5-1-2010

Estimations on the Generalized Exponential Distribution Using Grouped Data

Hassan Pazira
Payame Noor University of Tehran, Tehran, Iran, pazira.b@gmail.com

Parviz Nasiri
Payame Noor University of Tehran, Tehran, Iran, pnasiri@hotmail.com

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

 Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Estimations on the Generalized Exponential Distribution Using Grouped Data

Hassan Pazira   Parviz Nasiri
Payame Noor University of Tehran
Tehran, Iran

Classical and Bayesian estimators are obtained for the shape parameter of the Generalized-Exponential distribution under grouped data. In Bayesian estimation, three types of loss functions are considered: the Squared Error loss function which is classified as a symmetric function, the LINEX and Precautionary loss functions which are asymmetric. These estimators are compared with the corresponding estimators derived from un-grouped data empirically using Monte-Carlo simulation.

Key words: Generalized-Exponential; Bayesian estimation; Grouped data; LINEX loss; Precautionary loss; Newton-Raphson method; Fisher's information number; Monte-Carlo simulation.

Introduction

In various fields of science such as biology, engineering, and medicine it is not possible to obtain the measurements of a statistical experiment exactly, but is possible to classify them into intervals, rectangles or disjoint subsets (Alodat & Al-Saleh, 2002; Heitjan, 1989; Surles & Padgett, 2001; Wu & Perloff, 2005; Pipper & Ritz, 2006). For example, in life testing experiments, the failure time of a component may be observed to the nearest hour, day or month. Data for which true values are known only up to subsets of the sample space are called grouped data. In general, grouped data can be formulated as follows: Let $X_1, X_2, ..., X_n$ be a random sample from the density $f(x; \theta)$, $x \in \chi$, $\theta \in \Theta$ and $\chi_1, \chi_2, ..., \chi_{k+1}$ be a partition of the sample space χ and $N_j =$ the number of X_j's that fall in χ_j for $j = 1, 2, ..., k+1$. The set of pairs $\{(\chi_1, N_1), ..., (\chi_{k+1}, N_{k+1})\}$ is

Hassan Pazira is a Professor in the Department of Statistics. Email him at: pazira.b@gmail.com. Parviz Nasiri is an Associate Professor in the Department of Statistics. Email him at: pnasiri@hotmail.com.

called grouped data, and these data are used to draw inferences about the parameter θ. Due to a lack of complete information about the sample, there is a loss in information due to the grouping. Schervish (1995, p. 114), shows the following

$$I_{\underline{X}}(\theta) = I_{\underline{Y}}(\theta) + E_\theta \left[I_{\underline{X}|\underline{Y}}(\theta | \underline{Y}) \right]$$

where $I_{\underline{X}}(\theta)$ and $I_{\underline{Y}}(\theta)$ are Fisher's information numbers obtained from $\underline{X}$ and $\underline{Y}$, respectively, and $E_\theta \left[I_{\underline{X}|\underline{Y}}(\theta | \underline{Y}) \right]$ is the conditional score function. If $\underline{Y}$ is replaced by the grouped sample $\underline{n} = (N_1, N_2, ..., N_{k+1})$, then $I_{\underline{X}}(\theta) \geq I_{\underline{n}}(\theta)$ for all θ, meaning that the information in the sample $\underline{X}$ about θ is reduced to $I_{\underline{n}}(\theta)$ due to grouping.

Kuldorff (1961) considered non-Bayesian estimation from grouped data when the data were from normal and exponential distributions. Alodat and Al-Saleh (2000) considered the Bayesian estimation from grouped data when the underlying distribution is exponential. Alodat, et al. (2007) obtained Bayesian prediction intervals from grouped data when the underlying distribution is exponential. Aludaat, et al. (2008) obtained the Bayesian and non-Bayesian estimation from grouped data when the underlying distribution is Burr type X.

Recently a new distribution, called the Generalized-Exponential (GE) distribution, has been introduced. This distribution can be used quite effectively in situations where a skewed distribution is needed. It has been studied extensively (Gupta & Kundu, 1999, 2001a, 2001b, 2002, 2003a, 2003b, 2004; Raqab, 2002; Raqab & Ahsanullah, 2001; Zheng, 2002; Kundu, Gupta & Manglick, 2004). Singh, et al. (2008) estimated the parameters of this distribution under some symmetric and asymmetric loss functions using Lindley's approximation technique. Note that the generalized exponential distribution is a sub-model of the exponentiated Weibull distribution introduced by Mudholkar and Srivastava (1993) and later studied by Mudholkar, Srivastava and Freimer (1995) and Mudholkar and Hutson (1996). Also recently, Nasiri and Pazira (2010) conducted Bayesian and non-Bayesian estimations on the Generalized Exponential distribution in the presence of outliers.

This articles considers the group data problem when the density $f(x;\theta)$ is Generalized-Exponential, that is,

$$f(x;\theta) = \theta e^{-x}(1-e^{-x})^{\theta-1}; \; x > 0, \; \theta > 0, \quad (1)$$

and the corresponding distribution function is

$$F(x;\theta) = (1-e^{-x})^{\theta}; \; x > 0, \; \theta > 0. \quad (2)$$

The density functions of the generalized exponential distribution can take different shapes. For $\theta \leq 1$, it is a decreasing function and for $\theta > 1$, it is a unimodal, skewed, right tailed similar to the Weibull, or gamma density, function. It is observed that even for very large shape parameter (θ), it is not symmetric. For this density function (1), the mode is at $\log\theta$ for both $\theta > 1$ and $\theta \leq 1$, the mode is at 0 and the median is at $-\ln(1-\sqrt[\theta]{0.5})$. The mean, median and mode are non-linear functions of the shape parameter θ and, as this parameter goes to infinity, all of them tend to infinity. For large values of θ, the mean, median and mode are approximately equal to θ but they converge at different rates. Figure 1 shows the shape of $f(x;\theta)$ for different values of θ.

