

网络出版时间:2016-05-03 14:05

DOI:10.13207/j.cnki.jnwfufu.2016.06.028

网络出版地址: <http://www.cnki.net/kcms/detail/61.1390.S.20160503.1405.056.html>

基于交互熵原理的水文分布参数估计研究

牛林森, 宋松柏

(西北农林科技大学 水利与建筑工程学院, 陕西 杨凌 712100)

[摘要] **【目的】**研究交互熵法进行水文分布参数估计的普适性。**【方法】**应用最小交互熵原理研究 Gumbel 分布参数估计,在此基础上应用蒙特卡洛试验检验交互熵法统计性能,然后结合矩法和线性矩法等传统参数估计方法,以陕西省关中地区周至、武功、蒲城、礼泉、白水、潼关 6 个水文站年降水序列为例,计算年降水量设计值并拟合实测值序列,利用累积相对偏差平方和评价理论年降水量频率曲线对实测值序列的拟合效果。**【结果】**蒙特卡洛试验检验表明,交互熵法所求设计值的有效性指标估计量标准偏差(SE)和均方根误差(RMSE)小于矩法和线性矩法,偏差指标控制在 7% 以内;交互熵法估计周至、武功、蒲城、礼泉、白水、潼关 6 个水文站的累积相对偏差平方和分别为 0.000 017 68, 0.000 065 62, 0.000 019 66, 0.000 063 00, 0.000 014 12 和 0.000 016 61, 线性矩法估计上述 6 站的累积相对偏差平方和分别为 0.000 087 62, 0.000 093 55, 0.000 086 52, 0.000 101 39, 0.000 065 15 和 0.000 069 05, 矩法估计上述 6 站的累积相对偏差平方和分别为 0.000 108 74, 0.000 125 40, 0.000 092 41, 0.000 127 65, 0.000 085 49 和 0.000 093 57。由此可知,交互熵法不仅具有较好的有效性与合理的不偏性,而且与实测序列的拟合效果也明显优于传统方法。**【结论】**交互熵法是一种可行的水文分布参数估计方法,能有效提高 Gumbel 分布参数的估计精度。

[关键词] 水文分布;参数估计;交互熵法;Gumbel 分布;分位数约束

[中图分类号] P333.9

[文献标志码] A

[文章编号] 1671-9387(2016)06-0203-07

Application of cross entropy method in estimation of Gumbel distribution parameters

NIU Lin-sen, SONG Song-bai

(College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: **【Objective】** This paper studied the universal applicability of the cross entropy method in estimation of hydrology frequency distribution parameters. **【Method】** Kullback minimum cross entropy principle was used to estimate the parameters of Gumbel distribution. Then, Monte Carlo experiments were performed to verify its statistical performance. Observed precipitations at 6 hydrologic stations, Zhouzhi, Wugong, Pucheng, Liquan, Baishui and Tongguan in Central Shaanxi, were compared with predictions and the fitting results were evaluated using cumulative square error. **【Result】** Monte Carlo experiments indicated that the SE and RMSE of cross entropy method were less than those of moments method and L-moments method. The bias of cross entropy method was less than 7%. The cumulative of square errors of cross entropy method for the 6 hydrologic stations of Zhouzhi, Wugong, Pucheng, Liquan, Baishui, and Tongguan were 0.000 017 68, 0.000 065 62, 0.000 019 66, 0.000 063 00, 0.000 014 12, and 0.000 016 61, those of L-moments method were 0.000 087 62, 0.000 093 55, 0.000 086 52, 0.000 101 39, 0.000 065 15,

[收稿日期] 2014-11-03

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

[作者简介] 牛林森(1990—),女,河南驻马店人,硕士,主要从事流域水文模拟及水文预报研究。E-mail: nls1990@163.com

[通信作者] 宋松柏(1965—),男,陕西咸阳人,教授,博士,主要从事水文学水资源研究。E-mail: ssb6533@nwsuaf.edu.cn

and 0.000 069 05, while those of moments method were 0.000 108 74, 0.000 125 40, 0.000 092 41, 0.000 127 65, 0.000 085 49, and 0.000 093 57, respectively. The cross entropy method not only possessed good effectiveness and reasonable non-biasedness, but also fitted better with the observed data series than traditional methods. 【Conclusion】 Cross entropy method is a feasible method for parameter estimation, and it can effectively improve the estimation precision of Gumbel distribution parameters.

