Thailand Statistician
July 2020; 18(3): 333-339
http://statassoc.or.th
Contributed paper

An Alternative Estimator with Appropriate Plotting Position Estimates for the Generalized Exponential Distribution

Theerapong Kaewprasert [a] and Manad Khamkong*[b]
[a] Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand.
[b] Data Science Research Center, Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand.
*Corresponding author; e-mail: manad.k@cmu.ac.th

Received: 10 March 2019
Revised: 26 May 2019
Accepted: 8 July 2019

Abstract

In this paper, we propose an alternative method for estimating generalized exponential distributions by applying plotting positions of modified percentile estimates. We compared its efficiency with the classical maximum likelihood estimator and percentile estimator in terms of root mean square errors. Simulation results show that the percentile estimator outperformed the others for small sample sizes while the proposed estimator was most effective for medium to large sample sizes. This finding was supported by applying the proposed estimator to deduce whether Thai rainy season rainfall data followed a generalized exponential distribution from a rain gauging station in Fang district, Chiang Mai Province, Thailand.

Keywords: Modified percentile method, maximum likelihood estimation, rainfall data, root mean square error, simulation.

1. Introduction

For some time now, the study of lifetime data has been applied to research in various fields such as engineering (failure of a mechanical system) and medical research (patient survival times, etc.) These data can be arranged in a number of appropriate distribution such as exponential, Weibull, and gamma, among others. Exponential distributions are used in the study of time left until an incident occurs, such as the waiting period until the first customer comes into an office, the waiting period until the first phone call, the duration of a lamp until the bulb fails, and so on. However, the properties of an exponential distribution limit it to only one parameter: the scale parameter. Subsequently, Gupta and Kundu (1999) proposed a distribution that resolves this limitation: the generalized exponential (GE) distribution consisting of an additional parameter representing the shape as well as the original scale parameter, which is more appropriate for real world situations.

Later, Gupta and Kundu (2001) reported on parameter estimation methods for the GE distribution and compared their performances. They found that the percentile estimator (PCE) method was appropriate for small sample sizes whereas the maximum likelihood estimator (MLE) method was better for medium to large sample sizes. Although the PCE is more efficient than the MLE method with small sample sizes, medium to large sample sizes caused a problem. Therefore, our interest is in modifying the PCE method based on an appropriate plotting position estimator to overcome this by following the guidelines selecting a suitable parametric estimation method for a generalized exponential distribution. Furthermore, the proposed estimator was applied to estimate whether Thai rainy season rainfall data from Fang, Chiang Mai province, Thailand follows a GE distribution.

2. Generalized Exponential Distribution

Suppose that a random variable X has an exponential distribution with scale parameter λ and a cumulative distribution function (cdf) as following

$$F(x) = 1-e^{-\lambda x}, \quad x>0, \lambda>0. \tag{1}$$

A generalization of the exponential distribution can be obtained by introducing a shape parameter α. Thus, X is said to have a generalized exponential distribution with parameters (α,λ), or X is distributed as $\text{GE}(\alpha,\lambda)$, if it has the cdf

$$F(x;\alpha,\lambda) = \left(1-e^{-\lambda x}\right)^{\alpha}, \quad x>0, \alpha, \lambda>0, \tag{2}$$

and a probability density function (pdf)

$$f(x;\alpha,\lambda) = \alpha\lambda\left(1-e^{-\lambda x}\right)^{\alpha-1} e^{-\lambda x}, \quad x>0, \alpha, \lambda>0. \tag{3}$$

The mean and variance of the GE distribution are given by

$$E(X)=\frac{1}{\lambda}\left(\psi(\alpha+1)-\psi(1)\right) \text{ and } Var(X)=\frac{1}{\lambda^2}\left(\psi'(1)-\psi'(\alpha+1)\right),$$

where $\psi(x)=\frac{d}{dx}\ln\left(\Gamma(x)\right)=\frac{\Gamma'(x)}{\Gamma(x)}$ and $\psi'(x)$ denotes the derivative of $\psi(x)$.

