

水文频率分析中的多项式正态变换方法研究

梁忠民 戴昌军

(河海大学水资源环境学院,江苏 南京 210098)

摘要 采用多项式正态变换(PNT)方法,把偏态分布转换为正态分布,推求指定频率下正态分布的设计值(分位数),然后通过一一变换转换成原偏态分布对应频率的设计值.以 P-Ⅲ 型分布为例,用 Monte-Carlo 法进行模拟分析,并对 PNT 法的统计性能与线性矩(L-M)法的统计性能进行了比较.结果表明,PNT 法的精度略差于 L-M 法的精度.由于 PNT 法本质上是非参数估计方法,不需要假设水文变量的分布线型,因此该法理论上是一种较为可行的水文极值分析方法.

关键词 参数估计;线性矩法;多项式正态变换法;蒙特卡罗法;分位数;稳健性;污染样本

中图分类号 P333.9 **文献标识码** A **文章编号** 1000-198X(2004)04-0363-04

水文设计值的估计方法通常有参数方法和非参数方法之分.参数方法先假定总体分布形式,然后估计分布参数,再根据分布函数推求设计值.非参数方法不需要假定总体分布形式,就可直接根据样本估计设计值.

国内外对参数估计方法进行了大量研究.如丛树铮等^[1]对 MOM、ML、FIT 等方法进行了大量统计试验,认为在一般情况下绝对值适线准则结合 Weibull 绘点公式的适线法的精度最好;Greenwood 等^[2]于 1979 年提出了概率权重矩(PWM)法;Ding Jing 等^[3]将 PWM 法推广到了 P-Ⅲ 型分布中;马秀峰^[4]于 1984 年提出了通过减少端矩差来提高估计精度的权函数法;Hosking 等^[5]于 1990 年提出了线性矩(L-M)法. L-M 法理论上与 PWM 法等价,但更容易解释,并且应用方便,因而它是一种统计性能优良的参数估计方法.

1981 年, Tang 等首次将非参数统计方法应用于洪水频率分析中.此后人们对利用非参数统计方法推求水文变量设计值问题进行了深入研究^[6].由于非参数统计方法在统计推断时不需要假设样本的总体分布,因而往往具有优良的统计性能.特别地,与参数方法相比,该法稳健性更好.当用于分析计算的水文样本组成复杂时,非参数方法更加适用.

PNT 法在统计推断时先将样本变换成正态分布样本,再对正态分布进行估计,然后通过一一变换得到对原始分布的估计.因为该法不需要假设原始水文变量的分布线型,所以该法虽然具有参数方法的特点,但本质上属于广义的非参数方法.因此,本文针对 P-Ⅲ 型分布研究了 PNT 法在估计总体设计值时的统计性能.研究过程中分别考虑了有历史洪水加入和有污染样本两种情况的估计问题.同时,将 PNT 法与 L-M 法进行了对比,并提供了定量化的研究结果.

1 多项式正态变换(PNT)法

一个偏态分布的变量 X , 可通过三阶多项式进行正态变换^①:

$$X = a_0 + a_1 Z + a_2 Z^2 + a_3 Z^3 \quad (1)$$

式中: X ——偏态分布的变量; Z ——标准正态分布的变量; a_0, a_1, a_2, a_3 ——多项式系数.

式(1)在满足式

$$a_2^2 - 3a_1a_3 > 0 \quad a_3 > 0 \quad (2)$$

约束条件下是一一变换的.多项式系数 a_0, a_1, a_2, a_3 可用 L-M 法进行估计.

按照概率权重矩的定义,

$$\beta_m = \int_0^1 \chi(u) u^m du \quad (3)$$

将式(1)代入式(3),得

$$\beta_m(x) = a_0 C_{m,0} + a_1 C_{m,1} + a_2 C_{m,2} + a_3 C_{m,3} \quad (4)$$

$$C_{m,n} = \int_{-\infty}^{\infty} z^n \Phi^n(z) \varphi(z) dz$$

式中 $\Phi(z)$, $\varphi(z)$ 分别为正态分布的分布函数与概率密度函数.