Likelihood Function and M.L.E from Grouped Data

The likelihood density based on the grouped data is derived as follows. Let $X_1, X_2,...,X_n$ be a random sample from GE(θ). Assume that the sample space of $f(x;\theta)$ is partitioned into $k+1$ equally-spaced intervals as follows. Let $I_j = [(j-1)\delta, j\delta)$, $j = 1,...,k$ and $I_{k+1} = [k\delta, \infty)$, $\delta > 0$. If N_j denotes the number of
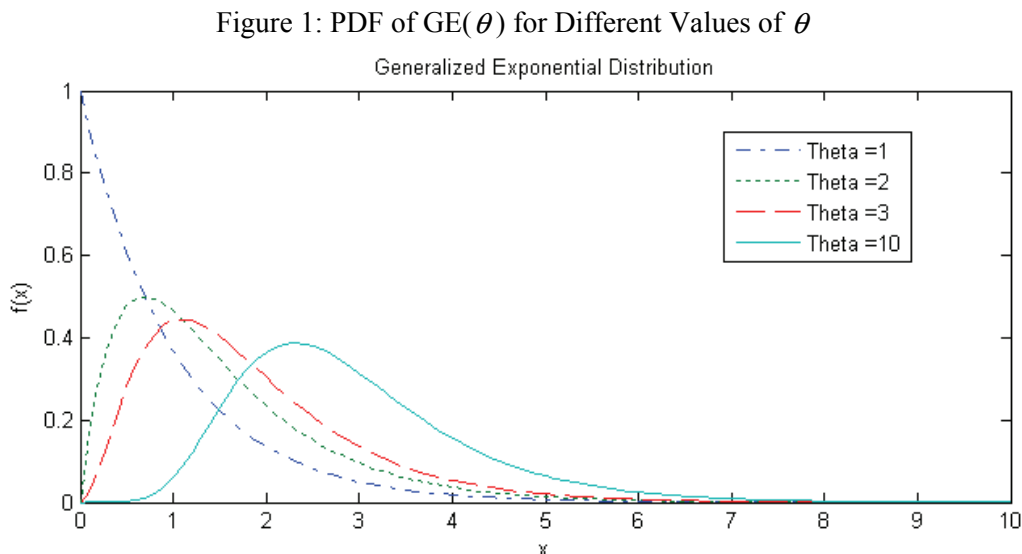
Figure 1: PDF of GE(θ) for Different Values of θ

X_j's that fall in I_j, $j = 1, 2, ..., k+1$, then $n = N_1 + ... + N_{k+1}$. Let

$$P_j = P_j(\theta)$$
$$= P(X \in I_j)$$
$$= P((j-1)\delta < X < j\delta)$$
$$= (1 - e^{-j\delta})^\theta - (1 - e^{-(j-1)\delta})^\theta,$$

for $j = 1, ..., k$ and

$$P_{k+1} = P_{k+1}(\theta) = P(X > k\delta) = 1 - (1 - e^{-k\delta})^\theta.$$

If

$$A_j = \log(1 - e^{-(j-1)\delta}),$$

then

$$P_j = e^{\theta A_{j+1}} - e^{\theta A_j},$$

for $j = 1, ..., k$ and

$$P_{k+1} = 1 - e^{\theta A_{k+1}}.$$

Thus, the density of $\underline{n} = (N_1, N_2, ..., N_{k+1})$ is given by the multinomial distribution as follows:

$$f(\underline{n}; \theta) = \frac{n!}{n_1! ... n_{k+1}!} P_1^{n_1} ... P_{k+1}^{n_{k+1}}$$
$$= C(1 - e^{\theta A_{k+1}})^{n_{k+1}} \prod_{j=1}^{k}(e^{\theta A_{j+1}} - e^{\theta A_j})^{n_j} \quad (3)$$

where C is a normalizing constant.

Next, find the MLE of θ based on the density (3) by maximizing the log-likelihood function

$$\log f(\underline{n}; \theta) = \text{constant} + \sum_{j=1}^{k} n_j \log(e^{\theta A_{j+1}} - e^{\theta A_j})$$
$$+ n_{k+1} \log(1 - e^{\theta A_{k+1}}).$$

The first derivative of the log-likelihood is

$$\frac{\partial \log f(\underline{n}; \theta)}{\partial \theta} =$$

$$\sum_{j=1}^{k} n_j \frac{A_{j+1} e^{\theta A_{j+1}} - A_j e^{\theta A_j}}{e^{\theta A_{j+1}} - e^{\theta A_j}} - n_{k+1} \frac{A_{k+1} e^{\theta A_{k+1}}}{1 - e^{\theta A_{k+1}}}.$$
$$(4)$$

The M.L.E for θ is the solution of $\partial \log f(\underline{n}; \theta)/\partial \theta = 0$, so the M.L.E is $\hat{\theta}$ such that

$$\sum_{j=1}^{k} n_j \frac{A_{j+1} e^{\hat{\theta} A_{j+1}} - A_j e^{\hat{\theta} A_j}}{e^{\hat{\theta} A_{j+1}} - e^{\hat{\theta} A_j}} = n_{k+1} \frac{A_{k+1} e^{\hat{\theta} A_{k+1}}}{1 - e^{\hat{\theta} A_{k+1}}}.$$
$$(5)$$