However, it is often difficult to fit an appropriate distribution to observed data. In particular, many authors have studied the goodness of fit to select an appropriate model. Stephens (1974) found that the Anderson-Darling (AD) test statistic as the power of a test was better than other tests and recommended its use when selecting models. AD is a test statistic that uses the goodness of fit for data on an ordinal scale when the distribution of information is continuous. The Anderson and Darling (1954) test is derived as

$$\text{AD} = -n-\frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\ln F(x_j)+\ln(1-F(x_{n-j+1}))\right], \tag{4}$$

where $F(.)$ is the expected cdf, x_j are the ordered data and n is the sample size.

3. Methods of Estimation

3.1. The maximum likelihood estimate (MLE) method

Let $X_1, X_2, ..., X_n$ be a random sample of size n from a $\text{GE}(\alpha,\lambda)$ distribution with pdf as (3). The likelihood function of the $\text{GE}(\alpha,\lambda)$ distribution is given by

$$L(\alpha,\lambda) = \prod_{i=1}^{n}\alpha\lambda\left(1-e^{-\lambda x_i}\right)^{\alpha-1} e^{-\lambda x_i}. \tag{5}$$

Subsequently, the log-likelihood function can be written as

$$\ln L(\alpha,\lambda) = n\ln(\alpha)+n\ln(\lambda)+(\alpha-1)\sum_{i=1}^{n}\ln(1-e^{-\lambda x_i})-\sum_{i=1}^{n}\lambda x_i. \tag{6}$$

The MLE estimates for the GE parameters are obtained by taking the partial derivatives of the log-likelihood function with respect to α and λ, respectively. Subsequently, the MLE estimators of α and λ are obtained by equating the resulting expressions to zero as follows:

$$\frac{\partial}{\partial\alpha}\ln L(\alpha,\lambda) = \frac{n}{\alpha}+\sum_{i=1}^{n}\ln\left(1-e^{-\lambda x_i}\right) = 0 \tag{7}$$

$$\frac{\partial}{\partial\lambda}\ln L(\alpha,\lambda) = \frac{n}{\lambda}+(\alpha-1)\sum_{i=1}^{n}\frac{x_i e^{-\lambda x_i}}{\left(1-e^{-\lambda x_i}\right)}-\sum_{i=1}^{n}x_i = 0. \tag{8}$$

In the paper by Gupta and Kundu (2001), the MLE of λ by applying the fixed point solution and obtained by using an iterative procedure is obtained as

$$\hat{\lambda}_{MLE} = \left[\frac{\sum_{i=1}^{n}\frac{x_i e^{-\hat{\lambda}_{MLE}x_i}}{\left(1-e^{-\hat{\lambda}_{MLE}x_i}\right)}}{\sum_{i=1}^{n}\ln\left(1-e^{-\hat{\lambda}_{MLE}x_i}\right)}+\frac{1}{n}\sum_{i=1}^{n}\frac{x_i e^{-\hat{\lambda}_{MLE}x_i}}{\left(1-e^{-\hat{\lambda}_{MLE}x_i}\right)}+\frac{1}{n}\sum_{i=1}^{n}x_i\right]^{-1} \tag{9}$$

and the MLE with respect to α can be written as

$$\hat{\alpha}_{MLE} = -\frac{n}{\sum_{i=1}^{n}\ln\left(1-e^{-\hat{\lambda}_{MLE}x_i}\right)}. \tag{10}$$

