

Exponentiated Gumbel Model for Software Reliability Data Analysis using MCMC Simulation Method

Raj Kumar

National Institute of Electronics
and Information Technology,
Gorakhpur, U.P., India

Ashwini Kumar Srivastava

Dept. of Computer Application,
Shivharsh.Kisan P.G. College,
Basti, U.P., India

Vijay Kumar

Dept. of Maths & Statistics,
D.D.U. Gorakhpur University,
Gorakhpur, U.P., India

ABSTRACT

In this paper, the Markov chain Monte Carlo (MCMC) method has been used to estimate the parameters of Exponentiated Gumbel(EG) model based on a complete sample. A procedure is developed to obtain Bayes estimates of the parameters of the Exponentiated Gumbel model using MCMC simulation method in OpenBUGS, an established software for Bayesian analysis using Markov Chain Monte Carlo (MCMC) method. The MCMC methods have been shown to be easy to implement computationally, the estimates always exist and are statistically consistent, and their probability intervals are convenient to construct. The R functions are developed to study the statistical properties, model validation and comparison tools of the proposed model and the output analysis of MCMC samples generated from OpenBUGS. The proposed methodology is suitable for empirical modeling. A simulated data set is considered for illustration under uniform and gamma sets of priors.

Keywords

Exponentiated Gumbel(EG) model, Parameter estimation, Maximum likelihood estimate (MLE), Bayes estimates, Markov Chain Monte Carlo (MCMC), OpenBUGS.

1. INTRODUCTION

The Exponentiated Gumbel (EG) model has been proposed as a modification of the classical Gumbel model[1]. Since the Gumbel model[2] yields narrower confidence intervals than the some other extreme value models but has also the risk of under-estimating the certain important characteristic of the model. Hence, the choice of model is not trivial.

Recently, Nadarajah in 2006[3,4] introduced the Exponentiated Gumbel (EG) model as

$$F(x; \alpha, \sigma) = \exp(-\alpha \exp(u)); (\alpha, \sigma) > 0, -\infty < x < \infty.$$

Where $u = -(x/\sigma)$, Moreover, hazard-rate functions, moments, asymptotics and maximum likelihood functions were presented.

In this paper, we have proposed the EG model for modeling the software reliability data. We have obtained the ML estimates and associated probability intervals. The Bayes estimation of the EG model is considered when both parameters are unknown. It is observed that the Bayes estimates cannot be computed explicitly under the assumption of independent uniform and gamma priors for the parameters. We have developed the procedure to generate MCMC samples using Gibbs sampling technique from the posterior density function in OpenBUGS, Based on the generated posterior samples, we can compute the Bayes estimates of the unknown parameters and also can construct highest posterior

density credible intervals. We have also estimated the reliability function.

One real data set has been analyzed to demonstrate how the proposed method can be used in practice for software reliability data.

2. MODEL ANALYSIS

2.1 Probability density function(pdf)

The probability density function is given by

$$f(x) = \frac{\alpha}{\sigma} \exp\left\{-\left(\frac{x}{\sigma}\right)\right\} \exp\left[-\alpha \exp\left\{-\left(\frac{x}{\sigma}\right)\right\}\right]; \quad (2.1)$$

where $-\infty < x < \infty, \alpha > 0, \sigma > 0$.

The R functions *dexpo.gumbel()* and *pexpo.gumbel()* given in [5] can be used for the computation of pdf and cdf, respectively. Some of the typical EG density functions for different values of α and for $\sigma = 1$ are depicted in Figure 1.

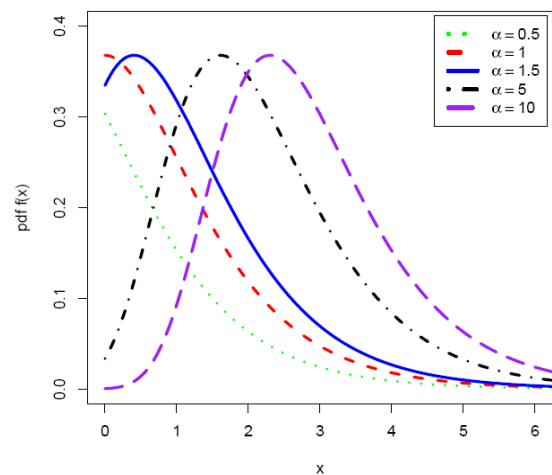


Fig 1. The PDF of EG model for $\sigma = 1$ and different values of α .

It is clear from the above figure that the density function of the Exponentiated Gumbel model can take different shapes

2.2 Cumulative density function(cdf)

The distribution function of Exponentiated Gumbel model with two parameters is given by

$$F(x) = \exp\left[-\alpha \exp\left\{-\left(\frac{x}{\sigma}\right)\right\}\right]; -\infty < x < \infty, \alpha > 0, \sigma > 0, \quad (2.2)$$

where, $\alpha > 0$ is the shape and $\sigma > 0$ is the scale parameter. The two-parameter Exponentiated Gumbel model will be denoted by $EG(\alpha, \sigma)$.

2.3 The Reliability function

The reliability/survival function is

$$R(x) = 1 - \exp \left[-\alpha \exp \left\{ -\left(\frac{x}{\sigma} \right) \right\} \right] \quad (2.3)$$

Here, $-\infty < x < \infty$, $\alpha > 0$ and $\sigma > 0$. The R function `sexpo.gumbel()` given in [5], computes the reliability function.