对于给定的一组样本 $\beta_0, \beta_1, \beta_2, \beta_3$ 按下式进行估计:

$$\begin{cases} b_0 = \frac{1}{n} \sum_{j=1}^n x_{j:n} & b_1 = \frac{1}{n} \sum_{j=2}^n \frac{j-1}{n-1} x_{j:n} \\ b_2 = \frac{1}{n} \sum_{j=3}^n \frac{(j-1)(j-2)}{(n-1)(n-2)} x_{j:n} & b_3 = \frac{1}{n} \sum_{j=4}^n \frac{(j-1)(j-2)(j-3)}{(n-1)(n-2)(n-3)} x_{j:n} \end{cases} \quad (5)$$

估计出 $\beta_0, \beta_1, \beta_2, \beta_3$ 后,将其代入式(4)即可求出 a_0, a_1, a_2, a_3 . 当系数 a_0, a_1, a_2, a_3 已知时,即可利用两种分布变换后的分位点对应相等的原则给定一个正态分布的分位点(分位数),代入式(1)就可以推出偏态分布的对应设计值.

2 统计试验研究

2.1 试验方案设计

2.1.1 对比方案

针对非简单样本和污染样本,将 PNT 法与目前流行的 L-M 法^[5]进行对比,并用对比结果来考察 PNT 法的统计性能. 非简单样本情况下,式(4)参数 β 的估计采用文献[3]的研究结果. 根据污染样本的概念,本文考虑了未能准确估计历史洪水重现期时样本受到的“污染”:假设历史洪水的真实重现期为 NN ,而由于估计误差,实际计算重现期为 N . 加入污染样本主要是为了考察估计方法的稳健性. 当 $NN = N$ 时,污染样本就变成一般的历史洪水样本;当 $NN = N = n, a = l = 0$ 时(n 为连续样本容量, a 为特大值个数, l 为连续样本中特大值个数),污染样本就转换为简单样本. 本文在构造污染样本时考虑了 20% 的重现期误差.

2.1.2 总体参数及模拟组数的选取

总体参数 本文考虑了如下 8 种情况: $EX = 1000, C_v = 0.5, C_s/C_v = 3.4.5$; $EX = 1000, C_v = 0.8, C_s/C_v = 3.4.5$; $EX = 1000, C_v = 1.0, C_s/C_v = 3.4$.

模拟组数 $K = 1000$.

2.1.3 样本选取

非简单样本: $n = 30, a = 2, N = 100, 500, 1000$; $n = 50, a = 3, N = 100, 500, 1000$.

污染样本: $n = 30, a = 2, NN = 100, N = 80, 120$; $n = 30, a = 2, NN = 500, N = 400, 600$; $n = 30, a = 2, NN = 1000, N = 800, 1200$; $n = 50, a = 3, NN = 100, N = 80, 120$; $n = 50, a = 3, NN = 500, N = 400, 600$; $n = 50, a = 3, NN = 1000, N = 800, 1200$.

2.1.4 计算方案

将总体参数与样本组合构成 144 种计算方案.

2.2 比较准则

统计量(记为 θ)的比较准则为

$$\begin{cases} b_\theta = \left(\frac{1}{K} \sum_{i=1}^K \hat{\theta}_i / \theta_0 - 1 \right) \times 100\% \\ s_\theta = \sqrt{\left[\sum_{i=1}^K (\hat{\theta}_i - \theta_0)^2 \right] / K} / \theta_0 \times 100\% \end{cases} \quad (6)$$

本文就 L-M 法与 PNT 法相应概率的设计值(x_p)进行比较.

2.3 统计试验结果分析

共分析计算了 $p = 0.1\%, 0.2\%, 0.5\%, 1\%$ 4 种情况下的无偏性指标 b_{x_p} 与有效性指标 s_{x_p} .

2.3.1 非简单样本

样本数为 30 时 ,各种方案平均 b_{x_p} ,PNT 法为 1.9% ,L-M 法为 1.7% ;各种方案平均 s_{x_p} ,PNT 法为 27.6% ,L-M 法为 25.1% .当样本数增加到 50 时 ,各种方案平均 b_{x_p} ,PNT 法为 2.1% ,L-M 法为 0.8% ;各种方案 s_{x_p} ,PNT 法为 19.2% ,L-M 法为 20.9% .样本 30 和 50 时 ,所有方案平均 b_{x_p} ,PNT 法为 2.0% ,L-M 法为 1.3% ;所有方案平均 s_{x_p} ,PNT 法为 24.3% ,L-M 法为 22.1% .

