

Use of beta- P distribution for modeling hydrologic events

Md. Sharwar Murshed^a, Yun Am Seo^b, Jeong-Soo Park^c, Youngsaeng Lee^{1,c}

^aSchool of Science and Engineering at Rajshahi Science and Technology University, Bangladesh;

^bApplied Meteorology Research Division, National Institute of Meteorological Science, Korea;

^cDepartment of Statistics, Chonnam National University, Korea

Abstract

Parametric method of flood frequency analysis involves fitting of a probability distribution to observed flood data. When record length at a given site is relatively shorter and hard to apply the asymptotic theory, an alternative distribution to the generalized extreme value (GEV) distribution is often used. In this study, we consider the beta- P distribution (BPD) as an alternative to the GEV and other well-known distributions for modeling extreme events of small or moderate samples as well as highly skewed or heavy tailed data. The L-moments ratio diagram shows that special cases of the BPD include the generalized logistic, three-parameter log-normal, and GEV distributions. To estimate the parameters in the distribution, the method of moments, L-moments, and maximum likelihood estimation methods are considered. A Monte-Carlo study is then conducted to compare these three estimation methods. Our result suggests that the L-moments estimator works better than the other estimators for this model of small or moderate samples. Two applications to the annual maximum stream flow of Colorado and the rainfall data from cloud seeding experiments in Southern Florida are reported to show the usefulness of the BPD for modeling hydrologic events. In these examples, BPD turns out to work better than beta- κ , Gumbel, and GEV distributions.

Keywords: extreme events, L-kurtosis, L-skewness, rainfall data, return level, stream flow data

1. Introduction

Estimations of the upper extreme quantiles corresponding to low probabilities of exceedance are needed in risk analysis, extreme value analysis, and reliability engineering (Castillo *et al.*, 2005). Some well-known distributions have been used to model the extreme values associated with the tail distribution. These include the generalized extreme value (GEV), generalized Pareto, generalized logistic (GLO), Pearson type-3 (PE3), beta- κ , and three-parameter log-normal (LN3) distributions. GEV distribution is widely used for modeling extreme events because it is supported by asymptotic theory; however, it has limitations when applied to small or moderate sample data (for example, Hosking and Wallis (1997)). In real flood frequency analysis in the United Kingdom, the aforementioned (GEV, GLO, PE3, LN3) distributions were selected most often as the best fitted model (Institute of Hydrology, 1999), which was confirmed by Kjeldsen and Prosdocimi (2015). A problem with this approach is that different distributions may be selected and applied for nearby sites, which may produce non-smooth quantile estimates over a region. Thus, to overcome this problem, it is desirable to use one flexible distribution that covers the above four models. The beta- P distribution (BPD) studied here is such a flexible model. Actually the L-moments ratio diagram (Hosking and Wallis, 1997) of

¹ Corresponding author: Department of Statistics, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju 61186, Korea. E-mail: yslee82@ejnu.net

the BPD turns out in this study to include those of GEV, GLO, LN3, and (part of) PE3 distributions as special cases.

The BPD and beta- κ distribution were introduced by Mielke and Johnson (1974) as special cases of the generalized second kind beta distribution. Wilks (1993) evaluated the performance of nine three-parameter distributions, including the aforementioned models, and found that the beta- κ distribution and BPD worked the best for extreme daily rainfall data in the United States. A detailed study of the beta- κ distribution was already reported by Murshed *et al.* (2011). However, no systematic study on the BPD has been reported. In this study, we focus on the systematic study of this distribution and its applicability for modeling hydrologic events. To estimate the parameters of the BPD, we use three methods: method of moments estimation (MME), L-moments estimation (L-ME), and maximum likelihood estimation (MLE). While the MLE is sensitive to outliers from the assumed population distribution especially for small samples, it is optimal for large samples. The L-ME proposed by Hosking (1990) has been shown to have robust performance compared with MME and MLE, particularly for small samples. The detailed derivations of the moments and L-moments for the BPD are provided in this study.

One goal of the present study is to illustrate the feasibility of the BPD in dealing with hydrologic events. Two real applications using the annual maximum stream flow data of Colorado and rainfall data from cloud seeding experiments in Southern Florida are illustrated. The comparison with other models such as the beta- κ , GEV, and Gumbel distributions shows that the BPD with the L-ME performs better than those distributions, particularly in small and skewed samples. This suggests that the BPD can be a useful alternative to GEV and other well-known distributions.

The remainder of the paper is organized as follows. Section 2 defines the BPD and presents its properties including moments. Section 3 describes the estimation methods. The simulation study is presented in Section 4. Section 5 illustrates the effectiveness of the distribution using real data examples. The discussion and conclusion are in Section 6.

2. Beta- P distribution

2.1. Definition

The BPD was given as a special case of the generalized second kind beta distribution (Mielke and Johnson, 1974). The cumulative distribution function (cdf) and the probability density function (pdf) of the BPD are

$$F(x) = 1 - \left[1 + \left(\frac{x}{\beta} \right)^\theta \right]^{-\alpha}, \quad x \geq 0, \quad (2.1)$$

$$f(x) = \frac{\alpha\theta}{\beta} \left(\frac{x}{\beta} \right)^{\theta-1} \left[1 + \left(\frac{x}{\beta} \right)^\theta \right]^{-(\alpha+1)}, \quad x > 0, \quad (2.2)$$

respectively, where $\alpha, \beta, \theta > 0$. Here, α and θ are the two shape parameters, and β is the scale parameter. There is no location parameter. The quantile function of the BPD is

$$x(F) = \beta \left[(1 - F)^{-\frac{1}{\alpha}} - 1 \right]^{\frac{1}{\theta}}, \quad 0 < F < 1. \quad (2.3)$$