The notation $\hat{\theta}_{MG}$ is used to denote the M.L.E of θ obtained from the grouped data. The Newton-Raphson method can be used to solve (5), thus, the solution of the equation is

$$\theta_{i+1} = \theta_i - \frac{h(\theta_i)}{h'(\theta_i)}; \quad i = 1, 2, 3, ... \quad (6)$$

where

$$h(\theta_i) = \sum_{j=1}^{k} n_j \frac{A_{j+1} e^{\theta_i A_{j+1}} - A_j e^{\theta_i A_j}}{e^{\theta_i A_{j+1}} - e^{\theta_i A_j}}$$
$$- n_{k+1} \frac{A_{k+1} e^{\theta_i A_{k+1}}}{1 - e^{\theta_i A_{k+1}}},$$

and

$$h'(\theta_i) = -\sum_{j=1}^{k} n_j \frac{(A_{j+1} - A_j)^2 e^{\theta_i (A_j + A_{j+1})}}{(e^{\theta_i A_{j+1}} - e^{\theta_i A_j})^2}$$
$$- n_{k+1} \frac{A_{k+1}^2 e^{\theta_i A_{k+1}}}{(1 - e^{\theta_i A_{k+1}})^2}.$$

Here, the initial solution θ_0 should be selected from the M.L.E of θ based on the un-grouped data. Gupta and Kundu (1999 & 2001a) showed

237

that the M.L.E of θ based on the un-grouped sample $X_1, X_2, ..., X_n$ is

$$\hat{\theta}_M = \frac{-n}{\sum\limits_{i=1}^{n} \log(1 - e^{-X_i})} . \qquad (7)$$

Fisher's Information Number

To find the Fisher's information number contained in the grouped sample about θ, find the expectation of the second derivative of the log-likelihood

$$\frac{\partial^2 \log f(\underline{n};\theta)}{\partial \theta^2} = -\sum_{j=1}^{k} n_j \, \Psi_{1j}(\theta) - n_{k+1} \, \Psi_2(\theta)$$
$$(8)$$

where

$$\Psi_{1j}(\theta) = \frac{(A_{j+1} - A_j)^2 \cdot e^{\theta(A_j + A_{j+1})}}{\left(e^{\theta A_{j+1}} - e^{\theta A_j}\right)^2}$$

and

$$\Psi_2(\theta) = \frac{A_{k+1}^2 \cdot e^{\theta A_{k+1}}}{\left(1 - e^{\theta A_{k+1}}\right)^2} .$$

If $I_G(\theta)$ denotes the Fisher's information number from grouped data, then

$$I_G(\theta) = -E\left[\frac{\partial^2 \log f(\underline{n};\theta)}{\partial \theta^2}\right],$$

and, because $E[N_j] = nP_j$,

$$I_G(\theta) = E\left[\sum_{j=1}^{k} N_j \, \Psi_{1j}(\theta) + N_{k+1} \, \Psi_2(\theta)\right]$$

$$= n\sum_{j=1}^{k} \frac{(A_{j+1} - A_j)^2 \, e^{\theta(A_j + A_{j+1})}}{e^{\theta A_{j+1}} - e^{\theta A_j}} + n\frac{A_{k+1}^2 \, e^{\theta A_{k+1}}}{1 - e^{\theta A_{k+1}}}$$
$$(9)$$

Using $I_G(\theta)$, a large sample $(1-\alpha)100\%$ confidence interval for θ can be found as follows:

$$\hat{\theta}_{MG} \pm Z_{1-\alpha/2}\sqrt{I_G(\hat{\theta}_{MG})^{-1}} . \qquad (10)$$

Simple calculations can show that the Fisher's information number about θ in a random sample $X_1, X_2, ..., X_n$ from (1), that is, from un-grouped data, is $I(\theta) = n/\theta^2$.

Bayesian Estimation

In classical statistics, the likelihood of the data, $L(\theta) = f(\underline{n};\theta)$, is used to estimate the parameter θ. In Bayesian statistics, it is assumed that θ has a prior distribution, for example, $\pi(\theta)$. The likelihood data is then combined with the prior distribution to obtain the posterior distribution $\pi(\theta|\underline{n})$. Herein, three loss functions based on the grouped data are used.

First the squared error loss function, that is a symmetric loss function, and with respect to it, the posterior mean is used to estimate θ, for example $\hat{\theta}_{BSG}$, because it minimizes the posterior expected loss, $E[(\hat{\theta} - \theta)^2 | \underline{n}]$. The second is the asymmetric LINEX loss function, which was introduced by Varian (1975). These loss functions were widely used by several authors; among of them Basu and Ebrahimi (1991), Pandey (1997), Soliman (2000) and Nassar and Eissa (2004). The LINEX loss function may be expressed as

$$L(\Delta) = b(e^{c\Delta} - c\Delta - 1), \, c \neq 0, \, b > 0, \quad (11)$$

where b and c are the scale and shape parameters, respectively, and $\Delta = \hat{\theta} - \theta$. The sign and magnitude of the shape parameter c reflects the direction and degree of asymmetry, respectively. (If $c > 0$, the overestimation is more serious than underestimation, and vice-versa). For c close to zero, the LINEX loss is approximately squared error loss and therefore almost symmetric. The posterior expectation of the LINEX loss function equation (11) is

