

网络出版时间:2017-08-21 13:50 DOI:10.13207/j.cnki.jnwafu.2017.10.018
网络出版地址:http://kns.cnki.net/kcms/detail/61.1390.S.20170821.1350.036.html

P-Ⅲ型分布序列无偏绘点位置计算研究

王誉杰,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 P-Ⅲ型分布序列无偏绘点位置计算公式。【方法】应用次序统计量数学期望值原理、双指数变换数值积分方法,求解 P-Ⅲ型分布绘点位置的数值解。在此基础上采用线性回归法概化出近似无偏绘点位置公式,并通过蒙特卡洛统计试验将文中近似无偏绘点位置公式与现有绘点位置公式的计算结果进行比较。【结果】用双指数变换数值积分可计算出 P-Ⅲ型分布任意绘点位置的高精度数值解。由此推导的近似无偏绘点位置公式具有较好的描述能力和预测能力。在陕西关中地区 36 个测站年降水数据频率计算和拟合优度评估中的应用表明,在概率点据相关系数、均方根误差 2 种评价标准下,推导公式拟合最优站数分别占总站数的 72.2% 和 69.4%。【结论】文中近似无偏绘点位置公式可应用于陕西关中地区降水频率的计算,且拟合度较好。

【关键词】 P-Ⅲ型分布;无偏绘点位置;蒙特卡洛统计试验;水文频率;关中地区
【中图分类号】 P333.9 **【文献标志码】** A **【文章编号】** 1671-9387(2017)10-0146-09

Calculation of unbiased plotting position for Pearson type-Ⅲ distribution

WANG Yujie, SONG Songbai

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

Abstract: 【Objective】 This study investigated unbiased plotting position formula for Pearson type-Ⅲ distribution. 【Method】 Expected values of order statistics and double exponential transform method for numerical integration were used to calculate the numerical value of plotting positions for Pearson type-Ⅲ. Then, linear regression was used to develop an approximate unbiased plotting position formula. The derived formula and existing formulas using Monte Carlo experiment were also compared. 【Result】 Numerical solutions of plotting positions of Pearson type-Ⅲ distribution can be calculated using the double exponential transform method for numerical integration. The unbiased plotting position formula was examined to be highly descriptive and predictive through statistical tests. Under the two evaluation criteria of PPCC and RMSE, annual precipitation data from 36 stations in Guanzhong was chosen for frequency calculation and the goodness of fit was evaluated. The best formula fitting stations were 72.2% and 69.4% out of the total stations, respectively. 【Conclusion】 The developed formula can be applied to hydrological frequency analysis of precipitation data in Guanzhong with good fitting.

Key words: P-Ⅲ distribution; unbiased plotting position; Monte Carlo experiment; hydrologic frequency; Guanzhong area

〔收稿日期〕 2016-09-02
〔基金项目〕 国家自然科学基金项目(51479171, 51179160, 50879070); 高等学校博士学科点专项科研基金项目(20110204110017)
〔作者简介〕 王誉杰(1991—), 女, 辽宁庄河人, 硕士, 主要从事流域水文模拟及水文预报研究。E-mail: 851526290@qq.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

水文频率计算是确定水利工程规模和运行管理的依据。水文频率分布参数估计有许多方法,如矩法、适线法和极大似然法等,实际工作中以适线法应用最为广泛。适线法计算成果的好坏决定于三个方面:一是总体分布线型的选取是否适当;二是绘点位置(也称经验频率)的计算是否合理;三是理论频率曲线与经验点据拟合好坏的判断准则选择是否恰当^[1-4]。因此,绘点位置的计算在适线法中具有至关重要的作用。

1946年, Kimball 首次推荐以数学期望作为特征值推求绘点位置公式,即无偏绘点位置公式^[5-6]。无偏绘点位置是指次序统计量 x_m 期望值 $E(x_m)$ 对应的概率 $P(E(x_m))$, 将 $P(E(x_m))$ 作为绘点位置计算得到的设计值具有无偏和方差最小属性, 因此通常称 $P(E(x_m))$ 为无偏绘点位置, 其与分布函数及分布参数有关。由于次序统计量期望值计算复杂, 不利于实际应用, 目前主要有两种获取途径: 其一是采用近似无偏绘点位置公式, 如 Blom^[7] 推出的正态分布的近似无偏绘点位置公式、Gringorten^[8] 推出的极值分布和指数分布的近似无偏绘点位置公式、Cunnane^[9] 推出的任意分布的近似无偏绘点位置公式、Nguyen 等^[10] 和 Zn-Na^[11] 推出的 P-Ⅲ分布偏态系数小于 3 的近似无偏绘点位置公式、Goel 等^[12] 推出的广义极值分布的近似无偏绘点位置公式; 其二是次序统计量数学期望值查数表, 如 Harter 等^[13] 给出的 11 个偏态系数、样本长度小于 40 的查数表, 谢自银等^[14] 给出的基于 P-Ⅲ型分布样本长度为 1 000、偏态系数 1 015 个的次序统计量期望值高精度计算结果, 刘九夫等^[15] 基于次序统计量不同样本长度期望值之间关系, 提出的样本长度 ≤ 1000 、偏态系数小于 10 的次序统计量期望值高精度计算框架。

从 20 世纪 50 年代开始, 我国一直沿用数学期望值公式计算绘点位置并进行 P-Ⅲ型分布参数适线法估算。一些学者认为, 在纵向离差 ($\sum \Delta x_m^2$ 或 $\sum |\Delta x_m|$) 最小的准则下适线, 根据期望值公式计算的绘点位置是有偏差的, 尤其对适线影响较大的前几项来说, 其偏差较一般公式大得多。期望值公式计算偏于安全, 实际工程设计中, 一般采用设计洪水值加上安全保证值, 这种计算导致设计洪水值偏大^[16-18]。因此, 在这种不合适的绘点位置上进行适线尚有待商榷, 研究 P-Ⅲ型无偏绘点位置公式对于保证水利工程安全运营、降低工程投资等具有重要的意义^[19]。

