MedCrave
Step into the World of Research
Biometrics & Biostatistics International Journal

Research Article

Open Access
CrossMark

Gumbel - Pareto distribution and it's applications in modeling COVID data

Abstract

A new distribution namely Gumbel- Pareto from Gumbel -X family[1] is introduced. Some properties including moments and order statistics are studied. A reliability measure for stress - strength analysis is derived. The method of maximum likelihood is proposed for estimating the distribution parameters.The flexibility of the new model is illustrated using two examples including Covid data.

Keywords: Gumbel distribution, gumbel - X family, gumbel – Pareto, order statistics, pareto distribution, stress - strength reliabilty, T - X family

Volume 10 Issue 3 - 2021

Jeena Joseph,[1] KK Jose[2]
[1]Department of Statistics, St.Thomas' College, Thrissur, India
[2]School of Mathematics and Statistics and Data Analytics, Mahatma Gandhi University, India

Correspondence: KK Jose, School of Mathematics and Statistics and Data Analytics, Mahatma Gandhi University, Kottayam, Kerala, India, Email kkjstc@gmail.com, kkj.smsda.mgu@gmail.com

Received: September 15, 2021 | **Published:** September 30, 2021

Introduction

Statistical distributions play an important role in parametric inference and are commonly applied to model real life data. In practical situations, existing standard distributions do not provide good fit to all types of real data sets. Hence statisticians are developing many new distributions which are flexible than standard distributions for the analysis of real data. New distributions are developed either by combining two or more existing distributions or by adding extra parameters to the existing distributions.

The beta generated family of distributions and Kumaraswamy generated family of distributions are generated by using distributions with support between 0 and 1 as the generator. As an extension, Alzaatreh et al.[2] proposed a general method by replacing the beta pdf with any non - negative continuous random variable T as the generator and another function $U(F(x))$ which satisfies the following conditions:

i. $U(F(x)) \in [a,b]$
ii. $U(F(x))$ is differentiable and monotonically non - decreasing.
iii. $U(F(x)) \to a$ as $x \to -\infty$ and $U(F(x)) \to b$ as $x \to \infty$

The new class of distributions is defined by

$$G(x) = \int_a^{U(F(x))} r(t)\,dt = R\left[U(F(x))\right]; \tag{1.1}$$

where $R(t)$ is the cdf and $r(t)$ is the pdf of the random variable T. Here, the cdf in (1.1) is a composite function of $(R,U,F)(x)$. The corresponding pdf is

$$g(x) = \left\{\frac{d}{dx}U(F(x))\right\} r\left\{U(F(x))\right\} \tag{1.2}$$

The p.d.f. $r(t)$ in (1.2) is "transformed" into a new pdf $g(x)$ through the function $U(F(x))$, which acts as a "transformer". That is, a random variable T, "the transformer", is used to transform another random variable T, "the transformed". The resulting family is known as "Transformed - Transformer" or "$T-X$" family of distributions. A large number of distributions, continuous and discrete, can be generated by applying any two existing univariate distributions based on this method. Alzaatreh et al.[2] gave several choices of $U(F(x))$ depending upon the support of the random variable T.

When the support of T is bounded or support of T is [0,1]: In this case $U(F(x))$ can be taken as $F(x)$ or $F^{\alpha}(x)$. This leads to the beta - generated family of distributions.

When the support of T is $[0,\infty)$, $a \geq 0$: Without loss of generality, we assume $a = 0$. Then $U(F(x))$ can be defined as $-log(1-F(x))$, $F(x)/(1-F(x))$, $-log(1-F^{\alpha}(x))$ and $F^{\alpha}(x)/(1-F^{\alpha}(x))$, where $\alpha > 0$.

When the support of T is $(-\infty,\infty)$: Then $U(F(x))$ can be taken as $log\left[-log(1-F(x))\right]$, $log\left[F(x)/(1-F(x))\right]$, $log\left[-log(1-F^{\alpha}(x))\right]$ and $log\left[F^{\alpha}(x)/(1-F^{\alpha}(x))\right]$.

In this paper, we are considering the third case, that is, the support of T is $(-\infty,\infty)$. For that, we consider T as the most important extreme value Type I distribution known as Gumbel distribution. This distribution has many applications including, to describe extreme wind spreads, sea wave heights, floods,rainfall during droughts, electrical strength of materials, air pollution problems, geological problems, naval engineering etc. Recently, the Gumbel distribution is used for modelling covid 19 data[4,5] also.