$$E[L(\hat{\theta} - \theta) | \underline{n}] \propto e^{c\hat{\theta}}E[e^{-c\theta} | \underline{n}] - c(\hat{\theta} - E[\theta | \underline{n}]) - 1$$
$$(12)$$

where $E[. | \underline{n}]$ denotes posterior expectation with respect to the posterior density of θ. Based

on results from Zellner (1986), the (unique) Bayes estimator of θ, denoted by $\hat{\theta}_{BLG}$ under the LINEX loss is the value $\hat{\theta}$ which minimizes (12), is given by

$$\hat{\theta}_{BLG} = -\frac{1}{c}\ln\left(E[e^{-c\theta}\,|\,\underline{n}]\right), \qquad (13)$$

provided that the expectation $E[e^{-c\theta}\,|\,\underline{n}]$ exists and is finite (see Calabria & Pulcini, 1996), then the precautionary loss function introduced by Norstrom (1996) is used. Norstrom introduced an alternative asymmetric precautionary loss function and also presented a general class of precautionary loss function with the quadratic loss function as a special case. These loss functions approach infinitely near the origin to prevent underestimation, thus giving conservative estimators - especially when low failure rates are being estimated. These estimators are very useful when underestimation may lead to serious consequences. A useful and simple asymmetric precautionary loss function is

$$L(\hat{\theta},\theta) = \frac{(\hat{\theta}-\theta)^2}{\hat{\theta}\,\theta}. \qquad (14)$$

The Bayes estimator under the above asymmetric loss function is denoted by $\hat{\theta}_{BPG}$ and may be obtained by solving the following equation,

$$\hat{\theta}_{BPG}^2 = \frac{E(\theta\,|\,\underline{n})}{E(\theta^{-1}\,|\,\underline{n})}. \qquad (15)$$

Note that the special case of the precautionary loss function (14) is the same as the entropy loss function (for details see Norstrom, 1996).

The following prior distribution for θ is used to derive an estimate for θ,

$$\pi(\theta) = \frac{\theta^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}}\exp(-\frac{\theta}{\beta}); \quad \theta > 0.$$

Using the Binomial theorem, the likelihood function of the grouped data is re-written as follows:

$$f(\underline{n};\theta) =$$

$$= C\prod_{j=1}^{k}\sum_{r_j=0}^{n_j}\binom{n_j}{r_j}(-1)^{r_j}\left(e^{\theta A_{j+1}}\right)^{n_j-r_j}\left(e^{\theta A_j}\right)^{r_j}$$

$$\times \sum_{r_{k+1}=0}^{n_{k+1}}\binom{n_{k+1}}{r_{k+1}}(-1)^{r_{k+1}}\left(e^{\theta A_{k+1}}\right)^{r_{k+1}}$$

$$= C\sum_{r_1=1}^{n_1}\cdots\sum_{r_{k+1}=0}^{n_{k+1}}\binom{n_1}{r_1}\cdots\binom{n_{k+1}}{r_{k+1}}(-1)^{r_1+\ldots+r_{k+1}}e^{\theta H(r_1,\ldots,r_{k+1})}$$

$$(16)$$

where

$$H(r_1,\ldots,r_{k+1}) = r_{k+1}A_{k+1} + \sum_{j=1}^{k}(n_j-r_j)A_{j+1} + \sum_{j=1}^{k}r_j A_j.$$

Combining the likelihood information with the prior information yields the posterior distribution of θ given $\underline{n}$,

$$\pi(\theta\,|\,\underline{n}) =$$

$$= \frac{f(\underline{n};\theta)\cdot\pi(\theta)}{\int_0^{\infty} f(\underline{n};\theta)\cdot\pi(\theta)\,d\theta}$$

$$\propto \sum_{r_1=0}^{n_1}\cdots\sum_{r_{k+1}=0}^{n_{k+1}}\binom{n_1}{r_1}\cdots\binom{n_{k+1}}{r_{k+1}}(-1)^{r_1+\ldots+r_{k+1}}\theta^{\alpha-1}e^{-\theta\left(\frac{1}{\beta}-H(r_1,\ldots,r_{k+1})\right)}$$

resulting in

$$\pi(\theta\,|\,\underline{n}) =$$

$$\frac{\sum_{r_1=0}^{n_1}\cdots\sum_{r_{k+1}=0}^{n_{k+1}}\binom{n_1}{r_1}\cdots\binom{n_{k+1}}{r_{k+1}}(-1)^{r_1+\ldots+r_{k+1}}\theta^{\alpha-1}e^{-\theta\left(\frac{1}{\beta}-H\right)}}{\Gamma(\alpha)\sum_{r_1=0}^{n_1}\cdots\sum_{r_{k+1}=0}^{n_{k+1}}\binom{n_1}{r_1}\cdots\binom{n_{k+1}}{r_{k+1}}(-1)^{r_1+\ldots+r_{k+1}}\left(\frac{1}{\beta}-H\right)^{-\alpha}}$$

$$(17)$$

where $H = H(r_1,\ldots,r_{k+1})$.