Figure 1 illustrates the shapes of the beta- P pdf for some specified values of (α, θ) and $\beta = 1$ (because β is an invariant scale parameter). When $\alpha > 1$ and $\theta > 3$ hold, the pdf exhibits light tails (Figure 1(a)), but heavy tails (Figure 1(b)) when $0.5 \leq \alpha < 1.0$ and $0 < \theta \leq 3.5$ hold. When $\alpha = 1$,

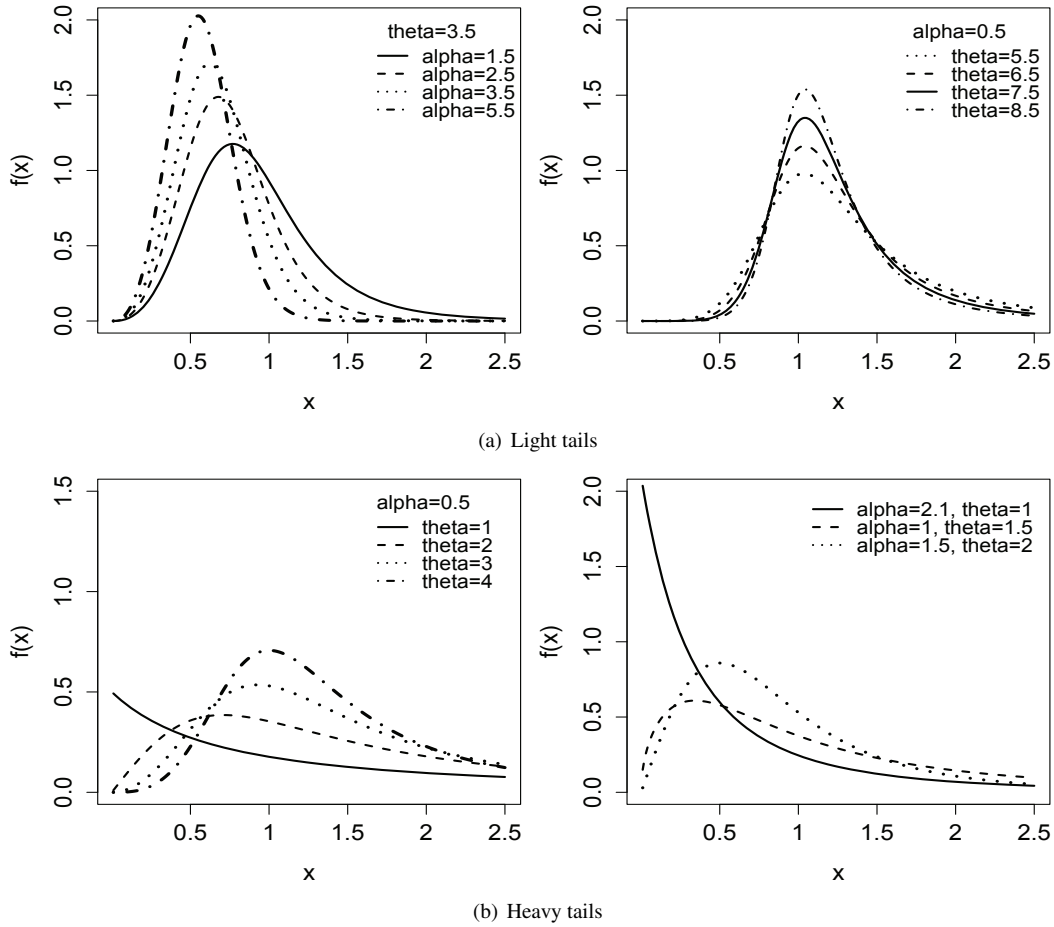


Figure 1: Shapes of the probability density functions of the beta-P distribution as the shape parameters α and θ change, but $\beta = 1$.

the BPD is the same as the GLO distribution (Ahmad *et al.*, 1988). The CDF of the GLO distribution is $F(x) = [1 + (x/\beta)^{-\theta}]^{-1}$, which is the same as that in (2.1) when $\alpha = 1$. The GLO distribution has been used widely for flood frequency analysis (Hosking and Wallis, 1997). As an generalized model of the GLO distribution, the BPD is also widely usable for flood frequency analysis. The closed form of the BPD provides better computational facility with the order statistic distribution than the gamma and log-normal distributions do. Sometimes, it is possible to be interested in the asymptotics of the sample extreme order statistics of the distribution. Identification of the feasible limiting distribution of the maxima (minima) is given in the Appendix.

2.2. Moments

Let μ'_r be the r^{th} order population moment about zero for the BPD. Then, μ'_r is given by

$$\mu'_r = E(X^r) = \alpha\beta^r B\left(1 + \frac{r}{\theta}, \alpha - \frac{r}{\theta}\right) = \alpha\beta^r h_r(\alpha, \theta), \quad \alpha\theta > r, \quad (2.4)$$

where $h_r(\alpha, \theta) = B(1 + r/\theta, \alpha - r/\theta)$, for $r = 1, 2, \dots$, and $B(\cdot)$ is a beta function. Note that the moments exist when $-\theta < r < \alpha\theta$.

