

Two-Sided Generalized Gumbel Distribution with Application to Air Pollution Data

Mustafa Ç. Korkmaz

ArtvinÇoruh University, Department of Statistics and Computer Sciences, Artvin/TURKEY

Email address:

mcagatay@artvin.edu.tr

To cite this article:

Mustafa Ç. Korkmaz. Two-Sided Generalized Gumbel Distribution with Application to Air Pollution Data. *International Journal of Statistical Distributions and Applications*. Vol. 1, No. 1, 2015, pp. 19-26. doi: 10.11648/j.ijdsd.20150101.14

Abstract: We introduce a univariate generalized form of the Gumbel distribution via two-sided distribution structure. We obtain its some properties such as special cases, density shapes, hazard rate function and moments. We give the maximum likelihood estimators of this two-sided generalized Gumbel distribution with an algorithm. Finally, a real data application based on air pollution data is given to demonstrate that it has real data modeling potential.

Keywords: Gumbel Distribution, Two-Sided Distribution, Generalized Gumbel Distribution, Exponentiated Gumbel Distribution

1. Introduction

The Gumbel distribution, denoted by *Gu*, is introduced by German statistician Emil J. Gumbel (1958) and it is introduced by the following cumulative distribution function (cdf) and probability density function (pdf)

$$G(x; \mu, \sigma) = \exp \left[-\exp \left(-\frac{x - \mu}{\sigma} \right) \right] \quad (1)$$

$$g(x; \mu, \sigma) = \frac{1}{\sigma} \exp \left[-\left\{ \frac{x - \mu}{\sigma} + \exp \left(-\frac{x - \mu}{\sigma} \right) \right\} \right] \quad (2)$$

respectively and where $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$.

Gu distribution is frequently used for modelling in many areas such as environmental, engineering and actuarial

$$F(x; \alpha, \beta, \xi) = \begin{cases} \beta^{1-\alpha} (G(x; \xi))^\alpha, & -\infty < x \leq G_{(x;\xi)}^{-1}(\beta) \\ 1 - (1 - \beta)^{1-\alpha} (1 - G(x; \xi))^\alpha, & G_{(x;\xi)}^{-1}(\beta) \leq x < \infty \end{cases} \quad (3)$$

where $-\infty < x < \infty$; $\alpha > 0$ shape parameter, $\beta \in (0, 1)$ reflection parameter, ξ is parameter vector, $G(x; \xi)$ is the cdf of the base distribution and $G_{(x;\xi)}^{-1}(\cdot)$ its inverse. The corresponding pdf is

sciences. It is also known as the extreme value distribution of type I. Also, the *Gu* distribution is a limit distribution of the generalized extreme value distribution (Von Mises, 1954). Kotz and Nadarajah (2000) explain this distribution in detail and with its applications. To increasing model flexibility of the *Gu* distribution there are several generalization of the *Gu* distribution in the literature such as the beta-*Gu* distribution (Nadarajah and Kotz, 2004), the generalized *Gu* distribution (Cooray, 2010), the Kumaraswamy-*Gu* distribution (Cordeiro et al., 2012), the *Gu*-Weibull distribution (Al-Aqtash et al, 2014) and the exponentiated generalized *Gu* distribution (Andrade et al. 2015). For more information on Gumbel and extreme value distributions, see Gumbel (1958), Johnson et al. (1995), Kotz and Nadarajah (2000), and Beirlant et al. (2006).

On the other hand, two-sided generalized a class of the distributions is introduced by Korkmaz and Genç (2015) by following cdf

$$f(x; \alpha, \beta, \xi) = \begin{cases} \alpha \beta^{1-\alpha} g(x; \xi) (G(x; \xi))^{\alpha-1}, & -\infty < x \leq G_{(x; \xi)}^{-1}(\beta) \\ \alpha (1-\beta)^{1-\alpha} g(x; \xi) (1-G(x; \xi))^{\alpha-1}, & G_{(x; \xi)}^{-1}(\beta) \leq x < \infty. \end{cases} \quad (4)$$

Since the standard two-sided power distribution (Van dorp and Kotz, 2002) is applied as a distribution class generator, the standard two-sided power distribution underlies of this generalized two-sided class. Since α comes from the standard two-sided power distribution, this parameter also controls the kurtosis and tail of the distribution. This general two-sided class can contain alternative distributions for modeling not only positive data but also negative data with high kurtosis. Korkmaz and Genç (2015) also define some members of this class using ordinary distributions such as exponential, Weibull, normal, Fréchet, half logistic, Pareto, Gumbel and Kumaraswamy. Two-sided generalized normal distribution is studied in detail by the authors. In this paper we obtain the some properties of the generalized form of the Gumbel distribution, referred to as the two-sided generalized Gumbel (*TSGG*) distribution, defined by Korkmaz and Genç (2015) via two-sided distribution structure.