The Bayesian estimate of θ with respect to the squared error loss function, based on the grouped data, is the posterior mean,

$$\hat{\theta}_{BSG} = E(\theta | \underline{n})$$

$$= \frac{\alpha \sum_{r_1=0}^{n_1} \cdots \sum_{r_{k+1}=0}^{n_{k+1}} \binom{n_1}{r_1} \cdots \binom{n_{k+1}}{r_{k+1}} (-1)^{r_1+\ldots+r_{k+1}} \left(\frac{1}{\beta}-H\right)^{-\alpha-1}}{\sum_{r_1=0}^{n_1} \cdots \sum_{r_{k+1}=0}^{n_{k+1}} \binom{n_1}{r_1} \cdots \binom{n_{k+1}}{r_{k+1}} (-1)^{r_1+\ldots+r_{k+1}} \left(\frac{1}{\beta}-H\right)^{-\alpha}}$$

$$(18)$$

Kundu and Gupta (2008) obtained the Bayes estimator of θ, based on the un-grouped data, only under square error loss function as follows

$$\hat{\theta}_{BS} = \frac{\alpha+n}{\beta^{-1} - \sum_{i=1}^{n} \log(1-e^{-x_i})}, \qquad (19)$$

Note that, for the non-informative prior - when $\alpha = \beta^{-1} = 0$ - the above Bayes estimator ($\hat{\theta}_{BS}$) and MLE of θ from the un-grouped data ($\hat{\theta}_M$), are identical.

The Bayesian estimate of θ with respect to the LINEX loss function, based on the grouped data, is obtained as follows:

$$\hat{\theta}_{BLG} =$$

$$= -\frac{1}{c} \ln\left(E[e^{-c\theta} | \underline{n}]\right)$$

$$= -\frac{1}{c} \ln\left(\frac{\sum_{r_1=0}^{n_1} \cdots \sum_{r_{k+1}=0}^{n_{k+1}} \binom{n_1}{r_1} \cdots \binom{n_{k+1}}{r_{k+1}} (-1)^{r_1+\ldots+r_{k+1}} \left(c+\frac{1}{\beta}-H\right)^{-\alpha}}{\sum_{r_1=0}^{n_1} \cdots \sum_{r_{k+1}=0}^{n_{k+1}} \binom{n_1}{r_1} \cdots \binom{n_{k+1}}{r_{k+1}} (-1)^{r_1+\ldots+r_{k+1}} \left(\frac{1}{\beta}-H\right)^{-\alpha}}\right)$$

$$(20)$$

Here, the Bayes estimator of θ is obtained under the LINEX loss function based on the un-grouped data. It may be shown that the posterior density of θ based on the un-grouped data is gamma with the shape and scale parameters as

$\alpha+n$ and $\beta^{-1} - \sum_{i=1}^{n} \log(1-e^{-x_i})$, respectively. Whereas,

$$E[e^{-c\theta} | \underline{x}] = \left(1 + \frac{c}{\beta^{-1} - \sum_{i=1}^{n} \log(1-e^{-x_i})}\right)^{-\alpha-n}$$

thus, the Bayes estimator of θ under the LINEX loss function based on the un-grouped data is as follows

$$\hat{\theta}_{BL} = \frac{\alpha+n}{c} \ln\left(1 + \frac{c}{\beta^{-1} - \sum_{i=1}^{n} \log(1-e^{-x_i})}\right).$$

$$(21)$$

The Bayesian estimate of θ with respect to the precautionary loss function, based on the grouped data, is obtained as follows:

$$\hat{\theta}_{BPG} = \sqrt{\frac{E(\theta | \underline{n})}{E(\theta^{-1} | \underline{n})}}$$

$$= \left(\frac{\alpha(\alpha-1) \sum_{r_1=0}^{n_1} \cdots \sum_{r_{k+1}=0}^{n_{k+1}} \binom{n_1}{r_1} \cdots \binom{n_{k+1}}{r_{k+1}} (-1)^{r_1+\ldots+r_{k+1}} \left(\frac{1}{\beta}-H\right)^{-\alpha-1}}{\sum_{r_1=0}^{n_1} \cdots \sum_{r_{k+1}=0}^{n_{k+1}} \binom{n_1}{r_1} \cdots \binom{n_{k+1}}{r_{k+1}} (-1)^{r_1+\ldots+r_{k+1}} \left(\frac{1}{\beta}-H\right)^{-\alpha+1}}\right)^{\frac{1}{2}}$$

$$(22)$$

Also, the Bayes estimator of θ is obtained under the precautionary loss function based on the un-grouped data. Whereas

$$\theta | \underline{x} \sim \Gamma(\alpha+n, \beta^{-1} - \sum_{i=1}^{n} \log(1-e^{-x_i})),$$

it may be shown that

$$E(\frac{1}{\theta} | \underline{x}) = \frac{\beta^{-1} - \sum_{i=1}^{n} \log(1-e^{-x_i})}{(\alpha+n-1)},$$

thus, the Bayes estimator of θ under the precautionary loss function based on the un-grouped data is as follows

$$\hat{\theta}_{BP} = \frac{\sqrt{(\alpha+n)(\alpha+n-1)}}{\beta^{-1} - \sum_{i=1}^{n}\log(1-e^{-x_i})}. \qquad (23)$$

Methodology
Simulation Study

The estimator $\hat{\theta}_M$ is the Maximum Likelihood Estimator (MLE) of the shape parameter of the generalized exponential distribution obtained from the un-grouped data; whereas, $\hat{\theta}_{MG}$ is the MLE of θ obtained from the grouped data. $\hat{\theta}_{BS}$, $\hat{\theta}_{BP}$ and $\hat{\theta}_{BL}$ are Bayes estimators under squared-error, precautionary and LINEX loss functions, respectively, based on un-grouped data. Also, $\hat{\theta}_{BSG}$, $\hat{\theta}_{BPG}$ and $\hat{\theta}_{BLG}$ are Bayes estimators under squared-error, precautionary and LINEX loss functions, respectively, based on grouped data. The notation CLG is used to denote the confidence length for θ based on the grouped data. Because

the large sample properties of these estimators are unknown, the bootstrap method can be used to assess the precision of estimates, but construction of bootstrap confidence interval is computationally more demanding than asymptotic confidence interval. Therefore, the 95% confidence interval is computed based on the MLE's. The main goal is to compare the estimators in terms of biases and MSE's.

