Journal of Hydrology 527 (2015) 471–481

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Development of an unbiased plotting position formula considering the coefficient of skewness for the generalized logistic distribution

Sooyoung Kim [1], Hongjoon Shin [1], Hyunjun Ahn [1], Jun-Haeng Heo *

School of Civil and Environmental Engineering, Yonsei University, Republic of Korea

ARTICLE INFO

Article history:
Received 12 January 2015
Received in revised form 23 April 2015
Accepted 1 May 2015
Available online 9 May 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Generalized logistic distribution
Order statistics
Plotting position formula
Genetic algorithms

SUMMARY

This study considered the plotting position formula with a coefficient of skewness for the generalized logistic distribution. For the development of the plotting position formula, the theoretical reduced variates were derived with consideration of the shape parameter of the generalized logistic distribution. The parameters of the plotting position formula were estimated using genetic algorithms. The accuracy of derived plotting position formula was examined using the error values between the theoretical and the calculated reduced variates from the derived and existing formulas. The error values from the derived plotting position formula were smaller than those from the existing formulas for $-0.30 \leqslant \beta < -0.05$ and $+0.05 < \beta \leqslant +0.30$. For $-0.05 \leqslant \beta \leqslant +0.05$, the error values from Gringorten's plotting position formula were smaller than those of other methods, but the differences were notably small, i.e., 0.0001–0.0008. As a result, the derived plotting position formula could be applied to the generalized logistic distribution with a shape parameter range of $-0.30 \leqslant \beta \leqslant +0.30$. In addition, the theoretical reduced variate shows a straighter line for sample data plotted on probability paper. And then, the coefficients of determination by the derived plotting position formula were higher than those by Gringorten's one for applied annual maximum rainfall data in Korea. Therefore, more reliable quantiles can be estimated using the derived plotting position formula.

1. Introduction

In conventional frequency analysis, the Gumbel, general extreme value (GEV), and log-Pearson type III distributions are generally used for annual maximum data. Many plotting position formulas have been developed for these probability distributions to graphically estimate the exceedance probabilities using sample data (Gringorten, 1963; Cunnane, 1978; Adamowski, 1981; Arnell et al., 1986; Nguyen et al., 1989; Nguyen and In-na, 1992; Goel and De, 1993; De, 2000; Kim et al., 2012).

Since Gringorten (1963) used the concept of mean square to develop the plotting position for extreme probability paper, Gringorten's plotting position formula has been applied generally to analyze the extreme values. In particular, this method is known as the best fit plotting position for extreme data based on the Gumbel distribution. Cunnane (1978) proposed the unbiased plotting positions, which represent the means of the order statistics from the reduced variate population. This concept had an impact on various future works for developing the unbiased plotting positions.

In addition to the commonly used probability distributions (i.e., the Gumbel, GEV, and log-Pearson type III distributions), various other probability distributions (e.g., generalized logistic, generalized Pareto, Wakeby, and Kappa) are commonly used as alternatives for frequency analysis under various conditions. For these probability distributions, the existing formulas (i.e., Gringorten) are applied to the illustrated estimation using the plotting positions. However, the existing plotting position formulas are not appropriate for other probability distributions because these existing formulas were developed for each specific distribution. Therefore, it is necessary to develop new plotting position formulas for other probability distributions.

Among the many probability distributions, the generalized logistic distribution is often used for frequency analysis of annual maximum hydrologic data, i.e., flood data. The generalized logistic distribution is an extended model of the two-parameter logistic distribution based on the generalization by Hosking and Wallis (1997). In Korea, Kim et al. (2004) applied various cluster techniques for regional frequency analysis and showed that the generalized logistic distribution was an appropriate fit for the annual

* Corresponding author. Tel.: +82 2 2123 2805.
E-mail addresses: sykim79@yonsei.ac.kr (S. Kim), sinong@yonsei.ac.kr (H. Shin), kamjakang@yonsei.ac.kr (H. Ahn), jhheo@yonsei.ac.kr (J.-H. Heo).
[1] Tel.: +82 2 393 1597.

http://dx.doi.org/10.1016/j.jhydrol.2015.05.002
0022-1694/

maximum rainfall data. In addition, the generalized logistic distribution was recommended for flood frequency analysis in UK by the Flood Estimation Handbook (Institute of Hydrology, 1999), and the plotting position formula proposed by Gringorten (1963) was also recommended for the generalized logistic distribution in the Flood Estimation Handbook (Institute of Hydrology, 1999). However, it is necessary to develop an appropriate plotting position formula for the generalized logistic model because the plotting position formula by Gringorten was originally developed for the Gumbel distribution.

In this study, a plotting position formula is developed for the generalized logistic distribution. For this purpose, the plotting position formula contains the coefficient of skewness related to the shape parameter and the mean of the order statistics from the theoretical reduced variates. In addition, the theoretical reduced variates are developed with consideration of the shape parameter. The parameters of the derived plotting position formula are estimated using the real-coded genetic algorithm (RGA), a type of genetic algorithm. To evaluate the accuracy of derived plotting position formula, the root mean square error, relative root mean square error, and relative bias are calculated for the theoretical and calculated reduced variates from the derived and existing formulas. In addition, a graphical display is constructed based on the theoretical reduced variates compared with the logistic reduced variates recommended by the Flood Estimation Handbook (Institute of Hydrology, 1999) using the derived and existing plotting position formulas. And then, the applicability of derived plotting position formula is investigated using the annual maximum rainfall data in Korea.