2.3. L-moments

The parameter estimates obtained from L-moments are sometimes more accurate in small samples than using MME or MLE (Hosking, 1990). As a result, L-ME has received considerable attention for analyzing skewed data or extreme data, such as extreme rainfall, flood frequency, and extreme wind speed data. See Hosking (1990) for a definition of L-moments as well of its properties and applications. Let λ_r be the r^{th} L-moment of the BPD. To calculate λ_r , we need a special case of the r^{th} probability weighted moment (Greenwood *et al.*, 1979):

$$p_r = E[X\{1 - F(X)\}^r] = \alpha\beta B\left(r\alpha + \alpha - \frac{1}{\theta}, 1 + \frac{1}{\theta}\right), \quad (2.5)$$

for $r = 0, 1, 2, \dots$. Then, λ_r is obtained by using the relation between λ_r and p_r (Hosking, 1990), for $\alpha\theta > 1$ as,

$$\lambda_1 = p_0 = \alpha\beta k_1, \quad \lambda_2 = p_0 - 2p_1 = \alpha\beta(k_1 - 2k_2), \quad (2.6)$$

$$\lambda_3 = p_0 - 6p_1 + 6p_2 = \alpha\beta(k_1 - 6k_2 + 6k_3), \quad (2.7)$$

$$\lambda_4 = p_0 - 12p_1 + 30p_2 - 20p_3 = \alpha\beta(k_1 - 12k_2 + 30k_3 - 20k_4), \quad (2.8)$$

where

$$k_j = \Gamma\left(-\frac{1}{\theta} + j\alpha\right)\Gamma\left(1 + \frac{1}{\theta}\right) / \Gamma(j\alpha + 1), \quad (2.9)$$

where $\Gamma(\cdot)$ is a gamma function, for $j = 1, 2, 3, 4$.

A convenient way of representing the L-moments of different distributions is the L-moments ratio diagram, exemplified by Figure 2. This diagram shows the L-moments on a graph whose axes are L-skewness and L-kurtosis. A two-parameter distribution with a location and a scale parameter plots as a single point, while a three-parameter distribution with location, scale and shape parameters plots a line. Distributions with more than one shape parameters generally cover two-dimensional areas on the graph (Hosking and Wallis, 1997). The shaded area of Figure 2 represents the L-skewness and L-kurtosis values attained from the BPD for all possible combinations of α and θ . This figure was drawn by using the “lmrd” function of the R package “lmom” developed by Hosking (2014). Note that in Figure 2, the GEV and LN3 distributions are located inside the shaded area. The upper part of the shaded area includes the GLO distribution line achieved when $\alpha = 1$. The PE3 distribution is partly included in the area. This suggests that the beta- P model has wider applications than the aforementioned four (GEV, GLO, PE3, LN3) distributions have.

The L-moment ratio diagram is used as a tool for aiding in the identification of a suitable frequency distribution to model the available samples. It typically plots the sample L-kurtosis against the sample L-skewness values and compare these to equivalent theoretical relationships derived for a range of candidate distributions. The closeness of the sample values to the theoretical lines or areas can then be used as a selection criterion for the most appropriate type of distribution (Peel *et al.*, 2001). The points in Figure 2 are the sample L-skewness and sample L-kurtosis of the annual maximum daily precipitation for 57 weather stations of South Korea (Park *et al.*, 2011; Seo *et al.*, 2015). Most points lie within or close to the shaded region drawn by the BPD. To fit these data well, one may recommend using the BPD instead of using the GEV, GLO, PE3, or LN3 distributions separately.

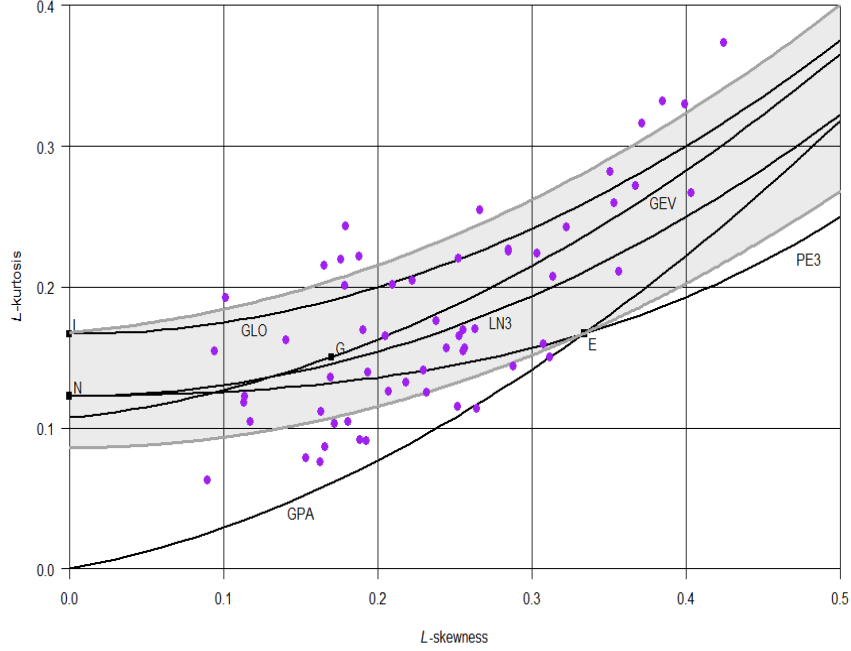


Figure 2: The L-moment ratio diagram for the BPD. The shaded area shows the L-skewness and L-kurtosis values attained from the BPD, in which the GEV and LN3 distributions are inside the area. The upper part of the shaded area includes the line for the GLO distribution, which is achieved when $\alpha = 1$ in the BPD. The points are the sample L-skewness and sample L-kurtosis of annual maximum daily rainfall for 57 weather stations in South Korea. BPD = beta-P distribution; GEV = generalized extreme value; LN3 = three-parameter log-normal; GLO = generalized logistic; E = exponential; G = Gumbel; L = logistic; N = normal; U = uniform; GPA = generalized Pareto; PE3 = Pearson type 3.

The higher-order L-moments, known as LH-moments (Wang, 1997), provide another way in which to estimate the parameters. This method has been used by Murshed *et al.* (2014), and Busabodhin *et al.* (2016). The LH-moments for the BPD are given in Murshed *et al.* (2009).

