文章编号:1001 - 4179(2011)19 - 0021 - 03

对数正态分布参数估计的积分变换矩法应用

张 明[1], 柏 绍 光[2]

(1. 云南省水利厅,云南 昆明 650021; 2. 云南省水文水资源局 昆明分局,云南 昆明 650051)

摘要:针对对数正态分布参数估计矩法存在的缺陷,探讨了梅林变换的应用,推导了参数对应的一般矩和对数矩,提出了一种对数正态分布参数估计方法 - 积分变换矩法,包括二参数对数正态分布和三参数对数正态分布。以西南诸河 16 个水文站为例,利用三参数对数正态分布作为水文频率线型,对年平均流量系列进行了拟合。结果表明,与矩法相比,积分变换矩法能明显提高 C_s 的估算精度。

关 键 词:对数正态分布; 梅林变换; 积分变换矩法; 矩法

中图法分类号: TV12 **文献标志码**: A

在水文频率计算中,对数正态分布是国际上普遍使用的水文频率线型,也是我国广泛使用的水文频率线型。其参数估计方法在国内外已进行了不少研究,并提出了矩法(MOM)、极大似然法(ML)、熵矩估计法(POMEMOM)和线性矩法(L - M)等方法[1-4]。其中,矩法是一种既快速又简单的方法,但是估计 C_s 值时误差偏大;极大似然法较矩法估计精度要高,但是由于似然函数的不规则性,其值有时不甚理想;熵矩估计法与极大似然法效果相当;线性矩法具有许多良好的统计特性,尤其是无偏性和有效性,但是敏感性较差。张明等研究了梅林变换在 P—Ⅲ型分布、指数 Γ 分布等水文频率线型参数估计中的应用,提出了一种积分变换矩法,与矩法相比显著提高了估算精度[5-6]。鉴此,为克服对数正态分布参数估计传统方法的局限性,更有效地处理估算精度问题,本文探讨了梅林变换在对数正态分布中的应用,提出了一种对数正态分布参数估计的积分变换矩法,在估算中引入对数矩迭代求解以提高计算准确性。以三参数对数正态分布为水文频率线型,通过对西南诸河 16 个水文站的年平均流量系列拟合,对这种方法的估算精度与矩法作了分析对比。

1 对数正态分布的积分变换矩法

1.1 二参数对数正态分布的积分变换矩法

二参数对数正态分布密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right] \tag{1}$$

式中,$f(x)$ 为密度函数;σ 为形状参数,$\sigma>0$;μ 为比例参数,$\mu\in(-\infty,\infty)$。

对式(1)进行梅林变换为

$$\mathrm{M}[f;s] = \exp\left\{\frac{[\mu+(s-1)\sigma^2]^2-\mu^2}{2\sigma^2}\right\} \tag{2}$$

式中,$\mathrm{M}[f;s]$ 为 $f(x)$ 的梅林变换。

查梅林变换公式得:

$$E[x^\alpha] = \int_0^\infty f(x)x^\alpha\mathrm{d}x = \mathrm{M}[f;s]\big|_{s=1+\alpha} \tag{3}$$

$$E[\ln^\alpha x] = \int_0^\infty f(x)\ln^\alpha x\mathrm{d}x = \mathrm{M}^{(\alpha)}[f;s]\big|_{s=1} \tag{4}$$

式中,$E[*]$ 为均值;α 为实数,$\alpha=1,2,\cdots,n$;(α) 为 α 阶导数。

将式(2) 代入式(3) 和式(4),并取 $\alpha=1$ 得:

$$E[x] = \exp(\mu+\frac{\sigma^2}{2}) \tag{5}$$

$$E[\ln x] = \mu \tag{6}$$

由于式(1)仅有 2 个参数,联解式(5)、(6)即可求出。考虑水文实测资料系列分为简单样本系列和非简单样本系列两种情况,分别求解如下。

(1) 简单样本系列。对于简单样本系列 $x_i(i=1,2,\cdots,n)$, 则式(5)、(6)左边为

收稿日期:2011 - 06 - 09

作者简介:张 明,男,硕士,主要从事水文水资源管理工作。E - mail:ynzm2006@163.com

$$E[x] = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7}$$

$$E[\ln x] = \frac{1}{n}\sum_{i=1}^{n} \ln x_i \tag{8}$$

（2）非简单样本系列。x_j 为特大变量（$j = 1,2,\cdots,b$）；N 为样本最大重现期；l 为实测期历史洪水个数，n 为实测洪水个数。则式（5）、（6）左边又变为