Al-Aqtash[1] proposed the Gumbel - X family by taking T as the Gumbel random variable

$$G(x) = e^{-e^{\frac{\mu}{\sigma}\left(\frac{F(x)}{\overline{F}(x)}\right)^{-1/\sigma}}} \tag{1.3}$$

By setting $\lambda = e^{\mu/\sigma}$ the cdf reduces to

$$G(x) = e^{-\lambda\left(\frac{F(x)}{\overline{F}(x)}\right)^{-1/\sigma}} \tag{1.4}$$

and the pdf is

$$g(x)=\frac{\lambda}{\sigma}f(x)\frac{\left(F(x)\right)^{\frac{1}{\sigma}-1}}{\left(\bar{F}(x)\right)^{\frac{1}{\sigma}+1}}e^{-\lambda\left(\frac{F(x)}{\bar{F}(x)}\right)^{-1/\sigma}} \quad (1.5)$$

The support of the random variable associated with (1.5) and f(.) are the same.

The paper is designed as follows. In section 2, we define the Gumbel-Pareto distribution. Some structural properties including moments, quantile function and order statistics are discussed in section 3.The maximum likelihood estimation of the model parameters is discussed in section 4.The application of this distribution to two real data sets are presented in section 5. In section 6, stress - strength analysis is discussed. Finally section 7 offers some concluding remarks.

Gumbel- Pareto distribution

Pareto distribution is a well known distribution for its capability in modeling heavy tailed data sets especially income and wealth data. Kochanczyk and Lipniack[6] has conducted a Pareto based evaluation of national responses to Covid - 19.

If the parent distribution is Pareto with parameters k and θ , with pdf

$$f(x)=\frac{k}{\theta}(\frac{x}{\theta})^{-k-1}, x>\theta \quad (2.1)$$

then the cdf of the four parameter Gumbel - Pareto distribution, denoted by $GuP(x;\lambda,\sigma,k,\theta)$ is given by

$$G_{GuP}(x;\lambda,\sigma,k,\theta)=e^{-\lambda[(\frac{x}{\theta})^k-1]^{-1/\sigma}}, \ x>\theta. \quad (2.2)$$

The corresponding pdf is given by

$$g_{GuP}(x;\lambda,\sigma,k,\theta)=\frac{\lambda k}{\sigma\theta}e^{-\lambda[(\frac{x}{\theta})^k-1]^{-1/\sigma}}[(\frac{x}{\theta})^k-1]^{-1/\sigma-1}(\frac{x}{\theta})^{k-1}. \quad (2.3)$$

The hazard function (hf) is obtained as

$$h(x;\lambda,\sigma,k,\theta)=\frac{\frac{\lambda k}{\sigma\theta}e^{-\lambda[(\frac{x}{\theta})^k-1]^{-1/\sigma}}[(\frac{x}{\theta})^k-1]^{-1/\sigma-1}(\frac{x}{\theta})^{k-1}}{1-e^{-\lambda[(\frac{x}{\theta})^k-1]^{-1/\sigma}}}, \quad (2.4)$$

Some structural properties

Transformation

Lemma 3.1:

If $Y\sim Gu(\mu,\sigma)$ *then* $X=\theta(e^Y+1)^{1/k}\sim Gu\ P$ *distribution*

The proof is done by using transformation technique.

Quantile function and simulation

The quantile function of Gumbel-Pareto is obtained by inverting (2.2) as

$$x=Q(u)=\theta[1+(-\frac{1}{\lambda}\log u)^{-\sigma}]^{1/k}$$

If $u\sim U(0,1)$, then X=Q(u) has pdf g(x).

By using Q(u), one can obtain the Galton skewness and Moor's Kurtosis which is defined as

$$S=\frac{Q(6/8)-2Q(4/8)+Q(2/8)}{Q(6/8)-Q(2/8)}$$

$$K=\frac{Q(7/8)-Q(5/8)+Q(3/8)-Q(1/8)}{Q(6/8)-Q(2/8)}$$

Moments

Theorem 3.1 The r^{th} raw moment of Gumbel Pareto distribution is

$$\mu_r^{'}=\theta^r\sum_{i=0}^{\infty}\lambda^{i\sigma}\binom{r/k}{i}\Gamma(1-i\sigma)$$

where $\Gamma(a)=\int_0^{\infty}t^{a-1}e^{-t}dt$ is the gamma function.

The skewness and kurtosis can also be calculated from ordinary moments using well-known relationships.

