Journal of Modern Applied Statistical Methods

Volume 14 | Issue 2 Article 12

11-1-2015

Monte Carlo Comparison of the Parameter Estimation Methods for the Two-Parameter Gumbel Distribution

Demet Aydin
Sinop University, demethanaydin@gmail.com

Birdal Şenoğlu
Ankara University, senoglu@science.ankara.edu.tr

Recommended Citation

Aydin, Demet and Şenoğlu, Birdal (2015) "Monte Carlo Comparison of the Parameter Estimation Methods for the Two-Parameter Gumbel Distribution," *Journal of Modern Applied Statistical Methods*: Vol. 14 : Iss. 2 , Article 12.
DOI: 10.22237/jmasm/1446351060

Journal of Modern Applied Statistical Methods
November 2015, Vol. 14, No. 2, 123-140.

ISSN 1538 − 9472

Monte Carlo Comparison of the Parameter Estimation Methods for the Two-Parameter Gumbel Distribution

Demet Aydin
Sinop University
Sinop, Turkey

Birdal Şenoğlu
Ankara University
Ankara, Turkey

The performances of the seven different parameter estimation methods for the Gumbel distribution are compared with numerical simulations. Estimation methods used in this study are the method of moments (ME), the method of maximum likelihood (ML), the method of modified maximum likelihood (MML), the method of least squares (LS), the method of weighted least squares (WLS), the method of percentile (PE) and the method of probability weighted moments (PWM). Performance of the estimators is compared with respect to their biases, MSE and deficiency (Def) values via Monte-Carlo simulation. A Monte Carlo Simulation study showed that the method of PWM was the best performance the other methods of bias criterion and the method of ML outperforms the other methods in terms of Def criterion. A real life example taken from the hydrology literature is given at the end of the paper.

Keywords: Gumbel distribution, estimation methods, Monte Carlo simulation, efficiency

Introduction

The Gumbel distribution was first proposed by E. J. Gumbel in 1941. It is a special case of the Generalized Extreme Value (GEV) distribution and is sometimes referred to as Extreme value type I distribution or just the log-Weibull distribution. It is widely used for modeling extreme events, or extreme order statistics. It has two forms, one for “minimum order statistics” and the other for “maximum order statistics.” In this study, we focus on the second form.

The Gumbel distribution has many applications in practice, such as annual maximum flow of river, floods, rainfalls, earthquake magnitudes, annual sea-level prediction and so on. It is of considerable importance in many areas of

Demet Aydın is Assistant Professor in the Department of Statistics. Email at demethanaydin@gmail.com. Birdal Şenoğlu is Professor in the Department of Statistics. Email at senoglu@science.ankara.edu.tr.

environmental sciences, e.g., hydrology, see Wallis and Wood (1985). Mathematical modeling of natural phenomena is becoming more and more important in this age of global warming, especially for public safety and economic issues. Therefore, estimating the model parameters precisely and efficiently is very important. There are various different estimation methods in the literature for estimating the parameters of the Gumbel distribution. The method of moments and the method of maximum likelihood (ML) are the most well known among them. There exist various studies in the literature identifying the most efficient method of estimation for the Gumbel distribution via Monte Carlo simulation study, see for example Landwehr et al. (1979) and Mahdi and Cenac (2004).

In the present work, these studies were extended by including four other estimation methods, namely, modified maximum likelihood (MML), least squares (LS), weighted least squares (WLS) and method of percentile. This is the first study comparing these seven different methods of estimation in the same study.

Gumbel distribution

The probability density function (PDF) and the cumulative density function (CDF) of the two-parameter Gumbel distribution with the location parameter μ and the scale parameter σ are defined as follows:

$$f(x)=\frac{1}{\sigma}\exp\left(-\frac{(x-\mu)}{\sigma}\right)\exp\left(-\exp\left(-\frac{(x-\mu)}{\sigma}\right)\right),\ -\infty<x<\infty,\ \mu\in\Re,\ \sigma\in\Re^{+} \tag{1}$$

and

$$F(x)=\exp\left(-\exp\left(-\frac{(x-\mu)}{\sigma}\right)\right) \tag{2}$$

respectively.

To understand the basic characteristics of the Gumbel distribution, the mean, the variance, the skewness and the kurtosis values are given as follows:

$$E(x)=\mu+\sigma\gamma,\ V(x)=\frac{\pi^{2}}{6}\sigma^{2},\ \beta_1:1.14\ \text{and}\ \beta_2:5.40 \tag{3}$$

respectively. Here, γ is the Euler's constant, with approximate value 0.5772.