From (1), (2), (3) and (4) the cdf and pdf of the *TSGG* distribution are easily obtained as

$$F(x; \theta) = \begin{cases} \beta^{1-\alpha} e^{-\alpha u}, & -\infty < x \leq \eta \\ 1 - (1-\beta)^{1-\alpha} (1-e^{-u})^\alpha, & \eta \leq x < \infty \end{cases} \quad (5)$$

and

$$f(x; \theta) = \begin{cases} \frac{\alpha \beta^{1-\alpha} u}{\sigma} e^{-u\alpha}, & -\infty < x \leq \eta \\ \frac{\alpha u}{\sigma} \left(\frac{1-e^{-u}}{1-\beta} \right)^{\alpha-1} e^{-u}, & \eta \leq x < \infty \end{cases} \quad (6)$$

respectively and where $u = \exp(-(x-\mu)/\sigma)$,

$$TSGG(\theta) = \beta LT1_{(-\infty, \eta)}(\alpha, \mu, \sigma) + (1-\beta) LT2_{(\eta, \infty)}(\alpha, \mu, \sigma) \quad (7)$$

where, $LT1_{(a,b)}$ denotes the doubly truncated $LT1(\alpha, \mu, \sigma)$ distribution with truncation points a and b , and similarly for $LT2(\alpha, \mu, \sigma)$. Hence *TSGG* has properties of both $LT1(\alpha, \mu, \sigma)$ distribution and the $LT2(\alpha, \mu, \sigma)$ distribution.

First and second derivatives of $\log f(x)$ for the *TSGG* distribution

$$\frac{d \log f(x)}{dx} = \begin{cases} -\frac{1}{\sigma} + \frac{\alpha u}{\sigma}, & x < \eta \\ -\frac{1}{\sigma} + \frac{u(1-\alpha e^{-u})}{\sigma(1-e^{-u})}, & x > \eta \end{cases} \quad (8)$$

$\theta = (\alpha, \beta, \mu, \sigma)$ parameter vector, $-\infty < x, \mu < \infty$, $\alpha, \sigma > 0$, $\eta = \mu - \sigma \log(-\log \beta)$ and $\beta \in (0, 1)$. We also note that η is the reflection point of the distribution.

In the rest of this paper, we obtain the special cases and explore density shape of (6). We examine hazard rate function. We derive formulas for the r th moment. We also consider the maximum likelihood estimates of parameters. Finally we end the paper with a real data application.

2. Special Cases and Shapes

When $\alpha = 1$ and $\alpha = 2$, the *TSGG* distribution reduced to Gumbel distribution and triangular-Gumbel distribution respectively. When $\beta \rightarrow 1$, the distribution consists of the *exponentiated-Gumbel* distribution with cdf $\exp(-\alpha u)$, that is Lehmann type 1-Gumbel distribution denoted by $LT1(\alpha, \mu, \sigma)$. When $\beta \rightarrow 0$, the distribution converges to distribution with cdf $1 - [1 - \exp(-u)]^\alpha$, that is Lehmann type 2-Gumbel distribution denoted by $LT2(\alpha, \mu, \sigma)$. We note that $LT2(\alpha, \mu, \sigma)$ distribution is also introduced by Nadarajah (2006). Further, the *TSGG* distribution is in fact a mixture of the $LT1(\alpha, \mu, \sigma)$ distribution truncated above at η and the distribution of the $LT2(\alpha, \mu, \sigma)$ distribution truncated below at the same point, with the mixing parameter β , that is,

$$\frac{d^2 \log f(x)}{dx^2} = \begin{cases} \frac{-u\alpha}{\sigma^2}, & x < \eta \\ \frac{u \left[1 + e^{-u} (\alpha e^{-u} + u(\alpha - 1) - 1 - \alpha) \right]}{\sigma^2 (1 - e^{-u})^2}, & x > \eta \end{cases} \quad (9)$$

Respectively and where $u = \exp(-(x - \mu) / \sigma)$.

