

一种修改的双权函数法

梁忠民¹, 张 华²

(1. 河海大学水文水资源及环境学院, 江苏 南京 210098;
2. 南昌水利水电高等专科学校, 江西 南昌 330046)

摘要 将两个负指数型函数之差所构成的函数类取作权函数. 在确定权函数的参数时, 主要考虑了两个因素: (a) 权函数与样本容量的关系, 以反映样本信息量大小在选择权函数中的作用; (b) 对本系列中排位在中上部的点据给予较大的权重. 对 P-Ⅲ型总体分布, 本文提出的权函数能使 σ 和 C_s 的估计由原方法的二阶加权中心矩降低为一阶加权中心矩估计. 统计试验结果表明, 修改后的双权函数法无偏性方面优于矩法, 有效性与矩法相当; 与 PWM 法的部分结果对比表明, 两者统计性能接近.

关键词 权函数法 样本矩 统计试验, 矩法 概率权重矩法 线性矩法

中图分类号 O212.1 **文献标识码** A **文章编号** 1000-1980(2001)06-0020-04

在水文频率分析计算中, 近年来线性矩法(L-矩)得到了广泛的研究和应用^[1]. L-矩法在数学上与概率权重矩(PWM)法等价, 但其更容易解释, 使用更方便. L-矩法(及 PWM 法)之所以具有优良的统计性能, 是因为 L-矩估计是样本次序统计量的线性组合. 与传统估计方法(如矩法)相比, 在 L-矩的估计中不包含样本值的平方、立方或更高次方的运算, 而仅是一阶样本矩的线性组合. 研究及应用表明, 这种在样本矩估计中的“降阶”作用是 L-矩法和 PWM 法获得成功的主要原因.

在研究 P-Ⅲ型分布的参数估计问题时, 马秀峰提出了权函数法^[2]. 从另外一个角度取得了这种“降阶”效果. 通过引入正态分布作为权函数, 使 C_s 的估计由原来矩法的三阶矩运算降低为一阶和二阶加权矩运算, 从而提高了 C_s 的估计精度. 刘光文在此基础上提出了双权函数法^[3], 旨在用权函数法同时解决 C_s 和 σ (或 C_v) 的估计问题, 但其方法并未能取得进一步“降阶”效果. 与 L-矩法及 PWM 法相比, 目前的权函数法仍需处理样本的二阶矩运算.

本文通过引入不同的权函数, 使 C_s 及 σ 的估计由原方法的二阶矩均变为一阶矩估计, 达到进一步“降阶”的目的.

1 权函数法(WF)

1.1 基本公式

令 $\Phi(x)$ 是一阶可导的连续函数, 对 P-Ⅲ型分布总体,

$$C_s = \frac{2}{\sigma} \frac{E[(x - \mu)\phi(x)] - \sigma^2 E[\phi'(x)]}{E[(x - \mu)\phi'(x)]} \tag{1}$$

式中 $E(\cdot)$ 代表数学期望运算. 若将 $\phi(x)$ 取成正态分布函数, 则式(1)为

$$C_s = -4\sigma \frac{H}{G} \tag{2}$$

其中

$$H = E[(x - \mu)\phi(x)] \qquad G = E[(x - \mu)^2\phi(x)]$$

显然, H, G 分别是以 $\Phi(x)$ 为权函数的一阶和二阶加权中心矩统计量, 因此 C_s 的样本估计已由矩法的三阶矩降低为由式(2)表示的二阶加权中心矩估计.

1.2 修改的双权函数

本文将权函数取成两个负指数型函数的组合^①：

$$\phi(x) = c(e^{-\frac{x}{m_1k}} - e^{-\frac{x}{m_2k}}) \quad x > 0 \tag{3}$$

式中： m_1, m_2, k, c 为系数， $m_1 \neq m_2$ 。如果限定 $\phi(x)$ 为概率密度函数，则 $c = 1/[(m_1 - m_2)k]$ 。不失一般性，可设定 m_1, m_2 为正整数且 $m_1 > m_2, k > 0$ 。显然 $\phi(x)$ 为凸函数。将 $\phi(x)$ 代入式(1)整理得