3. Estimation methods

3.1. Method of moments estimation

The MME can be obtained by equating the first three population moments of the BPD with the corresponding sample moments. Let m'_r be the r^{th} order sample moment about zero, given by $m'_r = n^{-1} \sum_{i=1}^n x_i^r$; $r = 1, 2, \dots$. To derive the estimates, we consider the following two functions of the first three population moments;

$$\frac{\mu'_2}{(\mu'_1)^2} = \frac{\Gamma(\alpha) \Gamma(1 + 2/\theta) \Gamma(\alpha - 2/\theta)}{[\Gamma(1 + 1/\theta) \Gamma(\alpha - 1/\theta)]^2} = g(\alpha, \theta), \quad \text{for } \alpha\theta > 2, \quad (3.1)$$

$$\frac{\mu'_3}{(\mu'_1)^3} = \frac{[\Gamma(\alpha)]^2 \Gamma(1 + 3/\theta) \Gamma(\alpha - 3/\theta)}{[\Gamma(1 + 1/\theta) \Gamma(\alpha - 1/\theta)]^3} = h(\alpha, \theta), \quad \text{for } \alpha\theta > 3. \quad (3.2)$$

The MME of α and β , say $\tilde{\alpha}$ and $\tilde{\beta}$, are obtained by equating these functions to the corresponding sample moment functions and thus by solving the following system of equations; $g(\tilde{\alpha}, \tilde{\theta}) = m'_2/(m'_1)^2$ and $h(\tilde{\alpha}, \tilde{\theta}) = m'_3/(m'_1)^3$. These equations are the functions of the two shape parameters, α and θ . It is solved by using a two-dimensional Newton-Raphson method. An estimate of the scale parameter ($\tilde{\beta}$) is obtained by substituting $\tilde{\alpha}$ and $\tilde{\theta}$ into another equation, which is

$$\tilde{\beta} = \frac{m'_1 \Gamma(\tilde{\alpha})}{\Gamma(1 + 1/\tilde{\theta}) \Gamma(\tilde{\alpha} - 1/\tilde{\theta})}.$$

3.2. Method of L-moments estimation

Let l_r be the r^{th} sample L-moment. Similar to the MME, the L-ME of α and β , say $\hat{\alpha}$ and $\hat{\beta}$, are obtained by equating the population L-moments (λ_r) to the corresponding sample L-moments. The sample L-skewness and sample L-kurtosis are written as $t_3 = l_3/l_2$ and $t_4 = l_4/l_2$, respectively. To obtain the L-ME, we solve the following system of equations with respect to α and θ : $\tau_3 = t_3$ and $\tau_4 = t_4$, under the restriction $\alpha\theta > 1$, where τ_3 and τ_4 are the population L-skewness and L-kurtosis, respectively. We again use a two-dimensional Newton-Raphson algorithm. An estimate of the scale parameter ($\hat{\beta}$) is obtained by substituting $(\hat{\alpha}, \hat{\theta})$ into (2.6), which is $\hat{\beta} = l_2/[\hat{\alpha}(k_1 - 2k_2)]$, where $k_1 = k_1(\hat{\alpha}, \hat{\theta})$ and $k_2 = k_2(\hat{\alpha}, \hat{\theta})$ are given as in (2.9).

3.3. Maximum likelihood estimation

For a given sample (x_1, x_2, \dots, x_n) from a BPD, the log-likelihood function is

$$\log L = n \log \alpha + n \log \theta + (\theta - 1) \left(\sum_{i=1}^n \log x_i \right) - n\theta \log \beta - (1 + \alpha) \sum_{i=1}^n \log \left[1 + \left(\frac{x_i}{\beta} \right)^\alpha \right], \quad (3.3)$$

for $\alpha > 0$, $\beta > 0$, and $\theta > 0$ as well as for $\min(x_1, \dots, x_n) > 0$. To compute the MLE, we maximize (3.3) numerically with respect to parameters. We used the Fortran program supplied by Professor Wilks. It uses the Levenberg-Marquardt method, which may be viewed as a generalization of the Newton-Raphson algorithm (Wilks, 1993).

4. Simulation study

We conducted Monte-Carlo simulations to investigate the performance of the three estimation methods (MME, L-ME, MLE) for high-quantile estimation. We generated random samples from the BPD by using the inverse transformation method. Sample sizes are $n = 20, 30, 50, 100$, and 200 . For each sample size, 2,000 trials were conducted.

To examine the performance of the estimation methods, we used two measures: the relative bias (Rbias) and relative root mean squared error (RRMSE). These are

$$\text{Rbias} = \frac{1}{M} \sum_{j=1}^M \left(\frac{q_j^e - q_j^t}{q_j^t} \right), \quad (4.1)$$

$$\text{RRMSE} = \sqrt{\frac{1}{M} \sum_{j=1}^M \left(\frac{q_j^e - q_j^t}{q_j^t} \right)^2}, \quad (4.2)$$

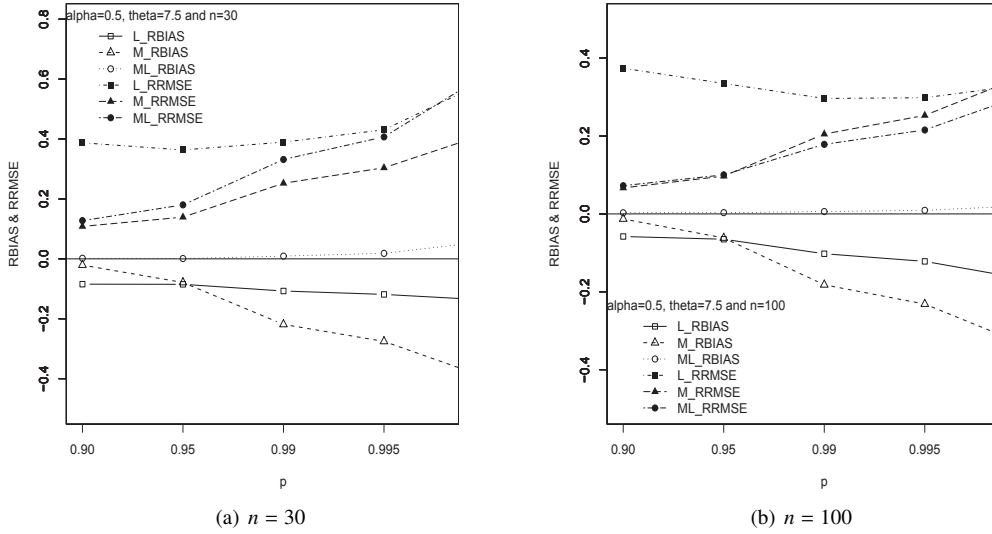


Figure 3: The Rbias and RRMSE plots of L-ME and MLE for different combinations of the shape parameters and sample sizes. Rbias = relative bias; RRMSE = relative root mean squared error; L-ME = L-moments estimator; MLE = maximum likelihood estimator.