It is seen that Gumbel distribution is positively skewed and moderately long tailed. See Figure 1 for the plot of the Gumbel distribution.

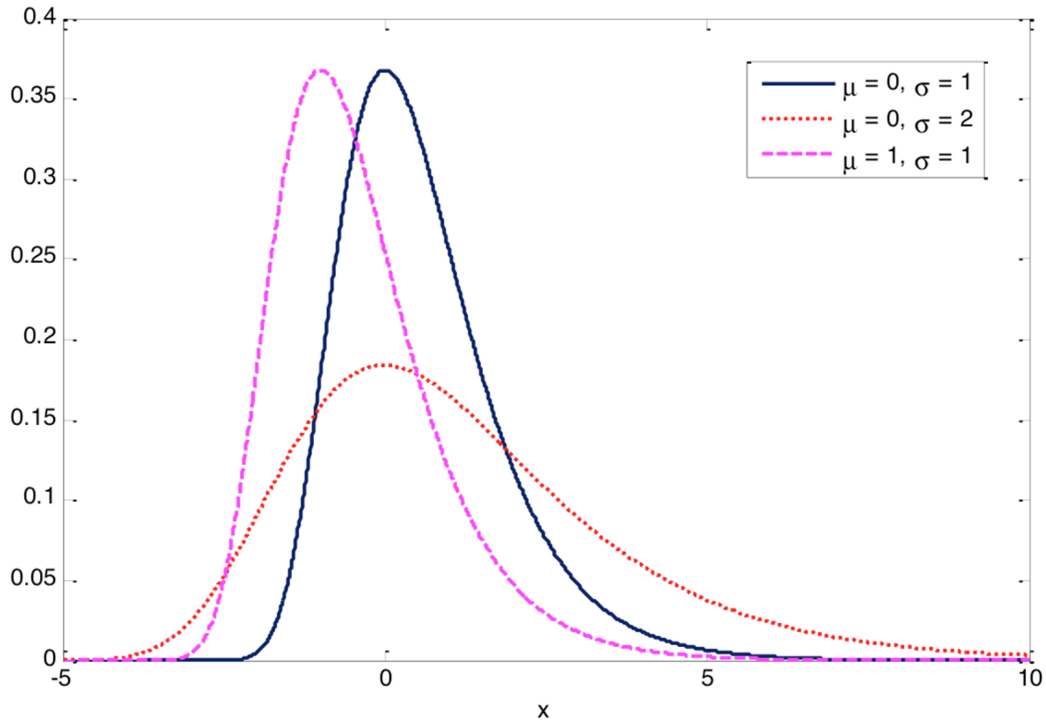

Figure 1. Plot of the Gumbel distribution for various μ and σ values.

The methods of estimation

In this section, we briefly describe the methods of estimation for the Gumbel distribution used in this study.

The method of moments

Moment estimators of the location parameter μ and the scale parameter σ of the Gumbel distribution are found by equating the sample moments to the corresponding theoretical moments.

In other words, they are the solutions of the following equalities

$$\bar{X} = \mu + \gamma\sigma \text{ and } S^2 = \frac{\pi^2}{6}\sigma^2 \tag{4}$$

ME of μ and σ are then obtained as

$$\tilde{\mu}_{ME} = \bar{X} - \gamma\tilde{\sigma}_{ME} \text{ and } \tilde{\sigma}_{ME} = \frac{\sqrt{6}}{\pi}S \tag{5}$$

respectively.

The method of Maximum Likelihood

ML estimators of the two-parameter Gumbel distribution in (1) are found by maximizing the following log-likelihood function with respect to the parameters of interest (i.e., with respect to μ and σ),

$$\ln L = -n\ln\sigma - \sum_{i=1}^{n} z_i - \sum_{i=1}^{n} g(z_i) \tag{6}$$

where

$$g(z_i) = \exp(-z_i),\ z_i = \frac{(x_i - \mu)}{\sigma}$$

First, we obtain the likelihood functions given below:

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{n}\frac{1}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{n} g(z_i) = 0 \tag{7}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_i - \frac{1}{\sigma}\sum_{i=1}^{n} z_i g(z_i) = 0 \tag{8}$$

It is clear that likelihood equations do not have explicit solutions. Therefore, we apply numerical methods to solve the equations (7) and (8). Iterative solutions of these equations are the ML estimates of the location parameter μ and the scale parameter σ.