These derivatives show that $f(x)$ has a mode at $x^* = \mu + \sigma \log \alpha$ on the $(-\infty, \eta)$ support. On the other support it also may have a mode at solution point of the

$u + (1 - u\alpha)e^{-u} - 1 = 0$. As a result we can say that *TSGG* distribution can be bimodal. Further

$$\lim_{x \rightarrow -\infty} f(x) = \lim_{x \rightarrow \infty} f(x) = 0$$

Plots of the pdf (6) for some parameter values are given in Figure 1.

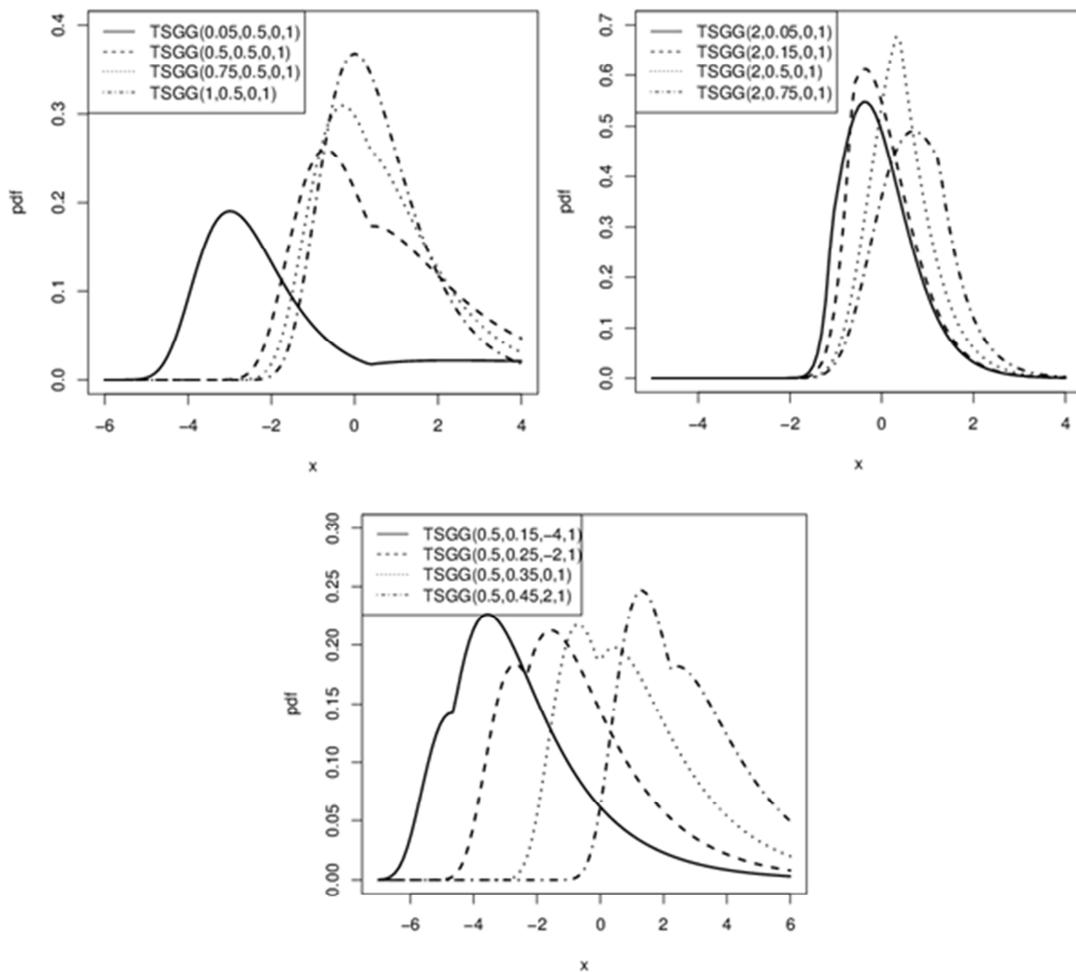


Figure 1. The pdf of the TSGG distribution for selected parameters value.

