

几种频率曲线稳健性初步研究

王春霞¹, 黄振平², 马军建², 侯云青², 林小丽²

(1. 广东省水利电力勘测设计研究院, 广东 广州 510170; 2. 河海大学水资源环境学院, 江苏 南京 210098)

摘要 采用 Monte-Carlo 方法, 以无偏性与有效性为依据, 确定评判指标并研究几种频率曲线的稳健性. 为了使研究成果具有代表性, 对各种分布的 3 组不同总体分别抽取 1 000 组样本(每组样本容量为 60)进行计算与评判. 结果表明, 在目前常用的分布曲线中, P-Ⅲ型分布具有较强的稳健性.

关键词 水文频率计算 分布线型 随机抽样 稳健性分析

中图分类号 P338 **文献标识码** A **文章编号** 1000-1980(2004)04-0380-04

建设与规划水电工程, 首先必须为其提供满足一定设计标准的水文设计值 x_p . x_p 估计的精度直接关系到工程的投资与安全. 因此, 长期以来, 各国水文工作者对此做了大量的分析与研究, 并取得了一定的成果. 但由于问题的复杂性, 水文频率分析中的许多问题, 特别是线型的选择、参数的估计等影响到设计值的关键问题至今仍未得到普遍令人满意的结论. 为此, 本文以 Monte-Carlo 方法为基础^[1], 对几种频率曲线的稳健性进行研究.

1 几种分布线型

水文研究和实践中曾使用过许多分布线型, 分析表明, 水文资料的分布是偏态的, 应选取偏态分布类线型^[2]. 各国学者对线型选择、各类曲线的比较和水文应用等做了研究^[3,4]. 从国内外有关研究和动态来看, 皮尔逊Ⅲ型(P-Ⅲ型)、皮尔逊Ⅴ型(P-Ⅴ型)、皮尔逊Ⅰ型(P-Ⅰ型)、对数正态型(L-N型)、克里茨基-闵凯里型(K-M型)和耿贝尔型(Gumbel型)最为常见和通用. 本文对这 6 种分布线型的稳健性进行研究.

我国李松仕^[4]采用两种拟合优度准则, 对 72 组具有较长实测系列的洪峰流量资料, 分析比较了几种频率分布, 结论是 P-Ⅲ型分布对我国大部分地区资料的拟合优于其他线型. 徐向阳^[5]对我国相对具有较长系列的 100 个站的暴雨资料采用期望值公式和绝对值准则进行适线, 综合比较了 6 条分布曲线, 结论是对数正态型是一种较好的暴雨频率分布线型. 我国其他水文学者也曾就几种常用的水文分布采用实测资料进行过对比分析, 大部分结论认为 P-Ⅲ型分布最适合我国水文资料的分布线型. 然而, 水文线型拟合鉴别的主要困难在于: 资料的中间部分, 各种线型的拟合都能调整得较好, 以致难以区分优劣; 但在曲线的两端, 尤其是高水部分, 不同的线型又相差很大. 众所周知, 我国的实测水文资料系列短, 在曲线中表现为关键的小概率部分无数据, 从而无法判定拟合的好坏. 本文以此为出发点, 用 Monte-Carlo 法对以上几种分布进行抽样, 继而研究了 6 种分布的稳健性.

2 研究方法及相关问题

2.1 研究途径

本文运用 Monte-Carlo 法, 先从某已知总体中抽取样本(总体已知, 则该分布某种设计频率的真值 x_p 可以推求出), 再选用其他某种分布线型, 然后用适线法估计未知参数并推求出各设计频率的设计值 $\hat{x}_p$. 将 $\hat{x}_p$ 与设计频率的真值进行比较, 分析估计设计值的各类误差, 再根据无偏性与有效性, 确定若干评判准则, 分析各种分布的稳健性.

研究步骤如下:

a. 从每一个分布的总体中抽取长度为 N 的 m 组样本系列, 用某种分布及适线法估计参数;

b. 根据估算的参数,求出相应设计频率的设计值;

c. 分析设计值的各类误差并研究该分布的稳健性.