Key words: hydrological distribution; parameter estimation; cross entropy method; Gumbel distribution; fractile-pair constraints

水文频率分析是水文学研究的重要内容,其通过研究和分析水文随机现象,揭示其中蕴含的统计规律,并对未来可能的情势做出预估,以满足水利工程规划、设计、管理以及水资源利用等工作的需要^[1]。水文频率分析包括分布线型选择和参数估计^[2],其中水文分布参数估计方法是水文频率分析的重要内容之一。现有的水文分布参数估计方法有矩法、极大似然法、概率权重矩法、线性矩法等^[3]。除矩法之外,其他方法在求解参数时都会受到线型种类的影响,即对于不同分布线型需要分别推导参数求解表达式和(或)方程(组),计算较为复杂。矩法虽然不受分布线型影响,但是精度较差(尤其是 Cs 值),一般用于初估参数值^[4]。寻求参数求解简便、精确度高的参数估计方法是水文频率分析领域的重要课题。交互熵(Cross entropy)又称相对熵(Relative entropy)、差异信息(Discrimination information)等,由 Good^[5]于 1950 年提出,表示两个分布之间的概率距离,其是 Shannon 信息熵的一般化,能够有效避免 Shannon 信息熵的缺点^[6]。Kullback^[7]于 1959 年提出了最小交互熵原理(Minimum cross entropy principle),该原理借助了 Jaynes 的最大熵原理(Maximum entropy principle),即所求分布即是不确定度最大的分布^[8-9],并采用类似贝叶斯法中先验分布和后验分布的概念,若先验分布(亦称参考分布)已知,且约束条件一定,则候选分布中能使交互熵函数值最小的分布即为所求分布,也称后验分布^[10-14]。交互熵法广泛应用于结构分析和光谱分析等领域^[15-16]。1988 年, Lind 等^[17]首次将交互熵法应用于水文领域,其以分位数对为约束条件,依据最小交互熵原理,采用 Gamma 分布和 Gumbel 分布,以加拿大 2 条河流的年最大洪峰流量数据为例,运用交互熵法估计分布参数并绘制出拟合效果图,然后将交互熵法所求熵值与传统的矩法和极大似然法相对比,得出了交互熵法较优的结论。美国学者 Hosking 于 1990 年提出的线性矩法是概率权重矩的线性组合,与概率权重矩法结果接近,并且具有良

好的统计性能^[18]。目前,国内尚缺少将交互熵法应用于水文领域的研究报道^[19]。为此,本研究采用 Gumbel 分布作为参考分布,以矩法和线性矩法等传统方法为对比,应用蒙特卡洛试验分析交互熵法的统计性能,并以陕西省关中地区 6 个水文站年降水序列为例,对交互熵法应用于水文领域的普适性进行了分析,以期对陕西关中地区水利工程设计中设计洪水的准确计算提供参考。

1 交互熵法原理

1.1 交互熵函数

设随机变量 X 的参考分布函数为 $P(x)$, 密度函数为 $p(x)$, 同时假设后验分布为 $Q(x)$, 其对应密度函数为 $q(x)$, 则交互熵函数 D 的形式为:

$$D = \int q(x) \log \frac{q(x)}{p(x)} dx. \quad (1)$$

式中: D 为交互熵函数, x 为随机变量 X 的实际观测值。

1.2 约束条件

选取分位数对为约束条件。将随机变量 X 的观测序列按升序排列, 得到新序列 $S = \{x_i\}$, $i = 1, 2, \dots, r$, 其中 $x_i \in R$, r 为观测序列中观测值的个数。假设 x 是一个可能发生的未知量, 那么其落在 X 由 x_i 分成的 $r+1$ 个子区间 $[x_0, x_1)$, $[x_1, x_2)$, $\dots$, $[x_r, x_{r+1})$ 内的机率相等。根据样本规则, 对应的分位数概率为 $i/(r+1)$, 假设 $Q(x|x_1, x_2, \dots, x_r)$ 是由序列 S 推导出的 X 的分布函数, $q(x|x_1, x_2, \dots, x_r)$ 为相应密度函数。则有:

$$(x; Q(x|x_1, x_2, \dots, x_r))_i = \left(x_i, \frac{i}{r+1}\right), i = 1, 2, \dots, r. \quad (2)$$

式中: r 为 X 观测序列中观测值的个数。

1.3 Kullback 最小交互熵原理

给定一个参考分布函数 $P(x)$ 和密度函数 $p(x)$, 在分位数 $x = x_i$ 上, $P(x)$ 值记为 P_i 。通过寻求后验分布函数 $Q(x|x_1, x_2, \dots, x_r)$, 使交互熵函数

为:

$$D(q,p)=\int q(x|x_1,x_2,\cdots,x_r)\cdot \log \frac{q(x|x_1,x_2,\cdots,x_r)}{p(x)}dx. \tag{3}$$

令交互熵函数在满足分位数对约束的条件下最小,则式(2)可写为期望值的形式:

$$g_i:\int f_i(x)q(x|x_1,x_2,\cdots,x_r)dx-\frac{1}{r+1}=0,i=1,2,\cdots,r. \tag{4}$$

式中: $f_i(x)$ 为指示函数, $f_i(x)=\begin{cases} 1,x\in I_i \\ 0,其他 \end{cases}$,其中 I_i 为随机变量 X 的第 i 个子区间。根据极值原理,引入拉格朗日乘子 λ_i ,得到受分位数对约束的交互熵函数 L ,其形式为:

$$L=D(q,p)+\sum_{i=1}^r\lambda_i g_i=\int q(x|x_1,x_2,\cdots,x_r)[\log q(x|x_1,x_2,\cdots,x_r)-\log p(x)]dx+\sum_{i=1}^r\lambda_i\left[\int f_i(x)q(x|x_1,x_2,\cdots,x_r)dx-\frac{1}{r+1}\right]. \tag{5}$$