3. Hazard Rate Function

The hazard rate function defined by

$$h(t) = f(t) / (1 - F(t))$$

and its important issue in the lifetime modeling. For *TSGG* distribution hazard rate function is given by

$$h(t; \theta) = \begin{cases} \frac{\alpha u}{\sigma (\beta^{\alpha-1} e^{\alpha u} - 1)}, & t \leq \eta \\ \frac{\alpha u e^{-u}}{\sigma (1 - e^{-u})}, & \eta \leq t, \end{cases} \quad (10)$$

where $u = \exp(-(t - \mu) / \sigma)$. We plot the hazard rate function of the *TSGG* distribution in Figure 2. From Figure 2 we see that *TSGG* distribution has increasing hazard rate as ordinary *Gu* distribution. We also note that contrary to the *Gu* distribution, we observe that the hazard rate function of the *TSGG* can be firstly unimodal then increasing shaped for some

selected values of α and β . With this property, the *TSGG* distribution is much more advantageous than ordinary *Gu* distribution. Moreover

$$\lim_{t \rightarrow -\infty} h(t) = 0 \text{ and } \lim_{t \rightarrow \infty} h(t) = \alpha / \sigma.$$

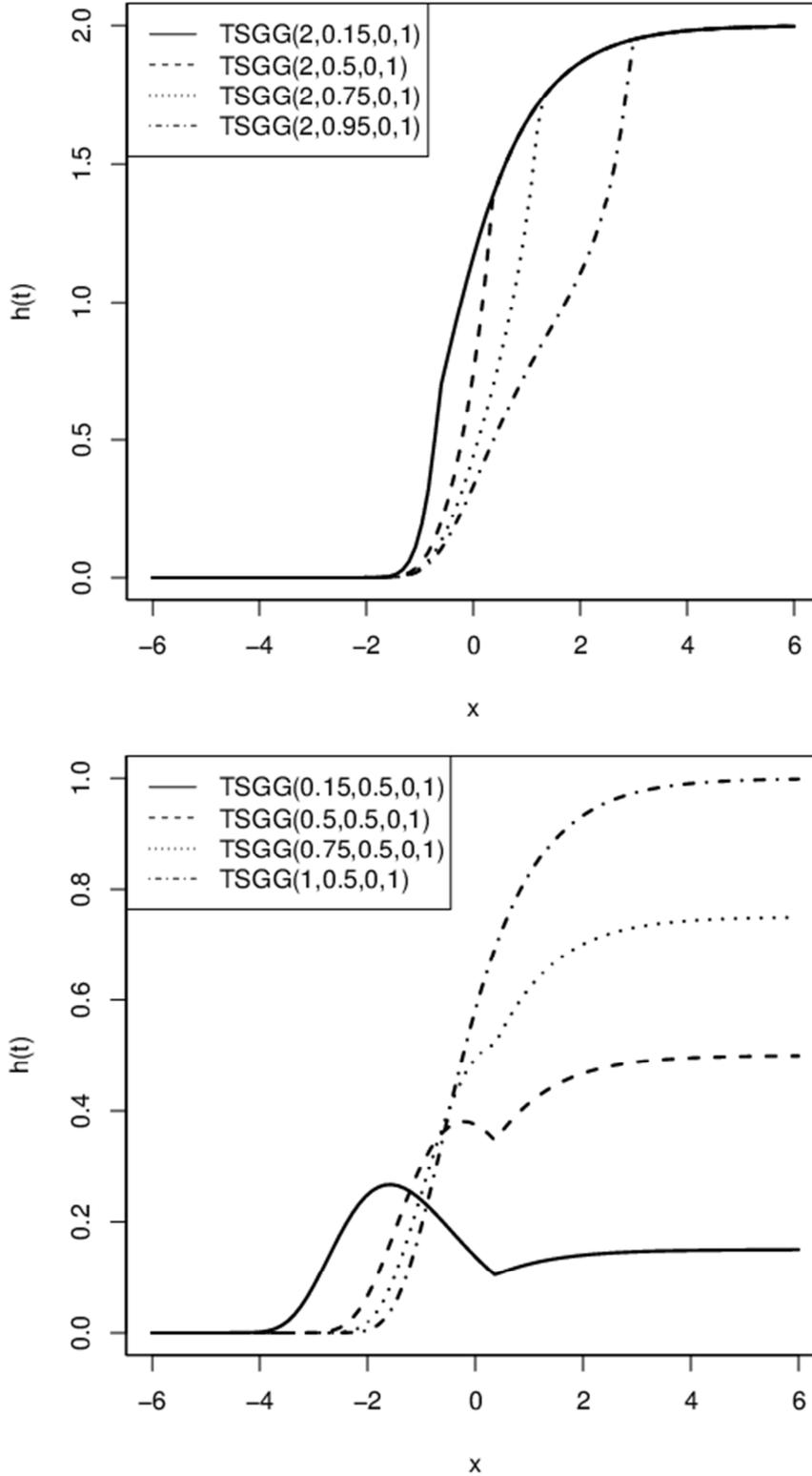


Figure 2. The hazard rate of the *TSGG* distribution for selected parameters value.

