

具有历史洪水时 P-Ⅲ 分布线性矩法的研究

陈元芳¹ 沙志贵² 陈剑池² 陈 民³

(1. 河海大学水文水资源及环境学院 江苏 南京 210098; 2. 长江水利委员会水文局 湖北 武汉 430010;
3. 水利部天津勘测设计院 天津 300222)

摘要 :为了填补国内空白 ,基于美国学者 Hosking1990 年提出的线性矩法(L-moment) ,专门就线性矩法在 P-Ⅲ 分布下参数估计的算法 ,与传统参数估计方法在统计性能方面的差异进行介绍与分析计算 ,提出了具有历史洪水时该法的计算公式 .大量统计试验结果表明 ,线性矩法确实具有良好性能 ,较矩法好得多 ,与概率权重矩法(PWM)结果很接近 ,所提出的具有历史洪水情况下的线性矩公式是合理有效的 .

关键词 参数估计 线性矩 历史洪水 不偏性 有效性

中图分类号 :P333.9 **文献标识码** :A **文章编号** :1000-1980(2001)04-0076-05

1 问题的提出

参数估计方法是洪水频率计算中的重要内容 ,国内学者在此方面做了大量研究 ,取得一系列研究成果 .如丛树铮、谭维炎等^[1]首次在国内应用统计试验方法研究了不同的适线准则及绘点公式下适线法的统计性能 ,并与极大似然法及矩法作了比较 ,得出期望值绘点公式结合纵向离差绝对值准则的适线法(简称绝对值准则适线法)精度最高 .丁晶、宋德敦等^[2]在美国学者 Greenwood1979 年定义概率权重矩的基础上 ,提出了 P-Ⅲ 分布的参数估计新方法——概率权重矩法 ,经统计试验方法论证 ,该法具有良好性能 ,尤其是不偏性能特别好 .国内学者马秀峰^[3]在经过长期思考后 ,为克服矩法具有较大端矩差的缺点 ,于 1984 年提出了权函数方法 ,这是一种由国内学者独立提出的 P-Ⅲ 分布参数估计方法 ,因此 ,引起了国内学者的关注 ,刘光文、刘治中^[4]分别就该法的改进进行了探索 .陈元芳等^[5,6]提出了具有历史洪水时马秀峰权函数法估计新公式 ,经论证该公式是有效的、合理的 ,从而拓宽了该法的使用范围 .但由于马秀峰权函数法中的参数 C_V 仍用矩法估计 ,所以该法仍然比 PWM 及绝对值准则适线法略差 .

Hosking 在 Greenwood1979 年定义概率权重矩之后 ,于 1990 年定义了线性矩(L-moment) .这引起国外水文学者的高度兴趣 ,并作了一系列研究工作^[7~9] ,主要工作是在简单样本、不同总体分布下 ,其总体统计性能与矩法比较 ,线性矩法如何用于地区综合 线型鉴别等 .相比较 ,国内水文学者对此方法的研究相对迟缓多了 ,至今未看到在国内刊物上有此方面的文献发表 .为了尽快弥补我国在这方面与国际上的差距 ,促进我国水文频率计算水平的提高 ,本文就 P-Ⅲ 分布总体下 ,如何用线性矩对给定水文系列估计其分布参数作了介绍 ,并就其统计性能与现有估计方法利用统计试验方法作了比较 ,此外 ,还提出了具有历史洪水情况下的 P-Ⅲ 分布线性矩法估计公式 .经过统计试验论证 ,该公式是合理有效的 .

2 概率分布的线性矩(L-moment)

2.1 线性矩的定义

设随机变量为 X ,其取值为 x ,分布函数为 $F(x)$,密度函数 $f(x)$.

1979 年 ,Greenwood 定义的概率权重矩是

$$\alpha_r = \int_0^1 x(1 - F(x))f(x)dx \quad (1)$$

$$\beta_r = \int_0^1 x F(x) dF(x) \tag{2}$$

概率权重矩往往难于直接用于解释概率分布几种参数,如尺度、形状参数的含义.事实上,以上信息可以用概率权重矩线性组合来表示,如分布尺度参数可以用 $\alpha_0 - 2\alpha_1$ 或 $2\beta_1 - \beta_0$ 的倍数表示(Landwehr, Matalas, Hosking 等),而形状参数可用 $6\beta_2 - 6\beta_1 + \beta_0$ 表示(Stedinger).在此基础上, Hosking 等定义了线性矩 λ_r (L-Moment).

