International Journal of Applied Operational Research
Vol. 4, No. 1, pp. 61-68, Winter 2014

Journal homepage: www.ijorlu.ir

Confidence interval for the two-parameter exponentiated Gumbel distribution based on record values

M. Abdi*

Received: 31 July 2013 ; **Accepted:** 4 December 2013

Abstract In this paper, we study the estimation problems for the two-parameter exponentiated Gumbel distribution based on lower record values. An exact confidence interval and an exact joint confidence region for the parameters are constructed. A simulation study is conducted to study the performance of the proposed confidence interval and region. Finally, a numerical example with real data set is given to illustrate the proposed procedures.

Keywords Confidence Interval, Exponentiated Gumbel Distribution, Exact Joint Confidence Region, Record Values.

1 Introduction

The Gumbel distribution is perhaps the most widely applied statistical distribution for climate modeling. Some of its application areas in climate modeling include: global warming problems, flood frequency analysis, rainfall modeling, and wind speed modeling. A recent book by Kotz and Nadarajah [1], which describes this distribution, lists over 50 applications, ranging from accelerated life testing through to earthquakes, floods, horse racing, rainfall, sea currents, wind speeds and track race records (to mention just a few).

In literature, exponentiated family of distributions defined in two ways. If $G(x;\sigma)$ is cumulative distribution function (cdf) of a baseline distribution then by adding one more parameter (say θ), the cdf of exponentiated baseline distribution $F(x;\theta,\sigma)$ is given by

a) $F(x;\sigma,\theta)=\left[G(x;\sigma)\right]^{\theta}, \quad \theta>0, \sigma\in\mathbb{C}, x\in R,$

b) $F(x;\sigma,\theta)=1-\left[1-G(x;\sigma)\right]^{\theta}, \quad \theta>0, \sigma\in\mathbb{C}, x\in R,$

where $\mathbb{C}$ is parameter space.

Gupta et al. [2] introduced the exponentiated exponential (EE) distribution as a generalization of the exponential distribution. The two parameters EE distribution associated with definition (a), have been studied in detail by Gupta and Kundu [3] which is a sub-model of the exponentiated Weibull distribution, introduced by Mudholkar and Shrivastava [4]. Nadarajah [5] introduced exponentiated Gumbel distribution using (b). Some of its application areas in climate modeling include global warming problem, flood frequency analysis, offshore modeling, rainfall modeling and wind speed modeling. Persson and Rydén [6], discussed

* **Corresponding Author.** (✉)
E-mail: me.abdi@bam.ac.ir. (M. Abdi)

M. Abdi
M.Sc, Department of Mathematics and Computing, Higher Education Complex of Bam, Bam, Iran.

estimation of T-year return values for significant wave height in a case study and compare point estimates and their uncertainties to the results given by alternative approaches using Gumbel or Generalized Extreme Value distributions.

A random variable X is said to have Gumbel distribution, if its cdf is

$$G(x;\sigma)=e^{-e^{-\frac{x}{\sigma}}}, \quad x>0,\sigma>0,$$

By introducing a shape parameter $\theta>0$ and using definition (a), the cdf of the exponentiated Gumbel distribution is

$$F(x;\theta,\sigma)=\left[G(x;\sigma)\right]^{\theta}=\left[e^{-e^{-\frac{x}{\sigma}}}\right]^{\theta}, \quad x>0,\theta>0,\sigma>0, \tag{1}$$

which is simply the θ^{th} power of cdf of the Gumbel distribution. The probability density function (pdf) corresponding to (1) is

$$f(x;\theta,\sigma)=\frac{\theta}{\sigma}e^{-\frac{x}{\sigma}-\theta e^{-\frac{x}{\sigma}}}, \quad x>0,\theta>0,\sigma>0. \tag{2}$$

We shall write $X\sim EG(\theta,\sigma)$ to denote an absolutely continuous random variable X having the two-parameter exponentiated Gumbel distribution with shape and scale parameters θ and σ respectively, whose pdf is given by (2).