4. Moments

Using the moment definition and setting $x = \mu - \sigma \log u$ we write the r th moment of the *TSGG* distribution.

$$E(X^r) = \frac{\alpha}{(1-\beta)^{\alpha-1}} \int_0^{-\log \beta} (\mu - \sigma \log u)^r (1 - e^{-u})^{\alpha-1} e^{-u} du + \frac{\alpha}{\beta^{\alpha-1}} \int_{-\log \beta}^{\infty} (\mu - \sigma \log u)^r e^{-u\alpha} du \quad (11)$$

Using the binomial expansion for $(\mu - \sigma \log u)^r$ and $(1 - e^{-u})^{\alpha-1}$ r th moment can be obtained by

$$E(X^r) = \frac{\alpha \mu^r}{(1-\beta)^{\alpha-1}} \sum_{j=0}^r \sum_{k=0}^r (-1)^{k+j} \binom{\alpha-1}{j} \binom{r}{k} \left(\frac{\sigma}{\mu}\right)^k I(j, k) + \frac{\alpha \mu^r}{\beta^{\alpha-1}} \sum_{k=0}^r (-1)^k \binom{r}{k} \left(\frac{\sigma}{\mu}\right)^k I(k) \quad (12)$$

where $I(j, k)$ and $I(k)$ denotes these integrals with

$$I(j, k) = \int_0^{-\log \beta} (\log u)^k e^{-u(j+1)} du \quad (13)$$

$$I(k) = \int_{-\log \beta}^{\infty} (\log u)^k e^{-u\alpha} du \quad (14)$$

respectively. Especially for $r=1$, using equations (1.6.10.2) and (1.6.10.3) in Prudnikov et al. (1986) we obtain the following cases for the calculating expected value

$$I(j, 0) = \frac{1 - \beta^{j+1}}{j+1}, \quad (15)$$

$$I(j, 1) = \frac{1}{j+1} \left[Ei((j+1) \log \beta) - \beta^{j+1} \log(-\log \beta) - \log(j+1) \right] \quad (16)$$

$$I(0) = \alpha^{-1} \beta^\alpha, \quad (17)$$

$$I(1) = \frac{1}{\alpha} \left[\beta^\alpha \log(-\log \beta) - Ei(\alpha \log \beta) \right], \quad (18)$$

where $Ei(\cdot)$ denotes the exponential integral and is $Ei(-ax) = -\int_x^\infty t^{-1} e^{-at} dt, a > 0$ (Prudnikov et al., Eq. 1.3.2.14, 1986). Thus,

$$E(X) = \frac{\alpha}{(1-\beta)^{\alpha-1}} \sum_{j=0}^{\infty} (-1)^j \binom{\alpha-1}{j} \left[\frac{\mu(1-\beta^{j+1})}{j+1} - \sigma I(j, 1) \right] + \frac{\alpha}{\beta^{\alpha-1}} \left[\frac{\mu \beta^\alpha}{\alpha} - \sigma I(1) \right]. \quad (19)$$

We sketch the skewness, $\delta_1 = (\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3) / (\mu_2 - \mu_1^2)^{3/2}$, and kurtosis, $\delta_2 = (\mu_4 - 4\mu_1\mu_3 + 6\mu_2\mu_1^2 - 3\mu_1^4) / (\mu_2 - \mu_1^2)^2$, measurement in Figure 3 where $\mu_r \equiv E(X^r)$.

We can observe empirically that the *TSGG* distribution can be left skewed, symmetric or right skewed. So it is much more flexible than the *Gu* distribution.