为使交互熵最小,根据变分法中的 Euler-Lagrange 方程,将式(5)仅对 $q(x)$ 求偏导,并令其为 0,有:

$$\frac{\partial L}{\partial q(x)}=\log \frac{q(x|x_1,x_2,\cdots,x_r)}{p(x)}+1+\sum_{i=1}^r\lambda_i f_i(x)=0. \tag{6}$$

整理,有:

$$q(x|x_1,x_2,\cdots,x_r)=p(x)\exp(-1-\sum_{i=1}^r\lambda_i f_i(x))=\mu(x)p(x). \tag{7}$$

其中:

$$\mu(x)=\exp(-1-\sum_{i=1}^r\lambda_i f_i(x)). \tag{8}$$

当 $x\in I_i$ 时,有:

$$\mu(x)=\mu_i=\exp(-1-\lambda_i). \tag{9}$$

将式(7)代入式(4)的左边,考虑 $x\in I_i$ 时 $f_i(x)=1$,故有:

$$\int f_i(x)q(x|x_1,x_2,\cdots,x_r)dx-\frac{1}{r+1}=\int \mu_i p(x)dx-\frac{1}{r+1}. \tag{10}$$

根据式(4)可知,式(10)等于 0,故有:

$$\mu_i=\frac{1}{(r+1)(P_{i+1}-P_i)}. \tag{11}$$

式中: P_i 和 P_{i+1} 分别为参考分布在 $x=x_i$ 和 $x=x_{i+1}$ 时的概率分布函数。因此,当 $x\in I_i$,有:

$$q(x|x_1,x_2,\cdots,x_r)=\mu_i p(x). \tag{12}$$

将式(9)和式(10)代入式(2)中,有:

$$D(q,p)=\sum_{i=0}^r\mu_i p(x)\log \frac{\mu_i p(x)}{p(x)}. \tag{13}$$

化简整理,得:

$$D(q,p)=-\log(r+1)-\frac{1}{r+1}\sum_{i=0}^r\log(P_{i+1}-P_i). \tag{14}$$

令 $c=\log(r+1)$, c 为非负数。再引入中间变量 $S(P)$:

$$S(P)=-\sum_{i=0}^r\log(P_{i+1}-P_i). \tag{15}$$

则有:

$$D_{\min}(q,p)=-c+S(P)/(r+1). \tag{16}$$

式中: $D_{\min}(q,p)$ 可以称为最小交互熵函数。

从式(16)可以看出,最小交互熵函数可以写成参考分布的函数,并且 $S(P)$ 值越小,最小交互熵函数值越小。在参考分布曲线类型选定的条件下,最小交互熵原理可解释为最优分布的最小交互熵函数值最小。

1.4 参考分布

选择 Gumbel 分布作为参考分布,其密度函数与分布函数分别为:

$$p(x)=\alpha\exp(-\alpha(x-u)-e^{-\alpha(x-u)}), \tag{17}$$

$$P(x)=\exp(-e^{-\alpha(x-u)}). \tag{18}$$

式中: α 为尺度参数, μ 为位置参数。

2 蒙特卡洛试验

本研究利用蒙特卡洛试验进行交互熵法统计性能研究,并与传统的参数估计方法,即矩法、线性矩法的估计结果进行比较。

2.1 评价标准

依据文献[20]中的评价标准,以偏差(Bias)和标准误差(SE)分别评判估计量($\hat{\alpha},\hat{u}$)及设计值($\hat{x}_P$)的不偏性和有效性,用估计量的均方根误差(RMSE)综合评判参数估计方法的有效性。

若估计量的偏差越接近 0,表明估计量的不偏性越好,在 M 次统计试验中,其计算公式为:

$$\text{Bias}(\hat{x}_P)=[x_P-\hat{\mu}(\hat{x}_P)]. \tag{19}$$

式中: x_P 为重现期 T 的总体设计值, $\hat{x}_P$ 为重现期 T 的估计值, $\hat{\mu}(\hat{x}_P)$ 为 M 次统计试验估计量 $\hat{x}_P$ 的均值。标符号“ $\hat{\cdot}$ ”表示估计值,下同。

本研究所选重现期有 4 个,分别是 50,100,200 和 1 000 年。由于本研究采用不及制累积概率,因

此上述 4 个重现期对应的频率 P 分别为 0.980, 0.990, 0.995, 0.999。

估计量标准误差(SE)越小,表明估计量的有效性越好,在 M 次统计试验中,其计算公式为:

$$SE(\hat{x}_P)=\sqrt{\frac{1}{M-1}\sum_{i=0}^M(\hat{x}_P^i-\hat{\mu}(\hat{x}_P))^2}。 \quad (20)$$