3.2. The percentile (PCE) method

Suppose that $X_1, X_2, ..., X_n$ are a random sample from a GE distribution population, then the p^{th} quantile of X is given by

$$x_p = -\frac{1}{\lambda}\ln\left(1-p^{(1/\alpha)}\right). \tag{11}$$

Furthermore, the plotting position has the following formula (Gringorten 1963) as

$$p_i = \frac{i-a}{n+1-2a}, \tag{12}$$

where p_i is the plotting probability of the i^{th} order statistic, n is the sample size and parameter a of the plotting position is a constant depending on the distribution. p_i is an approximation of $F(x_{(i)};\alpha,\lambda)$ where $i=1,2,...,n,$ thus PCE uses p_i in the case where $a=0$ (Weibull 1939) so that

$$p_{i:PCE} = \frac{i}{n+1}. \tag{13}$$

The PCE estimator of λ and α is derived by minimizing

$$\sum_{i=1}^{n}\left[x_{(i)}+\lambda^{-1}\ln\left(1-p_{i:PCE}^{(1/\alpha)}\right)\right]^2. \tag{14}$$

Subsequently, the estimate of the parameter λ becomes

$$\hat{\lambda}_{PCE} = \frac{\sum_{i=1}^{n}\left[\ln\left(1-p_{i:PCE}^{(1/\hat{\alpha}_{PCE})}\right)\right]^2}{\sum_{i=1}^{n}x_{(i)}\ln\left(1-p_{i:PCE}^{(1/\hat{\alpha}_{PCE})}\right)}, \tag{15}$$

and with respect to parameter $\alpha,$ becomes

$$\hat{\alpha}_{PCE} = \frac{\sum_{i=1}^{n} \ln\left(p_{i:PCE}\right) \ln\left(1-e^{-x_{(i)}}\right)}{\sum_{i=1}^{n}\left[\ln\left(1-e^{-x_{(i)}}\right)\right]^2}. \tag{16}$$

Gu and Yue (2013) proposed a method for estimating GE distribution by the inverse moment estimator (IME) method and comparing the MLE method. For small sample size, the IME method is more appropriate than the MLE method, while the PCE estimator outperformed the others for small sample size.

Although the PCE and IME methods were appropriate for small sample size, but the IME methodology has a more complex and complicated procedure than the PCE method. Thus, our interest is in modifying the PCE estimator based on the plotting position.

3.3. Proposed method: the modified percentile (MPE) method

The MPE estimator is developed by changing plotting position estimator p_i in the PCE estimator from case $a = 0$ to $a = 0.25$ (Adamowski 1981), i.e.

$$p_{i:MPE} = \frac{i-0.25}{n+0.5}. \tag{17}$$

Consequently, the estimate of the parameter λ becomes

$$\hat{\lambda}_{MPE} = \frac{\sum_{i=1}^{n}\left[\ln\left(1-p_{i:MPE}^{(1/\hat{\alpha}_{MPE})}\right)\right]^2}{\sum_{i=1}^{n} x_{(i)} \ln\left(1-p_{i:MPE}^{(1/\hat{\alpha}_{MPE})}\right)}, \tag{18}$$

and with respect to parameter α, becomes

$$\hat{\alpha}_{MPE} = \frac{\sum_{i=1}^{n} \ln(p_{i:MPE}) \ln\left(1-e^{-x_{(i)}}\right)}{\sum_{i=1}^{n}\left[\ln\left(1-e^{-x_{(i)}}\right)\right]^2}. \tag{19}$$

4. Numerical Analysis

4.1. Simulated data

To assess the performance of the estimators: MLE, PCE, and the proposed MPE, a simulation study to generate random samples from a $\text{GE}(\alpha, \lambda)$ distribution with different parameter values $(\alpha, \lambda) = \{(0.5,1), (1.0,1), (1.5,1), (2.0,1)\}$ was conducted using the R statistical program (R Core Team 2016). All of the experiments were performed for several sample sizes (n = 10, 20, 30, 50, and 100) based on 10,000 replications. The criteria for comparing the performance of the estimators was the sample root mean square error (RMSE) and absolute bias ($|\text{BIAS}|$), with the smallest RMSE and $|\text{BIAS}|$ indicating the best performance. We calculated RMSE and $|\text{BIAS}|$ as

$$RMSE = \sqrt{\frac{1}{10{,}000} \sum_{i=1}^{10{,}000} \left[(\alpha_i - \hat{\alpha}_i)^2 + (\lambda_i - \hat{\lambda}_i)^2\right]}, \tag{20}$$

$$|\text{BIAS}| = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} \left(\left|\alpha_i - \hat{\alpha}_i\right| + \left|\lambda_i - \hat{\lambda}_i\right|\right). \tag{21}$$