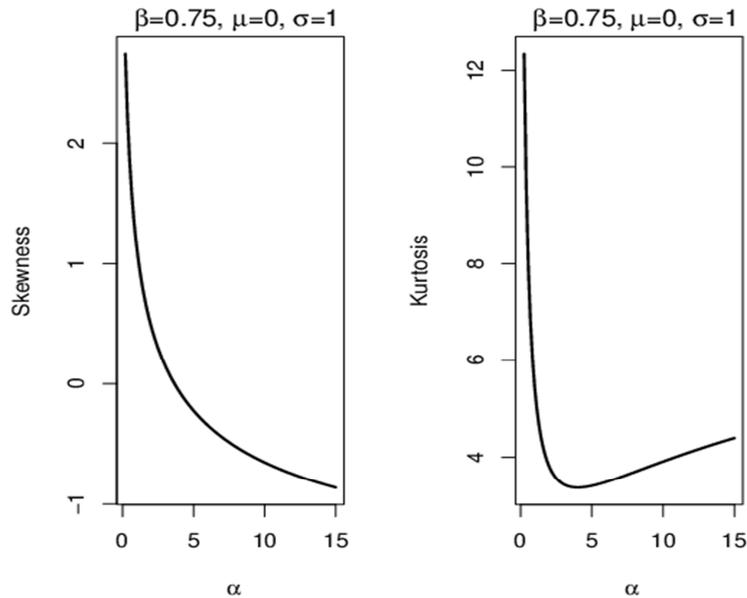


Figure 3. Skewness and kurtosis of the TSGG distribution for selected parameters value.

5. Maximum Likelihood Estimation

Let X_1, X_2, \dots, X_n be a random sample of size n from the TSGG distribution and let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ denote the corresponding order statistics. Then the log-likelihood function is given by

$$\begin{aligned} \ell(\boldsymbol{\theta}) = & n \log \alpha - n \log \sigma - \sum_{i=1}^n \frac{x_i - \mu}{\sigma} - \sum_{i=1}^n \exp\left(-\frac{x_i - \mu}{\sigma}\right) \\ & + (\alpha - 1) \log \left[\frac{\prod_{i=1}^r \exp\left\{-\exp\left(-\frac{x_{(i)} - \mu}{\sigma}\right)\right\} \prod_{i=r+1}^n \left[1 - \exp\left\{-\exp\left(-\frac{x_{(i)} - \mu}{\sigma}\right)\right\}\right]}{\beta^r (1 - \beta)^{n-r}} \right] \end{aligned} \quad (20)$$

where $x_{(r)} \leq \eta \leq x_{(r+1)}$ for $r = 1, 2, \dots, n$ and $x_{(0)} = -\infty$, $x_{(n+1)} = \infty$.

From Van dorp and Kotz (2002) and Korkmaz and Genç (2015), the estimating of the reflection point η is one of the $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ order statistics. Accordingly, we have the maximum likelihood estimate (MLE) of the β and α as

$$\hat{\beta} = \exp\left\{-\exp\left[-\left(x_{(\hat{r})} - \mu\right) / \sigma\right]\right\} \quad (21)$$

and by equating to zero the first derivative of the (20) respect to α

$$\hat{\alpha} = \frac{-n}{\log M(\hat{r}, \mu, \sigma)} \quad (22)$$

where $\hat{r} = \arg \max M(r, \mu, \sigma)$, $r \in (1, 2, \dots, n)$ with

$$M(r, \mu, \sigma) = \prod_{i=1}^{r-1} \frac{\exp\left\{-\exp\left[-(x_{(i)} - \mu) / \sigma\right]\right\}}{\exp\left\{-\exp\left[-(x_{(r)} - \mu) / \sigma\right]\right\}} \prod_{i=r+1}^n \frac{1 - \exp\left\{-\exp\left[-(x_{(i)} - \mu) / \sigma\right]\right\}}{1 - \exp\left\{-\exp\left[-(x_{(r)} - \mu) / \sigma\right]\right\}}. \quad (23)$$

For μ and σ , the associated likelihood estimating equations are found

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu} = \frac{n}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^n u_i - \frac{(\alpha-1)}{\sigma} \left[\sum_{i=1}^r u_i - \sum_{i=r+1}^n \frac{u_i e^{-u_i}}{1 - e^{-u_i}} \right] \quad (24)$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^n (u_i - 1) \log u_i + \frac{(\alpha-1)}{\sigma} \left[\sum_{i=1}^r u_i \log u_i - \sum_{i=r+1}^n \frac{u_i e^{-u_i} \log u_i}{1 - e^{-u_i}} \right] \quad (25)$$

where $u_i = \exp\left[-(x_i - \mu) / \sigma\right]$, $i = 1, 2, \dots, n$. We need some iterative procedure to find the estimates for μ and σ parameters. We may explain this procedure with an algorithm:

Step1: Set $k = 0$ and put an initial values $\hat{\mu}^{(0)}$ and $\hat{\sigma}^{(0)}$ for γ, θ in the log likelihood.