本文用下述方法分析设计值的误差,从而多方位地研究线型的稳健性.

a. 计算设计值均值的相对误差,分析设计值对真值的平均偏差程度. 设计值均值的相对误差为

$$\theta = \frac{\bar{x}_p - x_p}{x_p} \times 100\% \tag{1}$$

$$\bar{x}_p = \frac{1}{m} \sum_{i=1}^m \hat{x}_{pi}$$

式中 $\bar{x}_p$ ——设计频率的设计值均值; $\hat{x}_{pi}$ ——估计的设计频率的设计值; x_p ——已知分布的设计频率的真值.

b. 计算设计值对真值的相对均方误差. 设计值对真值的均方误差为

$$\sigma = \sqrt{\sum_{i=1}^m (\hat{x}_{pi} - x_p)^2 / m} \tag{2}$$

设计值对真值的相对均方误差为

$$\delta = \frac{\sigma}{x_p} \times 100\% \tag{3}$$

c. 计算设计值落在 ω 区间的频率. ω 区间的范围为

$$\omega = (x_p - \sigma, x_p + \sigma) \tag{4}$$

为便于对计算结果进行比较,各种分布都采用相同参数的总体. 本文取 3 组总体参数进行研究. 这 3 组参数分别为: $EX = 1000, C_v = 0.5, C_s = 2C_v$; $EX = 1000, C_v = 1.0, C_s = 2C_v$; $EX = 1000, C_v = 0.5, C_s = 3C_v$. 样本组数 m 取为 1000. 根据我国实测资料长度情况,样本系列长度取为 60. 设计频率取 $p_1 = 0.01, p_2 = 0.001$.

2.2 抽样方法

本文利用 Monte-Carlo 法,生成 $[0,1]$ 上的随机数,然后利用随机数生成各种分布的随机样本. 各种分布的抽样方法如下:

a. P-Ⅲ型、P-V型、P-I型、Gumbel 型分布的抽样. 首先生成 $[0,1]$ 均匀分布的随机数,然后根据相应的 Φ 值表内插求得 Φ_i ,再由式

$$x_i = EX(C_v\Phi_i + 1) \quad (i = 1, 2, \dots, 60) \tag{5}$$

求得抽样. 其中 Gumbel 分布的总体参数 C_s 为固定值 ($C_s = 1.1395471$).

b. L-N 型分布的抽样. 首先由 $[0,1]$ 上的随机数生成正态分布的抽样 $y_1, \dots, y_n$, 然后利用公式 $x = e^y$ 求得对数正态分布的随机抽样.

c. K-M 型分布的抽样. 首先生成 $[0,1]$ 上的随机数 u (作为概率), 由 K 值表内插得 K_i , 根据公式 $x_i = K_i EX$ 求出 K-M 分布的随机抽样.

2.3 参数估计

参数估计方法很多,我国常用适线法^[6]. 本文参数估计也采用我国普遍使用的适线法,并据两种准则(即绝对值准则和平方和准则)由计算机进行优化适线.

2.4 经验频率公式的选用

我国在计算经验频率时广泛采用期望值公式. 较多研究表明,用期望值公式适线易使设计值平均偏大. 尽管如此,从生产实际要求和设计安全角度出发,仍喜欢用使设计结果偏安全的期望值公式. 因此,本文从实际工作要求出发,也用期望值公式作为经验频率曲线的绘点公式.

3 曲线稳健性综合比较

在对线型进行比较时,借用无偏性和有效性概念,比较了设计值均值与真值的误差以及设计值与真值的均方误差. 为便于各种线型的对比,采用无量纲设计值均值的相对误差和无量纲相对均方误差进行分析.

为分析各种线型的稳定性,还计算了设计值落在区间 $\sigma = (x_p - \sigma, x_p + \sigma)$ 内的频率. 这里 σ 为设计值对真值的均方误差. 频率越大,则设计值处在真值附近的可能性越大,该曲线的适应性也越好.

假设每种曲线的权重一样,在对比各种曲线的优劣时,先对各种曲线分布稳健性的评价指标取平均值,再对各种曲线进行对比. 由于设计值均值的相对误差有正有负,对其平均正负会相互抵消,因此本文先对设计值均值的相对误差取绝对值,然后再取平均值. 计算结果见表 1~表 6.