Order statistics

Order statistics deals with the properties and applications of ordered random samples and their functions. Suppose $X_1,X_2,....X_n$ be a random sample from Gumbel Pareto distribution. Let $X_{r:n}$ denote the r^{th} order statistic. Then the pdf of $X_{r:n}$ can be expressed as

$$g_{r:n}(x)=\frac{n!}{(r-1)!(n-r)!}\sum_{j=0}^{n-r}(-1)^j\binom{n-r}{j}g(x)G(x)^{j+r-1} \quad (3.1)$$

Inserting $g(x)$ and $G(x)$ in (3.1)and after some algebra we get,

$$g_{r:n}(x)=\sum_{j=0}^{n-r}\left[\frac{(-1)^j n!}{(r-1)!(n-r)!}\binom{n-r}{j}\frac{\lambda k}{\sigma\theta}[\left(\frac{x}{\theta}\right)^k-1]^{-1/\sigma-1}\left(\frac{x}{\theta}\right)^{k-1}exp\left\{-(r+j)(\lambda[\left(\left(\frac{x}{\theta}\right)^k-1\right]^{-1/\sigma}\right\}\right.$$

$$=\sum_{j=0}^{n-r}\xi_j g(x;\lambda,\sigma,k,\theta) \quad (3.2)$$

where $\xi_j=\frac{(-1)^j n!}{(r-1)!(n-r)!}\binom{n-r}{j}$ and $g(x;\lambda,\sigma,k,\theta)$ is the Gumbel Pareto density function with parameters λ , σ , θ and θ .

It reveals that the pdf of Gumbel Pareto order statistics is the mixture of Gumbel Pareto densities.

Maximum likelihood estimation

The maximum likelihood method is applied for estimating the parameters of Gumbel-Pareto distribution. Let $X_1,X_2,....X_n$ be a random sample from Gumbel Pareto(GuP) distribution. Also let $\grave{E}=(\lambda,\sigma,k,\theta)$ The likelihood function for the GuP distribution is given by

$$L(\Theta)=\left(\frac{\lambda k}{\sigma\theta}\right)^n exp\left\{\left\{-\lambda\sum_{i=1}^n\left[\left(\frac{x}{\theta}\right)^k-1\right]^{-1/\sigma}\right\}\prod_{i=1}^n\left[\left(\frac{x}{\theta}\right)^k-1\right]\right\}^{-1/\sigma-1}\prod_{i=1}^n\left(\frac{x}{\theta}\right)^{k-1}$$

The components of the score vector $U(\grave{E})$ are given by

$$U_\lambda=\frac{n}{\lambda}-\sum_{i=1}^n[(\frac{x_i}{\theta})^k-1]^{-1/\sigma}$$

Citation: Joseph J, Jose KK. Gumbel - Pareto distribution and it's applications in modeling COVID data. *Biom Biostat Int J.* 2021;10(3):125–128. DOI: 10.15406/bbij.2021.10.00338

$$U_\sigma = -\frac{n}{\sigma} - \lambda\sum_{i=1}^{n}\left\{\left[\left(\frac{x_i}{\theta}\right)^k - 1\right]^{-1/\sigma}\log\left[\left(\frac{x_i}{\theta}\right)^k - 1\right]\right\} + \frac{1}{\sigma^2}\sum_{i=1}^{n}\log\left[\left(\frac{x_i}{\theta}\right)^k - 1\right]$$

$$U_k = \frac{n}{k} + \frac{\lambda}{\sigma}\sum_{i=1}^{n}\left\{\left[\left(\frac{x_i}{\theta}\right)^k - 1\right]^{-1/\sigma-1}\left(\frac{x_i}{\theta}\right)^k log\left(\frac{x_i}{\theta}\right)\right\} + \left(-\frac{1}{\sigma} - 1\right)\sum_{i=1}^{n}\frac{\left[\left(\frac{x_i}{\theta}\right)^k log\left(\frac{x_i}{\theta}\right)\right]}{\left[\left(\frac{x_i}{\theta}\right)^k - 1\right]} + nlog\left(\frac{x_i}{\theta}\right)$$

$$U_\theta = -\frac{nk}{\theta} - \frac{\lambda k}{\sigma\theta}\sum_{i=1}^{n}\left\{[\left(\frac{x_i}{\theta}\right)^k - 1]^{-1/\sigma-1}\left(\frac{x_i}{\theta}\right)^k\right\} + \frac{k}{\theta}\left(\frac{1}{\sigma} + 1\right)\sum_{i=1}^{n}\frac{\left(\frac{x_i}{\theta}\right)^k}{\left[\left(\frac{x_i}{\theta}\right)^k - 1\right]}$$

The parameters can be estimated by equating these nonlinear equations to zero and solving them using the *nlm* function in R program.