式中: $\hat{x}_P^i$ 为 M 次统计试验中第 i 个试验样本的估计值。

估计量均方根误差(RMSE)越小,表明参数估计方法的有效性越好,其计算公式为:

$$RMSE(\hat{x}_P)=\sqrt{[\text{Bias}(\hat{x}_P)]^2+[\text{SE}(\hat{x}_P)]^2}。 \quad (21)$$

RMSE 表示的是估计值与真实值偏差平方的算术平均值再开方。 α 、 μ 2 个参数值的误差计算原理类似,此处不再赘述。

2.2 试验方案设计与试验结果

为不失一般性,总体分布参数值取为 $\alpha=1, u=0$ 。样本容量用 n 表示,取 $n=50, 70, 100, 150, 200, 300, 700, 1\ 000$ 共 8 组方案,各组的模拟次数为 1 000^[21-22]。按照上述方案进行蒙特卡洛模拟。因篇幅原因试验结果不再全部列出,现选取具有代表性的 $n=50, 70, 100, 200$ 时的试验结果列于表 1。

表 1 不同样本容量时蒙特卡洛模拟的不偏性和有效性计算结果

Table 1 Non-biasedness and effectiveness of Monte Carlo simulations with different sample sizes

样本容量 Sample size	评价指标 Evaluation index	参数估计方法 Parameter estimation method	$\hat{\alpha}$	$\hat{u}$	$\hat{x}_{50}/x_{50}$	$\hat{x}_{100}/x_{100}$	$\hat{x}_{200}/x_{200}$	$\hat{x}_{1\ 000}/x_{1\ 000}$
50	偏差 Bias	矩法 Moments	−0.031 6	−0.006 2	0.008 9	0.009 1	0.009 3	0.009 6
		线性矩法 L-Moments	−0.016 6	−0.000 2	0.000 1	0.000 1	0.000 1	0.000 1
		交互熵法 Cross entropy	−0.090 7	−0.017 7	0.066 4	0.067 1	0.067 6	0.068 4
	标准误差 SE	矩法 Moments	0.148 8	0.152 8	0.152 4	0.150 3	0.149 0	0.147 1
		线性矩法 L-Moments	0.131 6	0.149 7	0.141 2	0.138 5	0.136 6	0.133 9
		交互熵法 Cross entropy	0.127 0	0.150 3	0.123 5	0.120 2	0.118 0	0.114 7
	均方根误差 RMSE	矩法 Moments	0.152 1	0.152 9	0.152 6	0.150 6	0.149 3	0.147 4
		线性矩法 L-Moments	0.132 6	0.149 7	0.141 2	0.138 5	0.136 6	0.133 9
		交互熵法 Cross entropy	0.156 1	0.151 3	0.140 2	0.137 7	0.136 0	0.133 6
70	偏差 Bias	矩法 Moments	−0.023 7	0.004 6	0.007 3	0.007 5	0.007 6	0.007 8
		线性矩法 L-Moments	−0.012 3	0.000 0	0.000 5	0.000 5	0.000 5	0.000 5
		交互熵法 Cross entropy	−0.068 1	−0.012 8	0.051 7	0.052 2	0.052 6	0.053 1
	标准误差 SE	矩法 Moments	0.125 5	0.129 2	0.129 4	0.127 7	0.126 6	0.125 0
		线性矩法 L-Moments	0.110 7	0.126 6	0.119 4	0.117 2	0.115 6	0.113 4
		交互熵法 Cross entropy	0.104 2	0.126 8	0.105 2	0.102 5	0.100 6	0.097 9
	均方根误差 RMSE	矩法 Moments	0.127 7	0.129 3	0.129 6	0.127 9	0.126 8	0.125 3
		线性矩法 L-Moments	0.111 4	0.126 6	0.119 4	0.117 2	0.115 6	0.113 4
		交互熵法 Cross entropy	0.124 5	0.127 5	0.117 2	0.115 0	0.113 5	0.111 4
100	偏差 Bias	矩法 Moments	−0.016 5	−0.003 3	0.005 2	0.005 4	0.005 5	0.005 6
		线性矩法 L-Moments	−0.008 5	0.000 0	0.000 5	0.000 5	0.000 5	0.000 5
		交互熵法 Cross entropy	0.049 9	−0.009 2	0.039 1	0.039 4	0.039 7	0.040 1
	标准误差 SE	矩法 Moments	0.103 1	0.107 9	0.107 7	0.106 3	0.105 3	0.104 0
		线性矩法 L-Moments	0.090 5	0.105 4	0.099 3	0.097 3	0.096 0	0.094 1
		交互熵法 Cross entropy	0.084 4	0.105 4	0.088 7	0.086 4	0.084 8	0.082 5
	均方根误差 RMSE	矩法 Moments	0.104 4	0.107 9	0.107 8	0.106 4	0.105 5	0.104 2
		线性矩法 L-Moments	0.090 9	0.105 4	0.099 3	0.097 3	0.096 0	0.094 1
		交互熵法 Cross entropy	0.098 1	0.105 8	0.097 0	0.095 0	0.093 7	0.091 8
200	偏差 Bias	矩法 Moments	−0.007 7	−0.002 3	0.001 8	0.001 9	0.002 0	0.002 1
		线性矩法 L-Moments	−0.003 9	−0.000 8	−0.000 3	−0.000 3	−0.000 2	−0.000 2
		交互熵法 Cross entropy	−0.027 8	−0.005 7	0.022 6	0.022 8	0.023 0	0.023 2
	标准误差 SE	矩法 Moments	0.073 2	0.076 6	0.077 5	0.076 5	0.075 9	0.074 9
		线性矩法 L-Moments	0.063 4	0.074 8	0.070 5	0.069 0	0.068 1	0.066 7
		交互熵法 Cross entropy	0.057 4	0.074 9	0.063 1	0.061 4	0.060 3	0.058 6
	均方根误差 RMSE	矩法 Moments	0.073 6	0.076 6	0.077 5	0.076 5	0.075 9	0.075 0
		线性矩法 L-Moments	0.063 6	0.074 8	0.070 5	0.069 0	0.068 1	0.066 7
		交互熵法 Cross entropy	0.063 8	0.075 1	0.067 0	0.065 5	0.064 5	0.063 0