where q_j^e and q_j^t are the estimated and true quantiles, respectively, and M is the successful convergence number among the 2,000 trials for each of the three estimation algorithms. The Rbias and RRMSE were calculated for two combinations of the shape parameters, namely $\alpha = 0.5$, $\theta = 3.5$ and $\alpha = 0.5$, $\theta = 7.5$, with a unit scale parameter. That is, we set $\beta = 1.0$ without loss of generality, because all estimation methods are scale equivariant for β . The Rbias and RRMSE were calculated for five upper quantiles, $q_{0.90}$, $q_{0.95}$, $q_{0.99}$, $q_{0.995}$, and $q_{0.999}$, using the above three methods. In addition, we used the average scaled absolute error (ASAE) to judge the overall goodness of fit, as in Castillo *et al.* (2005):

$$ASAE = \frac{1}{n} \sum_{i=1}^n \frac{|x_{(i)} - \hat{x}_{(i)}|}{x_{(n)} - x_{(1)}}, \quad (4.3)$$

where $x_{(i)}$ is the ascending i^{th} order observation and $\hat{x}_{(i)}$ are the p_i -quantile estimates obtained from the quantile function (2.3), for $p_i = i/(n+1)$. To calculate $\hat{x}_{(i)}$, the MLE or L-ME of the parameters was plugged into the quantile equation (2.3), as in (5.1) in the next section.

Figure 3 displays the upper-quantile Rbias and RRMSE for the samples of size $n = 30$, and 100. Figure 3(a) shows that most of the upper-quantile Rbias values of L-ME are smaller than those of MLE. The upper-quantile RRMSE values of L-ME are smaller than those of MLE for $p > 0.95$, particularly in the case of $n = 30$. The Figure 3(b) indicates that no one method is superior to the others in terms of both the Rbias and the RRMSE. No results were obtained for MME in the Figure 3(a) because the multiplication of the two shape parameters α and θ yields $\alpha\theta = 0.5 \times 3.5 = 1.75$, which violates the restriction of (3.1), $\alpha\theta > 2$. In the Figure 3(b), MME achieves good performance for the combination of parameters ($\alpha = 0.5$, $\theta = 7.5$). The performance of L-ME is better than that of MLE and MME, particularly in small samples. However, it sometimes fails to reach a solution for the equations, particularly for small samples. This may be partially because there is no solution inside

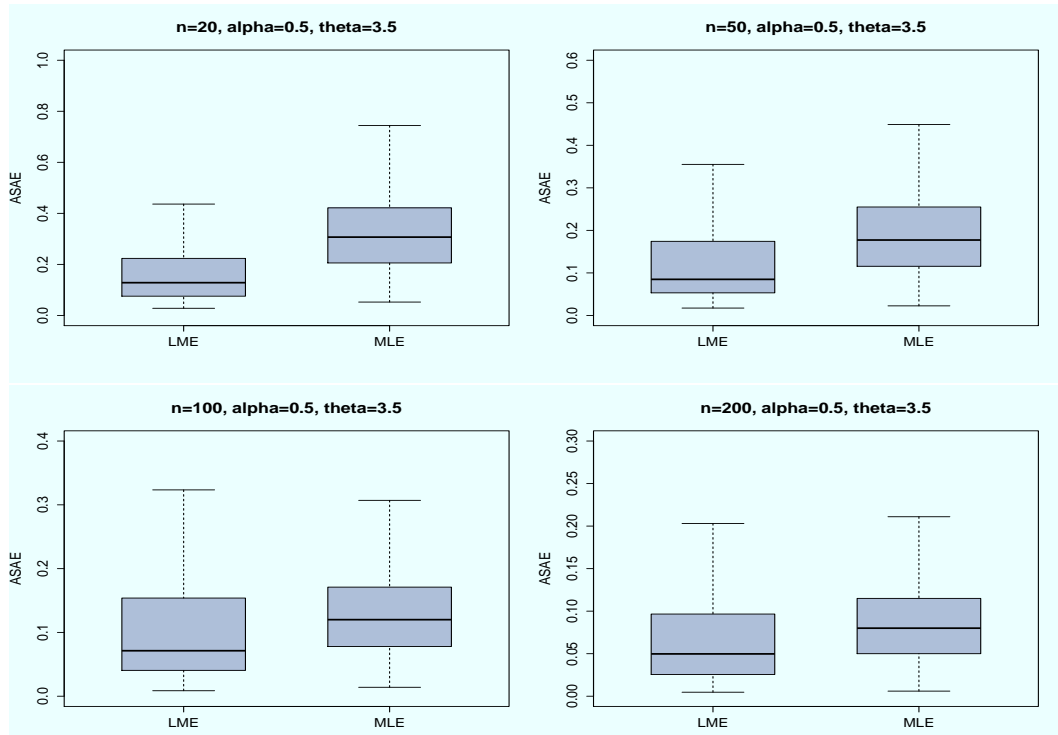


Figure 4: Boxplots of the ASAE values, using LME and MLE for various sample sizes ($n = 20, 50, 100, 200$) when the parameters are fixed as $\alpha = 0.5, \theta = 3.5$, and $\beta = 1.0$ for the heavy-tailed beta-P distribution. ASAE = average scaled absolute error; LME = L-moments estimator; MLE = maximum likelihood estimator.

the feasible region $\alpha\theta > 1$. For the moderate sample ($n = 100$), it is hard to see a global superiority between L-ME and MLE. For the large sample ($n = 200$, not presented in the figure), MLE works better than L-ME.