$$\lambda_r = \int_0^1 x P_{r-1}^*(F(x)) dF(x) \tag{3}$$

其中

$$P_r^*(u) = \sum_{k=0}^r \frac{(-1)^k (r+k)!}{(k!)(r-k)!} u^k$$

一般地,这种定义的线性矩与概率权重矩的关系如下:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^r \frac{(-1)^k (r+k)!}{(k!)(r-k)!} \alpha_k = \sum_{k=0}^r \frac{(-1)^k (r+k)!}{(k!)(r-k)!} \beta_k \tag{4}$$

具体地讲,两种矩的前四阶关系是

$$\begin{cases} \lambda_1 = \alpha_0 = \beta_0 \\ \lambda_2 = \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0 \\ \lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{cases} \tag{5}$$

从以上式子可知,线性矩是概率权重矩的线性组合,因此,它们之间应该具有非常密切的联系.

为了便于定义无因次的线性矩, Hosking 还提出了线性矩系数(L-Moment ratios):

$$\tau_r = \lambda_r / \lambda_2 \quad r = 3, 4 \tag{6}$$

此外,定义了反映尺度特征的系数 $L - C_V$, 即 τ_2 ,

$$\tau_2 = \lambda_2 / \lambda_1 \tag{7}$$

τ_3 反映形状(偏态)特征,故也称 L-skewness, τ_4 反映峰度特征,故称 L-kurtosis.

2.2 线性矩特性

- a. 存在性. 当分布期望值存在, 则各阶线性矩存在.
- b. 单一性. 当分布期望值存在, 则线性矩唯一, 即不存在两个分布有相同的线性矩.
- c. 数值范围. λ_1 可取任何值, λ_1 其实就是分布数学期望; $\lambda_2 \geq 0$, 对所有 $r \geq 3$, $|\tau_r| < 1$.

d. 线性性. 设随机变量 X, Y 为线性函数关系, 即 $Y = aX + b$, 则对应的两个随机变量线性矩 λ_r 及 λ_r^* 的关系如下:

$$\begin{aligned} \lambda_1^* &= a\lambda_1 + b, \lambda_2^* = |a|\lambda_2 \\ \tau_r^* &= (\text{Sign}(a))\tau_r \end{aligned}$$

其中 $r > 3$, Sign 为符号函数.

- e. 对称性. 若 X 是以均值 μ 为中心的对称分布, 则所有奇数阶线性矩系数, $\tau_r = 0$ ($r = 3, 5, 7, \dots$).

2.3 线性矩与次序统计量的关系

设样本为 $X_{i:n}$ ($i = 1, 2, \dots, n$), n 为样本容量, $X_{i:n}$ 是由小到大排列为第 i 个的样本值, 则可证明

$$\lambda_1 = E(X_{1:n}), \lambda_2 = \frac{1}{2} E(X_{2:n} - X_{1:n}), \lambda_3 = \frac{1}{3} E(X_{3:n} - 3X_{2:n} + X_{1:n}), \lambda_4 = \frac{1}{4} (X_{4:n} - 3X_{3:n} + 3X_{2:n} - X_{1:n})$$

3 线性矩与统计分布参数的关系

设 P-Ⅲ 分布密度函数如下:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} \quad (\alpha, \beta > 0; x > a_0) \tag{8}$$

由于 α, β, a_0 与线性矩关系复杂, 故 Hosking 等^[7]给出近似算法, 不过其误差可控制在 10^{-6} 以下, 因此具

有足够高的精度.

$$\lambda_1 = a_0 + \alpha/\beta \qquad \lambda_2 = \sqrt{\pi}\Gamma(\alpha + 0.5)/\Gamma(\alpha)/\beta$$

(9)

$$\tau_3 = \alpha^{-\frac{1}{2}} \frac{A_0 + A_1\alpha^{-1} + A_2\alpha^{-2} + A_3\alpha^{-3}}{1 + B_1\alpha^{-1} + B_2\alpha^{-2}} \qquad (\alpha \geq 1)$$

(10)

$$\tau_3 = \frac{1 + E_1\alpha + E_2\alpha^2 + E_3\alpha^3}{1 + F_1\alpha + F_2\alpha^2 + F_3\alpha^3} \qquad (\alpha < 1)$$

(11)

公式(10)(11)中常数 $A_0, A_1, A_2, A_3, B_1, B_2, E_1, E_2, E_3, F_1, F_2, F_3$ 取值分别是: 0.325 735 01, 0.168 691 50, 0.078 327 243, -0.002 912 053 9, 0.466 971 02, 0.242 554 06, 2.380 757 6, 1.593 179 2, 0.116 183 71, 5.153 329 9, 7.142 526 0, 1.974 505 6. τ_4 的近似公式见文献[7].