The method of Modified Maximum Likelihood

MML methodology was first introduced by Tiku (1967, 1968). It is used as an alternative to the well known ML methodology when the estimators of the parameters can not be obtained explicitly. Idea behind the MML methodology is based on the linearization of the nonlinear terms in the likelihood equations.

MML methodology is based on the following steps:

i) Likelihood equations given in (7) and (8) are written in terms of the order statistics, since complete sums are invariant to ordering, i.e., $\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} z_{(i)}$

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{n} \frac{1}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^{n} g\left(z_{(i)}\right) = 0 \tag{9}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^{n} z_{(i)} - \frac{1}{\sigma} \sum_{i=1}^{n} g\left(z_{(i)}\right) z_{(i)} = 0 \tag{10}$$

where

$$g\left(z_{(i)}\right) = \exp\left(-z_{(i)}\right) \text{ and } z_{(i)} = \frac{\left(x_{(i)} - \mu\right)}{\sigma},\ i = 1, 2, \ldots, n$$

ii) Linearize the nonlinear term in (9) and (10) by using the first two terms of the Taylor series expansion around the expected values of the order statistics

$$g\left(z_{(i)}\right) \cong g\left(t_{(i)}\right) + \left(z_{(i)} - t_{(i)}\right) \left\{ \frac{d}{dz} g\left(z_{(i)}\right) \right\}_{z_{(i)} = t_{(i)}}$$

or equivalently

$$g\left(z_{(i)}\right) \cong \alpha_i + \beta_i z_{(i)} \tag{11}$$

where

$$\alpha_i = \exp\left(-t_{(i)}\right) + \exp\left(-t_{(i)}\right)t_{(i)} \text{ and } \beta_i = -\exp\left(-t_{(i)}\right)$$

Here, $t_{(i)}$'s (i = 1, 2, …, n) are the expected values of the standardized order statistics $z_{(i)}$, i.e., $t_{(i)} = E(z_{(i)})$, and are obtained from the following equality:

$$\int_{-\infty}^{t_{(i)}} \exp(-z)\exp\left(-\exp(-z)\right)dz = \frac{i}{n+1} \quad (12)$$

Equation (12) gives

$$t_{(i)} = -\ln\left(-\ln\left(\frac{i}{n+1}\right)\right),\ i = 1, 2, \ldots, n$$

iii) By incorporating (11) into (9) and (10), we obtain the modified likelihood equations given below

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \sum_{i=1}^{n}\frac{1}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{n}\left(\alpha_i + \beta_i z_{(i)}\right) = 0 \quad (13)$$

and

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = \frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)} - \frac{1}{\sigma}\sum_{i=1}^{n}\left(\alpha_i + \beta_i z_{(i)}\right)z_{(i)} = 0 \quad (14)$$

iv) Solutions of the modified likelihood equations in (13) and (14) with respect to the unknown parameters are the following MML estimators

$$\hat{\mu}_{MML} = K_{MML} + L_{MML}\hat{\sigma}_{MML} \text{ and } \hat{\sigma}_{MML} = \frac{-B + \sqrt{B^2 - 4nC}}{2\sqrt{n(n-1)}} \quad (15)$$

where

$$K_{MML} = \frac{1}{m}\sum_{i=1}^{n} \beta_i x_{(i)},\ L_{MML} = \frac{\Delta}{m}, = \sum_{i=1}^{n} \Delta_i,\ \Delta_i = \left(\alpha_i - 1\right),$$
$$m = \sum_{i=1}^{n} \beta_i,\ B = \sum_{i=1}^{n} \Delta_i \left(x_{(i)} - \hat{\mu}\right) \text{ and } C = \sum_{i=1}^{n} \beta_i \left(x_{(i)} - \hat{\mu}\right)^2$$

MML estimators are asymptotically equivalent to the ML estimators. Therefore, they are asymptotically unbiased and minimum variance bound (MVB) estimators under the regularity conditions. However, in contrast to ML estimators, they are the explicit functions of the sample observations and avoid the computational difficulties encountered in the numerical solutions, such as multiple roots, nonconvergence of iterations or convergence to wrong values, see for example Barnett (1966). It should be noted that MML estimators are nearly unbiased and MVB estimators even for small samples.

The method of Least Squares

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the distribution function $F(.)$. LS estimators of the unknown parameters of $F(.)$ are obtained by minimizing the following equation:

$$\sum_{i=1}^{n} \left\{ F\left(X_{(i)}\right) - \frac{i}{n+1} \right\}^2 \tag{16}$$

with respect to the parameters of interest. It is known that $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ are the ordered random variables.