2.4 The Hazard function

The hazard rate function is

$$h(x; \alpha, \sigma) = \frac{\frac{\alpha}{\sigma} \exp \left\{ -\left(\frac{x}{\sigma} \right) \right\} \exp \left[-\alpha \exp \left\{ -\left(\frac{x}{\sigma} \right) \right\} \right]}{1 - \exp \left[-\alpha \exp \left\{ -\left(\frac{x}{\sigma} \right) \right\} \right]} \quad (2.4)$$

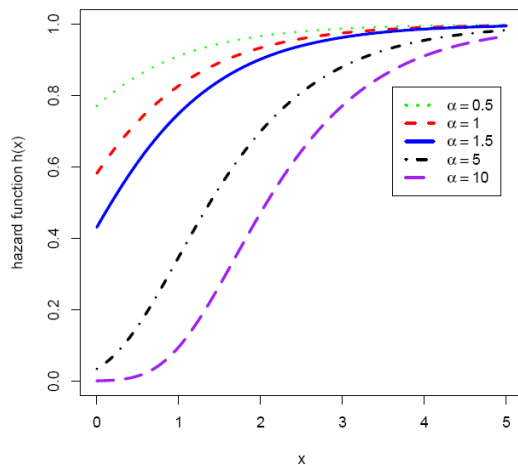


Fig 2. The hazard function of EG model for $\sigma=1$ and different values of α .

The hazard rate is an increasing function. It has been graphed in Figure 2 for scale parameter $\sigma=1$ and different values of shape parameter α .

The associated R function `hexpo.gumbel()` given in [5].

2.5 The cumulative hazard function

The cumulative hazard function $H(x)$ defined as

$$H(x) = -\{1 - \log F(x)\} \quad (2.5)$$

can be obtained with the help of `pexpo.gumbel()` function given in [5] by choosing arguments `lower.tail=FALSE` and `log.p=TRUE`, i.e.

`-pexpo.gumbel(x,alpha,sigma,lower.tail=FALSE,log.p=TRUE)`

2.6 The Failure rate average (fra) and Conditional survival cumulative hazard function(crf)

Two other relevant functions useful in reliability analysis are failure rate average (fra) and conditional survival function(crf). The failure rate average of X is given by

$$FRA(x) = \frac{H(x)}{x} = \frac{\int_0^x h(x) dx}{x}, \quad x > 0, \quad (2.6)$$

where $H(x)$ is the cumulative hazard function. An analysis for $FRA(x)$ on x permits to obtain the IFRA and DFRA classes.

The survival function (s.f.) and the conditional survival of X are defined by

$$R(x) = 1 - F(x)$$

and

$$R(x | t) = \frac{R(x+t)}{R(x)}, \quad t > 0, x > 0, R(\cdot) > 0, \quad (2.7)$$

respectively, where $F(\cdot)$ is the cdf of x . Similarly to $h(x)$ and $FRA(x)$, the distribution of x belongs to the new better than used (NBU), exponential, or new worse than used (NWU) classes, when $R(x | t) < R(x)$, $R(t | x) = R(x)$, or $R(x | t) > R(x)$, respectively.

The R functions `hra.expo.gumbel()` and `crf.expo.gumbel()` given in [5] can be used for the computation of failure rate average (fra) and conditional survival function(crf), respectively.

2.7 The Quantile function

The quantile function is given by

$$x_q = -\sigma \log \left\{ -\frac{1}{\alpha} (\log q) \right\}; \quad 0 < q < 1. \quad (2.8)$$

The computation of quantiles the R function `qexpo.gumbel()`, given in [5].

2.8 The random deviate generation function

Let U be the uniform (0,1) random variable and $F(\cdot)$ a cdf for which $F^{-1}(\cdot)$ exists. Then $F^{-1}(u)$ is a draw from distribution $F(\cdot)$. Therefore, the random deviate can be generated from $EG(\alpha, \sigma)$ by

$$x = -\sigma \log \left(-\frac{1}{\alpha} \log(u) \right); \quad 0 < u < 1. \quad (2.9)$$

where u has the $U(0, 1)$ distribution. The R function `rexpo.gumbel()`, given in [5], generates the random deviate from $EG(\alpha, \sigma)$.

3. MLE (MAXIMUM LIKELIHOOD ESTIMATION) AND INFORMATION MATRIX

For completeness purposes, in this section, we briefly discuss the maximum likelihood estimators (MLE's) of the two-parameter Exponentiated Gumbel model and discuss their asymptotic properties to obtain approximate confidence intervals based on MLE's[6].

Let $\underline{x}=(x_1, \dots, x_n)$ be a random sample of size n from $EG(\alpha, \sigma)$, then the log-likelihood function $L(\alpha, \sigma)$ can be written as;

$$\log L = n \log \alpha - n \log \sigma - \sum_{i=1}^n \frac{x_i}{\sigma} - \alpha \sum_{i=1}^n \exp \left\{ - \left(\frac{x_i}{\sigma} \right) \right\} \quad (3.1)$$

Therefore, to obtain the MLE's of α and σ [7], we can maximize (3.1) directly with respect to α and σ or we can solve the following two non-linear equations using Newton-Raphson method

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^n \exp \left\{ - \left(\frac{x_i}{\sigma} \right) \right\}$$

and,

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^n \left(\frac{x_i}{\sigma^2} \right) - \frac{\alpha}{\sigma^2} \sum_{i=1}^n x_i \exp \left\{ - \left(\frac{x_i}{\sigma} \right) \right\}$$