目前, 对于 P-Ⅲ型分布序列无偏绘点位置尚缺

乏细化和实用性研究, 本研究在吸收前人成果的基础上, 应用双指数变换数值积分方法求解出次序统计量期望值, 进而求解 P-Ⅲ分布序列绘点位置的高精度数值解, 概化出方便实际使用的无偏绘点位置公式, 并将文中模型应用于陕西关中地区降水资料的频率计算, 以期在水文计算中无偏绘点位置的计算提供理论支撑。

1 P-Ⅲ分布次序统计量期望值

1.1 P-Ⅲ型分布

P-Ⅲ分布的概率密度函数(PDF)为^[20]:

$$f(x) = \begin{cases} \frac{1}{\beta^\lambda \Gamma(\lambda)} e^{-(x-a_0)/\beta} (x-a_0)^{\lambda-1}, & a_0 \leq x < \infty \\ 0, & x < a_0 \end{cases} \quad (1)$$

式中: $f(x)$ 为概率密度函数, x 为随机变量, $\beta (\beta > 0)$ 为尺度参数, $\lambda (\lambda > 0)$ 为形状参数, $\Gamma(\lambda)$ 为伽马函数, $a_0 (-\infty < a_0 < \infty)$ 为位置参数。

P-Ⅲ分布的累计分布函数(CDF)为:

$$F(x) = \int_{a_0}^x f(x) dx \quad (2)$$

式中: $F(x)$ 为累计分布函数。

1.2 次序统计量的数学期望

假定随机变量为 x , 其累计分布函数为 $F(x)$, 概率密度函数为 $f(x)$, $x_1, x_2, \dots, x_m, \dots, x_n$ 为 x 的简单随机样本, 且独立同分布。若经过排序有 $x_1 \geq x_2 \geq \dots \geq x_m \geq \dots \geq x_n$, 显然, x_m 的取值完全由样本决定, 则 x_m 称为样本的第 m 项次序统计量^[21]。

次序统计量的密度函数 $g_m(x)$ 为:

$$g_m(x) = \frac{n!}{(m-1)! (n-m)!} [F(x)]^{n-m} [1-F(x)]^{m-1} f(x) \quad (3)$$

式中: n 为样本长度, m 为随机变量在样本中由大到小的排列次序。

次序统计量的分布函数 $G_m(x)$ 为:

$$G_m(x) = \frac{n!}{(m-1)! (n-m)!} \int_{-\infty}^x [F(x)]^{n-m} \times [1-F(x)]^{m-1} f(x) dx \quad (4)$$

次序统计量 x_m 的数学期望值 $E(x_m)$ 的表达式为:

$$E(x_m) = \int_{-\infty}^{+\infty} x g_m(x) dx = \frac{n!}{(m-1)! (n-m)!} \times \int_{-\infty}^{+\infty} x [F(x)]^{n-m} [1-F(x)]^{m-1} f(x) dx \quad (5)$$

将公式(5)进一步变换可以得到

$$E(x_m) = \int_{-\infty}^{+\infty} x g_m(x) dx = \frac{n!}{(m-1)! (n-m)!} \times$$

$$\int_0^1 x(F)[F]^{n-m}[1-F]^{m-1} dF. \quad (6)$$

2 次序统计量期望值数值计算

由于 P-III 分布次序统计量的期望值(如公式(6))涉及高次幂数值计算,用传统的高斯、辛普森等数值积分方法不能满足代数精度,计算结果偏差较大,因此本研究采用双指数变换数值积分方法进行计算。

2.1 双指数变换数值积分

设 (a, b) 为实直线上的一个区间(可以有限、无限或者半无限),定义在 (a, b) 上的函数 $f(x)$ 除过端点外是解析的,存在一个实数 c , 并令 $x = \varphi(t)$ 满足^[22-23], 则有:

$$a = \varphi(-\infty), b = \varphi(+\infty). \quad (7)$$

$$\frac{f(\varphi(t))\varphi'(t)}{\exp(-c\exp(|t|))} = O(|t| \rightarrow \infty). \quad (8)$$

式中: t 为中间变量; a, b 为实数; $\varphi(t)$ 为变换函数。则称该变换为双指数变换。

针对不同的积分区间 $[a, b]$, 有限、无限、半无限区间的变换函数 $\varphi(t)$ 分别为:

$$\int_a^b f(x) dx \Rightarrow x = \varphi(t) = \frac{a+b}{2} +$$

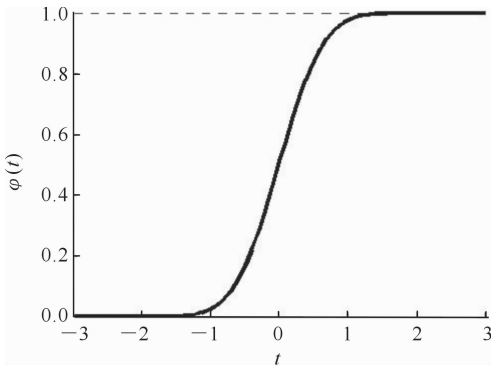


图 1 变换函数 $\varphi(t)$ 及其导数 $\varphi'(t)$ 的图像

Fig. 1 Transformation function $\varphi(t)$ and its derived function $\varphi'(t)$

由图 1 可知,当 $t \rightarrow -\infty$ 时, $\varphi(t)$ 很快收敛到 0; 当 $t \rightarrow \infty$ 时, $\varphi(t)$ 很快收敛到 1.0。当 $t \rightarrow \pm\infty$ 时, 导函数 $\varphi'(t)$ 以双指数的速度收敛到 0, 满足双指数变换数值积分的定义。

公式(6)经过双指数变换可以得到:

$$E(x_m) = \int_{-\infty}^{\infty} \{x[\varphi(t)]\} [1 - \varphi(t)]^{m-1} \varphi(t)^{n-m} \times \varphi'(t) dt. \quad (14)$$