Figure 4 presents the boxplots of the ASAE, which were computed from 2,000 Monte-Carlo trials for $n = 20, 50, 100$, and 200 . This figure shows that L-ME performs better than MLE, particularly for small samples, which is consistent to the previous result for different distributions (Murshed *et al.*, 2011; Park *et al.*, 2009). In the combination ($\alpha = 0.5, \theta = 3.5$), no MME is reported because of the constraint (3.1). The overall performance of L-ME is superior to the other two methods, particularly for small sample and for skewed data.

5. Real data examples

5.1. Stream flow data in Colorado

This example is based on time series data that consists of annual maximum stream flow amounts in cubic feet per second. Data were obtained from the U.S. Geological Survey (USGS) station 06714310 (Sand Creek tributary at Denver, Colorado). It covers 22 years from May 1 to August 31 between 1971 and 1992; however, the sample size is 21 years because the records for 1989 are missing. The data is accessed through the USGS website, nwis.waterdata.usgs.gov (U.S. Geological Survey, 2010).

The MLE and L-ME of the respective parameters for some distributions (i.e., GEV, Gumbel, beta-

Table 1: Parameter estimates of the Gumbel, GEV, beta- P , and beta- κ distributions, using the L-ME and the MLE methods for annual maximum stream flow data of Colorado and rainfall data of Southern Florida.

Distribution	Parameter	Stream flow data		Rainfall data	
		MLE	L-ME	MLE	L-ME
GEV	$\hat{\mu}$	121.50	128.70	113.80	135.70
	$\hat{\sigma}$	84.10	98.40	145.80	184.90
	$\hat{\xi}$	0.46	-0.24	0.91	-0.53
Gumbel	$\hat{\mu}$	145.40	140.90	211.50	200.60
	$\hat{\sigma}$	109.40	130.00	309.60	418.20
beta- P	$\hat{\beta}$	242.00	259.40	973.20	1290.60
	$\hat{\alpha}$	1.71	1.83	3.10	3.80
	$\hat{\theta}$	1.93	1.87	0.88	0.82
beta- κ	$\hat{\beta}$	109.59	266.16	382.19	767.66
	$\hat{\alpha}$	1.69	0.39	0.53	0.23
	$\hat{\theta}$	1.89	2.89	1.51	2.02

GEV = generalized extreme value; L-ME = L-moments estimation; MLE = maximum likelihood estimation.

Table 2: K-S and A-D goodness-of-fit test statistics with p-values in brackets, using the Gumbel, GEV, beta- P , and beta- κ distributions via the MLE and L-ME

Estimation	Criterion	Distribution	Stream flow data	Rainfall data
MLE	K-S	GEV	0.134 (0.809)	0.111 (0.905)
		Gumbel	0.178 (0.467)	0.221 (0.159)
		beta- P	0.133 (0.811)	0.081 (0.996)
		beta- κ	0.171 (0.516)	0.089 (0.985)
	A-D	GEV	0.504 (0.741)	0.319 (0.923)
		Gumbel	0.525 (0.719)	2.042 (0.088)
		beta- P	0.385 (0.862)	0.213 (0.986)
		beta- κ	0.479 (0.765)	0.233 (0.979)
L-ME	K-S	GEV	0.134 (0.800)	0.087 (0.989)
		Gumbel	0.143 (0.733)	0.247 (0.083)
		beta- P	0.127 (0.845)	0.116 (0.870)
		beta- κ	0.176 (0.481)	0.147 (0.631)
	A-D	GEV	0.380 (0.867)	0.326 (0.917)
		Gumbel	0.513 (0.731)	2.167 (0.075)
		beta- P	0.372 (0.875)	0.276 (0.953)
		beta- κ	0.637 (0.611)	0.669 (0.583)

K-S = Kolmogorov-Smirnov; A-D = Anderson-Darling; GEV = generalized extreme value; MLE = maximum likelihood estimation; L-ME = L-moments estimation.

P , and beta- κ) and dataset combinations are displayed in Table 1. Two goodness-of-fit measures, namely Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D), along with the p -values in Table 2. It indicates that the BPD fits the data better than the GEV, Gumbel, and beta- κ distributions do. In addition, two information criteria, namely the AIC and BIC (Table 3). For these statistics, a lower value represents a better fit. This table shows that the BPD has better capability to fit the data than the GEV, Gumbel, and beta- κ distributions do.

Figure 5 shows the estimated return levels (in cubic feet per second), using the L-ME method, with a 95% confidence interval. Here, the T -year return levels are obtained by substituting parameter estimates into the quantile function (2.3) as the $(1 - 1/T)$ quantile of the distribution:

$$\hat{R}_T = \hat{\beta} \left[\left\{ 1 - \left(1 - \frac{1}{T} \right) \right\}^{-\frac{1}{\alpha}} - 1 \right]^{\frac{1}{\theta}}. \quad (5.1)$$

Table 3: AIC and BIC values for the Gumbel, GEV, beta- κ , and beta- P distributions, calculated from stream flow data of Colorado and rainfall data of Southern Florida, where k is the number of parameters in distribution and $\ln L$ is the log likelihood function value.

Data	Distribution	k	$\ln L$	AIC	BIC
Stream flow	GEV	3	-131.48	268.96	272.09
	Gumbel	2	-133.14	270.28	272.37
	beta-P	3	-131.49	268.98	272.11
	beta- κ	3	-131.61	269.22	272.35
Rainfall	GEV	3	-183.35	372.70	370.95
	Gumbel	2	-194.48	392.96	391.79
	beta-P	3	-181.58	369.16	367.41
	beta- κ	3	-181.69	369.38	373.15

AIC = Akaike information criterion; BIC = Bayesian information criterion; GEV = generalized extreme value.