3.1 Information Matrix and Asymptotic Confidence Intervals

Let us denote the parameter vector by $\underline{\delta}=(\alpha, \sigma)$ and the corresponding MLE of $\underline{\theta}$ as $\hat{\underline{\delta}}=(\hat{\alpha}, \hat{\sigma})$ then the asymptotic normality results in

$$(\hat{\underline{\delta}} - \underline{\delta}) \rightarrow N_2 \left(0, (I(\underline{\delta}))^{-1} \right) \quad (3.2)$$

where $I(\underline{\delta})$ is the Fisher's information matrix given by

$$I(\underline{\delta}) = - \begin{bmatrix} E \left(\frac{\partial^2 \ln L}{\partial \alpha^2} \right) & E \left(\frac{\partial^2 \ln L}{\partial \alpha \partial \sigma} \right) \\ E \left(\frac{\partial^2 \ln L}{\partial \alpha \partial \sigma} \right) & E \left(\frac{\partial^2 \ln L}{\partial \sigma^2} \right) \end{bmatrix} \quad (3.3)$$

In practice, it is useless that the MLE has asymptotic variance $(I(\underline{\delta}))^{-1}$ because we do not know $\underline{\delta}$. Hence, we approximate the asymptotic variance by "plugging in" the estimated value of the parameters [8]. The common procedure is to use observed Fisher information matrix $O(\hat{\underline{\delta}})$ (as an estimate of the information matrix $I(\underline{\delta})$) given by

$$O(\hat{\underline{\delta}}) = - \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \alpha^2} & \frac{\partial^2 \ln L}{\partial \alpha \partial \sigma} \\ \frac{\partial^2 \ln L}{\partial \alpha \partial \sigma} & \frac{\partial^2 \ln L}{\partial \sigma^2} \end{bmatrix} \bigg|_{(\hat{\alpha}, \hat{\sigma})} = -H(\underline{\theta}) \bigg|_{\hat{\underline{\delta}}=\hat{\underline{\delta}}} \quad (3.4)$$

where H is the Hessian matrix, $\underline{\delta}=(\alpha, \sigma)$ and $\hat{\underline{\delta}}=(\hat{\alpha}, \hat{\sigma})$. The Newton-Raphson algorithm to maximize the likelihood produces the observed information matrix. Therefore, the variance-covariance matrix is given by

$$\left(-H(\underline{\theta}) \bigg|_{\hat{\underline{\delta}}=\hat{\underline{\delta}}} \right)^{-1} = \begin{bmatrix} \text{Var}(\hat{\alpha}) & \text{cov}(\hat{\alpha}, \hat{\sigma}) \\ \text{cov}(\hat{\alpha}, \hat{\sigma}) & \text{Var}(\hat{\sigma}) \end{bmatrix} \quad (3.5)$$

Hence, from the asymptotic normality of MLEs, approximate 100(1- γ)% confidence intervals for α and σ can be constructed as

$$\hat{\alpha} \pm z_{\gamma/2} \sqrt{\text{Var}(\hat{\alpha})} \quad \text{and} \quad \hat{\sigma} \pm z_{\gamma/2} \sqrt{\text{Var}(\hat{\sigma})}$$

where $z_{\gamma/2}$ is the upper percentile of standard normal variate.

3.2 Computation of Maximum Likelihood Estimation

We are using software reliability data set *SYS2.DAT* - 86 time-between-failures [9] is considered for illustration of the proposed methodology. In this real data set, Time-between-failures is converted to time to failures and scaled.

The Exponentiated Gumbel model is used to fit this data set. We have started the iterative procedure by maximizing the log-likelihood function given in (3.1) directly with an initial guess for $\alpha = 1.15$ and $\sigma = 130.0$, for away from the solution. We have used *optim()* function in R with option Newton-Raphson method [10]. The iterative process stopped only after 24 iterations. We obtain $\hat{\alpha} = 4.056801$, $\hat{\sigma} = 151.915039$ and the corresponding log-likelihood value = -734.5816. The similar results are obtained using *maxLik* package available in R. An estimate of variance-covariance matrix, using (3.4), is given by

$$\begin{bmatrix} \text{var}(\hat{\alpha}) & \text{cov}(\hat{\alpha}, \hat{\sigma}) \\ \text{cov}(\hat{\alpha}, \hat{\sigma}) & \text{var}(\hat{\sigma}) \end{bmatrix} = \begin{bmatrix} 0.1599290 & -0.3198579 \\ -0.3198579 & 9.4358089 \end{bmatrix}$$

Thus using (3.5), we can construct the approximate 95% confidence intervals for the parameters of Exponentiated Gumbel model based on MLE's. Table 1 shows the MLE's with their standard errors and approximate 95% confidence intervals for α and σ .

Table 1. Maximum likelihood estimate, standard error and 95% confidence interval

Parameter	MLE	Std. Error	95% Confidence Interval
alpha	4.056801	0.39632	(3.2800138, 4.8335882)
sigma	151.9150	3.06992	(145.897995, 157.932082)

4. MODEL VALIDATION

To study the goodness of fit of the EG model, we compute the Kolmogorov-Smirnov statistics between the empirical distribution function and the fitted distribution function when the parameters are obtained by method of maximum likelihood [11]. For this we have used the R function *ks.expo.gumbel()* given in [5]. The result of K-S test is $D=0.0708$ with the corresponding p-value = 0.634. Therefore, the high p-value clearly indicates that EG model can be used to analyze this data set.

We plot the empirical distribution function and the fitted distribution function in Figure 3, which indicates reasonable match between the empirical distribution function and the fitted distribution function.