$$C_s = \frac{2}{\sigma} \frac{U + \sigma^2 V}{W} \tag{4}$$

其中

$$U = E[(x - \mu)\phi(x)] \tag{5}$$

$$V = \frac{1}{m_1k}E[\phi(x)] - \frac{1}{m_1m_2k^2}E[e^{-\frac{x}{m_2k}}] \tag{6}$$

$$W = -\frac{1}{m_1k}E[(x - \mu)\phi(x)] + \frac{1}{m_1m_2k^2}E[(x - \mu)e^{-\frac{x}{m_2k}}] \tag{7}$$

对两个不同的 $\phi(x)$ 函数(对应两组不同的参数 (m_1, m_2, k))，由式(4)可得到两个对应的方程，其解为

$$\sigma^2 = -\frac{U_2}{V_2} \frac{\frac{W_1}{W_2} - \frac{U_1}{U_2}}{\frac{W_1}{W_2} - \frac{V_1}{V_2}} \tag{8}$$

$$C_s = \frac{2}{\sigma} (\frac{U_1}{W_1} + \sigma^2 \frac{V_1}{W_1}) \tag{9}$$

显然 σ^2 和 C_s 均是 $(x - \mu)^i (i = 0, 1)$ 的加权函数，所以，修改后的双权函数法，其估计为零阶和一阶样本矩运算，达到了进一步“降阶”的目的。

2 修改双权函数法样本估计的一般公式

令 $x_i (i = 1, 2, \cdots, n-l, n-l+1, \cdots, n-l+a)$ 代表一个由小到大排序的不连续水文极值序列(如洪水)。其中， a 为特大值个数， l 为实测系列中特大值个数($a \geq l \geq 0$)， n 为实测系列长度，并记 N 为样本的最大重现期。类似不连续样本矩法估计公式的推导过程，可以得到式(5)~(7)中 U, V 和 W 的样本估计如下：

$$\hat{U} = \frac{1}{N} [\frac{N-a}{n-l} \sum_{i=1}^{n-l} (x_i - \bar{x})\phi(x_i) + \sum_{i=n-l+1}^{n-l+a} (x_i - \bar{x})\phi(x_i)] \tag{10}$$

令

$$T_1 = E[\phi(x)] \quad T_2 = E[e^{-\frac{x}{m_2k}}]$$

则

$$\hat{T}_1 = \frac{1}{N} [\frac{N-a}{n-l} \sum_{i=1}^{n-l} \phi(x_i) + \sum_{i=n-l+1}^{n-l+a} \phi(x_i)] \tag{11}$$

$$\hat{T}_2 = \frac{1}{N} [\frac{N-a}{n-l} \sum_{i=1}^{n-l} e^{-\frac{x_i}{m_2k}} + \sum_{i=n-l+1}^{n-l+a} e^{-\frac{x_i}{m_2k}}] \tag{12}$$

$$\hat{V} = \frac{1}{m_1k} \hat{T}_1 - \frac{1}{m_1m_2k^2} \hat{T}_2 \tag{13}$$

$$\hat{W} = -\frac{1}{m_1k} \hat{U} + \frac{1}{m_1m_2k^2} \frac{1}{N} [\frac{N-a}{n-l} \sum_{i=1}^{n-l} (x_i - \bar{x})e^{-\frac{x_i}{m_2k}} + \sum_{i=n-l+1}^{n-l+a} (x_i - \bar{x})e^{-\frac{x_i}{m_2k}}] \tag{14}$$

对给定的两个 $\phi(x)$ ，由式(10)~(14)分别得到 U, V 和 W 的样本估计，再由式(8)和(9)得到 σ^2 (或 σ)和 C_s 的估计。

3 权函数 $\phi(x)$ 的确定

3.1 $\phi(x)$ 与样本信息量的关系

若将 m_1 取成样本容量，即 $m_1 = n' = n + a - l$ ，并分别取 $m_2 = 1$ 和 $m_2 = 2$ 则得到两个权函数

① 马秀峰等. 黄河中游暴雨洪水数学模型. 水利部黄河水利委员会水文局, 1998.

$$\phi_1(x) = \frac{1}{(n-1)k_1} (e^{-\frac{x}{n'k_1}} - e^{-\frac{x}{k_1}}) \qquad \phi_2(x) = \frac{1}{(n-2)k_2} (e^{-\frac{x}{n'k_2}} - e^{-\frac{x}{2k_2}})$$

显然, n' 越大, 则 $\phi_1(x), \phi_2(x)$ 越“矮胖”, 即各权函数值越接近; n' 大, 表明样本信息量大, 则权函数在估计中的作用应降低, 即应倾向于加等权重, 反之亦然.