Then the LS estimators of the parameters of the two-parameter Gumbel distribution are obtained by minimizing the function

$$G\left(\mu, \sigma\right) = \sum_{i=1}^{n} \left\{ \exp\left(-\exp\left(-z_{(i)}\right)\right) - \frac{i}{n+1} \right\}^2 \tag{17}$$

with respect to the parameters μ and σ.

The method of Weighted Least Squares

Let, $X_1, X_2, \ldots, X_n$ be a random sample of size n from the distribution function $F(.)$ and $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be the ordered random variables.

WLS estimators of the unknown parameters are obtained by minimizing the function

$$\sum_{i=1}^{n} w_i \left\{ F\left(X_{(i)}\right) - \frac{i}{n+1} \right\}^2 \tag{18}$$

with respect to the parameters of interest.

In case of the Gumbel distribution, the WLS estimators of the model parameters are obtained by minimizing the following function

$$G(\mu,\sigma) = \sum_{i=1}^{n} w_i \left\{ \exp\left(-\exp\left(-z_{(i)}\right)\right) - \frac{i}{n+1} \right\}^2 \tag{19}$$

with respect to the parameters μ and σ. Here,

$$z_{(i)} = \frac{x_{(i)} - \mu}{\sigma},\ w_i = \frac{1}{Var\left(F\left(z_{(i)}\right)\right)},\ \text{and } Var\left(F\left(z_{(i)}\right)\right) = \frac{(n+1)^2 (n+2)}{i(n-i+1)}$$

The method of percentile

Percentile estimators of the unknown parameters of the distribution function $F\left(\frac{x_{(i)} - \mu}{\sigma}\right)$ are found by minimizing the equation

$$\sum_{i=1}^{n} \left\{ x_{(i)} - F^{-1}\left(\frac{i}{n+1}\right) \right\}^2 \tag{20}$$

with respect to the unknown parameters. Here, $X_{(i)}$'s are defined as the i^{th} order statistics. For the Gumbel distribution, equation (20) reduces to

$$PE(\mu,\sigma) = \sum_{i=1}^{n} \left\{ x_{(i)} - \mu + \sigma \ln\left(-\ln\left(\frac{i}{n+1}\right)\right) \right\}^2 \tag{21}$$

Solutions of the equation (21) are the following percentile estimators of the location parameter μ and the scale parameter σ

$$\tilde{\mu}_{PE} = K_{PE} + L_{PE}\tilde{\sigma}_{PE} \text{ and } \tilde{\sigma}_{PE} = \frac{\sum_{i=1}^{n}\delta_i\left(\tilde{\mu}_{PE} - x_{(i)}\right)}{\sum_{i=1}^{n}\delta_i^2} \tag{22}$$

where

$$K_{PE} = \frac{1}{n}\sum_{i=1}^{n}x_{(i)},\ L_{PE} = \frac{\delta}{n},\ \delta_i = \ln\left(-\ln\left(\frac{i}{n+1}\right)\right) \text{ and } \delta = \sum_{i=1}^{n}\delta_i$$

The method of Probability Weighted Moments

The method of probability weighted moments has been defined by Greenwood et al. (1979). Similar to the traditional method of moments, parameter estimates are obtained by equating the analytical expressions for PWM to sample estimates.

They defined the PWM as follows

$$M_{i,j,k} = E\left[X^i F(X)^j \left(1-F(X)\right)^k\right] = \int_0^1 [x(F)]^i F^j (1-F)^k\, dF,\ i, j, k \in R \tag{23}$$

where $F(X)$ is the cdf of the random variable X and $x(F)$ is the inverse distribution function.

By adopting the convention $M_{1,0,k} = M_{(k)}$, the PWM estimators of μ and σ are obtained as

$$\hat{\mu}_{PWM} = \hat{M}_{(0)} - \gamma\hat{\sigma}_{PWM} \text{ and } \hat{\sigma}_{PWM} = \frac{\hat{M}_{(0)} - 2\hat{M}_{(1)}}{\ln 2} \tag{24}$$

respectively. $\hat{M}_{(k)}$ in (24) is an unbiased estimate of $M_{(k)}$ and is given by

$$\hat{M}_{(k)} = \frac{1}{n}\sum_{i=1}^{n}x_{(i)}\frac{C(n-i,k)}{C(n-1,k)} \tag{25}$$

where $x_{(i)}$ are the ordered observations and k is a nonnegative integer. See Landwehr et al. (1979) for more detailed information about the method of PWM.