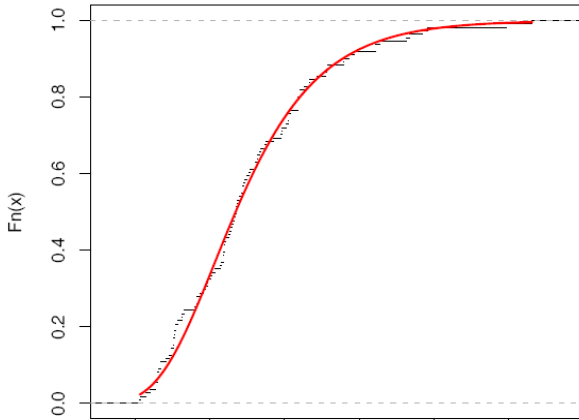


Fig 3. The graph of empirical distribution function and fitted distribution function.

Therefore, from above result and Figure 3, it is clear that the estimated EG model provides excellent fit to the given data.

The graphical methods widely used for checking whether a fitted model is in agreement with the data are Quantile-Quantile(Q-Q) and Probability-Probability (P-P) plots in model validation. The corresponding R functions are *qq.expo.gumbel()* and *pp.expo.gumbel()* given in [5].

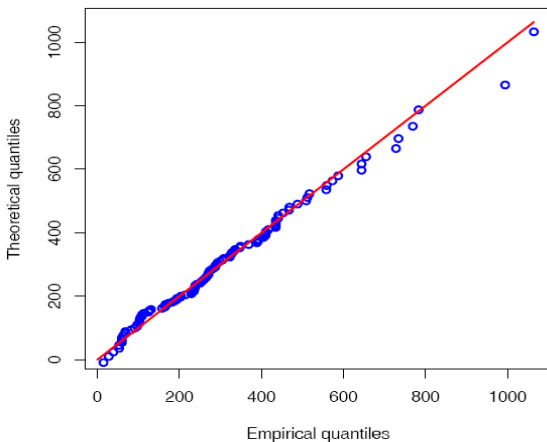


Fig 4. Quantile-Quantile(Q-Q) plot using MLEs as estimate.

Let $\hat{F}(x)$ be an estimate of $F(x)$ based on $x_1, x_2, \dots, x_n$. The scatter plot of the points

$\hat{F}^{-1}(p_{i:n})$ versus $x_{i:n}$, $i = 1, 2, \dots, n$, is called a Q-Q plot.

The Q-Q plot shows the estimated versus the observed quantiles. If the model fits the data well, the pattern of points on the Q-Q plot will exhibit a 45-degree straight line. As can be seen from the straight line pattern in Figure 4, the EG model fits the data very well. This is also supported by the Probability-Probability(P-P) plot in Figure 5.

Let $x_1, x_2, \dots, x_n$ be a sample from a given population with estimated cdf $\hat{F}(x)$. The scatter plot of the points

$\hat{F}(x_{i:n})$ versus $p_{i:n}$, $i = 1, 2, \dots, n$, is called a P-P plot. If the model fits the data well, the graph will be close to the 45-degree line[12]. Here we note that all the points in the P-P plot are inside the unit square $[0, 1] \times [0, 1]$.

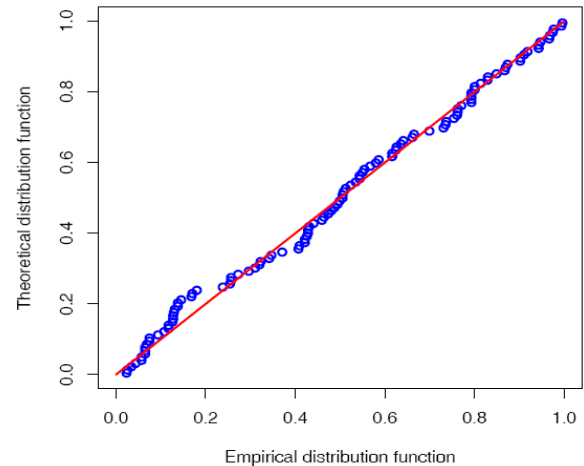


Fig 5. Probability-Probability(P-P) plot using MLEs as estimate.

As can be seen from Figure 4 and Figure 5 that the data do not deviate dramatically from the line.

5. BAYESIAN ESTIMATION USING MARKOV CHAIN MONTE CARLO (MCMC) METHOD

A Monte Carlo method is an algorithm that relies on repeated pseudo-random sampling for computation, and is therefore stochastic (as opposed to deterministic). Monte Carlo methods are often used for simulation. The union of Markov chains and Monte Carlo methods is called MCMC[13]. A Markov chain is a random process with a finite state-space and the Markov property, meaning that the next state depends only on the current state, not on the past[14].

The revival of Bayesian inference since the 1980s is due to MCMC algorithms and increased computing power. The most prevalent MCMC algorithms may be the simplest: random-walk Metropolis and Gibbs sampling. The quality of the marginal samples usually improves with the number of iterations. In Bayesian inference, the target distribution of each Markov chain is usually a marginal posterior distribution, such as each parameter. Each Markov chain begins with an initial value and the algorithm iterates, attempting to maximize the logarithm of the unnormalized joint posterior distribution and eventually arriving at each target distribution. Each iteration is considered a state.

The Gibbs algorithm starts by assuming some arbitrarily chosen initial values for the concerned variates and then generating the variate values from the various full conditionals in a cyclic order. That is, every time a variate value is generated from a full conditional, it is influenced by the most recent values of all other conditioning variables and, after each cycle of iteration, it is updated by sampling a new value from its full conditional. The entire generating scheme is repeated unless the generating chain achieves a systematic pattern of convergence. It can be shown that after a large number of iterations the generated variates can be regarded as

the random samples from the corresponding posteriors. Readers are referred to Gamerman [13, 14] for details of the procedure and the related convergence diagnostic issues.