根据梯形数值积分公式,式(14)无穷积分的等距网格离散公式为:

$$E(x_m) \approx \frac{n!}{(m-1)! (n-m)!} \times$$

$$\frac{1}{2} (b-a) \tanh\left(\frac{1}{2} \pi \sinh(t)\right). \quad (9)$$

$$\int_a^b f(x) dx \Rightarrow x = \varphi(t) = a + \exp\left(\frac{1}{2} \pi \sinh(t)\right). \quad (10)$$

$$\int_a^b f(x) dx \Rightarrow x = \varphi(t) = \sinh\left(\frac{1}{2} \pi \sinh(t)\right). \quad (11)$$

式中: $\sinh(t)$ 、 $\tanh(t)$ 分别表示双曲正弦函数和双曲正切函数。

2.2 次序统计量期望值的双指数变换数值积分

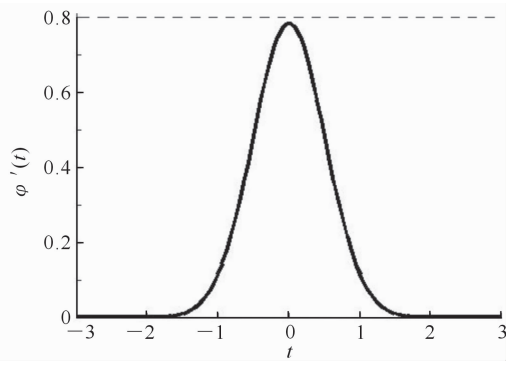
对公式(6)进行双指数积分变换,根据积分区间 $[0, 1]$, 选择第一种积分变换,得 $\varphi(t)$ 和 $\varphi'(t)$ 的函数式为:

$$\varphi(t) = \frac{1}{2} + \frac{1}{2} \tanh\left(\frac{\pi}{2} \sinh(t)\right). \quad (12)$$

$$\varphi'(t) = \frac{\pi \cosh(t)}{4 \cosh^2\left(\frac{\pi}{2} \sinh(t)\right)}. \quad (13)$$

式中: $\cosh(t)$ 表示双曲余弦函数。

变化函数 $\varphi(t)$ 和 $\varphi'(t)$ 与 t 的关系如图 1 所示。



$$h \sum_{k=-\infty}^{+\infty} \{x[\varphi(kh)]\} [1 - \varphi(kh)]^{m-1} \varphi(kh)^{n-m} \varphi'(kh). \quad (15)$$

式中: h 为网格尺度, k 为网格个数。

最后取 N 项截断,即计算到 N 项就可满足计算精度而停止计算,得到 x_m 的数学期望值的计算公式为:

$$E(x_m) \approx \frac{n!}{(m-1)! (n-m)!} \times h \sum_{k=-T}^T \{x[\varphi(kh)]\} [1 - \varphi(kh)]^{m-1} \times \varphi(kh)^{n-m} \varphi'(kh). \quad (16)$$

式中: $N=2T+1$,取 1,2,3…。

2.3 计算结果

为了验证双指数变换数值积分具有高精度的计算结果,同时利用高斯拉盖尔数值积分方法对 P-Ⅲ型分布次序统计量期望值进行数值计算,将 2 种数

值计算的结果与目前已被公认的 Harter 查数表结果进行对比,结果如表 1 所示。表 1 中误差(1)为高斯拉盖尔数值积分计算结果与 Harter 查数表数值的偏差值,误差(2)为双指数变换数值积分计算结果与 Harter 查数表数值的偏差值。

表 1 样本长度(n)为 100,偏态系数(C_s)为 2 时 2 种数值计算方法得到的次序统计量期望值与 Harter 查数表结果的对比

Table 1 Comparison of expected values of order statistics from two numerical integration methods and Harter table when n is 100 and C_s is 2

序号 Number	Harter	高斯拉盖尔 Gauss- Laguerre	双指数变换 Double- Exponential	误差 Error	
				(1)	(2)
1	5.187 38	5.185 57	5.187 38	0.001 81	0
2	4.187 38	4.192 61	4.187 38	-0.005 23	0
3	3.687 38	3.717 18	3.687 38	-0.029 80	0
4	3.354 04	3.335 53	3.354 04	0.018 51	0
5	3.104 04	2.988 74	3.104 04	0.115 30	0
10	2.358 41	2.448 67	2.358 41	-0.090 26	0
20	1.639 64	2.152 61	1.639 64	-0.512 97	0

序号 Number	Harter	高斯拉盖尔 Gauss- Laguerre	双指数变换 Double- Exponential	误差 Error	
				(1)	(2)
50	0.708 17	0.459 84	0.708 17	0.248 33	0
60	0.524 17	0.871 35	0.524 17	-0.347 17	0
70	0.368 83	0.064 47	0.368 83	0.304 36	0
80	0.234 40	0.483 44	0.234 40	-0.249 04	0
90	0.115 92	0.014 07	0.115 92	0.101 85	0
100	0.010 00	0.005 94	0.010 00	0.004 06	0

从表 1 可以看出,双指数变换数值积分方法计算精度较高,而用高斯拉盖尔数值积分方法计算的次序统计量期望值与 Harter 查数表偏差较大。因此,可用双指数变化数值积分方法计算 P-Ⅲ分布序列任意绘点位置的高精度数值解,即精确绘点位置。

3 无偏绘点位置通式概化

根据上述研究,可以通过双指数变化数值积分实现无偏绘点位置的数值计算,为了方便实际应用,本研究根据双指数变换法计算的绘点位置数值计算结果概化出一种新的公式,以近似替代绘点位置。Blom 提出了绘点位置公式的一般形式为:

$$P=\frac{m-a}{n-a'-a+1},a,a'<1。$$
 (17)