The purpose of this paper is to construct the interval estimation for the two-parameter exponentiated Gumbel distribution based on lower record values. The rest of this paper is organized as follows. Section 2 provides some preliminaries. In Section 3, we present an exact confidence interval for scale parameter σ, and an exact joint confidence region for the parameters (θ,σ) based on lower record values. In Section 4, a Monte Carlo simulation is conducted to study the performance of the proposed confidence interval and region. Finally in section 5, a numerical example with real data set is presented to illustrate the proposed methods.

2 Preliminaries

Let $X_1,X_2,\ldots$ be a sequence of independent and identically distributed (iid) continuous random variables with cdf $F(x)$ and pdf $f(x)$. An observation X_j is called an upper (lower) record value of this sequence its value exceeds (is lower than) that of all previous observations. Generally, let us define $T_1=1$, $U_1=X_1$, and for $n\geq 2$,

$$T_n=\min\left\{j>T_{n-1}:X_j>X_{T_{n-1}}\right\}, \quad U_n=X_{T_n}.$$

Then the sequence $\{U_n\}(\{T_n\})$ is known as upper record statistics (upper record times). Similarly, the lower record times S_n and the lower record values L_n are defined as follows:

$S_1 = 1, L_1 = X_1$, and for $n \geq 2$, $S_n = \min\left\{ j > S_{n-1} : X_j < X_{S_{n-1}} \right\}$, $L_n = X_{S_n}$. For more details on records and its applications, see Nevzorov[7], Ahsanullah[8] and Arnold et. al. [9]. The following lemmas are useful in this paper.

Lemma 2.1. Let $L_1 > L_2 > \ldots > L_m$ be the first m observed lower record values from a population with cdf $F(.)$. Define

$$U_i = -\ln[F(L_i)], \quad i = 1, 2, \ldots, m.$$

Then $U_1 < U_2 < \ldots < U_m$ are the first m upper record values from a standard exponential distribution.

Proof. From the joint pdf of $L_1, L_2, \ldots, L_m$ and using a simple Jacobian argument, we can easily obtain the joint pdf of $U_1, U_2, \ldots, U_m$ as

$$f_{U_1, U_2, \ldots, U_m}\left(u_1, u_2, \ldots, u_m\right) = e^{-u_m}, \qquad 0 < u_1 < u_2 < \ldots < u_m,$$

which is the joint pdf of the first m upper record values from a standard exponential distribution (see Arnold et al. [9]). The proof is thus obtained.

Lemma 2.2. If $U_1 < U_2 < \ldots < U_m$ are the first m upper record values from a standard exponential distribution. Then the spacings $U_1, U_2 - U_1, \ldots, U_m - U_{m-1}$ are iid random variables from a standard exponential distribution.

Proof. The proof can be found in Arnold et al. [9].

3 Main Result

Let $L_1 > L_2 > \ldots > L_m$ be the first m observed lower record values from the exponentiated Gumbel distribution. In this section, a $100(1-\alpha)\%$ confidence interval for scale parameter σ and a $100(1-\alpha)\%$ joint confidence region for (θ, σ) are constructed based on the observed lower records $L_1 > L_2 > \ldots > L_m$.

Let us define $Y_i = \theta \exp\left(-\dfrac{L_i}{\sigma}\right)$, $i = 1, 2, \ldots, m$. Then, by Lemma 2.1, $Y_1 < Y_2 < \ldots < Y_m$ are the first m upper record values from a standard exponential distribution. Moreover, by Lemma 2.2, we can observe that

$$\begin{cases} Z_1 = Y_1 \\ Z_2 = Y_2 - Y_1 \\ \vdots \\ Z_m = Y_m - Y_{m-1} \end{cases} \tag{3}$$

are iid random variables from a standard exponential distribution. Hence

$$V = 2Z_1 = 2Y_1, \tag{4}$$

has a chi-square distribution with 2j degrees of freedom and

$$U = 2\sum_{i=2}^{m} Z_i = 2\left(Y_m - Y_1\right), \tag{5}$$

has a chi-square distribution with $2(m-1)$ degrees of freedom. We can also find that U and V are independent random variables. Let

$$T = \frac{U/2(m-1)}{V/2} = \frac{U}{(m-1)V} = \frac{1}{m-1}\left(\frac{Y_m - Y_1}{Y_1}\right), \tag{6}$$

And

$$S = U + V = 2Y_m, \tag{7}$$

It is easy to show that T has an F distribution with $2(m-1)$ and 2 degrees of freedom and S has a chi-square distribution with $2m$ degrees of freedom. Furthermore, T and S are independent, see Johnson et al. ([10], P. 350).