The most widely used piece of software for applied Bayesian inference is the OpenBUGS[15]. The software offers a user-interface, based on dialogue boxes and menu commands, through which the model may then be analyzed using Markov Chain Monte Carlo techniques. It is a fully extensible modular framework for constructing and analyzing Bayesian probability models for the existing probability models, [16, 17]. As the EG model is not available in OpenBUGS, thus it requires incorporation of a module to estimate parameters of EG model. The Bayesian analysis of a probability model can be performed for the models defined in OpenBUGS. Recently, a number of probability models have been incorporated in OpenBUGS to facilitate the Bayesian analysis, [18]. The readers are referred to [2, 10-12] for implementation details of some models.

A module, *dexpo.gumbel(alpha, sigma)*, is written in component Pascal for EG model to perform full Bayesian analysis in OpenBUGS using the method described in Thomas [18], Lunn et al. [17] and Thomas et al. [16]. The module code can be obtained from the author.

5.1 Bayesian Analysis under Uniform Priors

The developed module is implemented to obtain the Bayes estimates of the EG model using MCMC method. The main function of the module is to generate MCMC sample from posterior distribution for non-informative set of priors, i.e. Uniform priors.

It frequently happens that the experimenter knows in advance that the probable values of θ lie over a finite range $[a, b]$ but has no strong opinion about any subset of values over this range. In such a case a uniform distribution over $[a, b]$ may be a good approximation of the prior distribution, its p.d.f. is given by

$$\pi(\theta) = \begin{cases} \frac{1}{b-a} & ; 0 < a \leq \theta \leq b \\ 0 & ; \text{otherwise} \end{cases}$$

We run the two parallel chains for sufficiently large number of iterations, say 5000 the burn-in, until convergence results. Final posterior sample of size 7000 is taken by choosing thinning interval five i.e. every fifth outcome is stored. Therefore, we have the posterior sample $\{\alpha_{1i}, \lambda_{1i}\}$, $i = 1, \dots, 7000$ from chain 1 and $\{\alpha_{2i}, \lambda_{2i}\}$, $i = 1, \dots, 7000$ from chain 2.

The chain 1 is considered for convergence diagnostics plots. The visual summary is based on posterior sample obtained from chain 2 whereas the numerical summary is presented for both the chains.

5.1.1 Convergence diagnostics

Before examining the parameter estimates or performing other inference, it is a good idea to look at plots of the sequential (dependent) realizations of the parameter estimates and plots thereof. We have found that if the Markov chain is not mixing well or is not sampling from the stationary distribution, this is usually apparent in sequential plots of one or more realizations. The sequential plot of parameters is the plot that most often exhibits difficulties in the Markov chain.

5.1.1.1 History(Trace) plot

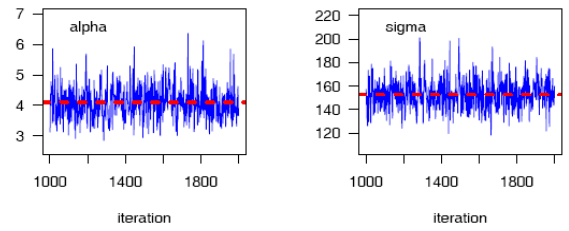


Fig 6. Sequential realization of the parameters α and σ

From the graph, we can conclude that the chain has converged as the plots show no long upward or downward trends, but look like a horizontal band.

5.1.1.2 Autocorrelation plot

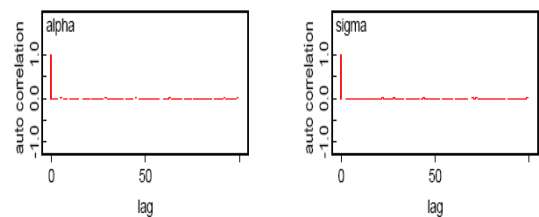


Fig 7. The autocorrelation plots for α and σ

5.1.2 Visual summary by using Kernel density estimates

Histograms can provide insights on skewness, behaviour in the tails, presence of multi-modal behaviour, and data outliers; histograms can be compared to the fundamental shapes associated with standard analytic distributions.

Histogram and kernel density estimate of α and σ based on MCMC samples, vertical lines represent the corresponding MLE and Bayes estimate.

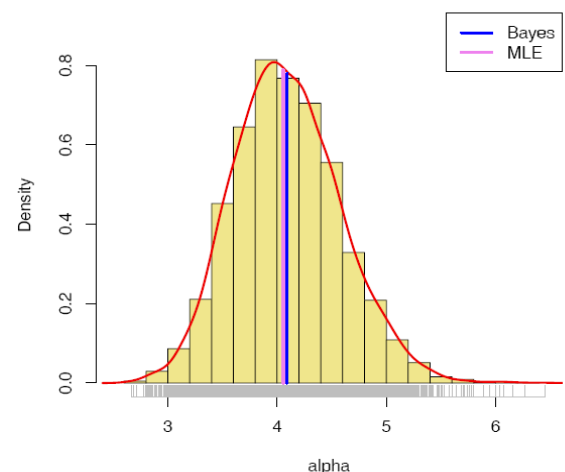


Fig 8. Histogram and kernel density estimate of α

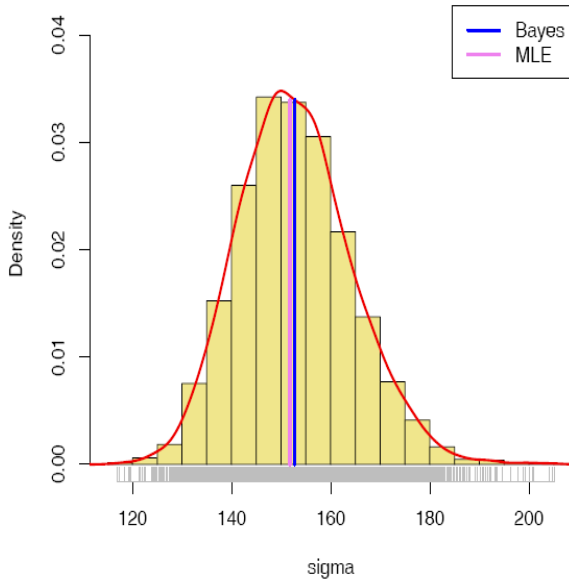


Fig 9. Histogram and kernel density estimate of σ

5.1.3 Numerical Summary

In Table 2, we have considered various quantities of interest and their numerical values based on MCMC sample of posterior characteristics for Exponentiated Gumbel model under uniform priors. The numerical summary is based on final posterior sample (MCMC output) of 7000 samples for alpha and sigma.