The results from Table 1 and Figure 1 show an efficiency comparison of the parameter estimation methods for a GE distribution. In all of the cases for the same sample size, the RMSE value increased as α increased. For small sample sizes (n = 10, 20), the PCE estimator outperformed the others. The MPE performed the best for medium and large sample sizes (n = 30, 50, 100), which is consistent with Gupta and Kundu (2001).

Moreover, when estimating the parameters of $\mathrm{GE}(\alpha,\lambda)$, the MLE estimator overestimated the actual parameters, PCE underestimated them, and MPE overestimated α and underestimated λ.

Table 1 Simulation results of the parameter estimation of $\mathrm{GE}(\alpha,\lambda)$

n	λ	α	Method MLE			Method PCE			Method MPE		
			$\hat{\lambda}$	$\hat{\alpha}$	RMSE	$\hat{\lambda}$	$\hat{\alpha}$	RMSE	$\hat{\lambda}$	$\hat{\alpha}$	RMSE
10	1	0.5	1.453	0.665	48.213	1.017	0.567	6.950*	1.207	0.656	25.944
		1.0	1.299	1.457	54.622	0.946	1.090	10.517*	1.090	1.295	30.871
		1.5	1.255	2.374	90.999	0.931	1.637	15.295*	1.059	1.984	48.796
		2.0	1.234	3.423	144.181	0.927	2.211	22.311*	1.047	2.732	73.303
20	1	0.5	1.191	0.567	20.198	0.927	0.506	7.333*	1.053	0.564	8.341
		1.0	1.132	1.171	21.564	0.915	0.984	8.612*	1.013	1.107	10.740
		1.5	1.113	1.802	32.287	0.917	1.469	8.845*	1.003	1.663	16.346
		2.0	1.104	2.457	46.886	0.919	1.959	9.041*	0.999	2.235	23.527
30	1	0.5	1.120	0.541	12.651	0.913	0.488	8.754	1.015	0.536	3.913*
		1.0	1.084	1.102	13.207	0.918	0.961	9.112	0.995	1.057	5.718*
		1.5	1.072	1.678	19.238	0.923	1.438	9.882	0.991	1.587	8.701*
		2.0	1.066	2.267	27.455	0.927	1.926	10.450*	0.990	2.126	12.657
50	1	0.5	1.067	0.523	7.017	0.914	0.476	8.978	0.990	0.513	1.619*
		1.0	1.047	1.056	7.316	0.925	0.946	9.224	0.983	1.017	2.403*
		1.5	1.040	1.598	10.548	0.932	1.417	10.756	0.983	1.526	3.095*
		2.0	1.037	2.145	14.932	0.936	1.892	12.504	0.983	2.045	4.761*
100	1	0.5	1.031	0.510	3.252	0.928	0.472	7.713	0.980	0.498	1.982*
		1.0	1.022	1.025	3.311	0.942	0.945	7.985	0.981	0.994	2.026*
		1.5	1.019	1.543	4.715	0.949	1.418	9.645	0.982	1.491	2.028*
		2.0	1.017	2.064	6.649	0.952	1.893	11.718	0.983	1.989	2.042*

Note: *The RMSE value was the smallest in this case.

4.2. Rainfall data

Thai rainy season rainfall data (in mm) for four months (February to May) from 1957 to 2015 (n = 59) were obtained from the Bureau of Water Management and Hydrology Royal Irrigation Department Thailand (2016). This real-life dataset was employed to compare the point estimates of the GE distribution parameters mentioned in the previous sections. The AD test statistic and corresponding p-values were used as criteria to fit the models and compare the performance of the estimators. The method of estimating $\mathrm{GE}(\alpha,\lambda)$ parameters in this way indicates that the method with the smallest AD value and highest p-value performed the best (Ghitany et al. 2017, Hussain et al. 2017).