Data analysis

In this section, we illustrate the effectiveness of Gumbel - Pareto distribution and compare the results with other existing models. To compare the distributions, we consider standardized goodness of fit measures like $-logL(\Theta)$, AIC (Akaike information criterion), CAIC (Consistent Akaike information criterion), BIC (Bayesian information criterion) and HQIC (Hannan - Quinn information criterion). Smaller these values, better is the fit.

Data set I: Number of deaths due to COVID-19 in China. This data is reported in (https://www.worldometers. info/coronavirus/country/china/) which represents daily deaths due to COVID-19 in China from 23 January to 28 March.

The data are: 8, 16, 15, 24, 26, 26, 38, 43, 46, 45, 57, 64, 65, 73, 73, 86, 89, 97, 108, 97, 146, 121, 143, 142, 105, 98, 136, 114, 118, 109, 97, 150, 71, 52, 29, 44, 47, 35, 42, 31, 38, 31, 30, 28, 27, 22, 17, 22, 11, 7, 13, 10, 14, 13, 11, 8, 3, 7, 6, 9, 7, 4, 6, 5, 3, 5.

Here we compare the new model with Exponentiated tranform of Gumbel type -II model (ETGT -II), Additive Gumbel type II (AGT -II) model and Gumbel type II model. The values of the statistics are given in Table 1.

From the table, we can see that the suggested model is suitable for real life applications.

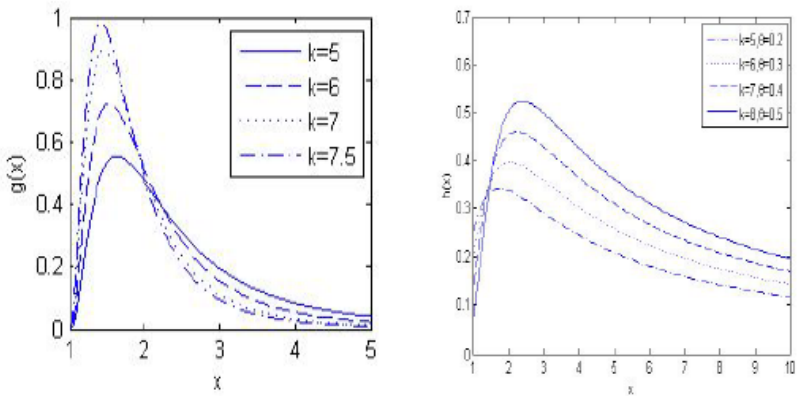

Figure 1 The graph of the pdf and hazard rate of Gumbel - Pareto distribution for various parameter values.

Table 1 The mles and the goodness of fit statistics $-logL$,AIC, CAIC, BIC and HQIC for the data set I

Distribution	mles	$-logL$	AIC	CAIC	BIC	HQIC
GuP	λ =2.527 σ =2.968 k=0.994	222.428	452.856	444.856	453.512	444.856
ETGT -II	θ =2.879 γ =1.086 δ =10.688 β =2.431	329.158	664.316	664.703	670.885	666.912
AGT -II	β =7.479 λ =13.432 δ =4.486 α =0.9137	331.081	670.162	670.818	678.921	673.623
GT -II	β =0.916 α =13.532	331.102	666.203	666.397	670.583	667.934

Citation: Joseph J, Jose KK. Gumbel - Pareto distribution and it's applications in modeling COVID data. *Biom Biostat Int J.* 2021;10(3):125–128. DOI: 10.15406/bbij.2021.10.00338

Data set II: The data set is a real data that consists of the number of successive failure for the air conditioning system reported of each member in a fleet of 13 Boeing 720 jet airplanes. The pooled data with 214 observations was considered by Proschan[7], Kus[8] and many others. Here we compare the model with existing Weibull Pareto model.