分析表明 ,无偏性指标 ,PNT 法与 L-M 法很接近 ,均略呈正偏 ,从水文安全角度考虑 ,是有利的 ,有效性指标 ,以 L-M 法为优 ,但差别很小 .两种方法部分估计结果如表 1、图 1 和图 2 所示 .

表 1 统计试验成果(部分方案)

方案		总体参数与样本数					估计方法	$b_{x_{p1}}$	$b_{x_{p2}}$	$b_{x_{p3}}$	$b_{x_{p4}}$	$s_{x_{p1}}$	$s_{x_{p2}}$	$s_{x_{p3}}$	$s_{x_{p4}}$
		C_v	C_s	NN	N	n									
加入历史洪水时	0.5	1.5	100	100	30	2	L-M	0.0129	0.0112	0.0100	0.0085	18	17	16	14
							PNT	0.0185	0.0118	0.0036	0.0061	24	21	17	15
	0.5	1.5	500	500	50	3	L-M	0.0049	0.0036	0.0037	0.0029	14	14	12	11
							PNT	0.0166	0.0101	0.0023	0.0051	19	17	14	12
	0.8	4.0	500	500	50	3	L-M	0.0073	0.0171	0.0021	-0.0010	24	25	22	21
							PNT	0.0536	0.0320	0.0054	0.0030	24	23	21	20
	1.0	3.0	100	100	30	2	L-M	0.0192	0.0174	0.0135	0.0103	25	25	23	22
							PNT	0.0509	0.0329	0.0330	0.0330	30	28	25	23
污染样本时	0.5	1.5	100	80	30	2	L-M	0.0335	0.0305	0.0271	0.0237	17	16	15	14
							PNT	0.0571	0.0466	0.0333	0.0323	25	22	18	15
	0.5	1.5	500	600	50	3	L-M	-0.0004	-0.0012	-0.0007	-0.0010	15	14	13	12
							PNT	-0.0037	-0.0085	-0.0140	-0.0096	20	17	14	13
	0.8	1.0	500	600	50	3	L-M	0.0026	0.0124	-0.0019	-0.0047	24	25	22	21
							PNT	0.0177	-0.0018	-0.0253	-0.0256	24	23	21	20
	1.0	3.0	100	120	30	2	L-M	-0.0189	-0.0192	-0.205	-0.0212	26	25	24	23
							PNT	0.0056	-0.0098	-0.0296	-0.0293	28	26	24	22

注 :EX=1000 , $p1=0.1\%$, $p2=0.2\%$, $p3=0.5\%$, $p4=1\%$.

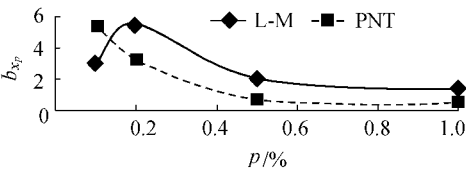


图 1 $N=100$, $n=30$, $a=2$, $C_v=0.8$,
 $C_s=3.2$ 时的无偏性指标

Fig.1 Non-bias index for $N=100$, $n=30$,
 $a=2$, $C_v=0.8$, and $C_s=3.2$

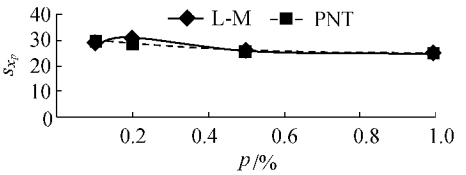


图 2 $N=100$, $n=30$, $a=2$, $C_v=0.8$,
 $C_s=3.2$ 时的有效性指标

Fig.2 Validity index for $N=100$, $n=30$,
 $a=2$, $C_v=0.8$, and $C_s=3.2$

2.3.2 污染样本

污染样本的情况分为两种 :一是计算重现期比实际重现期小 ;二是计算重现期比实际重现期大 .受篇幅限制 ,仅给出前者分析结果 .