Table 2 Comparison results of $\mathrm{GE}(\alpha,\lambda)$ parameter estimation from rainy season rainfall data, Fang district, Chiang Mai province

n	Method	AD	p-value	$\hat{\alpha}$	$\hat{\lambda}$
59	MLE	0.5639	0.6821	204.3866	0.0069
	PCE	0.5672	0.6789	203.0811	0.0069
	MPE	0.5348	0.7109	243.7547	0.0071

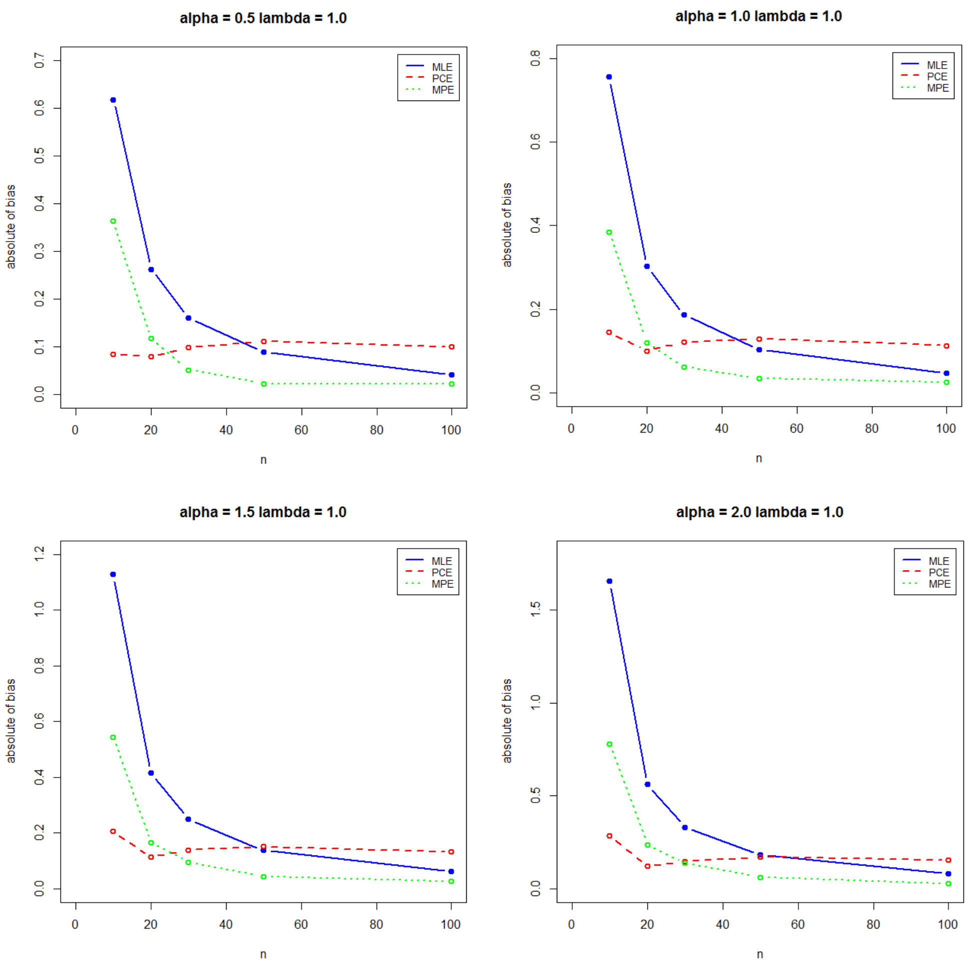

Figure 1 Absolute bias of the parameter estimation of $\text{GE}(\alpha,\lambda)$

We report the AD test statistic at 0.05 significance level for three methods: MLE, PCE and MPE and found that the GE distribution was appropriate for the Thai rainy season rainfall data in Table 2 and Figure 2. Furthermore, the MPE estimator had the lowest AD value (0.5348) and the highest p-value (0.7109) with $\hat{\alpha}$ = 243.7547 mm and $\hat{\lambda}$ = 0.0071 mm.