Step2: Compute the following estimates

$$\hat{\beta}^{(k+1)} = \exp\left\{-\exp\left[-(x_{(\hat{r})} - \mu^{(k)}) / \sigma^{(k)}\right]\right\}$$

$$\hat{\alpha}^{(k+1)} = -\frac{n}{\log M(\hat{r}, \mu^{(k)}, \sigma^{(k)})},$$

where $\hat{r} = \arg \max M(r, \mu^{(k)}, \sigma^{(k)})$, $r \in (1, 2, \dots, n)$ with

$$M(r, \mu^{(k)}, \sigma^{(k)}) = \prod_{i=1}^{r-1} \frac{\exp\left\{-\exp\left[-(x_i - \mu^{(k)}) / \sigma^{(k)}\right]\right\}}{\exp\left\{-\exp\left[-(x_r - \mu^{(k)}) / \sigma^{(k)}\right]\right\}} \prod_{i=r+1}^n \frac{1 - \exp\left\{-\exp\left[-(x_i - \mu^{(k)}) / \sigma^{(k)}\right]\right\}}{1 - \exp\left\{-\exp\left[-(x_r - \mu^{(k)}) / \sigma^{(k)}\right]\right\}}$$

Step3: Update μ and σ by using (25) and (26) to find $\hat{\mu}^{(k+1)}$ and $\hat{\sigma}^{(k+1)}$

Step4: If

$$\left| \ell(\hat{\alpha}^{(k+1)}, \hat{\beta}^{(k+1)}, \hat{\mu}^{(k+1)}, \hat{\sigma}^{(k+1)}) - \ell(\hat{\alpha}^{(k)}, \hat{\beta}^{(k)}, \hat{\mu}^{(k)}, \hat{\sigma}^{(k)}) \right|$$

is less than a given tolerance, say 10^{-2} , then stop. Else $k = k + 1$ and go to Step 2.

We note that the usual regularity conditions, which belong to the asymptotic normality of the MLEs, are not ensured for the *TSGG* distribution since the support of the pdf of the *TSGG* depends on parameters β, μ, σ and the pdf is not differentiable at η . In addition to the estimator of the β is based on the order statistics. So, the observed information matrix, which used to obtain the asymptotic variances of the MLEs, can be found numerically via optimization procedure in packet programme such as *R, Maple, Matlab*.

6. Data Analysis

In this section, we give a real data application. The

computations of the MLEs of all parameters for all the distributions are obtained by using the *optim* function in *R* program with *L-BFGS-B* method. This function also gives the numerically differentiated observed information matrix. The data are from the New York State Department of Conservation corresponding to the daily ozone level measurements in New York in May-September, 1973. Recently, Nadarajah (2008), Leiva et al.(2010), Cordeiro et al. (2013) and Korkmaz and Genç (2014) analyzed these data. To see the performance of the *TSGG*, we fit this data set to *TSGG*. After fitting the *TSGG* distribution to this data set, we find the following MLE results:

$$\hat{\alpha} = 0.1507(0.0141), \quad \hat{\beta} = 0.2443(0.0435), \\ \hat{\mu} = 18.3183(0.0274), \quad \hat{\sigma} = 4.7290(0.0278)$$

and $\ell(\hat{\boldsymbol{\theta}}) = -538.8174$ where standard errors are given in parentheses.

Also we give the value of the Kolmogorov-Smirnov goodness of fit test statistic as 0.0175 with a *p-value* 0.5995.

Hence we accept the null hypothesis that the data set is come from the *TSGG* distribution. We give the fitted *TSGG* density and empirical cdf plots in Figure. 4. These conclusions are

also supported by Figure 4. Therefore, we show that the *TSGG* distribution has the real data modeling potential.

