\Gamma(1+\hat{k})} \left[\frac{2b'_1}{1-F_0^2} - \frac{b'_0}{1-F_0} - \frac{P(1+\hat{k}, -2\log F_0)}{2^k(1-F_0^2)} + \frac{P(1+\hat{k}, -\log F_0)}{1-F_0} \right] \quad (15)$$

4. Monte Carlo 试验

给定样本容量为 $n=30$, $\xi=0.0$, $\alpha=1.0$, $k=-0.2$ 时, 六种不同低删失样本情况下统计试验计算结果见表 2。

当 $F_0 = 0.0$ 时, Monte Carlo 试验中的概率权重矩就是普通概率权重矩, 以该估计量的统计特性作为标准。

1) 无偏性(Bias)从表 2 可以看出, 随着 F_0 的不断增大, 统计参数估计量的无偏性, 相比较 $F_0 = 0.0$ 的情况, 在 $F_0 = 0.1, 0.2$ 的情况下仍较好, 在 $F_0 = 0.3, 0.4, 0.5$ 的情况下较差。总体来看, 随着 F_0 值的增大, 统计参数估计量的无偏性趋于劣化, 但在适度的 F_0 取值下, 仍能保持较好的无偏性。

对于设计值估计量, 相比较 $F_0 = 0.0$ 时的无偏性, 在 $F_0 = 0.1, 0.2$ 时的无偏性仍较好, 在 $F_0 = 0.3, 0.4, 0.5$ 时无偏性较差。设计值 x_{100} 的估计量 \hat{x}_{100} 在 $F_0 = 0.1$ 时其无偏性较 $F_0 = 0.0$ 时更好, 设计值 \hat{x}_{200} 的估计量 \hat{x}_{200} 在 $F_0 = 0.2$ 时其无偏性较 $F_0 = 0.0$ 时略有改善。

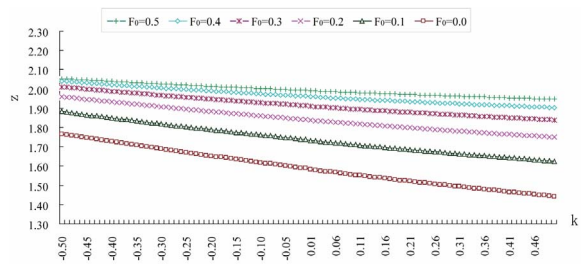


Figure 1. Relation curve between k and z in different F_0
图 1. 不同 F_0 下 k 、 z 关系曲线

Table 1. The parameters computational table of regression curve
表 1. 回归曲线参数计算表

F_0	α_0	α_1	α_2	R^2
0.5	133.1743	-123.9419	28.6582	0.9999
0.4	80.5019	-74.5182	17.0665	0.9999
0.3	50.7827	-47.0726	10.7286	0.9999
0.2	32.5646	-30.4975	6.9552	0.9999
0.1	20.6104	-19.7210	4.5207	0.9999
0.0	10.9480	-10.6606	2.3682	0.9999

Table 2. The achievement table of statistical tests on the method of PPWM from different lower bound censored samples
表 2. 不同低删失样本下部分概率权重矩法统计试验计算结果

评价指标	F_0	$\hat{\xi}$	$\hat{\alpha}$	\hat{k}	\hat{x}_{100}	\hat{x}_{200}
Bias	0.0	-0.0129	0.0025	-0.0365	0.0404	-0.1484
	0.1	-0.0030	-0.0198	-0.0402	-0.0197	-0.2505
	0.2	0.0150	-0.0481	-0.0508	0.0564	-0.1572
	0.3	-0.1438	0.3629	0.1976	-1.2209	-2.9633
	0.4	0.2004	-0.3505	-0.1523	0.3539	0.4048
	0.5	0.6469	-0.9943	-0.2966	0.5804	0.9036
SE	0.0	0.2120	0.1874	0.1634	0.4405	0.5402
	0.1	0.2097	0.2396	0.1870	0.4332	0.5326
	0.2	0.2089	0.3080	0.2090	0.4342	0.5420
	0.3	0.2183	0.2063	0.1549	0.4173	0.5610
	0.4	0.2595	0.6191	0.2929	0.4368	0.5417
	0.5	0.5351	1.3168	0.3905	0.4522	0.5532
RMSE	0.0	0.2124	0.1874	0.1674	0.4406	0.5405
	0.1	0.2097	0.2404	0.1913	0.4332	0.5333
	0.2	0.2095	0.3117	0.2151	0.4343	0.5422
	0.3	0.2615	0.4174	0.2511	0.4475	0.6431
	0.4	0.3279	0.7114	0.3302	0.4393	0.5434
	0.5	0.8395	1.6501	0.4903	0.4587	0.5614