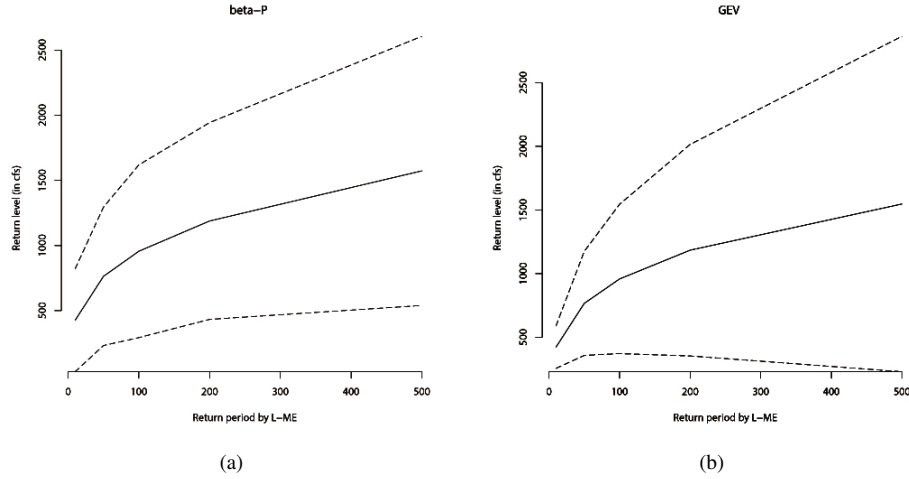


Figure 5: The approximated 95% confidence intervals (dotted lines) of the predictive return levels obtained by using L-ME for the beta- P distribution (a) and for the GEV distribution (b) for stream flow data of Colorado. The beta- P distribution provides a smaller confidence interval than the GEV distribution does. L-ME = L-moments estimator; GEV = generalized extreme value.

The T -year return level R_T is expected to be exceeded, on average, once every T years (Castillo *et al.*, 2005). The approximated 95% confidence interval of R_T is obtained using $(\hat{R}_T \pm 1.96 \times se_B)$, where se_B is the standard error of \hat{R}_T calculated by using the bootstrap method. In Figure 5, the BPD provides a shorter confidence interval than the GEV distribution does. Overall, we can say that the BPD describes the stream flow data better than the Gumbel, beta- κ and GEV distributions.

5.2. Rainfall data in Southern Florida

This example consists of 26 rainfall amounts (units in acre-feet) from seeded clouds of experiments conducted in Southern Florida (Simpson, 1972). It is based on radar-evaluated rainfall data from 52 Southern Florida cumulus clouds. Mielke and Johnson (1974) used this data to evaluate the performance of beta- P , beta- κ , gamma, and log-normal distributions.

Based on the K-S and A-D test criteria, the BPD is better in most cases (Table 2). In addition, the results from the AIC and BIC in Table 3 suggest the better fitting ability of the BPD to the data

compared with the GEV, beta- κ , and Gumbel distributions. The overall performance of the Gumbel distribution is worse than that of the other three distributions used in this study to describe both annual maximum stream flow and rainfall data.

6. Summary and discussion

In this study, we examined the beta- P distribution (BPD) for modeling hydrologic events. The BPD, which is a particular case of the generalized second kind beta distribution, seems to be a representative model to describe extreme rainfall data (Wilks, 1993). However, no systematic study of this topic has been conducted. We bridge this gap in the literature, using three estimation methods (MME, L-ME, and MLE), and find the following main results:

1. The presented simulation study suggests that the BPD via the L-ME is satisfactory.
2. The L-moment ratio diagram implies that the BPD has wider applications than the GLO, LN3, PE3, and GEV distributions.
3. Our findings suggest that the BPD can be a useful alternative to the GEV, GLO, PE3, and LN3 distributions for modeling (extreme) hydrologic events.

The BPD has no location parameters; rather, it has two shape parameters (α and θ) and one scale parameter (β). The lack of a location parameter can be a disadvantage of this distribution. Thus, adding a location parameter would form a four-parameter BPD, which might be an interesting topic of future research. In addition, the modified inverse moment estimation (Gui, 2016) can be applied to this distribution.

Acknowledgements

The authors gratefully acknowledge Professor Daniel Wilks at Cornell University for supplying the Fortran code for the MLE calculation. They also thank Tae Kyoung Kim for his help in computation. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2016R1A2B4014518). Seo's work was supported by the "Research and Development of KMA Weather Climate, and Earth System Services" of the National Institute of Meteorological Sciences (NIMS) of the Korea Meteorological Administration (KMA). Lee's work was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1A6A3A11032852).

Appendix: Asymptotics of extreme statistics of the beta- P distribution

If X_1, \dots, X_n are random samples from the beta- P distribution (BPD) and if $\bar{X}_n = (X_1 + \dots + X_n)/n$ denotes the sample mean, then, following the central limit theorem, $\sqrt{n}(\bar{X}_n - E(X))/\sqrt{\text{Var}(X)}$ converges in distribution to the standard normal distribution as $n \rightarrow \infty$, except when the population distribution possesses too heavy a tail. Sometimes, rather than the sample average, one would be interested in the corresponding asymptotics of the sample extreme statistics, $M_n = \max(X_1, \dots, X_n)$ and $m_n = \min(X_1, \dots, X_n)$. According to extreme value theory (e.g., Castillo *et al.*, 2005), only the feasible limit distribution for the maxima (minima) of any continuous distribution is the maximal (minimal) GEV distribution, which includes the well-known Weibull, Gumbel, and Fréchet distributions as special cases. Identifying the limiting distribution in the latter three distributions is known as the domain

of attraction problem, for the given cdf $F(x)$. To solve this problem for the BPD, we use Theorem 9.5 (for maxima) and Theorem 9.6 (for minima) in Castillo *et al.* (2005).