式中: P 为绘点位置, a,a' 分别为参数。

按式(17)进行近似绘点位置公式概化,即:

$$P=\frac{m+b_1}{n-a_1}。$$
 (18)

表 2 参数 $a_1、b_1$ 的线性回归结果

Table 2 Linear regression of a_1 and b_1

n	$C_s=0.1$		$C_s=3$	
	a_1	b_1	a_1	b_1
5	10.291 28	-5.654 54	10.983 63	-6.425 44
10	20.256 56	-10.637 23	21.005 06	-11.458 58
15	30.239 07	-15.628 52	31.017 84	-16.475 61
20	40.228 22	-20.623 11	41.026 49	-21.486 31
25	50.220 72	-25.619 37	51.032 80	-26.493 77
30	60.215 18	-30.616 61	61.037 66	-31.499 32
35	70.210 89	-35.614 47	71.041 53	-36.503 64
40	80.207 46	-40.612 76	81.044 70	-41.507 12
45	90.204 64	-45.611 35	91.047 36	-46.509 98
50	100.202 28	-50.610 17	101.049 63	-51.512 40

n	$C_s=0.1$		$C_s=3$	
	a_1	b_1	a_1	b_1
55	110.200 26	-55.609 17	111.051 59	-56.514 46
60	120.198 53	-60.608 30	121.053 30	-61.516 24
65	130.197 01	-65.607 54	131.054 81	-66.517 81
70	140.195 67	-70.606 87	141.056 16	-71.519 19
75	150.194 48	-75.606 28	151.057 37	-76.520 42
80	160.193 41	-80.605 74	161.058 46	-81.521 53
85	170.192 44	-85.605 26	171.059 46	-86.522 53
90	180.191 57	-90.604 83	181.060 36	-91.523 43
95	190.190 77	-95.604 43	191.061 19	-96.524 26
100	200.190 04	-100.604 06	201.061 96	-101.525 03

表 3 系数 c_1 、 c_2 、 c_3 、 c_4 的线性回归结果
Table 3 Linear regression of c_1 , c_2 , c_3 , and c_4

C_s	c_1	c_2	c_3	c_4	C_s	c_1	c_2	c_3	c_4
0.1	1.999 27	0.249 59	−0.999 63	−0.633 77	1.6	1.999 38	0.427 15	−0.999 75	−0.888 34
0.2	1.999 27	0.252 13	−0.999 64	−0.644 43	1.7	1.999 41	0.453 52	−0.999 78	−0.916 39
0.3	1.999 27	0.255 30	−0.999 64	−0.655 39	1.8	1.999 45	0.482 17	−0.999 81	−0.946 32
0.4	1.999 27	0.259 54	−0.999 64	−0.666 97	1.9	1.999 50	0.513 15	−0.999 84	−0.978 19
0.5	1.999 27	0.264 97	−0.999 65	−0.679 30	2.0	1.999 55	0.546 50	−0.999 89	−1.012 04
0.6	1.999 27	0.271 67	−0.999 65	−0.692 43	2.1	1.999 62	0.582 23	−0.999 94	−1.047 93
0.7	1.999 27	0.279 72	−0.999 65	−0.706 48	2.2	1.999 69	0.620 37	−1.000 00	−1.085 88
0.8	1.999 28	0.289 20	−0.999 66	−0.721 52	2.3	1.999 77	0.660 91	−1.000 06	−1.125 93
0.9	1.999 28	0.300 18	−0.999 66	−0.737 62	2.4	1.999 86	0.703 85	−1.000 13	−1.168 09
1.0	1.999 28	0.312 77	−0.999 67	−0.754 90	2.5	1.999 97	0.749 18	−1.000 21	−1.212 37
1.1	1.999 29	0.327 04	−0.999 67	−0.773 43	2.6	2.000 08	0.796 85	−1.000 30	−1.258 77
1.2	1.999 30	0.343 11	−0.999 68	−0.793 32	2.7	2.000 20	0.846 85	−1.000 39	−1.307 27
1.3	1.999 31	0.361 07	−0.999 69	−0.814 67	2.8	2.000 33	0.899 13	−1.000 50	−1.357 89
1.4	1.999 33	0.381 00	−0.999 71	−0.837 55	2.9	2.000 47	0.953 65	−1.000 61	−1.410 59
1.5	1.999 35	0.403 00	−0.999 73	−0.862 09	3.0	2.000 62	1.010 37	−1.000 73	−1.465 36

4 绘点位置公式优选

4.1 统计试验 1(检验公式描述能力)

比较本研究给出的近似无偏绘点位置公式与已有的 Weibull^[24]、Blom^[7]、Gringorten^[8]、Cunnane^[9]、Nguyen^[10] 公式,在相同的条件下筛选与绘点位置高精度数值解相对误差较小的公式。P-Ⅲ分布的统计参数给定为: $E(x)=100$ 、 $C_s=1$ 、 $C_s/C_v=2.5$ 和 $E(x)=100$ 、 $C_s=3$ 、 $C_s/C_v=3$ (其中 $E(x)$ 为数学期望, C_s 为偏态系数, C_v 为变差系数), n 分别取 40 和 60,计算结果见表 4 和表 5。

评价标准采用相对误差 $Bias$,有:

$$Bias_i=\left|\frac{P_i-Q_{i,j}}{P_i}\right|。$$

(20)

式中: P_i 为用双指数变换法数值积分计算出的第 i 个绘点位置高精度数值解, $Q_{i,j}$ 为用第 j 个绘点位置公式计算出的第 i 个绘点位置近似解。

由表 4 和表 5 可以看出,本研究推导的近似无偏绘点位置公式的计算结果与绘点位置高精度数值解的相对误差最小,其次为 Nguyen,我国长期使用的期望值公式与绘点位置高精度数值解的相对误差最大。

表 4 $E(x)=100$ 、 $C_s=1$ 、 $C_s/C_v=2.5$ 、 $n=40$ 时各公式与绘点位置数值解的相对偏差
Table 4 Bias between different formulas and exact plotting position for $E(x)=100$, $C_s=1$, $C_s/C_v=2.5$, $n=40$