Methodology 1

Monte Carlo simulation study

An extensive Monte Carlo simulation study was conducted to compare the performance of the different estimators proposed in the previous section. Performances of the different estimators are compared with respect to their biases, MSE and Def values. Def is the natural measure of the joint efficiency of the pair ($\hat{\mu}, \hat{\sigma}$), see Tiku and Akkaya (2004). It is defined as given below.

Definition: Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be the estimators of the parameters θ_1 and θ_2, respectively. Def is a MSE based measure of the joint efficiency of estimators of a set of parameters of a probability distribution. Then, the Def of the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ is defined as

$$Def\left(\hat{\theta}_1, \hat{\theta}_2\right) = MSE\left(\hat{\theta}_1\right) + MSE\left(\hat{\theta}_2\right) \tag{26}$$

where

$$MSE\left(\hat{\theta}\right) = Var\left(\hat{\theta}\right) + Bias^2\left(\hat{\theta}\right) \text{ and } Bias\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta.$$

Results 1

The Mean, MSE and Def values of the parameter estimators were computed based on ⟦100000/*n*⟧ Monte Carlo runs for various sample sizes ranging from 5 to 1000 (i.e., n = 5, 10, 50, 100 and 1000). Here, ⟦.⟧ shows the integer value function. The location parameter μ and the scale parameter σ are taken to be 0 and 1 without loss of generality throughout the study, since all the estimators are invariant under the linear transformations of the data. All the computations were conducted in MATLAB R2010a. Simulation results are presented in Table 1.

Table 1. Simulated Means, Variance, MSE and Def values for the different parameter estimators of μ and σ; $\mu = 0$, $\sigma = 1$

		μ			σ			
n		**Mean**	**Variance**	**MSE**	**Mean**	**Variance**	**MSE**	**Def**
5	*ML*	0.0876	0.2365	0.2441	0.8491	0.1221	0.1449	0.3890
	MML	0.1965	0.2508	0.2894	0.9785	0.1989	0.1994	0.4888
	LS	-0.0127	0.2869	0.2871	1.2366	0.5363	0.5923	0.8794
	WLS	-0.0238	0.5399	0.5404	1.2688	2.0432	2.1155	2.6559
	PE	0.0033	0.2394	0.2395	1.2715	0.3541	0.4278	0.6673
	ME	0.0569	0.2369	0.2401	0.9178	0.1747	0.1815	0.4216
	PWM	0.0057	0.2324	0.2324	1.0066	0.1976	0.1977	0.4301
10	*ML*	0.0358	0.1133	0.1146	0.9197	0.0611	0.0675	0.1821
	MML	0.0957	0.1168	0.1260	0.9741	0.0691	0.0697	0.1957
	LS	-0.0088	0.1205	0.1206	1.1031	0.1234	0.1341	0.2547
	WLS	-0.0170	0.1249	0.1252	1.1259	0.1432	0.1590	0.2842
	PE	-0.0069	0.1140	0.1140	1.1698	0.1395	0.1683	0.2823
	ME	0.0233	0.1154	0.1159	0.9512	0.0920	0.0944	0.2103
	PWM	-0.0031	0.1120	0.1120	0.9970	0.0872	0.0872	0.1992
50	*ML*	0.0097	0.0224	0.0224	0.9839	0.0122	0.0124	0.0349
	MML	0.0229	0.0226	0.0231	0.9915	0.0124	0.0125	0.0356
	LS	0.0011	0.0248	0.0248	1.0195	0.0190	0.0194	0.0442
	WLS	-0.0015	0.0261	0.0261	1.0273	0.0227	0.0235	0.0496
	PE	-0.0035	0.0231	0.0231	1.0617	0.0235	0.0273	0.0504
	ME	0.0084	0.0233	0.0233	0.9884	0.0207	0.0208	0.0442
	PWM	0.0022	0.0224	0.0225	0.9991	0.0164	0.0164	0.0389
100	*ML*	0.0037	0.0110	0.0110	0.9930	0.0060	0.0060	0.0170
	MML	0.0106	0.0110	0.0111	0.9962	0.0061	0.0061	0.0172
	LS	-0.0001	0.0121	0.0121	1.0119	0.0092	0.0093	0.0215
	WLS	-0.0016	0.0128	0.0128	1.0164	0.0111	0.0113	0.0242
	PE	-0.0044	0.0114	0.0114	1.0388	0.0112	0.0127	0.0241
	ME	0.0035	0.0115	0.0115	0.9941	0.0105	0.0105	0.0220
	PWM	0.0001	0.0110	0.0110	0.9999	0.0079	0.0079	0.0190
1000	*ML*	0.0000	0.0011	0.0011	0.9990	0.0006	0.0006	0.0017
	MML	0.0007	0.0011	0.0011	0.9992	0.0006	0.0006	0.0017
	LS	-0.0003	0.0012	0.0012	1.0008	0.0008	0.0008	0.0021
	WLS	-0.0004	0.0013	0.0013	1.0012	0.0011	0.0011	0.0024
	PE	-0.0019	0.0012	0.0012	1.0072	0.0011	0.0011	0.0023
	ME	-0.0002	0.0012	0.0012	0.9995	0.0011	0.0011	0.0022
	PWM	-0.0003	0.0011	0.0011	0.9997	0.0007	0.0007	0.0019