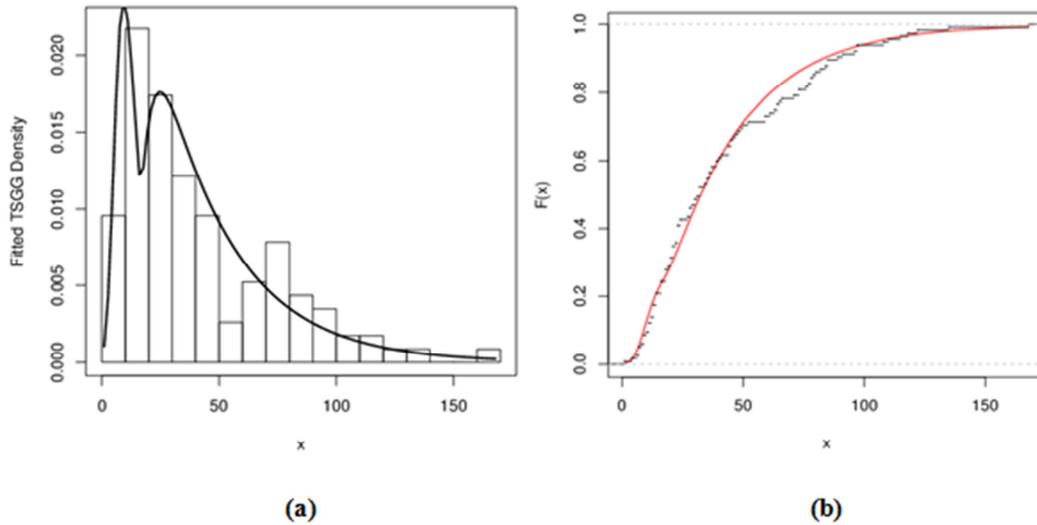


Figure 4. (a) Fitted the TSGG density of the ozone level data. (b) Empirical and fitted cdf's.

References

- [1] Al-Aqtash, R., Lee, C., Famoye, C. (2014). Gumbel-Weibull distribution: Properties and Applications. *Journal of Modern Applied Statistical Methods*, 13(2), 201-225.
- [2] Andrade, T., Rodrigues, H., Bourguignon, M., Cordeiro, G.M. (2015). The exponentiated generalized Gumbel Distribution. *Revista Colombiana de Estadística*, 38(1), 123-143.
- [3] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2006). *Statistics of Extremes: Theory and Applications*. West Sussex, England: John Wiley and Sons Ltd.
- [4] Cooray, K. (2010). Generalized Gumbel distribution. *Journal of Applied Statistics*, 37(1), 171-179.
- [5] Cordeiro, G.M., Nadarajah, S., Ortega, E.M.M. (2012). The Kumaraswamy Gumbel distribution. *Statistical Methods & Applications*, 21, 139-168.
- [6] Cordeiro, G.M., Silva, G.O., Ortega, E.M.M. (2013). The beta-Weibull geometric distribution. *Statistics: A Journal of Theoretical and Applied Statistics*, 47(4), 817-834.
- [7] Gumbel, E.J. (1958). *Statistics of Extremes*, Columbia University Press, New York.
- [8] Johnson, N. L., Kotz, S., Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2 (2nd ed.). New York: John Wiley and Sons, Inc.
- [9] Korkmaz, M. Ç., Genç, A. I. (2014). A lifetime distribution based on a transformation of a two-Sided power variate. *Journal of Statistical Theory and Applications*, (in press).
- [10] Korkmaz, M.Ç., Genç, A.I., (2015). A New Generalized Two-sided Class of Distributions with an Emphasis on Two-sided Generalized Normal Distribution. *Communications in Statistics Simulation and Computation*, DOI: 10.1080/03610918.2015.1005233.
- [11] Kotz, S., Nadarajah, S. (2000). *Extreme value distributions: theory and applications*. Imperial College Press, London.
- [12] Leiva, V., Vilca, F., Balakrishnan, N., Sanhueza, A. (2010). A skewed sinh-normal distribution and its properties and application to air pollution. *Communications in Statistics Theory and Methods*, 39, 426-443.
- [13] Nadarajah, S. (2008). A truncated inverted beta distribution with application to air pollution data. *Stochastic Environmental Research and Risk Assessment*, 22, 285-289.
- [14] Nadarajah, S. (2006). The exponentiated Gumbel distribution with climate application. *Environmetrics*, 17, 13-23.
- [15] Nadarajah, S., Kotz, S. (2004). The beta Gumbel Distribution. *Mathematical Problems in Engineering*, 4, 323-332.
- [16] Prudnikov, A. P., Brychkov, Y. A., Marichev, O. I. (1986). *Integrals and series*, vols 1, 2 and 3. Gordon and Breach Science Publishers, Amsterdam.
- [17] Van dorp, J. R., Kotz, S. (2002). The standard two-sided power distribution and its properties: With applications in financial engineering. *The American Statistician*, 56, 90-99.
- [18] Von Mises, R. (1954). La distribution de la grande de n valeurs, *American Mathematical Society*, 271-294.