序号 Number	数值解 Numerical solution	推导公式 Derived formula	Weibull	Blom	Gringorten	Cunnane	Nguyen
1	0.014 37	0.021 58	0.697 14	0.080 48	0.028 76	0.038 55	0.000 20
2	0.038 49	0.009 27	0.267 49	0.049 03	0.010 33	0.034 17	0.017 45
3	0.063 14	0.007 68	0.158 87	0.032 91	0.010 60	0.024 34	0.012 69
4	0.087 90	0.005 81	0.109 97	0.024 65	0.009 54	0.018 85	0.009 43
5	0.112 68	0.004 48	0.082 25	0.019 74	0.008 66	0.015 49	0.007 31
10	0.236 72	0.001 58	0.030 34	0.010 18	0.006 61	0.008 81	0.002 96
15	0.360 79	0.000 58	0.014 03	0.007 10	0.005 87	0.006 63	0.001 51
20	0.484 89	0.000 03	0.006 00	0.005 53	0.005 45	0.005 50	0.000 74
25	0.609 05	0.000 38	0.001 16	0.004 52	0.005 11	0.004 75	0.000 20
30	0.733 30	0.000 78	0.002 17	0.003 72	0.004 76	0.004 12	0.000 29
35	0.857 71	0.001 26	0.004 72	0.002 96	0.004 32	0.003 48	0.000 82
40	0.982 06	0.001 55	0.006 57	0.002 45	0.004 05	0.003 07	0.001 17
总和 Sum	19.900 70	0.002 10	0.043 79	0.010 41	0.006 40	0.008 32	0.002 37

表 5 $E(x)=100, C_s=3, C_s/C_v=3, n=60$ 时各公式与绘点位置数值解的相对偏差

Table 5 Bias between different formulas and exact plotting position, for $E(x)=100, C_s=3, C_s/C_v=3, n=60$

序号 Number	数值解 Numerical solution	推导公式 Derived formula	Weibull	Blom	Gringorten	Cunnane	Nguyen
1	0.009 08	0.057 40	0.804 58	0.141 90	0.025 36	0.097 14	0.047 51
2	0.024 88	0.004 47	0.317 70	0.083 96	0.042 85	0.068 17	0.041 84
3	0.041 13	0.007 12	0.195 72	0.059 28	0.035 28	0.050 06	0.029 16
4	0.057 48	0.006 51	0.140 83	0.046 74	0.030 20	0.040 39	0.021 88
5	0.073 87	0.005 66	0.109 68	0.039 23	0.026 84	0.034 47	0.017 30
10	0.155 95	0.003 12	0.051 21	0.024 39	0.019 67	0.022 57	0.007 89
15	0.238 07	0.002 18	0.032 91	0.019 63	0.017 29	0.018 73	0.004 81
20	0.320 15	0.001 82	0.024 10	0.017 40	0.016 23	0.016 95	0.003 41
25	0.402 19	0.001 74	0.019 01	0.016 22	0.015 73	0.016 03	0.002 72
30	0.484 16	0.001 82	0.015 79	0.015 58	0.015 54	0.015 56	0.002 39
35	0.566 05	0.002 01	0.013 64	0.015 26	0.015 54	0.015 37	0.002 29
40	0.647 86	0.002 29	0.012 16	0.015 16	0.015 68	0.015 36	0.002 36
45	0.729 57	0.002 64	0.011 15	0.015 21	0.015 92	0.015 48	0.002 54
50	0.811 19	0.003 03	0.010 46	0.015 37	0.016 23	0.015 70	0.002 80
55	0.892 77	0.003 38	0.009 93	0.015 53	0.016 51	0.015 91	0.003 04
60	0.975 39	0.002 62	0.008 42	0.014 59	0.015 68	0.015 01	0.002 18
总和 Sum	29.518 80	0.003 67	0.046 20	0.022 21	0.017 99	0.020 59	0.006 08

4.2 统计试验 2(检验公式的预测能力)

给定 P-Ⅲ型分布的参数值和重现期(T),利用 Matlab 中的 gaminv 语句可以计算出相应的理论值 Q_T 。给定样本容量,分别用不同绘点位置公式计算出对应的样本值 $X_{(i),j}$,利用概率权重矩法^[25]估计重现期的设计值 $\hat{Q}_{(T),j}$ 。P-Ⅲ型参数给定为: $E(x)=100, C_v=0.25, C_s/C_v=3, n=50$;重现期 T 为 50,100,200,500,1 000。

评价标准分别采用相对偏差、相对均方根误差。相对偏差(RB)。用来评价估计量的无偏性,若参数估计量偏大,则相对偏差为负值,反之则为正值。相对偏差的绝对值越小,表明该公式的无偏性

越好。相对偏差的表达式为:

$$RB_{(T)}=\frac{Q_T-\hat{Q}_{(T),j}}{Q_T}。$$

(21)

相对均方根误差。用来评价估计量的有效性,其值越小,则该公式的有效性越好。相对均方根误差的表达式为:

$$RMSRB_{(T)}=\sqrt{\frac{1}{M}\sum_{i=1}^M RB_{(T_i)}^2}。$$

(22)

式中: M 为统计试验所研究的重现期数量, T_i 表示第 i 个重现期。通过上述计算,在给定条件下各公式预测能力的评价结果见表 6。

表 6 给定条件下各公式对绘点位置的预测能力比较

Table 6 Predictive ability of different plotting position formulas under given conditions