反之,当已知 $\lambda_1, \lambda_2, \tau_3$, 则 EX, C_V, C_S 的计算公式如下:

当 $|\tau_3| < 1/3$ 时, 则

$$Z = 3\pi\tau_3^2 \qquad \alpha = \frac{1 + 0.2906Z}{Z + 0.1882Z^2 + 0.0442Z^3}$$

(12)

当 $1/3 < |\tau_3| < 1$ 时, 则

$$Z = 1 - |\tau_3| \qquad \alpha = \frac{0.36067Z - 0.59567Z^2 + 0.25361Z^3}{1 - 2.78861Z + 2.56096Z^2 - 0.77045Z^3}$$

(13)

$$EX = \lambda_1 \qquad C_V = \lambda_2\pi^{1/2}\alpha^{1/2}\Gamma(\alpha)/\Gamma(\alpha + 1/2)$$

(14)

$$C_S = 2\alpha^{-\frac{1}{2}}\text{Sign}(\tau_3)$$

(15)

实际计算时, $\Gamma(\alpha)$ 采用数值分析近似计算, 由于 α 取值可能很大, 会导致由(14)式计算 C_V 出现死循环, 故笔者建议采用以下公式推求 $\Gamma(\alpha)/\Gamma(\alpha + 1/2)$ [α] 表示小于 α 的最大整数):

$$\Gamma(\alpha)/\Gamma(\alpha + \frac{1}{2}) = \frac{(\alpha - 1)}{(\alpha + \frac{1}{2} - 1)} \frac{(\alpha - 2)}{(\alpha + \frac{1}{2} - 2)} \cdots \frac{(\alpha - [\alpha])}{(\alpha + \frac{1}{2} - [\alpha])} \frac{\Gamma(\alpha - [\alpha])}{\Gamma(\alpha + \frac{1}{2} - [\alpha])}$$

4 样本线性矩的计算

4.1 简单样本时的样本线性矩公式

设样本为 $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$, 则线性矩 $\lambda_1, \lambda_2, \lambda_3$ 对应的样本矩 l_1, l_2, l_3 的计算公式由 Hosking 建议如下:

$$l_1 = b_0 \qquad l_2 = 2b_1 - b_0 \qquad l_3 = 6b_2 - 6b_1 + b_0 \qquad \tau_3 = l_3/l_2$$

(16)

$$b_0 = \frac{1}{n} \sum_{j=1}^n x_{j:n} \qquad b_1 = \frac{1}{n} \sum_{j=2}^n \frac{(j-1)}{(n-1)} x_{j:n} \qquad b_2 = \frac{1}{n} \sum_{j=3}^n \frac{(j-1)(j-2)}{(n-1)(n-2)} x_{j:n}$$

(17)

这样, 只要给定样本 $x_{1:n}, x_{2:n}, \cdots, x_{n:n}$, 就可估算出 $\lambda_1, \lambda_2, \lambda_3, \tau_3$.

4.2 具有历史洪水情况下样本线性矩计算公式的提出

实际工作中, 常遇到的是有历史洪水的水文系列, 然而国外至今未提出 P-Ⅲ 分布具有历史洪水信息样本线性矩公式, 这就大大限制了该公式在实际中的应用, 为此, 本文尝试提出能考虑历史洪水的样本线性矩公式.

考虑 P-Ⅲ 分布具有历史洪水样本矩的思路较为简单, 要充分利用前人的研究成果. 根据以上分析, 线性矩是概率权重矩的线性组合, 因此, 两者关系应该非常密切, 而 P-Ⅲ 分布在具有历史洪水时的样本概率权重矩, 已由丁晶等人^[2]提出, 且具有较高精度, 故可尝试直接采用.