The following conclusions are drawn from the results of the Monte Carlo simulation study.

i) According to the bias comparisons of the estimators:

As far as the location parameter μ is concerned, MML did not perform well especially for small n values ($n = 5$ and 10). PE and PWM estimators show the best performance among the others, since they are more or less unbiased even for small sample sizes. It is observed in Table 1 that biases of the different estimators considered in this study decrease as the sample size n increases.

If our concern is the scale parameter σ, all the scale estimators (except PWM and MML) have substantial bias in cases where a small number of data samples ($n = 5$ and 10) are available. For these sample sizes, LS, WLS and PE overestimate σ while ML and ME underestimate. PWM shows the best performance and followed by the MML estimator for all the sample sizes. Similar to the comments made about the location estimators, bias of the scale estimators decreases as the sample size n increases.

ii) According to the efficiency comparisons of the estimators:

Simulation results show that the method of ML outperforms the other methods for estimating the location parameter μ in all cases except $n = 5$ and 10. For these sample sizes, the method of PWM shows the best performance among the other methods with the smallest MSE.

For estimating the scale parameter σ, it is observed that ML works the best for all sample sizes.

It should be noted that there is not much difference in the performances between ML and MML estimators especially for moderate ($n = 50$ and 100) and large ($n = 1000$) sample sizes as mentioned in the section on MML.

iii) According to the joint efficiency (Def) comparisons of the estimators:

It is clear from the simulation results presented in Table 1 that the method of ML provides the smallest Def values in all cases, therefore it is the best method for jointly estimating the location parameter μ and the scale parameter σ of the Gumbel distribution. Second best performance is shown by the method of MML for all values of n except $n = 5$. For $n = 5$, ME is the second most efficient method of the seven. Third place (in terms of the joint efficiency) was taken by the method of PWM.

Note that the simulation results presented in this study are in accordance with those of the Landwehr et al. (1979) who compared the methods of PWM, ME and ML.

Methodology 2

Asymptotic variances

In this part, obtain the exact variances of the ML estimators as

$$V(\hat{\mu}) \cong \frac{6}{\pi^2}\frac{\sigma^2}{n}\left(\frac{\pi^2}{6}+(1-\gamma)^2\right) \text{ and } V(\hat{\sigma}) \cong \frac{6}{\pi^2}\frac{\sigma^2}{n}$$

by using the diagonal elements of I^{-1} (where $I = \left[I_{ij}\right]_{i,j=1,2}$ is the Fisher information matrix), see Panjer (2006). These variances are also known as the Rao-Cramer Lower Bounds (RCLBs) for the parameters μ and σ. Elements of the symmetric matrix **I** are given by

$$I_{11} = -E\left(\frac{\partial^2 \ln L}{\partial \mu^2}\right) = \frac{n}{\sigma^2}$$

$$I_{12} = -E\left(\frac{\partial^2 \ln L}{\partial \mu \partial \sigma}\right) = -\frac{n}{\sigma^2}(1-\gamma)$$

$$I_{22} = -E\left(\frac{\partial^2 \ln L}{\partial \sigma^2}\right) = \frac{n}{\sigma^2}\left(\frac{\pi^2}{6}+(1-\gamma)^2\right)$$

Results 2

Table 2 shows that the RCLBs for the parameters and for various different sample sizes.

Table 2. RCLBs for the parameters μ and σ

n	$V(\hat{\mu})$	$V(\hat{\sigma})$
5	0.2217	0.1215
10	0.1108	0.0607
50	0.0221	0.0121
100	0.0110	0.0060
1000	0.0011	0.0006

It is seen that simulated variances of the ML estimators given in Table 1 are very close to the RCLBs even for small sample sizes. This is another indication of the fact that the ML estimators show the best performance for estimating the parameters of the Gumbel distribution.