由表 1 可以看出:①随着样本容量的增加,3 种方法对应设计值的 Bias、SE 和 RMSE 都逐渐增加;

②线性矩法的设计值偏差 Bias 最小,矩法次之,交互熵法最大。线性矩法设计值估计具有良好的不偏性,交互熵法的设计值偏差 Bias 控制在 7%以内,可以满足精度要求;③交互熵法的设计值标准误差 SE 最小,线性矩法次之,矩法最差。说明利用交互熵法估计设计值优势明显;④交互熵法的均方根误差 RMSE 最小,线性矩法次之,矩法最差。说明将交互熵法应用于 Gumbel 分布参数估计时,其有效性优于其余 2 种方法。

表 2 陕西省关中地区 6 个水文站年降水量资料系列长度

站名 Station	周至 Zhouzhi	武功 Wugong	蒲城 Pucheng	礼泉 Liquan	白水 Baishui	潼关 Tongguan
起止年份 Sample period	1957—2008	1955—2007	1959—2008	1959—2007	1962—2008	1957—2008
实测长度/年 Systematic record length	52	53	50	49	47	52

3.1 参数估计

以矩法所得参数作为初始值,利用 Matlab 2010b 中的优化函数确定最小交互熵函数最小值和分布参数,并将结果与矩法、线性矩法计算结果进行对比,结果见表 3。由表 3 可以看出:①矩法、线性

3 实例应用

本研究选取陕西省关中地区周至、武功、蒲城、礼泉、白水、潼关 6 个水文站年降水量资料(表 2),经过“三性”审查,所有资料均满足可靠性、代表性及一致性要求。利用这些资料研究交互熵法在 Gumbel 分布参数估计中的可行性,并结合矩法和线性矩法进行拟合效果评价。

矩法及交互熵法的 $S(P)$ 和 $D_{\min}(q,p)$ (最小交互熵函数值)依次减小;②按照 Kullback 最小交互熵原理,所用到的 3 种推求 Gumbel 分布参数的方法中,矩法最差,线性矩法次之,交互熵法最优。

表 3 陕西省关中地区 6 个水文站年降水量分布参数及熵值计算结果

站点 Station	参数估计方法 Method of parameter estimation	$S(P)$	$D_{\min}(q,p)$	α	μ
周至 Zhouzhi	矩法 Moments	237.245 0	0.506 0	0.008 078	563.766 150
	线性矩法 L-Moments	236.784 7	0.497 3	0.007 899	562.142 413
	交互熵法 Cross entropy	235.857 5	0.479 8	0.007 092	553.588 061
武功 Wugong	矩法 Moments	251.349 3	0.665 6	0.008 103	535.502 509
	线性矩法 L-Moments	250.882 2	0.657 0	0.007 868	533.376 963
	交互熵法 Cross entropy	250.372 7	0.647 5	0.007 289	525.845 028
蒲城 Pucheng	矩法 Moments	227.638 6	0.531 7	0.010 671	468.677 088
	线性矩法 L-Moments	227.478 9	0.528 5	0.010 587	468.243 045
	交互熵法 Cross entropy	226.450 1	0.508 4	0.009 440	461.631 738
礼泉 Liquan	矩法 Moments	226.137 0	0.610 7	0.009 417	469.552 847
	线性矩法 L-Moments	225.696 3	0.601 9	0.009 187	468.014 820
	交互熵法 Cross entropy	224.978 0	0.587 3	0.008 368	460.285 698
白水 Baishui	矩法 Moments	214.609 6	0.599 8	0.010 340	516.106 482
	线性矩法 L-Moments	213.853 7	0.584 0	0.010 001	514.210 874
	交互熵法 Cross entropy	212.751 5	0.561 1	0.008 852	508.516 273
潼关 Tongguan	矩法 Moments	235.881 9	0.480 3	0.008 939	540.920 207
	线性矩法 L-Moments	235.200 0	0.467 4	0.008 661	538.839 051
	交互熵法 Cross entropy	234.282 6	0.450 1	0.007 789	530.945 322