评价标准 Criterion	重现期 Return period	Weibull	Blom	Gringorten	Cunnane	Nguyen	推导公式 Derived formula
RB	50	-0.015 01	-0.004 45	-0.002 68	-0.003 77	0.001 15	0.001 28
	100	-0.015 85	-0.006 10	-0.004 59	-0.005 51	0.001 51	0.001 36
	200	-0.016 23	-0.007 76	-0.006 63	-0.007 31	0.001 98	0.001 51
	500	-0.016 16	-0.009 97	-0.009 45	-0.009 74	0.002 73	0.001 80
	1 000	-0.015 76	-0.011 64	-0.011 63	-0.011 60	0.003 37	0.002 07
RMSRB		0.015 81	0.008 39	0.007 70	0.008 09	0.002 29	0.001 63

从表 6 可以看出,本研究近似无偏绘点位置公式的预测能力最好,其次是 Nguyen 公式和 Gringorten 公式。同时也可以看出,不论是公式的描述能力还是预测能力,我国一直沿用的 Weibull 经验频率公式的偏差均最大。

5 实例应用

选用陕西关中地区主要测站(宝鸡、宝鸡县、凤县、凤翔、扶风、麟游、眉县、岐山、武功、千阳、太白、陇县、渭南、蒲城、富平、临潼、蓝田、咸阳、户县、华

县、合阳、白水、潼关、韩城、周至、大荔、西安、澄城、礼泉、乾县、兴平、永寿、彬县、长武、淳化、华阴)的年降水资料,数据来自于陕西省气象局。经审查资料满足可靠性、一致性、代表性要求。

在前文的统计试验中,Gringorten 公式、Nguyen 公式和本研究推导公式的描述能力和预测能力均较好,且计算偏差与其他公式相比明显偏小。因此,将 Gringorten 公式、Nguyen 公式和本文推导公式应用于研究区数据的分析计算,优选出拟合陕西关中地区降水数据的最好公式。

根据研究区测站资料,采用概率权重矩法进行 P-Ⅲ型分布参数估计,计算研究区降水数据绘点位置高精度数值解。最后,用概率点据相关系数检验法(PPCC 检测法)和均方根误差(RMSE)检验法对 3 个公式进行评价。统计结果如表 7 和表 8 所示。

概率点据相关系数检测法(PPCC)中概率点据相关系数的计算公式为:

$$r=\frac{\sum (x_i-\bar{v})\sum (x_{pi}-\bar{\omega})}{\sqrt{\sum (x_i-\bar{v})^2\sum (x_{pi}-\bar{\omega})^2}}。$$

(23)

式中: r 为概率点据相关系数, x_i 为实测样本序列排序后的各观测值, $\bar{v}$ 为序列 x_i 的平均值, x_{pi} 为在假定理论分布上近似绘点位置 p_i 所对应的理论值, $\bar{\omega}$ 为序列 x_{pi} 的平均值。求出两个序列的相关系数,相关性系数越高,说明该公式的拟合效果越优。

均方根误差(RMSE)的计算公式为:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^n(\frac{x_i-x_{pi}}{x_i})^2}。$$

(24)

表 7 研究区数据中不同无偏绘点位置公式的概率点据相关系数检验结果

Table 7 Correlation coefficient of different plotting position formulas in the study area

测站 Station	Gringorten	Nguyen	推导公式 Derived formula	测站 Station	Gringorten	Nguyen	推导公式 Derived formula
宝鸡站 Baojizhan	0.992 65	0.991 05	0.993 08	富平 Fuping	0.996 95	0.993 86	0.997 06
宝鸡县 Baojixian	0.971 59	0.969 12	0.971 75	大荔 Dali	0.993 85	0.992 32	0.993 88
凤县 Fengxian	0.993 23	0.992 66	0.992 95	澄城 Chengcheng	0.987 86	0.991 43	0.986 30
凤翔 Fengxiang	0.986 46	0.981 02	0.987 32	白水 Baishui	0.990 81	0.990 21	0.990 63
扶风 Fufeng	0.986 52	0.976 75	0.987 44	西安 Xian	0.994 90	0.994 81	0.994 93
麟游 Linyou	0.986 89	0.983 99	0.987 01	周至 Zhouzhi	0.991 22	0.990 65	0.991 55
陇县 Longxian	0.992 95	0.989 66	0.993 19	临潼 Lintong	0.993 23	0.992 61	0.993 25
眉县 Meixian	0.992 44	0.989 89	0.992 24	蓝田 Lantian	0.995 83	0.995 49	0.996 08
岐山 Qishan	0.991 04	0.988 50	0.991 48	户县 Huxian	0.995 24	0.995 19	0.995 50
千阳 Qianyang	0.990 92	0.990 07	0.991 33	彬县 Binxian	0.995 87	0.994 74	0.995 34
太白 Taibai	0.992 07	0.990 25	0.990 90	长武 Changwu	0.994 79	0.993 41	0.994 84
渭南 Weinan	0.994 70	0.993 04	0.994 80	礼泉 Liquan	0.990 30	0.989 11	0.990 46
潼关 Tongguan	0.993 35	0.993 48	0.993 06	乾县 Qianxian	0.983 49	0.982 38	0.983 39
蒲城 Pucheng	0.988 26	0.983 35	0.987 99	武功 Wugong	0.986 76	0.983 00	0.987 29
华阴 Huayin	0.993 15	0.990 38	0.993 72	咸阳 Xianyang	0.978 30	0.952 94	0.979 78
华县 Huaxian	0.989 36	0.987 42	0.989 55	兴平 Xingping	0.993 80	0.992 61	0.993 81
合阳 Heyang	0.976 83	0.979 42	0.975 67	永寿 Yongshou	0.987 61	0.985 49	0.987 90
韩城 Hancheng	0.990 47	0.990 43	0.988 61	淳化 Chunhua	0.975 62	0.973 46	0.975 70

表 8 研究区数据中不同无偏绘点位置公式的均方根误差检测结果

Table 8 Root mean square error of different plotting position formulas in the study area