计算重现期比实际重现期小 ,样本数为 30 时 ,各种方案平均 b_{x_p} ,PNT 法为 4.7% ,L-M 法为 4.3% ;各种方案平均 s_{x_p} ,PNT 法为 27.8% ,L-M 法为 24.8% .当样本数增加到 50 时 ,各种方案平均 b_{x_p} ,PNT 法为 5.6% ,L-M 法为 4.3% ;各种方案平均 s_{x_p} ,PNT 法为 21.4% ,L-M 法为 18.9% .样本 30 和 50 时 ,所有方案平均 b_{x_p} ,PNT 法为 5.1% ,L-M 法为 4.3% ;所有方案平均 s_{x_p} ,PNT 法为 24.6% ,L-M 法为 21.9% .

污染样本下 ,计算重现期小于实际重现期时 ,结论与非简单样本时一致 .两种方法部分估计结果如表 1、图 3 和图 4 所示 .

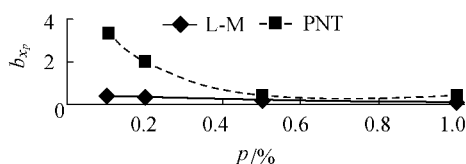


图3 $N = 1000$, $n = 50$, $a = 3$, $C_v = 0.5$, $C_s = 2.5$ 时的无偏性指标

Fig.3 Non-bias index for $N = 1000$, $n = 50$, $a = 3$, $C_v = 0.5$, and $C_s = 2.5$

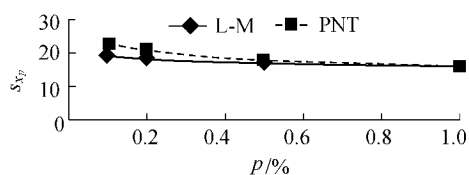


图4 $N = 1000$, $n = 50$, $a = 3$, $C_v = 0.5$, $C_s = 2.5$ 时的有效性指标

Fig.4 Validity index for $N = 1000$, $n = 50$, $a = 3$, $C_v = 0.5$, and $C_s = 2.5$

3 结 论

a. 对于所研究的有历史洪水加入的非简单样本和污染样本, PNT 法对总体设计值的估计, 无偏性和有效性与 L-M 法十分接近, 故 PNT 法也是一种性能优良的水文极值估计方法。

b. PNT 法采用变换技术可将任意分布的变量变换成正态分布变量。由于正态分布在水文学中应用广泛, 而且基于正态分布的各类处理方法已很成熟, 因此 PNT 法可望在水文学其他问题如误差分析、相关分析、多变量组合等方面发挥作用。但对于水文频率分析问题, 参数方法易使设计规范化。

参考文献:

- [1] 丛树铮, 谭维炎. 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] GREENWOOD J A, LANDWEHR J M, MATALAS N C, et al. Probability weighted moments: definition and relation to parameters of several distribution expressible in inverse form[J]. Water Resource Research, 1979, 15(5): 1049—1054.
- [3] DING Jing, YANG Rong-fu. The determination of probability weighted moments with the incorporation of extraordinary values into sample data and their application to estimating parameters for the Pearson type three distribution[J]. Journal of Hydrology, 1988, 101: 63—81.
- [4] 马秀峰. 计算水文频率参数的权函数法[J]. 水文, 1984 (3): 1—11.
- [5] HOSKING J R M, WALLIS J R. Regional frequency analysis—An approach based on L-Moments[M]. London: Cambridge University Press, 1997. 14—43.
- [6] ADAMOWSKI K. Nonparametric kernel estimation of flood frequencies[J]. Water Resource Research, 1985, 21(11): 1585—1590.

Polynomial normal transforms in flood frequency analysis

LIANG Zhong-min, DAI Chang-jun

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

Abstract: The polynomial normal transforms method (PNT) was applied to convert the skew distribution into the normal distribution. By derivation of the design value for the normal distribution under certain frequency, the value of the skew distribution under corresponding frequency was obtained through one-to-one transform. With the P-III distribution taken as an example, the statistical characteristics of the PNT method were investigated by means of the Monte-Carlo simulation. The comparison of the PNT method with the L-M method shows that the precision of the PNT method is compatible to that of the L-M method. Owing to the fact that the PNT method is a non-parametric estimation method in essence without requirement for an assumption of a statistical distribution for hydrological variables, the method is a potential method for flood frequency analysis when there is not adequate knowledge on the selection of the statistical model for the extreme value.

Key words: parameter estimation; linear moment method; polynomial normal transforms method; Monte-Carlo method; quantiles; robustness; contaminated sample