3.2 年降水量频率曲线图绘制

根据表 3 的参数值,推求年降水量设计值,并绘制年降雨量频率曲线,结果如图 1 所示。由图 1 可以看出:选用 Gumbel 分布在 3 种方法下的拟合效果均表现良好;在实测数据的中低值部分,运用矩法和线性矩法求得的分布曲线拟合效果相差不大,比

较周至、武功、蒲城、白水、潼关、礼泉 6 个水文站的拟合效果可以看出,运用交互熵法所求降水量频率曲线明显较前 2 种方法求得的曲线更加接近实测值;周至、潼关、武功、白水、礼泉 5 个站的拟合曲线表明:在实测数据的高值处,线性矩法优于矩法;蒲城站 2 种方法拟合效果差异不明显;在周至、武功、

蒲城、礼泉、白水、潼关 6 个站点,用交互熵法计算得到的分布曲线在高值处的拟合效果均明显优于矩法和线性矩法。

总的来说,无论在传统方法拟合效果较为接近的中低值部分,还是高值部分,交互熵法所求曲线的拟合效果最优,线性矩法次之,矩法最差。

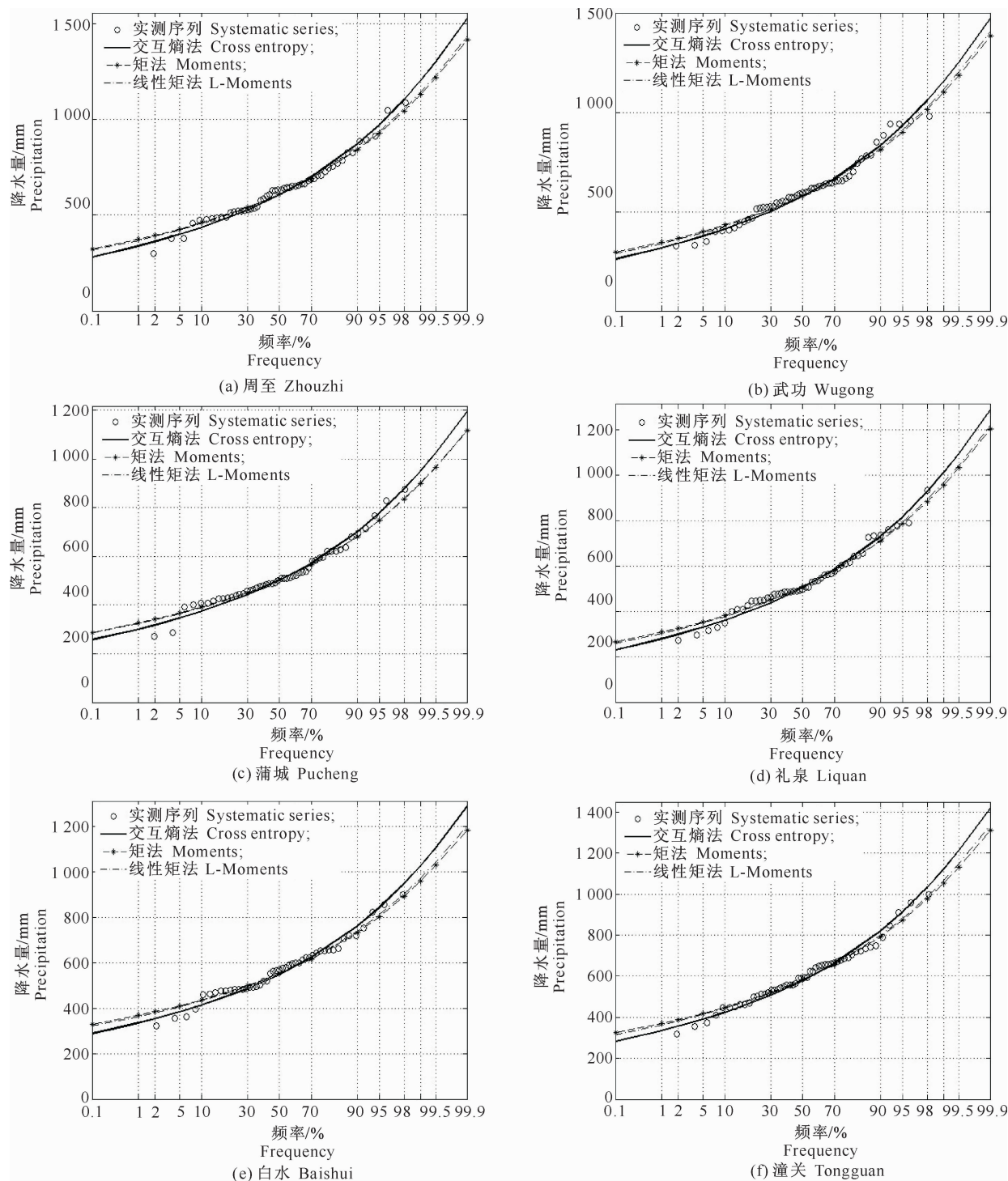


图 1 陕西省关中地区 6 个水文测站年降水量频率曲线

Fig. 1 Precipitation frequency at 6 stations in Central Shaanxi

3.3 拟合优度评价

应用累积相对偏差平方和(δ)分析上述参数估计方法的拟合效果。实测值与设计值累积偏差平方

和(δ)的计算公式为:

$$\delta=\sum_{i=1}^n\left(\frac{x_i-\hat{x}_i}{x_i}\right)^2。$$

(22)

式中: x_i 为实测值; $\hat{x}_i$ 为设计值, g 为实测样本容量。

不同参数估计方法在 6 个水文站点 δ 的计算结果如表 4 所示。

表 4 3 种参数估计方法的累积相对偏差平方和的比较

Table 4 Cumulative relative square errors of three different parameter estimation methods

参数估计方法 Method of parameter estimation	周至 Zhouzhi	武功 Wugong	蒲城 Pucheng	礼泉 Liquan	白水 Baishui	潼关 Tongguan
矩法 Moments	0.000 108 74	0.000 125 40	0.000 092 41	0.000 127 65	0.000 085 49	0.000 093 57