$$E[x] = \frac{1}{N}\left(\frac{N-b}{n-l}\sum_{i=1}^{n-1} x_i + \sum_{j=1}^{b} x_j\right) \tag{9}$$

$$E[\ln x] = \frac{1}{N}\left(\frac{N-b}{n-l}\sum_{i=1}^{n-1} \ln x_i + \sum_{j=1}^{b} \ln x_j\right) \tag{10}$$

1.2 三参数对数正态分布的积分变换矩法

三参数对数正态分布密度函数为

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma(y-\tau)}\exp\left\{-\frac{[\ln(y-\tau)-\mu]^2}{2\sigma^2}\right\} \quad (y \geqslant \tau) \tag{11}$$

式中，τ 为移位参数。可用均值 $\overline{Y}$、变差系数 C_v 和偏态系数 C_s 来确定，即

$$\begin{cases} \tau = \overline{Y}\left(1 - \dfrac{C_v}{\eta}\right) \\ \sigma = \sqrt{\ln(1-\eta^2)} \\ \mu = \dfrac{1}{2}\ln\left[\dfrac{1+\eta^2}{(\overline{Y}-\tau)^2}\right] \\ \eta = \left(\dfrac{C_s + \sqrt{C_s^2+4}}{2}\right)^{1/3} - \left(\dfrac{-C_s + \sqrt{C_s^2+4}}{2}\right)^{1/3} \end{cases} \tag{12}$$

比较式（11）、（1），将二参数对数正态分布密度函数式（5）、（6）的参数转换为三参数对数正态分布密度函数参数，得：

$$E[y] = \overline{Y} = \tau + \exp\left(\mu + \frac{\sigma^2}{2}\right) \tag{13}$$

$$E[\ln(y-\tau)] = \mu \tag{14}$$

由于式（11）比式（1）多一个参数，再将式（11）代入式（3），并取 $\alpha = 2$ 得

$$E[(y-\tau)^2] = \exp[2(\mu+\sigma^2)] \tag{15}$$

由式（13）和式（15）推导得

$$C_v = \frac{\sqrt{\exp(2\mu+\sigma^2)[\exp(\sigma^2)-1]}}{\tau + \exp\left(\mu + \dfrac{\sigma^2}{2}\right)} \tag{16}$$

将式（13）和式（16）联解，代入式（14）得

$$E[\ln(y - \overline{Y} + H)] = \ln H - \frac{\sigma^2}{2} \tag{17}$$

其中

$$H = \frac{\overline{Y}C_v}{\sqrt{\exp(\sigma^2)-1}} \tag{18}$$

利用式（17）迭代求解 σ，并由下式求出 C_s：

$$C_s = 3\varnothing + \varnothing^3 \tag{19}$$

其中

$$\varnothing = \sqrt{\exp(\sigma^2)-1} \tag{20}$$

至此，已求得 $\overline{Y}$、C_v 和 C_s，可依照传统方法通过离均系数 $\varnothing_p$ 来计算不同频率 P 时的设计值 Y_p。

仍考虑简单样本系列和非简单样本系列两种情况：

（1）简单样本系列。式（13）、（14）和（16）左边分别为：

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{21}$$

$$E[\ln(y - \overline{Y} + H)] = \frac{1}{n}\sum_{i=1}^{n} \ln(y_i - \overline{Y} + H) \tag{22}$$

$$C_v = \frac{1}{\overline{Y}}\sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{Y})^2}{n-1}} \tag{23}$$