表 1 $EX = 1000, C_v = 0.5, C_s/C_v = 2$ 时绝对值准则下各种曲线的比较

Table 1 Comparison of several curves under absolute value criterion for $EX = 1000, C_v = 0.5$ and $C_s/C_v = 2$

曲线类型	百年一遇			千年一遇		
	平均相对均 方误差/%	平均设计值均值 的相对误差/%	平均设计值落在 ω 区间的频率	平均相对均 方误差/%	平均设计值均值 的相对误差/%	平均设计值落在 ω 区间的频率
P-Ⅲ	10.71	3.12	0.923	15.47	7.62	0.681
P-V	11.38	3.30	0.909	20.23	8.26	0.613
P-I	12.11	6.51	0.882	20.26	16.24	0.484
L-N	26.71	21.14	0.511	48.13	40.84	0.201
K-M	12.97	6.10	0.855	18.51	8.39	0.586
Gumbel	11.85	7.08	0.897	18.01	13.88	0.590

表 2 $EX = 1000, C_v = 0.5, C_s/C_v = 2$ 时平方和准则下各种曲线的比较

Table 2 Comparison of several curves under square-sum criterion for $EX = 1000, C_v = 0.5$ and $C_s/C_v = 2$

曲线类型	百年一遇			千年一遇		
	平均相对均 方误差/%	平均设计值均值 的相对误差/%	平均设计值落在 ω 区间的频率	平均相对均 方误差/%	平均设计值均值 的相对误差/%	平均设计值落在 ω 区间的频率
P-Ⅲ	8.47	1.67	0.978	12.16	5.55	0.786
P-V	8.25	1.58	0.971	13.75	4.71	0.761
P-I	10.03	4.23	0.930	17.35	13.70	0.587
L-N	32.41	27.24	0.445	58.87	51.07	0.151
K-M	10.06	4.28	0.965	14.12	6.69	0.725
Gumbel	9.49	5.22	0.973	15.00	11.55	0.662

表 3 $EX = 1000, C_v = 1.0, C_s/C_v = 2$ 时绝对值准则下各种曲线的比较

Table 3 Comparison of several curves under absolute value criterion for $EX = 1000, C_v = 1.0$ and $C_s/C_v = 2$

曲线类型	百年一遇			千年一遇		
	平均相对均 方误差/%	平均设计值均值 的相对误差/%	平均设计值落在 ω 区间的频率	平均相对均 方误差/%	平均设计值均值 的相对误差/%	平均设计值落在 ω 区间的频率
P-Ⅲ	15.13	4.59	0.852	21.05	12.19	0.575
P-V	15.11	5.73	0.860	30.09	18.98	0.456
P-I	18.48	8.36	0.789	27.01	21.29	0.431
L-N	79.23	66.23	0.206	172.42	147.95	0.059
K-M	18.13	6.69	0.810	25.53	15.03	0.509
Gumbel	18.56	13.59	0.747	24.96	21.04	0.403

表 4 $EX = 1000, C_v = 1.0, C_s/C_v = 2$ 时平方和准则下各种曲线的比较

Table 4 Comparison of several curves under square-sum criterion for $EX = 1000, C_v = 1.0$ and $C_s/C_v = 2$

曲线类型	百年一遇			千年一遇		
	平均相对均 方误差/%	平均设计值均值 的相对误差/%	平均设计值落在 ω 区间的频率	平均相对均 方误差/%	平均设计值均值 的相对误差/%	平均设计值落在 ω 区间的频率
P-Ⅲ	13.06	2.90	0.907	17.98	9.26	0.648
P-V	11.78	2.89	0.923	22.05	11.53	0.612
P-I	16.52	6.20	0.852	24.43	19.27	0.490
L-N	93.65	81.80	0.192	202.65	178.89	0.058
K-M	16.90	5.01	0.871	22.69	12.45	0.579
Gumbel	15.38	10.14	0.835	21.20	17.47	0.499

表 5 $EX = 1000, C_v = 0.5, C_s/C_v = 3$ 时绝对值准则下各种曲线的比较

Table 5 Comparison of several curves under absolute value criterion for $EX = 1000, C_v = 0.5$ and $C_s/C_v = 3$