A real life example

Meriç (Maritsa or Evros) is the longest river of the Balkan Peninsula and the second longest river of in South-Eastern Europe. Its length is 530 km with a catchments area of more than 53,000 square kilometers, see Sezen et al. (2007). It is a highly industrialized, highly agricultural and highly populated area with approximately 2 million inhabitants. The Meriç River basin is distributed over the territories of three countries, namely, Bulgaria (66%), Turkey (28%) and Greece (6%). The Meriç River has four main tributaries known as Ardas (Bulgaria and Greece), Tundzha (Bulgaria and Turkey), Erythropotamos (mostly in Greece) and Ergene (in Turkey), see Skiyas and Kallioras (2007).

The main reason for analyzing the data belonging to the Meriç River is its high risk of flooding. It is known that one or two flooding events have occurred annually during the last decade. They have caused severe economic, socioeconomic and environmental impacts, see Skiyas and Kallioras (2007).

The maximum daily flood discharge (annual) is measured in cubic meters per second (m^3/s) for the Meriç River at Turkey, recorded during the period 1982-2006. These measurements have been taken from the Kirişhane station, Edirne (Turkey), see Sezen et al. (2007).

Discharge is defined as the volume of the water flowing through a specified point of a stream in a given interval of time. Therefore, especially in flood periods, identifying the distributional characteristics (such as mean and variance) of the maximum daily discharge data is extremely important for flood control, water resources planning, design of hydraulic structures, management and decision making (Chen & Chiu, 2004).

The aim is to fit a distribution to the maximum daily discharge (annual) data by using the Methods of Estimation described. To have an idea about the underlying distribution of the data, we use the Kolmogorov-Simirnov (KS) test. According to the KS test, we do not reject the null hypothesis

H_0: Distribution of the maximum daily discharge (annual) data is Gumbel since $KS_{cal} = 1.1349 < KS_{tab} = 0.2376$.

For the maximum daily discharge (annual) data, estimates of the parameters of the Gumbel distribution are obtained as reported in Table 3.

Table 3. Parameter estimates of the Gumbel distribution for the Meriç River during 1982-2006.

Estimator	$\hat{\mu}$	$\hat{\sigma}$
ML	539.8018	302.2066
MML	545.5504	303.8097
LS	504.1084	314.3558
WLS	497.1617	323.3498
PE	509.0342	430.4036
ME	509.4286	395.1687
PWM	527.4405	363.9631

See Figure 2 for the plots of the fitted densities based on these estimate values. It can be seen from the figure that the fitted densities based on the ML and the MML estimates provide better fit than the fitted densities based on the other estimates for the Meriç River data.

Conclusion

Seven estimation methods for estimating the parameters of the two-parameter Gumbel distribution were compared. Performance of the estimators is compared with respect to their biases, MSE and Def values.

Comparing all the seven methods, it is clear that as far as bias is concerned, the method of PWM outperforms the other methods for all sample sizes. It can also be seen from the simulation results that all the estimators of the location parameter μ and the scale parameter σ are asymptotically unbiased. In terms of the joint efficiency, the method of ML works the best for all sample sizes. However,