From Table 2, we can see that newly developed Gumbel Pareto distribution is suitable for the given data than the existing Weibull Pareto distribution.[2]

Table 2 The mles and their standard errors (SE) and the goodness of fit statistics $-logL$,AIC, CAIC, BIC and HQIC for the data set II

Distribution	mles	SE	$-logL$	AIC	CAIC	BIC	HQIC
GuP	λ =9.6166	1.213	1005.81	2017.62	2017.74	2027.72	2021.71
	σ =10.1333	2.281					
	k=7.233	1.678					
WP	θ =0.9981	0.0026	1459.62	2925.24	2925.35	2935.34	2929.32
	θ =9.8626	0.008					
	θ =0.9283	0.004					
	b=0.1267	0.00001					

Stress - strength analysis

The reliability is defined as the probability of not failing, denoted by R and is defined as $R = P(X < Y)$ where Y represents the stress and Y represents the strength of a component. For the evaluation of R , here we assume that both the random variables follow the distributions belonging to the same family and are independent. There are a number of applications in the literature including stress - strength model and breakdown of a system having two components. If X and Y are two independent random variables with cdf $F_1(x)$ and $F_2(y)$ and pdf $f_1(x)$ and $f_2(y)$ respectively. Then

$$R = P(X < Y) = \int_{-\infty}^{\infty} F_2(t) f_1(t) dt \tag{6.1}$$

Lemma 6.1 If X and Y are two independent random variables following Gumbel - X family of distributions with parameters (λ_1, σ_1) and (λ_2, σ_2) respectively. Then

$$R = \sum_{j=0}^{\infty} \frac{(-1)^j {\lambda_2}^j}{j! {\lambda_1}^{\frac{j\sigma_1}{\sigma_2}}} \Gamma\left(j\frac{\sigma_1}{\sigma_2} + 1\right) \tag{6.2}$$

A reliability test plan is developed when the life time of the items follow Gumbel - Pareto distribution. See Jeena and Jose[9] for more details.[10-14]

Conclusion

In this paper, we proposed the new Gumbel-Pareto distribution. We study some of its structural properties including moments, quantile functions and order statistics.The estimation of the model parameters is addressed by maximum likelihood method. We fit the new model to two real data sets to demonstrate the usefulness in practice. We conclude that GuP distribution provides consistently better fit than other competing models for the data set. We hope that the proposed model will attract wider applications in various areas such as engineering, survival and lifetime data, hydrology,economics, Biostatistical data on Cancer, Covid etc.

References

1. Al - Aqtash R. On generating a new family of distributions using the logit function. Ph.D. thesis, central michigan university, mount pleasant, michigan. 2013.
2. Alzaatreh A, Lee C, Famoye F. A new method for generating families of continuous distributions. Metron. 2013a;71(1):63–79.
3. Alzaatreh A, Lee C, Famoye F. Weibull pareto distribution and its applications. *Communications in statistics - theory and methods*. 2013b; 42(9): 1673–1691.
4. Yoo K, Arashi M, Bekker A. Pitting the Gumbel and logistic growth models against one another to model COVID-19 spread. *medRxiv*. 2020.
5. Sindhu TN, Shafiq A, Al-Mdallal QM. Exponentiated transformation of Gumbel Type-II distribution for modeling COVID-19 data. *Alexandria Engineering Journal*. 2021;60(1):671–689.
6. Kocha Åczyk M, Lipniacki T. Pareto-based evaluation of national responses to COVID-19 pandemic shows that saving lives and protecting economy are non-trade-off objectives. *Scientific reports*. 2021;11(1):1–9.
7. Proschan F. Theoretical explanation of observed decreasing failure rate. *Technometrics*. 1963;5(3):375–383.
8. Kus C. A New lifetime distribution. *Computational statistics and data analysis*. 2007;51(9):4497–4509.
9. Joseph J, Jose KK. Reliability test plan for gumbelâ€“pareto life time model. *International Journal of Statistics and Reliability Engineering*. 2021;8(1):121–131.
10. Beare BK, Toda. On the emergence of a power law in the distribution of COVID-19 cases. *Physica D: Nonlinear Phenomena*. 2020;412:132649.
11. EJ Gumbel. *Statistical theory of extreme values and some practical applications*. Applied Mathematics, 1st edn. vol. 33, U.S. Department of Commerce, National Bureau of Standards, ASIN B0007DSHG4, Gaithersburg, Md, USA. 1954.
12. Kotz S, Nadarajah S. *Extreme value distributions: theory and applications*. Imperial College Press, London. 2000.
13. S Nadarajah. The exponentiated Gumbel distribution with climate application, *Environmetrics*. 2006;17(1):13–23.
14. Wong F, Collins JJ. Evidence that coronavirus superspreading is fat-tailed. *Proceedings of the national academy of sciences*. 2020;117(47):29416–29418.

Citation: Joseph J, Jose KK. Gumbel - Pareto distribution and it's applications in modeling COVID data. *Biom Biostat Int J.* 2021;10(3):125–128. DOI: 10.15406/bbij.2021.10.00338