另一方面, 相同样本信息量下, 应强调权函数的加权作用, 即更应分配不等权重. 显然对固定的 n' 和 k , 取 $m_2 = 1$ 和 $m_2 = 2$ 所构成的两个权函数是所有可选方案中 (即取 $m_2 = 1, 2, \dots, n' - 1$) 能达到这一效果的两个权函数.

3.2 权重分配

按稳健估计的观点, 对分布中心位置的点据应分配较大权重, 否则给予较小权重. 在 PWM 法的估计中, 样本量值越大, 权重越大, 从 PWM 法的估计效果看 (无偏及有效性), 这在估计中似乎是一种有利的加权方式. 权衡以上考虑, 本文对 $\phi_1(x)$ 和 $\phi_2(x)$ 限定如下: $\phi_1(x)$ 的峰值出现在其均值位置 (众数, 均值合一) 并使 $k_2 = k_1$. 通过推导, 两个权函数的众数 x_1 和 x_2 分别为

$$x_1 = \frac{n' \ln(n') k_1}{n' - 1} \qquad x_2 = 2(1 + \frac{1}{n' - 2})(1 - \frac{\ln(2)}{\ln(n')}) \mu$$

对容量 $n' = 20, 30, 40, 50$ 的样本 (水文频率分析中常遇的样本容量范围), x_2 分别为 $1.62\mu, 1.65\mu, 1.67\mu, 1.68\mu$. 故总体看, 所取的两个权函数对排位在中上部的点据给予较大权重.

4 统计试验研究

4.1 统计试验方案及比较准则

式 (3) 中, $x \in (0, +\infty)$, 故无法对标准化 P-III 分布进行统计试验研究. 不失一般性, 本文限定 P-III 型分布总体均值 $\mu = 1$. 总体的其它参数分别取 $C_v = 0.4, 0.5, 0.6, 0.8, 1.0$; $C_s/C_v = 2, 3, 4, 5$; $N = 100, 150, 200, 250, 300, 500$; $a = 0, 1, 2, 3, 5$; $n = 20, 30, 40, 50$. 统计试验组数取 1000. 采用相对偏差 ($B\theta$) 和相对均方误差 ($S\theta$), 分别比较估计的无偏性和有效性.

4.2 结果及分析

限于篇幅, 图 1~2 仅显示了 WF 法、PWM 法与矩法估计 X_p ($p = 0.1\%$) 的部分结果. 其中, PWM 法的结果取自文献 [4], WF 法和矩法的结果是取本次 $C_v = 1.0$ 的计算方案. 图 1 是三种方法的 BX_p 和 SX_p 随 C_s 的变化关系. 对比可知, 矩法的负偏估计随 C_s 的增加而增加, WF 与 PWM 法提供的是较小的正偏或负偏估计; 但随着 C_s 的增加, 三种方法估计的有效性都有所降低 (即 SX_p 增加). 图 2 是 BX_p 和 SX_p 随样本容量 (n') 的变化关系, 结果再次显示矩法有不可忽视的负偏估计. 虽然通常矩法会有略优的有效性, 但 PWM 和 WF 法的有效性随着样本信息的增加将明显得到改善.