$\{\alpha_{1i}, \sigma_{1i}\}$, $i = 1, \dots, 7000$ from chain 1,

and,

$\{\alpha_{2i}, \sigma_{2i}\}$, $i = 1, \dots, 7000$ from chain 2.

Table 2 : Numerical summaries based on MCMC sample of posterior characteristics for Exponentiated Gumbel model under Uniform priors

Characteristics	Chain 1		Chain 2	
	alpha	sigma	alpha	sigma
Mean	4.09047	153.1463	4.091270	152.8067
Standard Deviation	0.4945715	11.88160	0.4877884	11.48320
Monte Carlo(MC) error	0.006463	0.1533	0.005338	0.1362
Minimum	2.588	118.1	2.66600	117.0
2.5th Percentile($P_{2.5}$)	3.184	132.4	3.20795	132.4
First Quartile (Q_1)	3.742	144.7	3.74975	144.8
Median	4.068	152.3	4.06450	152.2
Third Quartile (Q_3)	4.420	160.6	4.40925	160.1
97.5th Percentile($P_{97.5}$)	5.133	178.5	5.10105	176.8
Maximum	6.328	207.1	6.45000	205.1
Mode	3.953174	150.9671	3.965804	149.7628
95% Credible Interval	(3.184, 5.133)	(132.4, 178.5)	(3.20795, 5.10105)	(132.4, 176.8)
95% HPD Credible Interval	(3.137, 5.061)	(131.4, 177.0)	(3.189, 5.074)	(132.5, 176.8)

5.2 Bayesian Analysis under Gamma Priors

The developed module is implemented to obtain the Bayes estimates of the Exponentiated Gumbel model using MCMC method to generate MCMC sample from posterior distribution for given set of gamma priors, which is most widely used prior distribution of θ is the inverted gamma distribution with parameters a and b (>0) with p.d.f. given by

$$\pi(\theta) = \begin{cases} \frac{b^a}{\Gamma(a)} \theta^{-(a+1)} e^{-b/\theta} & ; \theta > 0 \quad (a, b) > 0 \\ 0 & ; \text{otherwise} \end{cases}$$

We also run the two parallel chains for sufficiently large number of iterations, say 5000 the burn-in, until convergence results. Final posterior sample of size 7000 is taken by choosing thinning interval five i.e. every fifth outcome is stored and same procedure is adopted for analysis as used in the case of uniform priors.

5.2.1 Convergence diagnostics

Simulation-based Bayesian inference requires using simulated draws to summarize the posterior distribution or calculate any relevant quantities of interest. We need to treat the simulation draws with care. The first step in making an inference from an MCMC analysis is to ensure that an equilibrium distribution has indeed been reached by the Markov chain, i.e., that the chain has converged. For each parameter, we started the chain at an arbitrary point (the initial value or init chosen for each parameter), and because successive draws are dependent on the previous values of each parameter, the actual values chosen for the inits will be noticeable for a while. Therefore, only after a while is the chain independent of the values with which it was started. These first draws ought to be discarded as a burn-in as they are unrepresentative of the equilibrium distribution of the Markov chain.

5.2.1.1 Running Mean (Ergodic mean) Plot

The convergence pattern can be studied by calculating the running mean which is the mean of all sampled values up to and including that at a given iteration. We thus generate a time series(Iteration number) graph of the running mean for each parameter in the chain. The Ergodic mean plots for the parameters shown in figure 10 depict the convergence pattern.

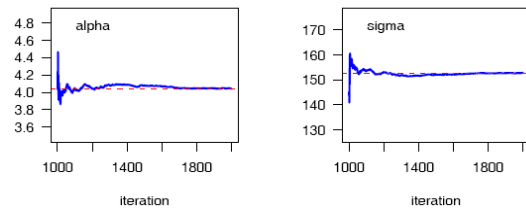


Fig 10. The Ergodic mean plots for alpha and sigma

5.2.1.2 Brooks-Gelman-Rubin Diagnostic

Evidence for convergence comes from the red line being close to 1 on the y-axis and from the blue and green lines being stable (horizontal) across the width of the plot.

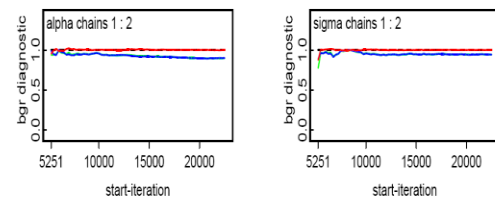


Fig 11. The BGR plots for alpha and sigma.

From the Figure 11, it is clear that convergence is achieved. Thus we can obtain the posterior summary statistics.

5.2.2 Visual summary by using Box plots

The boxes represent inter-quartile ranges and the solid black line at the (approximate) centre of each box is the mean; the arms of each box extend to cover the central 95 per cent of the distribution - their ends correspond, therefore, to the 2.5% and 97.5% quantiles. (Note that this representation differs somewhat from the traditional.)