Let $F_{\alpha(\upsilon_1,\upsilon_2)}$ be the percentile of F distribution with right-tail probability α and υ_1 and υ_2 degrees of freedom. Next theorem gives an exact confidence interval for the scale parameter σ base on lower record values.

Theorem 3.1. Suppose that $L_1 > L_2 > \ldots > L_m$ be the first m observed lower record values from EG distribution in (1). Then, for any $0 < \alpha < 1$,

$$\frac{L_1 - L_m}{\ln[1+\left(m-1\right)F_{\frac{\alpha}{2}}(2(m-1),2)]} < \sigma < \frac{L_1 - L_m}{\ln[1+\left(m-1\right)F_{1-\frac{\alpha}{2}}(2(m-1),2)]},$$

is a $100(1-\alpha)\%$ confidence interval for σ.

Proof. From (6), we know that the pivot

$$T\left(\sigma\right) = \frac{1}{m-1}\left[\frac{e^{-\frac{L_m}{\sigma}} - e^{-\frac{L_1}{\sigma}}}{e^{-\frac{L_1}{\sigma}}}\right] = \frac{1}{m-1}\left[e^{\frac{L_1 - L_m}{\sigma}} - 1\right],$$

has an F distribution with $2(m-1)$ and 2 degrees of freedom. We note that $T(\sigma)$ is strictly

decreasing function of σ. Hence for $0<\alpha<1$, we obtain
$F_{1-\frac{\alpha}{2}}\left(2(m-1),2\right)<T<F_{\frac{\alpha}{2}}\left(2(m-1),2\right)$,
is equivalent to the event

$$\frac{L_1-L_m}{\ln[1+(m-1)F_{\frac{\alpha}{2}}(2(m-1),2)]}<\sigma<\frac{L_1-L_m}{\ln[1+(m-1)F_{1-\frac{\alpha}{2}}(2(m-1),2)]},$$

this completes the proof.

It should be mentioned here that we can also use $T(\sigma)$ to test null hypothesis $H_0:\sigma=\sigma_0$. Let $\chi^2_\alpha(\upsilon)$ denote the percentile of χ^2 distribution with right-tail probability α and υ degrees of freedom. Next theorem gives an exact joint confidence region for the parameters θ and σ.

Theorem 3.2. Suppose that $L_1>L_2>\ldots>L_m$ be the first m observed lower record values from EG distribution. Then, the following inequalities determine $100(1-\alpha)\%$ joint confidence region for θ and σ:

$$\frac{L_1-L_m}{\ln[1+(m-1)F_{\frac{1-\sqrt{1-\alpha}}{2}}(2(m-1),2)]}<\sigma<\frac{L_1-L_m}{\ln[1+(m-1)F_{\frac{1+\sqrt{1-\alpha}}{2}}(2(m-1),2)]},$$

$$\frac{\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m)}{2e^{-\frac{L_m}{\sigma}}}<\theta<\frac{\chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m)}{2e^{-\frac{L_m}{\sigma}}}$$

Proof. From (7), we know that

$$S=2\theta e^{-\frac{L_m}{\sigma}},$$

has a χ^2 distribution with $2m$ degrees of freedom, and it is independent of T. Hence, for $0<\alpha<1$, we have

$$P[F_{\frac{1+\sqrt{1-\alpha}}{2}}(2(m-1),2)<T<F_{\frac{1-\sqrt{1-\alpha}}{2}}(2(m-1),2)]=\sqrt{1-\alpha},$$

and

$$P[\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m)<S<\chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m)]=\sqrt{1-\alpha}.$$