5. Conclusions

The classical PCE and the MLE methods are usually used to estimate the parameters from a GE distribution with small sample size and medium to large sample size, respectively. In this study, a new parameter estimation method for a GE distribution was derived and compared in a variety of ways. The proposed MPE estimator is a modification of the PCE method based on an appropriate plotting position estimator. In simulation studies, it outperformed the other methods for medium to large sample sizes and the parameter estimates α and λ were closer to the actual parameter values than the MLE method in all situations. In addition, parameter estimation of a $\text{GE}(\alpha,\lambda)$ distribution using the rainy season rainfall data from the Fang district, Chiang Mai Province, Thailand supported these findings. We recommend the proposed MPE estimator as a useful tool for determining GE distribution parameters using medium to large sample size.

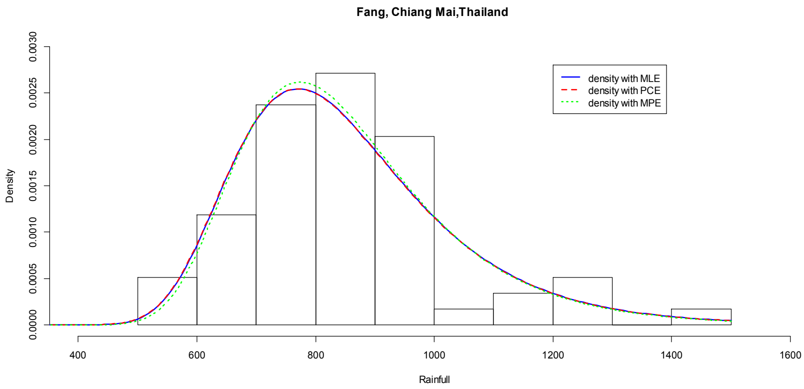

Figure 2 Comparison of histogram and theoretical densities of $\mathrm{GE}(\alpha, \lambda)$ parameter estimation from rainy season rainfall data, Fang district, Chiang Mai province

Acknowledgements

The authors would like to thank the two referees and the editor for their constructive comments and suggestions which improved of the manuscript. This research was supported by Chiang Mai University.

References

Adamowski K. Plotting formula for flood frequency. Water Resour Bull. 1981; 17: 197-202.

Anderson TW, Darling DA. A test of goodness of fit. J Am Stat Assoc. 1954; 49: 765-769.

Bureau of Water Management and Hydrology Royal Irrigation Department Thailand. Rainfall data from the Fang rain gauge station in Chiang Mai. 2016 [cited 2016 Jul 13]. Available from: http://hydro-1.net.

Ghitany ME, Song P, Wang S. New modified moment estimators for the two-parameter weighted lindley distribution. J Stat Comput Simul. 2017; 87: 3225-3240.

Gringorten II. A plotting rule for extreme probability paper. J Geophys Res. 1963; 68: 813-814.

Gu BQ, Yue RX. Some notes on parameter estimation for generalized exponential distribution. Commun Stat-Theory Methods. 2013; 42: 1787-1805.

Gupta RD, Kundu D. Generalized exponential distributions. Aust N Z J Stat. 1999; 41: 173-188.

Gupta RD, Kundu D. Generalized exponential distribution: different methods of estimations. J Stat Comput Simul. 2001; 69: 315-337.

Hussain T, Bakouch HS, Iqbal ZA. New probability model for hydrologic event: properties and applications. J Agric Biol Environ Stat. 2017; 23: 63-82.

R Core Team. R: language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2016.

Stephens MA. EDF statistics for goodness of fit and some comparisons. J Am Stat Assoc. 1974; 69: 730-737.

Weibull W. A statistical theory of the strength of materials. Stockholm: Generalstabens Litografiska Anstalts Förlag, 1939.