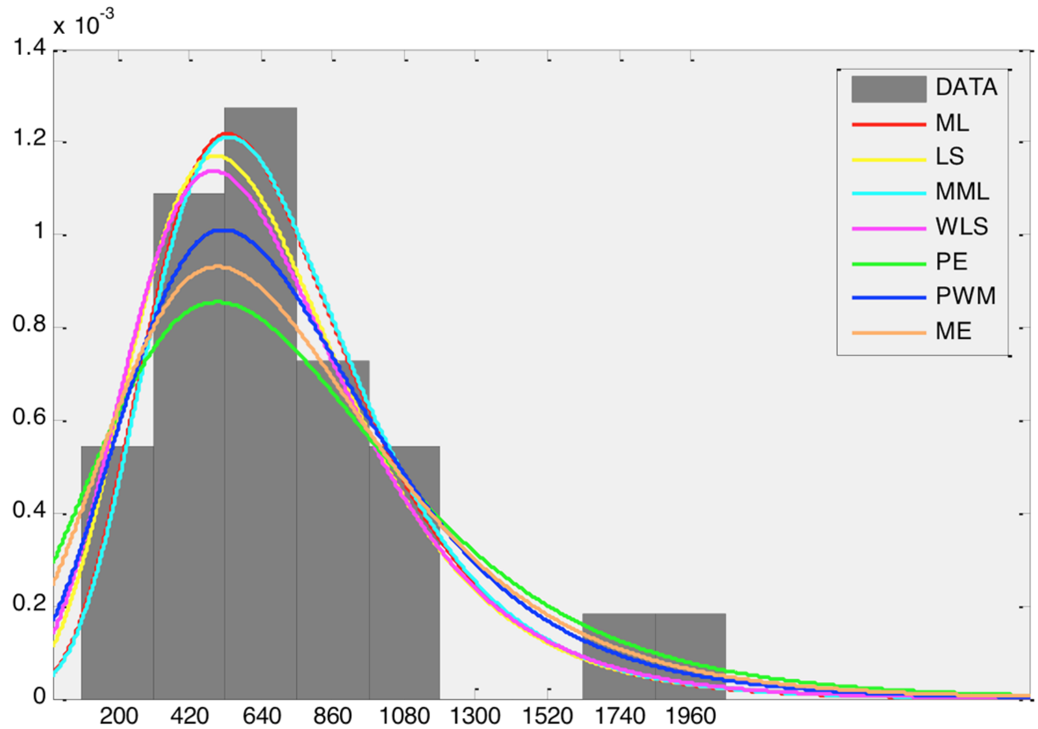

Figure 2. Histogram of the maximum daily flood discharges (annual) for the Meriç River data and the fitted densities.

the Def values of the MML estimators $(\hat{\mu}, \hat{\sigma})$ are quite close to that of ML estimators especially for moderate and large sample sizes as expected. As far as computation is concerned, MML estimators are easy to compute and do not have the computational complexities of ML estimators. Therefore, their computation takes very little CPU time, see Kantar and Şenoğlu (2008). If our consideration is both efficiency and the CPU time, then we recommend to use the MML estimators for estimating the pair (μ, σ) for moderate and large sample sizes.

References

Barnett, V. D. (1966). Order statistics estimators of the location of the Cauchy Distribution. *Journal of the American Statistical Association*, *61*(316), 1205–1218. Correction in *63*, 383–385. doi:10.1080/01621459.1966.10482205

Chen, Y. C., & Chiu C. L. (2004). A fast method of flood discharge estimation. *Hydrological Processes*, *18*(9), 1671–1684. doi:10.1002/hyp.1476

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, *15*(5), 1049–1054.

Gumbel, E. J. (1941). The return period of flood flows. *The Annals of Mathematical Statistics*, *12*(2), 163–190.

Kantar, Y. M., & Şenoğlu, B. (2008). A comparative study for the location and scale parameters of the Weibull distribution with given shape parameter. *Computers and Geosciences, 34*, 1900–1909. doı:10.1016/j.cageo.2008.04.004

Landwehr, J. M., Matalas, N. C., & Wallis J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, *15*(5), 1055–1064.

Mahdi, S., & Cenac, M. (2005). Estimating parameters of Gumbel distribution using the methods of moments, probability weighted moments and maximum likelihood. *Revista de Matemática: Teoría y Aplicaciones*, *12*(1-2), 151–156.

Panjer, H. H. (2006). *Operational risk modeling analytics*. Hoboken, NJ: John Wiley and Sons.

Sezen, N., Gündüz, N., & Malkaralı, S. (2007). *Meriç River Floods And Turkish–Bulgarian Cooperations*. Paper presented at the International Congress on River Basin Management, Antalya, Turkey.

Skiyas, S., & Kallioras, A. (2007). Cross border co-operation and the problem of flooding in the Evros delta. In J. Verwijmeren & M. Wiering (Eds), *Many rivers to cross: cross border co-operation in river management* (pp. 119–144). Delft, The Netherlands: Eburon Academic Publishers.

Tiku, M. L. (1967). Estimating the mean and standard deviation from censored normal samples. *Biometrika*, *54*(1-2), 155–165. doi:10.1093/biomet/54.1-2.155

Tiku, M. L. (1968). Estimating the parameters of log-normal distribution from censored samples. *Journal of the American Statistical Association*, *63*(321), 134–140. doi:10.1080/01621459.1968.11009228

Tiku, M. L., & Akkaya, A. D. (2004). *Robust estimation and hypothesis testing*. New Delhi: New Age International (P) Ltd. Publishers.

Wallis, J. R., & Wood, E. F. (1985). Relative accuracy of log Pearson III procedures. *Journal of Hydraulic Engineering, 111*(7), 1043-1056. doi:10.1061/(ASCE)0733-9429(1985)111:7(1043)