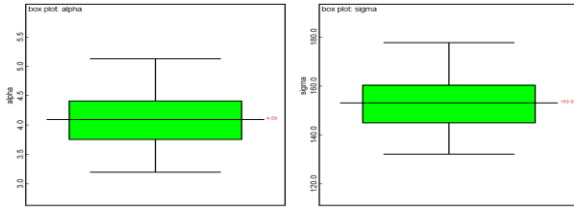


Fig 12. The boxplots for alpha and sigma.

5.2.3 Numerical Summary

In Table 3, we have considered various quantities of interest and their numerical values based on MCMC sample of posterior characteristics for EG model under Gamma priors. The Highest probability density (HPD) intervals are computed the algorithm described by Chen and Shao (1999) under the assumption of unimodal marginal posterior distribution

Table 3 : Numerical summaries based on MCMC sample of posterior characteristics for Exponented Gumbel model under Gamma priors

Characteristics	Chain 1		Chain 2	
	alpha	sigma	alpha	sigma
Mean	4.059488	152.9463	4.071116	152.7308
Standard Deviation	0.4994789	11.70947	0.4945235	11.59696
Monte Carlo(MC) error	0.005553	0.1308	0.005761	0.1386
Minimum	2.54100	119.6	2.30300	119.5000
2.5th Percentile($P_{2.5}$)	3.18395	131.5	3.16685	131.6000
First Quartile (Q_1)	3.71100	144.8	3.72700	144.8000
Median	4.02950	152.4	4.04400	152.3000
Third Quartile (Q_3)	4.37725	160.5	4.38500	160.1000
97.5th Percentile($P_{97.5}$)	5.12500	177.0	5.10500	177.0025
Maximum	6.38200	202.0	6.19700	200.7000
Mode	3.994693	152.3224	4.016151	153.0121
95% Credible Interval	(3.18395, 5.125)	(131.5, 177.0)	(3.16685, 5.1050)	(131.6, 177.0025)
95% HPD Credible Interval	(3.168, 5.089)	(130.5, 175.7)	(3.086, 5.006)	(131.1, 176.1)

6. COMPARISON WITH MLE UNDER UNIFORM PRIORS

For the comparison with MLE, we have plotted two graphs. In Figure 13, the density functions $f(x; \hat{\alpha}, \hat{\sigma})$ using MLEs and Bayesian estimates, computed via MCMC samples under uniform priors, are plotted.

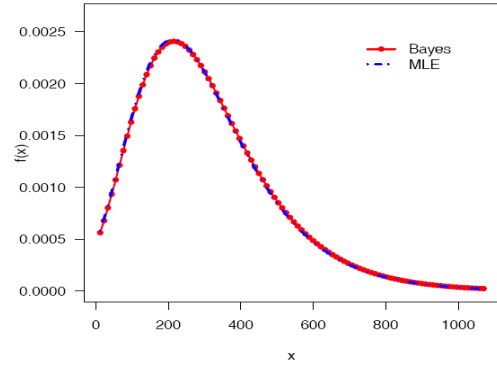


Fig 13. The density functions $f(x; \hat{\alpha}, \hat{\sigma})$ using MLEs and Bayesian estimates.

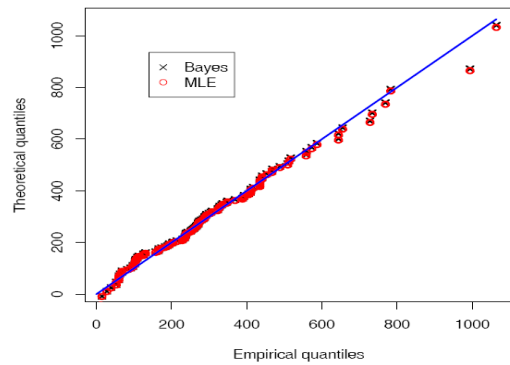


Fig 14. Q-Q plot of empirical quantiles and theoretical quantiles computed from MLE and Bayes estimates.

whereas, Fig.14 represents the Quantile-Quantile(Q-Q) plot of empirical quantiles and theoretical quantiles computed from MLE and Bayes estimates.

Thus, It is clear from the above figures, the MLEs and the Bayes estimates with respect to the uniform priors are quite close and fit the data very well.

7. COMPARISON WITH MLE UNDER GAMMA PRIORS

For the comparison with MLE, we have plotted a graph which exhibits the estimated reliability function which is shown by dashed line using Bayes estimate under gamma priors and the empirical reliability function which is shown by solid line. It is clear from Fig.15, the MLEs and the Bayes estimates with respect to the gamma priors are quite close and fit the data very well.

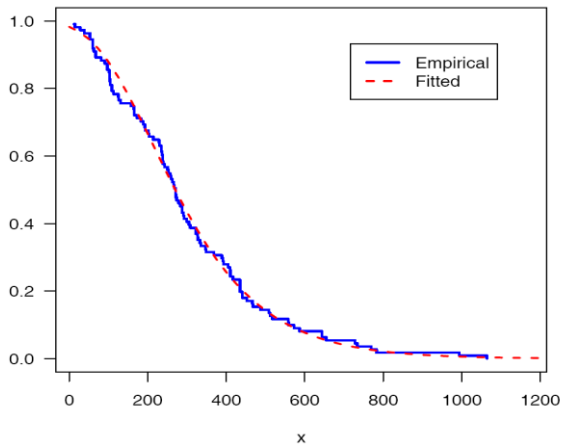


Fig 15. The estimated reliability function(dashed line) and the empirical reliability function (solid line).