As noted, $\hat{\theta}_{MG}$ and hence its MSE cannot be put in a convenient closed form. Therefore, MSE's of the estimators are empirically evaluated based on a Monte-Carlo simulation study of 1,000 samples by MATLAB mainly for small sample sizes. The simulation study was carried out for $\theta=1$ with sample sizes $n = 6, 9, 12, 15, 18$ and 20. These samples were placed into five intervals ($k=4$) with $\delta=1$. The loss and prior parameters are arbitrarily taken as $c = -1.5, -1, -0.5, 0.5, 1$ and 1.5, $\alpha=2$ and $\beta=0.5$. Results are summarized in Table 1-2 and Figures 1-5.

Table 1: Bias and MSE of the MLE's and Three Bayes Estimates from Un-grouped Data,
when $k=4$, $\delta=1$, $\theta=1$, $\alpha=2$ and $\beta=0.5$ (MSE in parenthesis)

n	$\hat{\theta}_M$	$\hat{\theta}_{BS}$	$\hat{\theta}_{BP}$	$\hat{\theta}_{BL}$					
				C					
				−1.5	−1	−0.5	0.5	1	1.5
6	0.2569	0.1297	0.0567	0.2896	0.2288	0.1761	0.0882	0.0508	0.0269
	(0.4697)	(0.1419)	(0.1126)	(0.3063)	(0.2314)	(0.1792)	(0.1150)	(0.0957)	(0.0821)
9	0.1551	0.1025	0.0512	0.2031	0.1666	0.1333	0.0740	0.0475	0.0227
	(0.1893)	(0.0963)	(0.0806)	(0.1655)	(0.1366)	(0.1139)	(0.0825)	(0.0719)	(0.0638)
12	0.1096	0.0850	0.0462	0.1357	0.1206	0.1071	0.0580	0.0412	0.0199
	(0.1205)	(0.0751)	(0.0665)	(0.1118)	(0.0970)	(0.0849)	(0.0672)	(0.0609)	(0.0559)
15	0.0929	0.0730	0.0409	0.1304	0.1103	0.0912	0.0506	0.0389	0.0187
	(0.0918)	(0.0642)	(0.0571)	(0.0902)	(0.0800)	(0.0714)	(0.0581)	(0.0531)	(0.0490)
18	0.0704	0.0570	0.0303	0.1036	0.0874	0.0719	0.0427	0.0289	0.0156
	(0.0728)	(0.0547)	(0.0497)	(0.0722)	(0.0654)	(0.0596)	(0.0504)	(0.0469)	(0.0439)
20	0.0629	0.0519	0.0277	0.0934	0.0791	0.0652	0.0390	0.0265	0.0144
	(0.0668)	(0.0510)	(0.0469)	(0.0658)	(0.0601)	(0.0553)	(0.0474)	(0.0443)	(0.0417)

Table 2: Bias and MSE of the MLE's and Three Bayes Estimates from the Grouped Data, when $k=4$, $\delta=1$, $\theta=1$, $\alpha=2$ and $\beta=0.5$ (MSE in parenthesis)

n	$\hat{\theta}_{MG}$	$\hat{\theta}_{BSG}$	$\hat{\theta}_{BPG}$	$\hat{\theta}_{BLG}$						CLG
				C						
				−1.5	−1	−0.5	0.5	1	1.5	
6	0.3274	0.1386	0.0079	0.4326	0.3127	0.2173	0.0727	0.0160	0.0335	3.1229
	(0.6874)	(0.1280)	(0.1038)	(0.3930)	(0.2575)	(0.1770)	(0.0985)	(0.0815)	(0.0729)	
9	0.1817	0.0983	-0.0033	0.3032	0.2237	0.1564	0.0475	0.0025	-0.0377	2.3723
	(0.3099)	(0.1025)	(0.0925)	(0.2379)	(0.1731)	(0.1305)	(0.0847)	(0.0749)	(0.0693)	
12	0.1205	0.0726	-0.0104	0.2299	0.1707	0.1187	0.0312	-0.0061	-0.0400	1.9877
	(0.2380)	(0.1010)	(0.0880)	(0.1899)	(0.1492)	(0.1208)	(0.0817)	(0.0722)	(0.0682)	
15	0.0896	0.0708	0.0338	0.1917	0.1588	-0.0130	-0.0423	0.0036	-0.0166	1.7430
	(0.1732)	(0.0868)	(0.0819)	(0.1446)	(0.1183)	(0.0993)	(0.0765)	(0.0704)	(0.0670)	
18	0.0885	0.0635	0.0016	0.1725	0.1327	0.0965	0.0285	0.0030	0.0232	1.5973
	(0.1560)	(0.0837)	(0.0800)	(0.1363)	(0.1152)	(0.0994)	(0.0764)	(0.0702)	(0.0668)	
20	0.0619	0.0443	-0.0133	0.1418	0.1066	0.0741	0.0154	-0.0049	-0.0343	1.4862
	(0.1312)	(0.0783)	(0.0729)	(0.1156)	(0.0996)	(0.0872)	(0.0705)	(0.0649)	(0.0578)	