（2）非简单样本系列。式（13）、（14）和（16）左边又分别为：

$$\overline{Y} = \frac{1}{N}\left(\frac{N-b}{n-l}\sum_{i=1}^{n-l} y_i + \sum_{j=1}^{b} y_j\right) \tag{24}$$

$$E[\ln(y - \overline{Y} + H)] = \frac{1}{N}\left[\frac{N-b}{n-l}\sum_{i=1}^{n-l}\ln(y_i - \overline{Y} + H) + \sum_{j=1}^{b}\ln(y_j - \overline{Y} + H)\right] \tag{25}$$

$$C_v = \frac{1}{\overline{Y}}\sqrt{\frac{1}{N-1}\left[\sum_{j=1}^{b}(y_j - \overline{Y})^2 + \frac{N-b}{n-l}\sum_{i=1}^{n-l}(y_i - \overline{Y})^2\right]} \tag{26}$$

2 算 例

现以三参数对数正态分布为水文频率分布线型，以云南省境内西南诸河的怒江、伊洛瓦底江、澜沧江、红河主要干支流 16 个水文站年平均流量系列为例，分析本文提出的积分变换矩法参数估算的精度及适应性。各个水文站年平均流量系列长度列于表 1 中（均为简单样本系列），已经过“三性”分析，符合水文频率分析计算要求。为与矩法进行比较，计算结果同列于表 1。从表 1 中可以看出，两种方法计算 $\overline{Y}$、C_v 值均相同，而 C_s 值各异，积分变换矩法估算的 C_s 比矩法增大。

限于篇幅，从怒江、伊洛瓦底江、澜沧江、红河水系中各选一个有代表性水文站的年平均流量频率曲线制作成图，见图 1 ~ 4。从图 1 ~ 4 可看出，按本文提出的积分变换矩法估计的统计参数适线，上部、中部、下部点线配合明显较矩法为好，表明积分变换矩法估计统计参数 C_s 值的精度有所提高。

表 1　不同估计方法的参数计算结果

流域	站名	面积/km²	年限/a	矩法 $\bar{Y}$/(m³·s⁻¹)	矩法 C_v	矩法 C_s	积分变换矩法 $\bar{Y}$/(m³·s⁻¹)	积分变换矩法 C_v	积分变换矩法 C_s
伊洛瓦	麻栗坝	294	42	11.67	0.199	0.460	11.67	0.199	0.680
底江	梁河	1525	37	44.61	0.197	0.637	44.61	0.197	0.673
	拉贺练	4225	48	186.33	0.194	0.702	186.33	0.194	1.007
	盏西	1548	44	90.79	0.184	0.419	90.79	0.184	0.742
怒江	姑老河	4185	44	92.80	0.208	0.410	92.80	0.208	0.627
	大文	657	42	14.32	0.214	0.468	14.32	0.214	0.561
红河	李仙渡	17273	46	380.07	0.199	0.680	380.07	0.199	0.744
	元江	21554	56	166.78	0.328	0.831	166.78	0.328	1.007
	蛮燕	1514	44	33.20	0.244	0.171	33.20	0.244	0.457
	忠爱桥	3562	46	85.02	0.231	0.668	85.02	0.231	0.856
澜沧江	太平关	2910	45	61.91	0.204	0.355	61.91	0.204	0.648
	勐省	1766	40	52.27	0.182	0.392	52.27	0.182	0.674
	甸南	918	40	8.14	0.276	0.533	8.14	0.276	0.674
	甸头	711	40	19.15	0.209	0.736	19.15	0.209	0.955
	曼中田	1133	43	33.31	0.215	0.503	33.31	0.215	0.768
	光明	390	36	7.74	0.267	0.682	7.74	0.267	1.020