Suppose X_1, X_2, \dots, X_n are random samples of size n , drawn from a BPD. For the quantile function (2.3), we have

$$\lim_{\epsilon \rightarrow 0} \frac{x(1-\epsilon) - x(1-2\epsilon)}{x(1-2\epsilon) - x(1-4\epsilon)} = 2^{\frac{1}{\alpha\theta}}, \quad \alpha\theta > 0, \quad (6.1)$$

$$\lim_{\epsilon \rightarrow 0} \frac{x(\epsilon) - x(2\epsilon)}{x(2\epsilon) - x(4\epsilon)} = 2^{-\frac{1}{\theta}}, \quad \theta > 0, \quad (6.2)$$

where α and θ are the shape parameters of the associated limit distribution. The above quantities (6.1) and (6.2) are crucial points for determining the domain of attraction. Hence, Theorems 9.5 and 9.6 and (6.1) and (6.2) imply that the BPD belongs to the Fréchet maximal and Weibull minimal domains of attraction. The maxima (minima) of the BPD asymptotically forms the Fréchet (Weibull) distribution. The form of the norming constants a_n, b_n, c_n , and d_n can also be determined using Theorems 9.5 and 9.6 in Castillo *et al.* (2005). The possible selections of constants are

$$a_n = 0 \quad \text{and} \quad b_n = F^{-1}\left(1 - \frac{1}{n}\right) = \beta \left[\left\{ 1 - \left(1 - \frac{1}{n}\right) \right\}^{-\frac{1}{\alpha}} - 1 \right]^{\frac{1}{\theta}},$$

$$c_n = \inf\{x | F(x) > 0\} = 0 \quad \text{and} \quad d_n = x\left(\frac{1}{n}\right) - c_n = \beta \left[\left(1 - \frac{1}{n}\right)^{-\frac{1}{\alpha}} - 1 \right]^{\frac{1}{\theta}}.$$

References

- Ahmad MI, Sinclair CD, and Werritty A (1988). Log-logistic flood frequency analysis. *Journal of Hydrology*, **98**, 205–224.
- Busababodhin P, Seo YA, Park JS, and Kumphon B (2016). LH-moment estimation of Wakeby distribution with hydrological applications. *Stochastic Environmental Research and Risk Assessment*, **30**, 1757–1767.
- Castillo E, Hadi AS, Balakrishnan N, and Sarabia JM (2005). *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley-Interscience, Hoboken, NJ.
- Greenwood JA, Landwehr JM, Matalas NC, and Wallis JR (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form, *Water Resources Research*, **15**, 1049–1054.
- Gui W (2016). Modified inverse moment estimation: its principle and applications, *Communications for Statistical Applications and Methods*, **23**, 479–496.
- Hosking JRM (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics, *Journal of the Royal Statistical Society Series B (Methodological)*, **52**, 105–124.
- Hosking JRM (2014). A R package ‘lmom’, Retrieved May 15, 2016, from: <http://cran.r-project.org/web/packages/lmom/index.html>
- Hosking JRM and Wallis JR (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press, Cambridge.
- Institute of Hydrology (1999). *Flood Estimation Handbook: Vol. 5. Catchment Descriptors*, Institute of Hydrology, Wallingford.

- Kjeldsen TR and Prosdocimi I (2015). A bivariate extension of the Hosking and Wallis goodness-of-fit measure for regional distributions, *Water Resources Research*, **51**, 896–907.
- Mielke PW and Johnson ES (1974). Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology, *Water Resources Research*, **10**, 223–226.
- Murshed MS, Kim S, and Park JS (2011). Beta- κ distribution and application to hydrologic events, *Stochastic Environmental Research and Risk Assessment*, **25**, 897–911.
- Murshed MS, Park BJ, Jeong BY, and Park JS (2009). LH-moments of some distributions useful in hydrology, *Communications for Statistical Applications and Methods*, **16**, 647–658
- Murshed MS, Seo YA, and Park JS (2014). LH-moment estimation of a four parameter kappa distribution with hydrologic applications, *Stochastic Environmental Research and Risk Assessment*, **28**, 253–262.
- Park JS, Kang HS, Lee Y, and Kim MK (2011). Changes in the extreme daily rainfall in South Korea, *International Journal of Climatology*, **31**, 2290–2299.
- Park JS, Seo SC, and Kim TY (2009). A kappa distribution with a hydrological application, *Stochastic Environmental Research and Risk Assessment*, **23**, 579–586.
- Peel MC, Wang QJ, Vogel RM, and McMahon TA (2001). The utility of L-moment ratio diagrams for selecting a regional probability distribution, *Hydrological Sciences Journal*, **46**, 147–155.
- Seo YA, Lee Y, Park JS, Kim MK, Cho CH, and Baek HJ (2015). Assessing changes in observed and future projected precipitation extremes in South Korea, *International Journal of Climatology*, **35**, 1069–1078.
- Simpson J (1972). Use of the gamma distribution in single-cloud rainfall analysis, *Monthly Weather Review*, **100**, 309–312.
- U.S. Geological Survey (2010). Stream flow data of Colorado from U.S. Geological Survey, Retrieved April 30, 2016, from: <http://nwis.waterdata.usgs.gov/co/nwis/>
- Wang QJ (1997). LH moments for statistical analysis of extreme events, *Water Resources Research*, **33**, 2841–2848.
- Wilks DS (1993). Comparison of three-parameter probability distributions for representing annual extreme and partial duration precipitation series, *Water Resources Research*, **29**, 3543–3549.

Received June 18, 2017; Revised August 4, 2017; Accepted August 21, 2017