Figure 1: MSE's of the Estimators $\hat{\theta}_M$, $\hat{\theta}_{BS}$, $\hat{\theta}_{BP}$ and $\hat{\theta}_{BL}$ from the Un-grouped Data, Based on Table 1
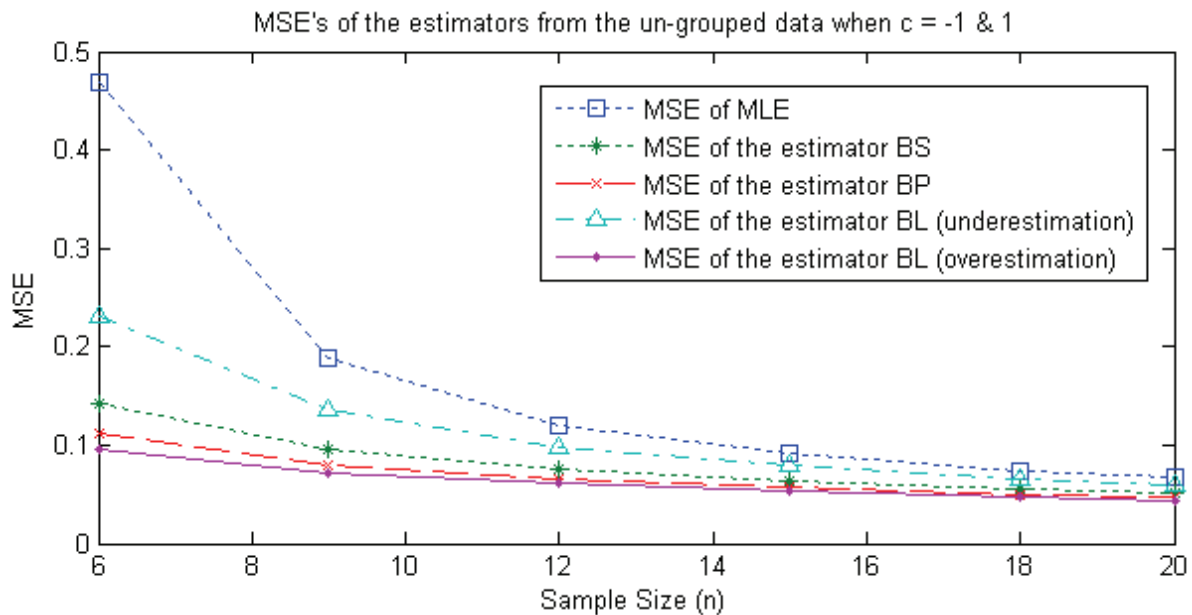
Figure 2: MSE's of the Estimators $\hat{\theta}_{MG}$, $\hat{\theta}_{BSG}$, $\hat{\theta}_{BPG}$ and $\hat{\theta}_{BLG}$ from the Grouped Data, Based on Table 2
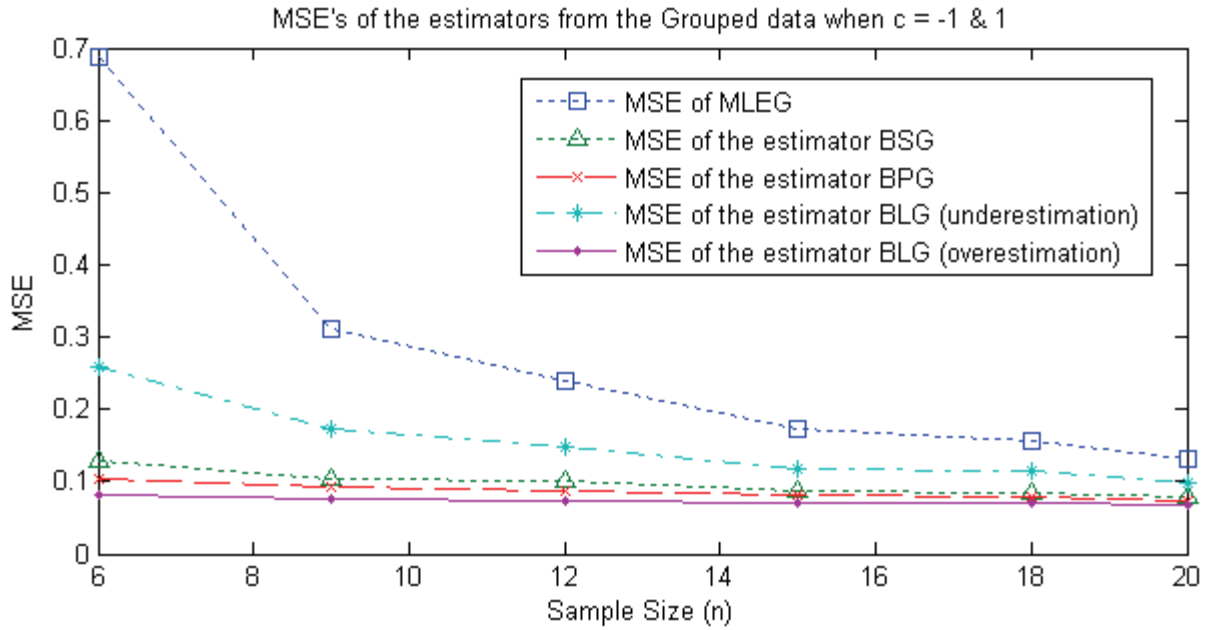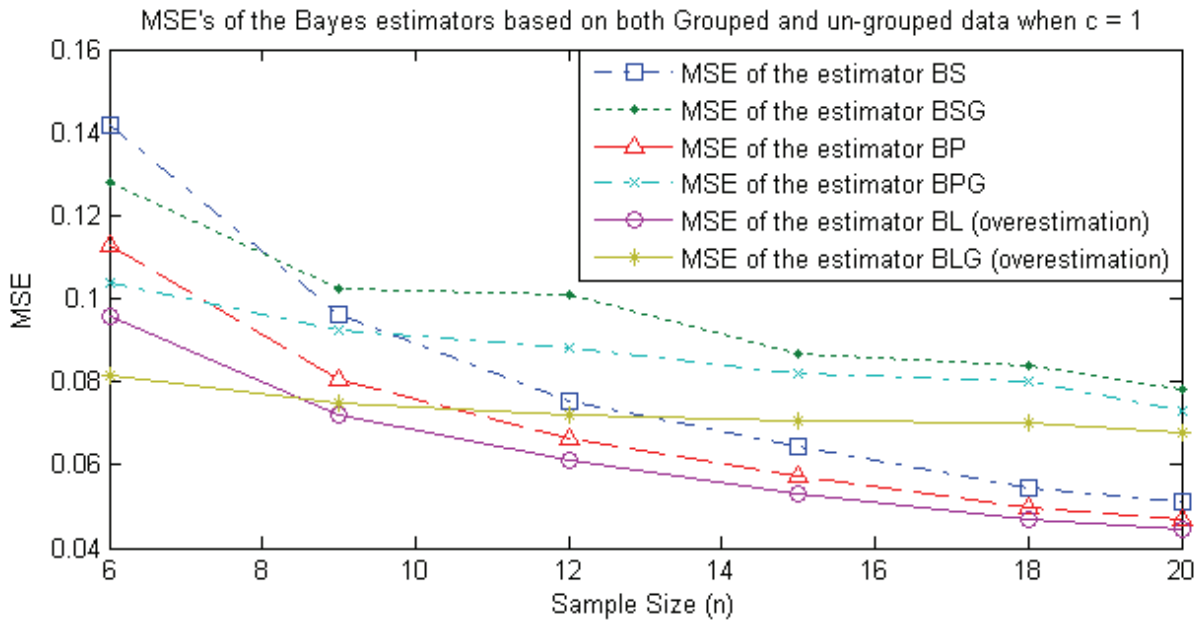