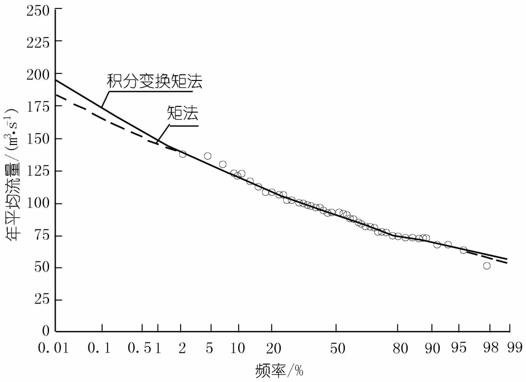

图 1　怒江姑老河水文站年平均流量曲线

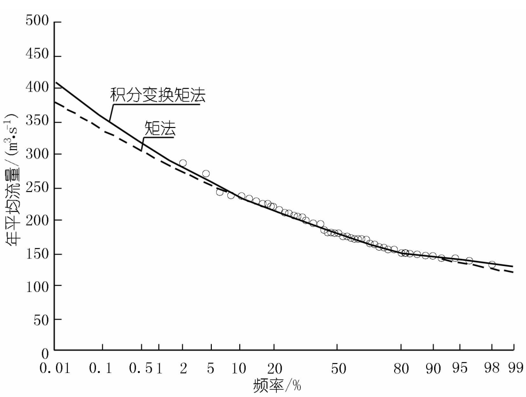

图 2　伊洛瓦底江拉贺练水文站年平均流量曲线

3　结论

本文从梅林变换理论出发，分别提出了二参数对数正态分布和三参数对数正态分布一种参数估计的积分变换矩法，该方法还分别考虑了简单样本系列和非

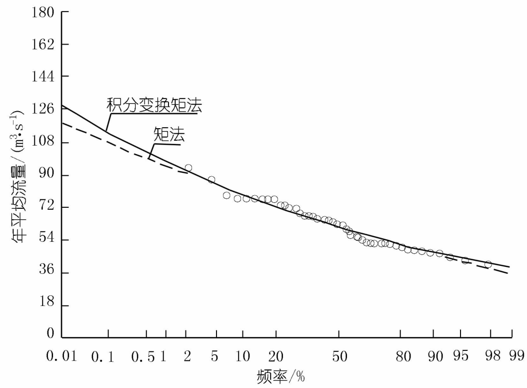

图 3　澜沧江太平关水文站年平均流量曲线

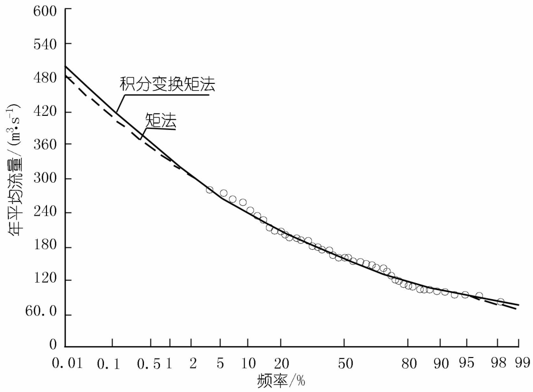

图 4　红河元江水文站年平均流量曲线

简单样本系列两种情况。对于三参数对数正态分布；$C_s = 3\emptyset + \emptyset^3$ 中的 α 采用对数矩迭代求出 。以三参数对数正态分布拟合西南诸河 16 个水文站的年平均流量样本系列为例，本文提出的积分变换矩法估算 C_s 的精度比矩法有明显提高。

参考文献：

[1]　Giesbrecht F, Kempthorne. Maximum likelihood estimation in the three - parameter log normal distribution [J]. J. R. Stat. Sec., 1976, 38(3): 257 - 264.

[2]　Singh V P. Entropy - based Parameter Estimation in Hydrology [M]. Boston/London: Kluwer Academic Publishers, 1998.

[3]　陈元芳，沙志贵. 可考虑历史洪水对数正态分布线性矩法的研究[J]. 河海大学学报：自然科学版，2003，31(1)：80 - 83.

[4]　Hosking J R M, Wallis J R. Regional Frequency Analysis, An Approach Based on L - moments [M]. Cambridge: Cambridge University Press, 1997.

[5]　张明，柏绍光. Mellin 变换在 P—Ⅲ型分布参数估计中的应用[J]. 水电能源科学，2010，28(6)：7 - 9.

[6]　张明，柏绍光. 基于 Mellin 变换的指数 Γ 分布参数估计方法[J]. 水电能源科学，2011，29(4)：8 - 10.

（编辑：李　慧）

（下转第 37 页）