线性矩法 L-Moments	0.000 087 62	0.000 093 55	0.000 086 52	0.000 101 39	0.000 065 15	0.000 069 05
交互熵法 Cross entropy	0.000 017 68	0.000 065 62	0.000 019 66	0.000 063 00	0.000 014 12	0.000 016 61

从表 4 可以看出,在陕西关中的 6 个水文站,用交互熵法(CE)所求曲线的累计相对偏差平方和(δ)最小,其次是线性矩法,矩法的累积相对偏差平方和(δ)最大,说明交互熵法拟合效果最优,这与图 1 的拟合效果一致,并且与表 3 计算所得的最小交互熵函数值的表现规律相同,表明交互熵法是一种可行的参数估计方法。

4 讨论与结论

1)介绍了受分位数对约束的交互熵,以及基于 Kullback 最小交互熵原理求 Gumbel 分布参数的方法。利用蒙特卡洛试验研究交互熵法的统计性能,并与矩法和线性矩法相对比,结果表明交互熵法具有良好的有效性,且其不偏性满足精度要求,证明交互熵法是一种统计性能良好的参数估计方法。

2)选择 Gumbel 分布曲线拟合陕西省关中地区 6 个水文站的年降水资料,所采用的 3 种方法的分布曲线拟合效果均可以达到要求,设计值的累积相对偏差平方和 δ 也很小。因此,Gumbel 分布是一种能较好拟合关中地区年降水变量的分布曲线,可在以后的工作中推广使用。

3)选定 Gumbel 分布曲线的前提下,以陕西省关中地区 6 个水文站为例,对交互熵法、矩法和线性矩法的拟合效果和累积相对偏差平方和进行对比分析。结果表明,交互熵法是一种简单且有效的水文分布参数估计方法。

4)交互熵法与分布类型无关,并且原理简单,计算简便。本研究仅是一个交互熵法进行水文分布参数估计的探索性研究,对于我国水文计算普遍使用的 P-Ⅲ分布,尚有待于进一步研究。

[参考文献]

[1] 李 扬. 水文频率新型计算理论与应用研究 [D]. 陕西杨凌:西北农林科技大学,2013.
Li Y. Research on new theory and application of hydrologic frequency analysis [D]. Yangling, Shaanxi: Northwest A&F University, 2013. (in Chinese)