From these relationships, we conclude that

$$P\left[F_{\frac{1+\sqrt{1-\alpha}}{2}}\left(2\left(m-1\right),2\right)<T<F_{\frac{1-\sqrt{1-\alpha}}{2}}\left(2\left(m-1\right),2\right),\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}\left(2m\right)<S<\chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}\left(2m\right)\right]=1-\alpha,$$

or equivalently

$$\frac{L_1-L_m}{\ln[1+\left(m-1\right)F_{\frac{1-\sqrt{1-\alpha}}{2}}(2(m-1),2)]}<\sigma<\frac{L_1-L_m}{\ln[1+\left(m-1\right)F_{\frac{1+\sqrt{1-\alpha}}{2}}(2(m-1),2)]},$$

$$\frac{\chi^2_{\frac{1+\sqrt{1-\alpha}}{2}}(2m)}{2e^{-\frac{L_m}{\sigma}}}<\theta<\frac{\chi^2_{\frac{1-\sqrt{1-\alpha}}{2}}(2m)}{2e^{-\frac{L_m}{\sigma}}}.$$

4 Simulation study

In this section, a Monte Carlo simulation is conducted to study the performance of the proposed confidence interval and joint confidence region. In this simulation, we randomly generated lower record sample $L_1,\ L_2,\ \ldots\ ,L_m$ from the Gumbel distribution with the value of parameters $\left(\theta=0.2,\sigma=0.01\right)$, and then computed 95% confidence intervals and regions using the results presented in Section 3. We then replicated the process $5,000$ times. We presented in Table 1, the simulated average confidence length for parameter σ, confidence area for the parameters (θ,σ) and the 95% coverage probabilities of the proposed confidence intervals and regions.

From Table 1, we observe when m increases, the average confidence length for σ, and the average confidence area for (θ,σ) are decreased. The simulation results shows that the coverage probabilities of the exact confidence intervals for parameter σ and joint confidence regions for parameters (θ,σ) are close to the desired level of 0.95 for different sample sizes. Hence, our proposed methods for constructing exact confidence intervals and joint confidence regions can becused reliably.

Table 1 The simulated average confidence length (CL), confidence area (CA) and 95% coverage probabilities (CP) for the parameters.

m	$CL\ (\sigma)$	$CA(\theta,\sigma)$	$CP(\sigma)$	$CP(\theta,\sigma)$
3	0.0819	56923.76	0.955	0.948
5	0.0367	0.1605	0.952	0.956
7	0.0268	0.0874	0.950	0.950
10	0.0210	0.0516	0.953	0.952
15	0.0171	0.0319	0.957	0.954
20	0.0149	0.0217	0.954	0.957
25	0.0137	0.0173	0.956	0.954
30	0.0130	0.0146	0.957	0.948

5 Numerical example

In this section, real example with climate record data are given to illustrate the proposed confidence intervals and joint confidence regions. We present a data analysis and illustrate application of the results in Section 3, to the seasonal (July 1-June 30) rainfall in inches recorded at Los Angeles Civic Center from 1962 to 2012 (see the website of Los Angeles Almanac: http://www.laalmanac.com/ weather/we13.htm). The data are as follows:

08.38,	07.93,	13.68,	20.44,	22.00,	16.58,	27.47,	07.74,	12.32,	07.17,
21.26,	14.92,	14.35,	07.21,	12.30,	33.44,	19.67,	26.98,	08.96,	10.71,
31.28,	10.43,	12.82,	17.86,	07.66,	12.48,	08.08,	07.35,	11.99,	21.00,
27.36,	08.11,	24.35,	12.44,	12.40,	31.01,	09.09,	11.57,	17.94,	04.42,
16.42,	09.25,	37.96,	13.19,	03.21,	13.53,	09.08,	16.36,	20.20,	08.69.