2. Derivation of the plotting position formula for the generalized logistic distribution

2.1. Generalized logistic distribution

The generalized logistic distribution is a special case of the Kappa distribution and a re-parameterized model of the log-logistic distribution (Ahmad et al., 1988). The cumulative distribution function (CDF) of the generalized logistic distribution is given by (Hosking and Wallis, 1997);

$$F(x) = \left[1 + \left\{1 - \frac{\beta}{\alpha}(x - \varepsilon)\right\}^{\frac{1}{\beta}}\right]^{-1} \tag{1}$$

where ε, α, and β are location, scale, and shape parameters, respectively, and the range of variable (x) for negative ($\beta < 0$) and positive ($\beta > 0$) shape parameters are given by $\varepsilon + \frac{\alpha}{\beta} \leqslant x < \infty$ and $-\infty < x \leqslant \varepsilon + \frac{\alpha}{\beta}$, respectively. The central moments for the generalized logistic distribution are defined by (Shin et al., 2010)

$$\mu_2 = \frac{\alpha^2}{\beta^2}\left\{\Gamma(1+2\beta)\Gamma(1-2\beta) - \Gamma^2(1+\beta)\Gamma^2(1-\beta)\right\} \tag{2}$$

$$\mu_3 = \frac{\alpha^3}{\beta^3}\left\{-\Gamma(1+3\beta)\Gamma(1-3\beta) + 3\Gamma(1+\beta)\Gamma(1-\beta)\Gamma(1+2\beta)\Gamma(1-2\beta) - 2\Gamma^3(1+\beta)\Gamma^3(1-\beta)\right\} \tag{3}$$

In addition, the relationship between the coefficient of skewness (γ) and shape parameter (β) is defined using the central moments in Eqs. (2) and (3) and is shown in Fig. 1.

$$\gamma = \frac{\mu_3}{\mu_2^{3/2}} \tag{4}$$

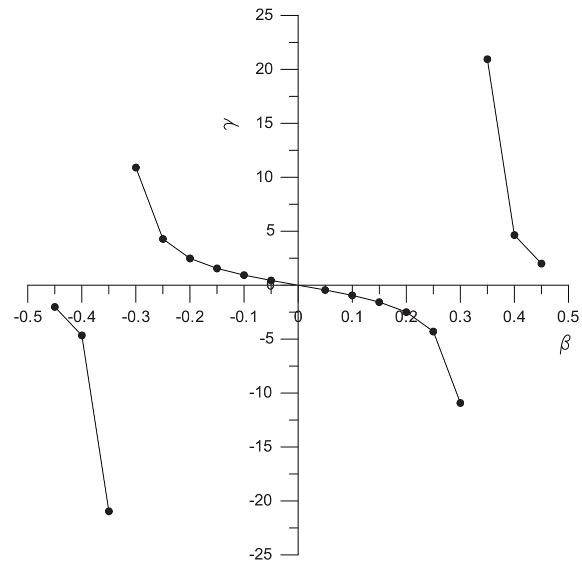

Fig. 1. Relationship between the coefficient of skewness (γ) and the shape parameter (β) for the generalized logistic distribution.

2.2. Theoretical derivation of the expected values of the order statistics

The probability density function (PDF) $g(x_r)$ of the rth order statistics (x_r) of sample (n) is defined as follows

$$g(x_r) = \frac{n!}{(r-1)!(n-r)!}F(x_r)^{r-1}\{1 - F(x_r)\}^{n-r}f(x_r) \tag{5}$$

where $f(x_r)$ and $F(x_r)$ are the density and distribution functions from the sample, respectively. The expected values of x_r are denoted as

$$E[x_r] = \frac{n!}{(r-1)!(n-r)!}\int_{-\infty}^{\infty} x_r F(x_r)^{r-1}\{1 - F(x_r)\}^{n-r}f(x_r)dx \tag{6}$$

If the shape parameter is negative ($\beta < 0$), the reduced variate y_1 is defined as follows

$$y_1 = 1 - \frac{\beta}{\alpha}(x - \varepsilon) \tag{7}$$

where the range of y_1 is $0 < y_1 < \infty$ for $\beta < 0$. The reduced variate in Eq. (7) substitutes into the CDF of Eq. (1), and the reduced variate for $\beta < 0$ is denoted as follows

$$y_1 = \left(\frac{1-F}{F}\right)^{\beta} \tag{8}$$

where F is the CDF for $\beta < 0$.

Table 1
Assumed plotting position formulas.

Case	Plotting position formula
1	$P_i = \frac{i+a\cdot\gamma+b}{n+c}$
2	$P_i = \frac{i+a}{n+b\cdot\gamma+c}$
3	$P_i = \frac{i+a\cdot\gamma+b}{n+c\cdot\gamma+d}$
4	$P_i = \frac{i+a\cdot\gamma^2+b\cdot\gamma+c}{n+d}$
5	$P_i = \frac{i+a\cdot\gamma^2+b\cdot\gamma+c}{n+d\cdot\gamma+e}$
6	$P_i = \frac{i+a}{n+b\cdot\gamma^2+c\cdot\gamma+d}$
7	$P_i = \frac{i+a\cdot\gamma+b}{n+c\cdot\gamma^2+d\cdot\gamma+e}$
8	$P_i = \frac{i+a\cdot\gamma^2+b\cdot\gamma+c}{n+d\cdot\gamma^2+e\cdot\gamma+f}$