[2] 詹道江,徐向阳,陈元芳. 工程水文学 [M]. 北京:中国水利水电出版社,2011:141.
Zhan D J, Xu X Y, Chen Y F. Engineering hydrology [M]. Beijing: China Water Power Press, 2011: 141. (in Chinese)
[3] Rao A R, Hamed K H. Flood frequency analysis [M]. Florida: CRC Press LLC, 2000: 73-81.
[4] 桑燕芳,王 栋,吴吉春. 水文频率分析中参数估计 SAGA-ML 方法的研究 [J]. 水文, 2009, 29(5): 23-29.
Sang Y F, Wang D, Wu J C. Research on SAGA-ML method for parameter optimation in hydrologic frequency analysis [J]. Journal of China Hydrology, 2009, 29(5): 23-29. (in Chinese)
[5] Good I J. Probability and the weighing of evidence [J]. Biometrika, 1950, 38(3): 170-171.
[6] Pandey M D. Extreme quantile estimation using order statistics with minimum cross-entropy principle [J]. Probabilistic Engineering Mechanics, 2001, 16(1): 31-42.
[7] Kullback S. Information theory and statistics [M]. New York: Dover Publications, INC, 1959.
[8] Jaynes E T. Information theory and statistical mechanics [J]. Physical Review, 1957, 106(4): 620-630.
[9] 俞礼军,严 海,严宝杰. 最大熵原理在交通流统计分布模型中的应用 [J]. 交通运输工程学报, 2001, 1(3): 91-94.
Yu L J, Yan H, Yan B J. Maximum entropy method and its application in probability density function of traffic flow [J]. Journal of Traffic and Transportation Engineering, 2001, 1(3): 91-94. (in Chinese)
[10] Shore J, Johnson R. Properties of cross-entropy minimization [J]. IEEE Transactions on Information Theory, 1981, 27(4): 472-482.
[11] Deng J, Pandey M D. Using partial probability weighted moments and partial maximum entropy to estimate quantiles from censored samples [J]. Probabilistic Engineering Mechanics, 2009, 24(3): 407-417.
[12] Deng J, Pandey M D, Gu D. Extreme quantile estimation from censored sample using partial cross-entropy and fractional partial probability weighted moments [J]. Structural Safety, 2009, 31(1): 43-54.
[13] Deng J, Pandey M D. Cross entropy quantile function estimation from censored samples using partial probability weighted moments [J]. Journal of Hydrology, 2008, 363(1): 18-31.

Chen X L, Lou P H. The application of improved hierarchical clustering algorithm to analyze literature [J]. Journal on Numerical Methods and Computer Applications, 2009, 30 (4): 277-287. (in Chinese)

[21] 陈彦光. 基于 Matlab 的地理数据分析 [M]. 北京: 高等教育出版社, 2012: 159-179.

Chen Y G. Geographic data analysis with Matlab [M]. Beijing: Higher Education Press, 2012: 59-179. (in Chinese)

[22] 郭爱军, 畅建霞, 黄 强, 等. 渭河流域气候变化与人类活动对径流影响的定量分析 [J]. 西北农林科技大学学报(自然科学版), 2014, 42(8): 212-220.

Guo A J, Chang J X, Huang Q, et al. Quantitative analysis of the impacts of climate change and human activities on runoff change in Weihe basin [J]. Journal of Northwest A&F University (Natural Science Edition), 2014, 42(8): 212-220. (in Chinese)

[23] Wang W, Kong J, Duan L, et al. Research on the conversion relationships between the river and groundwater in the Yellow River drainage area [J]. Science in China Series E: Technological Sciences, 2004, 14(1): 25-41.

[24] Wu W, Xu Z X, Liu X C. Impact of Baojixia water diversion works on the hydrologic regime in the Wei River basin [J]. Procedia Environmental Sciences, 2012, 13: 1653-1662.

(上接第 209 页)

[14] Pandey M D. Minimum cross-entropy method for extreme value estimation using peaks-over-threshold data [J]. Structural Safety, 2001, 23(4): 345-363.

[15] Shore J. Minimum cross-entropy spectral analysis [J]. IEEE Transactions on Acoustics Speech & Signal Processing, 1981, 29(2): 230-237.

[16] Lind N C, Solana V. Fractile constrained entropy estimation of distributions based on scarce data [J]. Civil Engineering Systems, 1990, 7(2): 87-93.

[17] Lind N C, Hong H P, Solana V. A cross entropy method for flood frequency analysis [J]. Stochastic Hydrology and Hydraulics, 1989, 3(3): 191-202.

[18] Hosking J R. M. L-moments: analysis and estimation of distributions using linear combinations of order statistics [J]. Journal Royal Statistical Society, 1990, 52(2): 105-124.

[19] 牛林森, 宋松柏. 交互熵法在洪水频率分布参数估计中的应用 [J]. 水资源研究, 2013, 2(6): 389-394.

Niu L S, Song S B. An application of cross entropy method to the parameters estimation in flood frequency analysis [J]. Journal of Water Resources Research, 2013, 2(6): 389-394. (in Chinese)

[20] Bhattarai K P. Partial L-moments for the analysis of censored flood samples [J]. Hydrological Sciences Journal, 2004, 49(5): 855-868.

[21] 董双林. Gumbel 分布的参数估计方法的统计分析 [J]. 水利学报, 1989(11): 35-42.

Dong S L. Statistical analysis of the methods in Gumbel distribution parameter estimation [J]. Journal of Hydraulic Engineering, 1989(11): 35-42. (in Chinese)

[22] 罗 纯, 王筑娟. Gumbel 分布参数估计及在水位资料分析中应用 [J]. 应用概率统计, 2005, 21(2): 169-175.

Luo C, Wang Z J. The estimates of the parameters of Gumbel distribution and their application to the analysis of the water level data [J]. Chinese Journal of Applied Probability and Statistics, 2005, 21(2): 169-175. (in Chinese)