设水文样本最大重现期为 N , 历史洪水个数为 a , 实测期历史洪水个数为 l , 实测期样本长度为 n , 且由小至大排列的样本为 $\{x'_m, m = 1, 2, \cdots, n-l+a\}$, 则计算公式为

$$b_0 = \frac{1}{N} \left[\frac{N-a}{n-l} \sum_{m=1}^{n-l} x'_m + \sum_{m=n-l+1}^{n-l+a} x'_m \right]$$

$$b_1 = \frac{1}{N} \left[\frac{N-a}{n-l} \sum_{m=1}^{n-l} \frac{m-1}{n-l-1} \frac{N-a-1}{N-1} x'_m + \sum_{m=n-l+1}^{n-l+a} \frac{N-n+l-a+m-1}{N-1} x'_m \right]$$
$$b_2 = \frac{1}{N} \left[\frac{N-a}{n-l} \sum_{m=1}^{n-l} \frac{(m-1)(m-2)}{(n-l-1)(n-l-2)} \frac{(N-a-1)(N-a-2)}{(N-1)(N-2)} x'_m + \sum_{m=n-l+1}^{n-l+a} \frac{(N-n+l-a+m-1)(N-n+l-a+m-2)}{(N-1)(N-2)} x'_m \right]$$

当 b_0, b_1, b_2 由样本估计后,由式(16)可求出 l_1, l_2, l_3, τ_3 这样得出的公式可望有较好的统计性质.

5 线性矩法的统计试验研究

线性矩法尤其是具有历史洪水情况下的该法统计性质如何,这是大家所关心的,需进行统计试验分析与计算才能弄清楚.因此,本文对线性矩法与现行常见的几种估计方法作了统计试验比较研究.

5.1 参数优劣评判标准

以参数及设计值的无偏性及有效性为标准,对于设计值采用相对偏差均值 B_{xp} 和相对根均方值 S_{xp} 表示,计算公式见文献[5]. P 取0.1%,1%两种.统计试验数 $N_s = 1000$ 次. B_{xp} 为正值表示偏大,为负值表示偏小; $|B_{xp}|$ 越大,偏差越大; S_{xp} 越小越有效.

5.2 试验方案设计

5.2.1 连序样本

实测序列长度 n 取30和50,总体参数 $EX_0 = 1.0, C_{v0}$ 取0.3,0.5,1.0, C_{s0}/C_{v0} 取3,4共12个方案.

5.2.2 非连序样本

最大重现期 N 取100, n 取30和50,历史洪水个数1,3,总体参数 $EX_0 = 1.0, C_{v0} = 0.3, 0.5, 1.0, C_{s0}/C_{v0}$ 取3,4倍,共24种方案.

5.3 试验结果分析

大量统计试验计算结果表明(部分结果列于表1):

a.在简单样本下,线性矩法比矩法在参数及设计值不偏性及有效性方面都有较大提高,线性矩法的计算结果与概率权重矩法的计算结果相当接近,总体上略优于概率权重矩法.与绝对值则适线法相比,其不偏性比适线法好,有效性时好时差.

表 1 参数、设计值不偏性有效性计算成果(部分方案)

总体参数与样本组成					估计方法	Table 1 Calculated results of unbiasedness and effectiveness of parameters and design flood (partial scheme)							
C_{v0}	C_{s0}	N	n	a		EC_e	EC_s	SC_e	SC_s	B_{xp1}	B_{xp2}	S_{xp1}	S_{xp2}
0.5	2.0	30	30	0	MOM	0.48	1.57	0.09	0.77	-5.41	-7.99	21.91	27.37
0.5	2.0	30	30	0	CFM	0.52	2.02	0.10	0.55	3.17	4.50	22.79	28.40
0.5	2.0	30	30	0	LM	0.50	1.99	0.09	0.55	1.22	2.02	22.38	27.80
0.5	2.0	30	30	0	PWM	0.50	1.99	0.09	0.56	1.12	1.91	22.42	27.86
0.5	1.5	30	30	0	MOM	0.49	1.23	0.08	0.66	-4.20	-5.80	19.02	24.47
0.5	1.5	30	30	0	CFM	0.52	1.49	0.09	0.55	2.18	3.12	19.71	25.17
0.5	1.5	30	30	0	LM	0.50	1.51	0.08	0.49	0.63	1.37	18.92	23.97
0.5	1.5	30	30	0	PWM	0.50	1.47	0.08	0.55	0.16	0.68	19.21	24.49
0.5	2.0	100	30	1	MOM	0.48	1.85	0.07	0.77	-3.90	-5.02	15.27	19.43
0.5	2.0	100	30	1	CFM	0.52	2.09	0.08	0.64	3.45	5.15	18.34	24.29
0.5	2.0	100	30	1	LM	0.50	1.99	0.08	0.49	0.22	0.64	17.21	21.28
0.5	2.0	100	30	1	PWM	0.50	1.99	0.08	0.50	0.13	0.54	17.28	21.31
1.0	4.0	100	50	3	MOM	0.97	3.36	0.17	1.23	-6.03	-9.13	23.84	29.68
1.0	4.0	100	50	3	CFM	1.05	4.23	0.20	0.89	5.31	8.15	25.65	32.52
1.0	4.0	100	50	3	LM	1.01	4.08	0.16	0.65	1.95	3.11	23.33	27.03
1.0	4.0	100	50	3	PWM	1.01	4.08	0.16	0.65	1.92	3.08	23.33	27.04