测站 Station	Gringorten	Nguyen	推导公式 Derived formula	测站 Station	Gringorten	Nguyen	推导公式 Derived formula
宝鸡站 Baojizhan	0.023 79	0.025 47	0.023 15	富平 Fuping	0.017 88	0.022 81	0.018 07
宝鸡县 Baojixian	0.231 13	0.228 64	0.225 21	大荔 Dali	0.035 86	0.037 32	0.035 49
凤县 Fengxian	0.021 47	0.023 73	0.021 70	澄城 Chengcheng	0.044 04	0.042 37	0.042 90
凤翔 Fengxiang	0.032 26	0.036 98	0.031 91	白水 Baishui	0.033 75	0.034 37	0.033 41
扶风 Fufeng	0.038 67	0.047 36	0.038 50	西安 Xian	0.023 56	0.023 61	0.023 29
麟游 Linyou	0.041 50	0.043 68	0.041 17	周至 Zhouzhi	0.042 56	0.043 27	0.041 94
陇县 Longxian	0.029 05	0.033 56	0.028 84	临潼 Lintong	0.035 91	0.036 73	0.035 21
眉县 Meixian	0.041 89	0.043 97	0.041 19	蓝田 Lantian	0.023 89	0.023 60	0.023 49
岐山 Qishan	0.029 96	0.032 78	0.029 83	户县 Huxian	0.019 16	0.018 75	0.019 49
千阳 Qianyang	0.031 30	0.031 98	0.030 54	彬县 Binxian	0.021 84	0.021 83	0.021 41
太白 Taibai	0.034 38	0.037 35	0.033 13	长武 Changwu	0.028 74	0.030 43	0.028 27
渭南 Weinan	0.033 30	0.034 76	0.033 05	礼泉 Liquan	0.042 01	0.043 83	0.042 36
潼关 Tongguan	0.025 72	0.025 76	0.025 23	乾县 Qianxian	0.050 73	0.052 11	0.050 98
蒲城 Pucheng	0.055 24	0.057 80	0.054 80	武功 Wugong	0.038 37	0.041 69	0.038 03
华阴 Huayin	0.038 14	0.041 57	0.037 17	咸阳 Xianyang	0.068 85	0.081 06	0.068 84
华县 Huaxian	0.045 36	0.046 95	0.045 83	兴平 Xingping	0.035 66	0.037 16	0.035 91
合阳 Heyang	0.041 53	0.041 06	0.042 14	永寿 Yongshou	0.042 39	0.043 97	0.042 63
韩城 Hancheng	0.050 30	0.052 55	0.051 50	淳化 Chunhua	0.158 23	0.160 03	0.160 53

表 7 和表 8 的统计结果表明:对于概率点据相关系数检验法,本文推导公式拟合效果最优,36 个测站中有 26 个测站的概率点据相关系数最高,占研究区测站数的 72.2%;对于均方根误差检验法而言,同样以本文推导公式的拟合效果最优,36 个测站中有 25 个测站的均方根误差最小,占研究区测站数的 69.4%。Gringorten 公式、Nguyen 公式在概率点据相关系数检验法和均方根误差检验法中,拟合效果均相对较差。

6 结 论

1)本研究分别用高斯拉盖尔数值积分和双指数变换数值积分计算了 P-Ⅲ型分布次序统计量期望值的数值解,通过与 Harter 数表的对比,表明与传统的高斯拉盖尔数值积分方法计算结果相比,双指数变换数值积分求解的 P-Ⅲ型分布次序统计量期望值具有更高的精度。

2)统计试验结果表明,现有的绘点位置计算公式中,Weibull 公式计算偏差较大,而且随偏态系数 C_s 的增大而加剧。Gringorten 公式是以 $E[F(x_m)]$ 为依据建立的极值 I 型分布的绘点位置公式,对于 P-Ⅲ型分布拟合效果不佳。Blom 公式为近似正态分布的无偏绘点位置公式,也不适用 P-Ⅲ型分布。Nguyen 公式是用概率加权矩和回归分析的方法得到的,这种推导方式使该公式在水文频率应用中存在一定的局限性,在本文统计试验中,虽然其比 Weibull 公式拟合优度稍好,但该公式在 C_s 较小时表现出的统计特性较好,而在 C_s 较大时产生的相对偏差较大。通过统计试验检验,本文推导的绘点位置公式在 $C_s=0.1\sim 3.0$ 时,其描述能力和预测能力均较好。

3)对陕西关中地区的实例应用计算结果表明,本文推导的近似无偏绘点位置公式计算结果的拟合效果较好。因此,本文近似无偏绘点位置公式可用于研究区降水经验频率的计算。

[参考文献]