2) 有效性(*SE*)统计参数估计量的有效性, 相比较 $F_0 = 0.0$ 的情况, 在 $F_0 = 0.1, 0.2, 0.3$ 时有效性仍较好, 在 $F_0 = 0.4, 0.5$ 时有效性较差。

设计值估计量的有效性, 相较 $F_0 = 0.0$ 的情况, 在 $F_0 = 0.1, 0.2, 0.3, 0.4, 0.5$ 的情况下仍较好, 且设计值 x_{100} 的估计量 \hat{x}_{100} 在 $F_0 = 0.1, 0.2, 0.3, 0.4$ 时有效性均较 $F_0 = 0.0$ 时更好, 设计值 \hat{x}_{200} 的估计量 \hat{x}_{200} 在 $F_0 = 0.1$ 时有效性较 $F_0 = 0.0$ 时有所改善。可见低删失样本下的 PPWM 法的统计参数估计量和设计值估计量具有较好的有效性。

3) 基于均方根误差(*RMSE*)的方法优劣评价随着 F_0 值的不断增加, 各统计参数估计量($\hat{\xi}$ 、 $\hat{\alpha}$ 和 \hat{k})

的 *RMSE* 值总体上在不断增大, 表明对于低删失样本, 部分概率权重矩法(PPWM)相比较概率权重矩法(PWM)在参数估计上较劣。

对于设计值估计量的 *RMSE* 值只是发生了微小的变化, 说明低删失样本的部分概率权重矩法(PPWM)法在设计值(高分位数)估计方面有较好的有效性。

5. 结论

低删失样本下部分概率权重矩法(PPWM)可避免在洪水频率分布参数估计中小洪水样本的滋扰行为。Monte Carlo 试验表明, 对于低删失样本, PPWM 在参数估计和设计值估计上的无偏性和有效性都较好, 而且适当的低删失门限值, 其统计特性相比较普通概率权重矩(PWM)更好。因此, 可以认为, 若一个年最大值系列中小值与高分位数的估计没有很大的关联(如年最大洪水峰量系列), 在以获取高分位数(大重现期的设计洪水)为目标的估计中, 应用低删失样本部分概率权重矩不会影响其估计的有效性。

参考文献 (References)

- [1] 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. *Journal of Hydrology*, 1990, 120(1-4): 115-124.
- [2] 丁晶, 宋德敦. 概率权重矩法及其在 P-III 分布中的应用[J]. *水利学报*, 1988, 3: 1-11. DING Jing, SONG Dedun. The application of probability weighted moments to estimating parameters for Pearson type-III distribution. *Journal of Hydraulic Engineering*, 1988, 3: 1-11. (in Chinese)
- [3] 秦大庸, 孙济良. 概率权重矩法在指数 Γ 分布中的应用[J]. *水利学报*, 1989, 11: 1-9. QIN Dayong, SUN Jiliang. The application of probability weighted moments to estimating parameters in exponential gamma distribution. *Journal of Hydraulic Engineering*, 1989, 11: 1-9. (in Chinese)
- [4] 丁晶, 杨荣富, 等. 概率权重矩法估计正态分布参数[J]. *成都科技大学学报*, 1991, 6: 35-70. DING Jing, YANG Rongfu, et al. Estimating parameters of normal distribution by the method of probability weighted moments. *Journal of Chengdu University of Science and Technology*, 1991, 6: 35-70. (in Chinese)
- [5] 李松仕. 关于线性矩法与概率权重矩法同解关系的分析研究[J]. *水文*, 2004, 24(3): 30-32. LI Songshi. Study on the relation of same solution between L moments and PWM. *Hydrology*, 2004, 24(3): 30-32. (in Chinese)
- [6] WANG, Q. J. Estimation of the GEV distribution from censored samples by method of partial probability weighted moments. *Journal of Hydrology*, 1990, 120(1-4): 103-114.
- [7] WANG, Q. J. Using partial probability weighted moments to fit the extreme value distributions to censored sample. *Water Re-*

- [8] source Research, 1996, 32(6): 1767-1771.
- HOSKING, J. R. M., WALLIS, J. R. and WOOD, E. F. Estimation of the generalized extreme value distribution by the method of probability weighted moments. Technometrics, 1985, 27(3): 251-261.