注: EC_e, EC_s 表示平均的样本 C_e, C_s 值; SC_e, SC_s 表示样本 C_e, C_s 均方误差; LM 表示线性矩法, CFM 表示适线法, MOM 表示矩法, PWM 表示概率权重矩法; $P1 = 1\%, P2 = 0.1\%$.

b.在有历史洪水情况下,线性矩法能保持在简单样本下的优良统计性能,结论与简单样本时一致.值得一提的是,线性矩法与概率权重矩几乎完全一致,其不偏性比适线法、矩法要好得更显著,有效性也好.总之,本文提出的有历史洪水下的线性矩公式看来是合理有效的.

6 结 论

a. 线性矩法是一种具有良好统计性能的参数估计方法,特别是具有良好的不偏性,多数情况下与 PWM 相当,在有些情况下还略优于 PWM 法.该法不仅可用于单站频率计算,其更大的优点是便于地区综合、线型鉴别等,有关这方面的研究尚待进一步深入分析.

b. 本文提出的具有历史洪水时样本线性矩的公式,从统计试验结果看是有效的,建议在实际中采用.

c. 本文对在 P-Ⅲ 分布下,线性矩法这一新的参数计算方法,就其性能等进行较为全面的分析与介绍,在国内看来还是第一次,希望能引起国内广大水文工作者的兴趣与重视,起到抛砖引玉之作用.

参考文献:

- [1] 丛树铮,谭维炎,黄守信,等.水文频率中参数估计方法的统计试验研究[J].水利学报,1980(3):1~11.
- [2] 丁晶,宋德敦,杨荣富.估计 P-Ⅲ 型分布新方法——概率权重矩法[J].成都科技大学学报,1980(2):12~18.
- [3] 马秀峰.计算水文频率参数的权函数法[J].水文,1980(3):1~11.
- [4] 刘治中.数值积分函数法推求 P-Ⅲ 型分布参数[J].水文,1987(7):11~14.
- [5] 陈元芳,侯玉.P-Ⅲ 型参数估计方法的研究[J].河海大学学报,1992,20(3):24~31.
- [6] 陈元芳,侯玉.现行几种权函数法的比较[A].见:梁瑞驹主编.2000 年中国水文展望研讨会文集[C].南京:河海大学出版社,1991.199~205.
- [7] Hosking J R M, Wallis J R. Regional frequency analysis—An approach based on L-moment[M]. London: Cambridge University Press, 1997. 1~280.
- [8] Pearson C P. New Zealand regional flood frequency analysis using L-moments[J]. Journal of Hydrology, 1991, 30: 53~64.
- [9] Wallis J R. Regional frequency studies using L-moments[A]. In: Young P C. In Concise Encyclopedia of Environmental Systems[C]. Pergamon, Oxford: 1993. 468~476.

Study on L-moment Estimation Method for P-Ⅲ Distribution with Historical Flood

CHEN Yuan-fang¹, SHA Zhi-gui², CHEN Jian-chi², CHEN Min³

(1. College of Water Resources and Environment, Hohai Univ., Nanjing 210098, China;

2. Yangtze River Commission, Wuhan 430010, China;

3. Investigation and Design Institute of Tianjin, Tianjin 300222, China)

Abstract: The L-moment method was proposed by Hosking in 1990. It is a new and good method for flood frequency analysis, on which many studies have been performed abroad. In this paper, the procedure of L-moment method for P-Ⅲ distribution is first introduced, and a formula of L-moment method for P-Ⅲ distribution with historical information is also proposed. A comparison with conventional estimation by the Monte-Carlo method shows that the L-moment method is of good statistical performance, that the result from it is almost the same as that from PWM, and that the new formula is effective.

Key words: parameter estimation method; L-moment; historical flood; unbiasedness; effectiveness