8. CONCLUSION

The Exponentiated Gumbel model with shape parameter α and scale parameter σ has been discussed and estimate of its parameters obtained based on a complete sample by using the Markov chain Monte Carlo (MCMC) method. The MCMC method has proven more effective as compared to the usual methods of estimation.

Bayesian analysis under different set of priors has been carried in to OpenBUGS to study the convergence pattern. A numerical summary based on MCMC samples of posterior characteristic for Exponentiated Gumbel model has been worked out under non-informative and informative set of priors. A visual summary under different set of priors which include box plot, kernel density estimation and comparison with MLE has been also attempted and it has been found that the proposed methodology is suitable for empirical modeling and best suited for data set which is considered for illustration under uniform and gamma sets of priors.

9. ACKNOWLEDGMENTS

The authors are thankful to the editor and the referees for their valuable suggestions, which improved the paper to a great extent.

10. REFERENCES

- [1]. Gumbel, E.J.(1954). *Statistical theory of extreme values and some practical applications*. Applied Mathematics Series, 33. U.S. Department of Commerce, National Bureau of Standards.
- [2]. Kumar, R., Srivastava, A.K. and Kumar, V. (2012). *Analysis of Gumbel Model for Software Reliability Using Bayesian Paradigm*, International Journal of Advanced Research in Artificial Intelligence, Vol. 1 (9), 39-45.
- [3]. Nadarajah S. (2006). *The Exponentiated Gumbel distribution with climate application*, Environmetrics 17:13-23
- [4]. Persson, K. and Rydén, J. (2010). *Exponentiated Gumbel Distribution for Estimation of Return Levels of Significant Wave Height*, Journal of Environmental Statistics, Vol. 1 (3), 01-12.
- [5]. Kumar, V. and Ligges, U. (2011). *reliaR : A package for some probability distributions*. <http://cran.r-project.org/web/packages/reliaR/index.html>.
- [6]. Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed., New York, Springer-Verlag.
- [7]. Chen, M., Shao, Q. and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*, Springer, New York.
- [8]. Lawless, J. F., (2003). *Statistical Models and Methods for Lifetime Data*, 2nd ed., John Wiley and Sons, New York.
- [9]. Lyu, M.R., (1996). *Handbook of Software Reliability Engineering*, IEEE Computer Society Press, McGraw Hill, 1996.
- [10]. Srivastava, A.K. and Kumar V. (2011). *Analysis of Software Reliability Data using Exponential Power Model*. International Journal of Advanced Computer Science and Applications, Vol. 2, No. 2, February 2011, 38-45.
- [11]. Srivastava, A.K. and Kumar V. (2011). *Markov Chain Monte Carlo methods for Bayesian inference of the Chen model*, International Journal of Computer Information Systems, Vol. 2 (2), 07-14.
- [12]. Srivastava, A.K. and Kumar V. (2011). *Software reliability data analysis with Marshall-Olkin Extended Weibull model using MCMC method for non-informative set of priors*, International Journal of Computer Applications, Vol. 18(4), 31-39.
- [13]. Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed., New York, Springer-Verlag.
- [14]. Chen, M., Shao, Q. and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*, Springer, New York.
- [15]. Thomas, A. (2010). *OpenBUGS Developer Manual*, Version 3.1.2, <http://www.openbugs.info/>.
- [16]. Thomas, A., O'Hara, B., Ligges, U. and Sturtz, S. (2006). *Making BUGS Open*, R News, 6, 12-17, <http://mathstat.helsinki.fi/openbugs/>.
- [17]. Lunn, D.J., Andrew, A., Best, N. and Spiegelhalter, D. (2000). *WinBUGS – A Bayesian modeling framework: Concepts, structure, and extensibility*, Statistics and Computing, 10, 325-337.
- [18]. Kumar, V., Ligges, U. and Thomas, A. (2010). *ReliaBUGS User Manual : A subsystem in OpenBUGS for some statistical models*, Version 1.0, OpenBUGS 3.2.1, <http://openbugs.info/w/Downloads/>.
- [19]. Ihaka, R. and Gentleman, R.R. (1996). *R: A language for data analysis and graphics*, Journal of Computational and Graphical Statistics, 5, 299–314.
- [20]. Venables, W. N., Smith, D. M. and R Development Core Team (2010). *An Introduction to R*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-12-7, <http://www.r-project.org>.

10. AUTHOR'S PROFILE

RAJ KUMAR received his MCA from M.M.M. Engineering College, Gorakhpur and pursuing Ph.D. in Computer Science from D.D.U. Gorakhpur University. Currently working in National Institute of Electronics and Information Technology (formerly known as DOEACC Society), Gorakhpur, Ministry of Communication and Information Technology, Government of India.

ASHWINI KUMAR SRIVASTAVA received his M.Sc in Mathematics from D.D.U.Gorakhpur University, MCA(Hons.) from U.P.Technical University, M. Phil in Computer Science from Allagappa University and Ph.D. in Computer Science from D.D.U.Gorakhpur University, Gorakhpur. Currently working as Assistant Professor in Department of Computer Application in Shivharsh Kisan P.G.

College, Basti, U.P. He has got 8 years of teaching experience as well as 4 years research experience. His main research interests are Software Reliability, Artificial Neural Networks, Bayesian methodology and Data Warehousing.

VIJAY KUMAR received his M.Sc and Ph.D. in Statistics from D.D.U. Gorakhpur University. Currently working as Associate Professor in Department of Maths. and Statistics in D.D.U Gorakhpur University, Gorakhpur. He has got 17 years of teaching/research experience. He is visiting Faculty of Max-Planck-Institute, Germany. His main research interests are Bayesian statistics, reliability models and computational statistics using OpenBUGS and R.