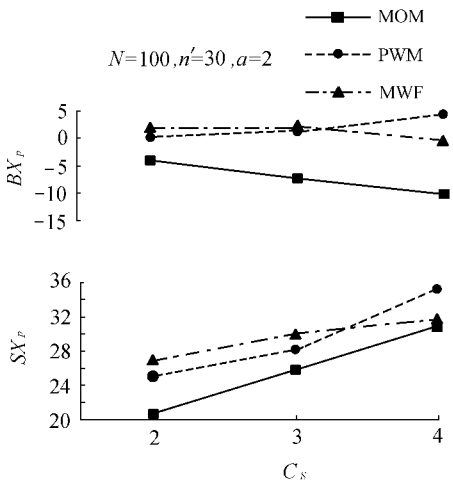


图 1 BX_p, SX_p 随 C_s 的变化关系

Fig.1 Variations of BX_p and SX_p with C_s

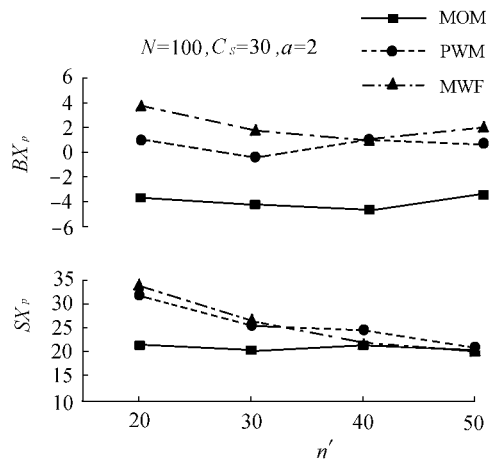


图 2 BX_p, SX_p 随样本容量 n' 的变化关系

Fig.2 Variations of BX_p and SX_p with n'

5 结 论

- a. 本文通过选取两个负指数型的组合函数作为权函数,对双权函数法作了修改,结果使其对 C_s 和 σ (或 C_v) 的估计均‘降阶’为样本(加权)一阶矩估计. 对 P-Ⅲ 型分布总体,修改后的 WF 法取得了与 L-矩法(或 PWM 法)相类似的理论结果.
- b. 权函数的选择主要考虑了两点,一是将样本容量直接作为权函数中的一个参数,以反映样本信息量大小与权函数加权作用之间的关系;二是对中上部点据予以较大权重,兼顾考虑估计的稳健性和其它统计性能如无偏性及有效性等.
- c. 统计试验结果表明,与矩法相比,修改后的 WF 法具有较优的无偏性,其有效性总体上与矩法较为接近;与 PWM 法部分结果的对比表明,无论无偏性还是有效性,两种方法的估计效果较为一致.

参考文献:

[1] Hosking J R M, Wallis J R. Regional frequency analysis[M]. Cambridge: Cambridge University Press, 1997. 14 ~ 141.

[2] 马秀峰. 计算水文频率参数的权函数法[J]. 水文, 1984, 21 (3): 1 ~ 8.

[3] 刘光文. 皮尔逊Ⅲ型分布参数估计[J]. 水文, 1990, 25 (4): 1 ~ 15.

[4] Ding J, Song D, Yang R. Further research on application of probability weighted moments in estimating the parameters of the Pearson type three distribution[J]. Water Resources Research, 1989, 110: 239 ~ 257.

A Modified Double-Weighted Function Method

LIANG Zhong-min¹, ZHANG Hua²

- (1. College of Water Resources and Environment, Hohai Univ., Nanjing 210098, China ;
2. College of Water Resources and Hydropower, Nanchang 230046, China)

Abstract : In this study, a weight function is constructed by the combination of two negative exponential functions, and the parameter determination for the weight function is based on two factors. One is the quantity of sample information ; the other is the distribution of more weight to the data at the middle and upper positions in a ranked sample. For the general P-III distribution, the present function may help simplify the estimation of the Coefficient of Variation (CV) and the Coefficient of Skewness (CS) from the second-order and the third-order moments of the conventional method to the first-order moment. The Monte Carlo simulation shows that the present method is better than the Method of Moment (MOM) in unbiasedness, and is generally comparable to MOM in efficiency. A comparison of the present method with the Probability Weighted Moments (PWM) indicates that the two methods are comparable to each other both in unbiasedness and in efficiency for the studied cases.

Key words : weighted function method ; sample moment ; statistical test ; method of moment ; probability weighted moment ; linear moment