曲线类型	百年一遇			千年一遇		
	平均相对均 方误差/%	平均设计值均值 的相对误差/%	平均设计值落在 ω 区间的频率	平均相对均 方误差/%	平均设计值均值 的相对误差/%	平均设计值落在 ω 区间的频率
P-Ⅲ	11.28	3.34	0.898	16.33	9.08	0.629
P-V	11.71	3.06	0.882	22.70	10.30	0.572
P-I	13.08	6.98	0.838	21.89	17.81	0.425
L-N	19.53	11.62	0.709	32.46	23.35	0.425
K-M	14.16	6.53	0.809	22.42	11.05	0.497
Gumbel	11.11	5.06	0.906	15.53	10.59	0.650

表 6 $EX = 1000, C_v = 0.5, C_s/C_v = 3$ 时平方和准则下各种曲线的比较

Table 6 Comparison of several curves under square-sum criterion for $EX = 1000, C_v = 0.5$ and $C_s/C_v = 3$

曲线类型	百年一遇			千年一遇		
	平均相对均方误差/%	平均设计值均值的相对误差/%	平均设计值落在 ω 区间的频率	平均相对均方误差/%	平均设计值均值的相对误差/%	平均设计值落在 ω 区间的频率
P-Ⅲ	9.24	1.77	0.956	13.27	6.93	0.731
P-V	8.86	1.76	0.949	16.17	6.14	0.711
P-Ⅰ	10.94	4.23	0.901	18.75	14.76	0.525
L-N	24.16	17.17	0.677	40.69	31.61	0.422
K-M	11.20	4.36	0.940	17.19	8.40	0.625
Gumbel	9.45	3.74	0.965	13.29	8.91	0.711

4 结 论

- a. 3 组参数情况下,从 3 种评价指标综合结果来看,P-Ⅲ 型的稳健性最好.
- b. 当偏态系数和离均系数增大时,设计值的平均相对均方误差和平均设计值均值的相对误差都相应增大,设计值落在 ω 区间的频率相应减少,说明曲线的稳健性在减弱.
- c. 两种适线准则,总体上讲,平方和准则优于绝对值准则.
- d. P-Ⅲ 型分布能较好地拟合水文资料系列.该型曲线既弹性足、精度高、稳定性好、抽样误差小,又使用方便,即可利用离均系数 Φ 值表计算频率 p 的设计值 x_p .从本文结果看,其稳健性优于其他几种曲线,因此 P-Ⅲ 型曲线是一种相对较好的用于水文频率计算的线型.

参考文献：

[1] 裴鹿成,王仲奇.蒙特卡罗方法及其应用[M].北京:海洋出版社,1998.15—57.

[2] 金光炎.水文频率分析述评[J].水科学进展,1999,10(3):319—327.

[3] 孙济良.关于洪水频率分析中的线型问题[J].水利水电技术,1987,(6):1—5.

[4] 李松仕.几种频率分布线型对我国洪水资料适应性的研究[J].水文,1984,(1):1—7.

[5] 徐向阳.暴雨频率线型的选择[D].南京:河海大学,1984.

[6] 金光炎.论水文频率计算中的适线法[J].水文,1990,(2):1—6.

Preliminary study on steadiness of several frequency curves

WANG Chun-xia¹, HUANG Zhen-ping², MA Jun-jian², HOU Yun-qing², LIN Xiao-li²
(1. Guangdong Provincial Investigation, Design and Research Institute
of Water Conservancy and Electric Power, Guangzhou 510170, China ;
2. College of Water Resources and Environment, Hohai Univ., Nanjing 210098, China)

Abstract :The steadiness of several frequency curves and their judgment criteria were studied based on the unbiasedness and effectiveness by use of the Monte-Carlo method. One thousand groups of samples (the capacity of each group being 60) extracted from 3 different populations of various distributions were calculated and judged to ensure the high universality of the research results. The research shows that, among the distributions commonly used, the P-III distribution is of high steadiness.

Key words :hydrological frequency calculation ;type of distribution, random sampling ;steadiness analysis