- [1] 鲍振鑫. 水文频率分析适线法参数估计研究 [D]. 南京:南京水利科学研究院,2010.
Bao Z X. Parameter estimation on curve-fitting methodology in hydrologic frequency analysis [D]. Nanjing: Nanjing Hydraulic Research Institute,2010.
- [2] 齐哲娴,宋松柏. 广义极值分布序列经验概率的计算 [J]. 西北农林科技大学学报(自然科学版),2016,44(12):219-225.
Qi Z X, Song S B. Plotting position formula for general extreme value distribution [J]. Journal of Northwest A&F University (Natural Science Edition),2016,44(12):219-225.
- [3] Lozanoaguilera E D, Estudillomartinez M D, Castillogutiérrez S. A proposal for plotting positions in probability plots [J]. Journal of Applied Statistics,2014,41(1):118-126.
- [4] Kim S, Shin H, Ahn H, et al. Development of an unbiased plotting position formula considering the coefficient of skewness for the generalized logistic distribution [J]. Journal of Hydrology, 2015,527(23):471-481.
- [5] Kimball B F. Assignment of frequencies to a completely ordered set of sample data [J]. Eos Transactions American Geophysical Union,1946,27(6):843-846.
- [6] Kimball B F. Discussion of "Assignment of frequencies to a completely ordered set of sample data" [J]. Eos Transactions American Geophysical Union,1947,28(6):951-953.
- [7] Blom G. Statistical estimates and transformed beta-variables [M]. Stockholm:Almqvist & Wiksell,1958:174-182.
- [8] Gringorten I I. A plotting rule for extreme probability paper [J]. Journal of Geophysical Research,1963,68(3):813-814.
- [9] Cunnane C. Unbiased plotting positions-a review [J]. Journal of Hydrology,1978,37(3):205-222.
- [10] Nguyen V T V, In-Na N, Bobee B. New plotting-position formula for Pearson type-III distribution [J]. Journal of Hydraulic Engineering,1989,115(6):709-730.
- [11] In-Na N. An unbiased plotting position formula for the general extreme value distribution [J]. Journal of Hydrology,1989, 106(3):193-209.
- [12] Goel N K, De M. Development of unbiased plotting position formula for general extreme value distributions [J]. Stochastic Hydrology and Hydraulics,1993,7(1):1-13.
- [13] Harter H L. CRC handbook of tables for the use of order statistics in estimation [M]. New York:Crc Press,1996.
- [14] 谢自银,刘九夫,张建云. P-Ⅲ型分布次序统计量期望值数值计算 [J]. 水文,2005,25(6):24-29.
Xie Z Y, Liu J F, Zhang J Y. A numerical arithmetic for computing expected values of Pearson type-III order statistics [J]. Hydrology,2005,25(6):24-29.
- [15] 刘九夫,张建云. 水文概率 P-Ⅲ型分布无偏绘点位置的数值计算方法 [J]. 水利学报,2006,37(8):938-944.
Liu J F, Zhang J Y. Numerical arithmetic for computing unbiased plotting positions of hydrological probability statistics with Pearson type-III distribution [J]. Journal of Hydraulic Engineering,2006,37(8):938-944.
- [16] 郭生练,叶守泽. 论水文计算中的经验频率公式 [J]. 武汉水利电力学院学报,1992,25(2):38-45.
Guo S L, Ye S Z. Another look at plotting position formulae in hydrology [J]. Journal of Wuhan University of Hydraulic & Electric Engineering,1992,25(2):38-45.
- [17] 郭生练,刘章君,熊立华. 设计洪水计算方法研究进展与评价 [J]. 水利学报,2016,47(3):302-314.
Guo S L, Liu Z J, Xiong L H. Advances and assessment on design flood estimation methods [J]. Journal of Hydraulic Engi-

neering,2016,47(3):302-314.

[18] Kim S,Shin H,Joo K,et al. Development of plotting position for the general extreme value distribution [J]. Journal of Hydrology,2012,475(12):259-269.

[19] 潘军宁. 关于适线法中经验频率公式的讨论 [J]. 海洋工程, 1996,14(3):66-74.

Pan J N. Discussion about the plotting position formula used in curve-fitting method [J]. The Ocean Engineering,1996,14 (3):66-74.

[20] Ji X W,Jing D,Shen H W,et al. Plotting positions for Pearson type-III distribution [J]. Journal of Hydrology,1984,74(1): 1-29.

[21] 金光炎. 水文统计理论与实践 [M]. 南京:东南大学出版社, 2012.

Jin G Y. Hydrological statistical theory and practice [M]. Nanjing:Southeast University Press,2012.

[22] 赵延刚,陈鉴业,洪平海. 双指数变换及其应用 [J]. 甘肃科技,2009,25(14):192-193.

Zhao Y G,Chen J Y,Hong P H. Double exponential transform and its application [J]. Gansu Science and Technology, 2009,25(14):192-193.

[23] Takahasi H,Mori M. Double exponential formulas for numerical integration [J]. Publications of the Research Institute for Mathematical Sciences,1974,9(3):721-741.

[24] Weibull W. A statistical theory of the strength of materials [M]. Stockholm:Generalstabens Litografiska Anstalts Forlag,1939.

[25] Greenwood J A,Landwehr J M,Matalas N C,et al. Probability weighted moments,definition and relation to parameters of several distributions expressable in inverse form [J]. Water Resources Research,1979,15(5):1049-1054.

.....

(上接第 145 页)

[20] 谢 华,罗 强,黄介生. 基于三维 Copula 函数的多水文区丰枯遭遇分析 [J]. 水科学进展,2012,23(2):186-193.

Xie H,Luo Q,Huang J S. Synchronous asynchronous encounter analysis of multiple hydrologic regions based on 3D Copula function [J]. Advances in Water Science,2012,23(2):186-193.

[21] 侯芸芸,宋松柏,赵丽娜,等. 基于 Copula 函数的 3 变量洪水频率研究 [J]. 西北农林科技大学学报(自然科学版),2010, 38(2):219-228.

Hou Y Y,Song S B,Zhao L N,et al. Research of three trivariate flood frequency based on Copula function [J]. Journal of Northwest A&F University (Nat Sci Ed),2010,38(2):219-228.

[22] 郭生练,闫宝伟,肖 义,等. Copula 函数在多变量水文分析计算中的应用及研究进展 [J]. 水文,2008,28(3):1-7.

Guo S L,Yan B W,Xiao Y,et al. Multivariate hydrological analysis and estimation [J]. Journal of China Hydrology,2008, 28(3):1-7

[23] 陈 璐. Copula 函数理论在多变量水文分析计算中的应用研究 [M]. 武汉:武汉大学出版社,2013:47-49.

Chen L. The application analysis of multivariate hydrological on Copula connect function theory [M]. Wuhan:Wuhan University Press,2013:47-49.

[24] 丁志宏,冯 平,张 永. 基于 Copula 模型的丰枯频率分析:以南水北调西线工程调水区径流与黄河上来水的丰枯遭遇研究为例 [J]. 长江流域资源与环境,2010,19(7):759-764.

Ding Z H,Feng P,Zhang Y. Rich-poor frequency analysis of runoff based on Copula model;a case study on encounter probability of rich-poor runoff between rivers in area of Western route of South to North water transfer project and that of the upper reach of the Yellow River [J]. Resources and Environment in the Yangtze Basin,2010,19(7):759-764.