Here, we checked the validity of the exponentiated Gumbel Model based on the parameters $\hat{\theta} = 7.3840, \hat{\sigma} = 5.8332$, using the Kolmogorov Smirnov (K-S) test. It is observed that the K-S distance is $K - S = 0.0996$ with a corresponding $P - Value = 0.6818$. So, the exponentiated Gumbel model provides a good fit to the above data. If only the lower record values of the seasonal rainfall have been observed, these are

8.38, 7.93, 7.74, 7.17, 4.42, 3.21.

To find a 95% confidence interval for σ, and a joint confidence region for θ and σ, we need the following percentiles:

$$F_{0.025}\left(10,2\right) = 39.39797, F_{0.975}(10,2) = 0.1832712,$$
$$F_{0.0127}\left(10,2\right) = 78.13913, F_{0.9873}(10,2) = 0.1434067,$$

and

$$\chi^2_{0.0127}\left(12\right) = 25.4812, \chi^2_{0.9873}(12) = 3.765882.$$

By Theorem 3.1, the 95% CI for σ is (1.1277, 6.4876), with confidence length 5.3599. By Theorem 3.2, the 95% JCR for θ and σ is determined by the following inequalities:

$$0.8659 < \sigma < 9.5635, \qquad \frac{3.7656}{2e^{-\frac{3.21}{\sigma}}} < \theta < \frac{25.4812}{2e^{-\frac{3.21}{\sigma}}}.$$

with area 114285453. Figure 1 shows the above joint confidence region for the parameters.

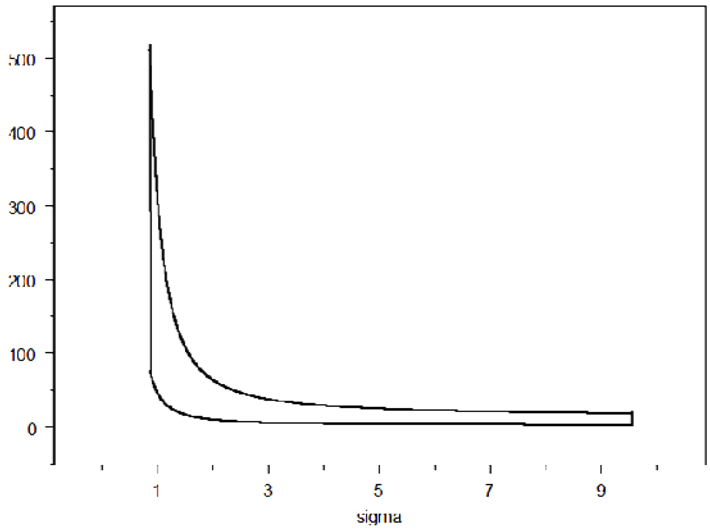

Fig. 1 Joint confidence region for parameters θ and σ .

Acknowledgments

The authors would like to thank the Editor and the referees for their useful comments which improved the paper.

References

1. Kotz, S., Nadarajah, S. (2000). Extreme Value Distributions: Theory and Applications. London: Imperial College Pres:.
2. Gupta, R. C., Gupta, P. L., Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. Communications in Statistics-Theory and Methods, 27: 887-904.
3. Mudholkar, G. S., Srivastava, D. (1993). The exponentiated Weibull family for analyzing bathtub failure-rate data. IEEE Trans. Reliability. 42: 2, 299-302.
4. Nadarajah, S., (2005). The exponentiated Gumbel distribution with climate application. Environmetrics. 17 (1), 13-23.
5. Persson, K., Rydén, J. (2010). Exponentiated Gumbel Distribution for Estimation of Return Levels of Significant Wave Height, Journal of Environmental Statistics. 1(3), 1-12.
6. Nevzorov, V. B., (1988). Records. Theory of Probability and its Application. 32, 201-228.
7. Ahsanullah, M. (1995). Record Statistics. New York : Nova Science Publishers, Inc. Commack.
8. Arnold, B. C., Balakrishnan, N., Nagaraja, H. N. (1998). Records. New York: John Wiley & Sons.
9. Gupta R.D., Kundu, D. (2001). Exponentiated exponential family: an alternative to gamma and Weibull distributions. Biometrical Journal, 43, 117-130.
10. Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). Continuous Univariate Distributions. New York: John Wiley & Sons.