Figure 3: MSE's of the Bayes Estimators under Three Loss Functions Based on the Un-grouped and Grouped Data (when the overestimation is more serious than underestimation, $c = 1$)

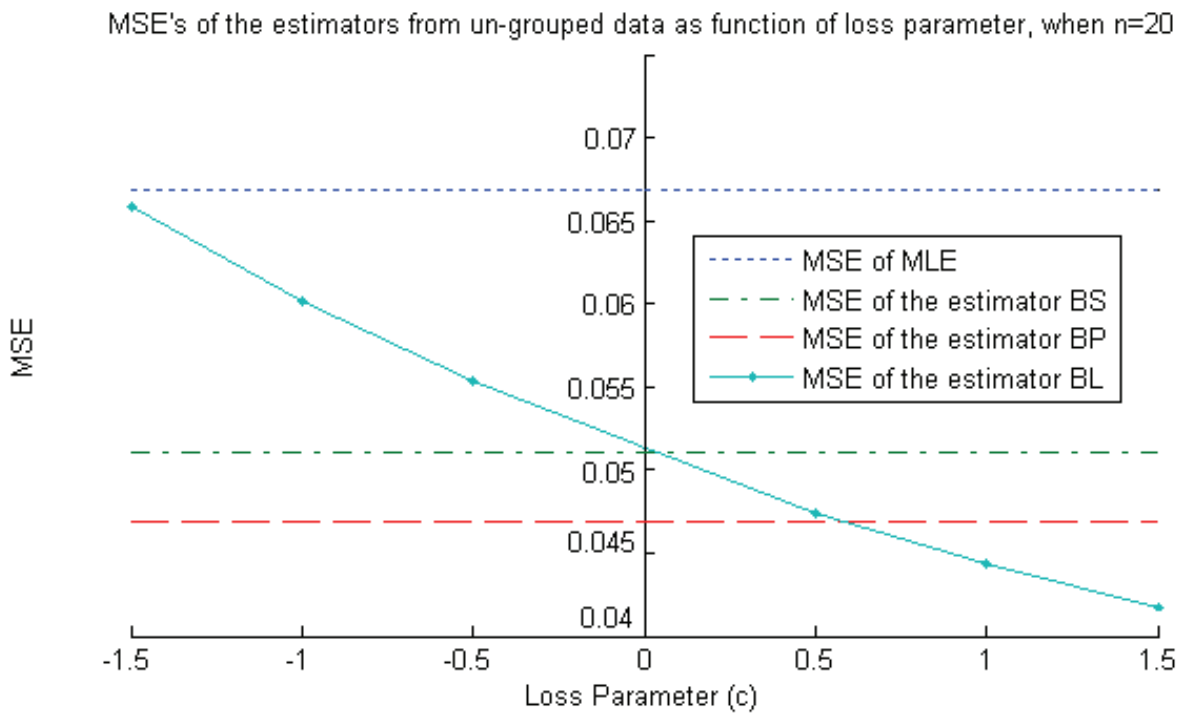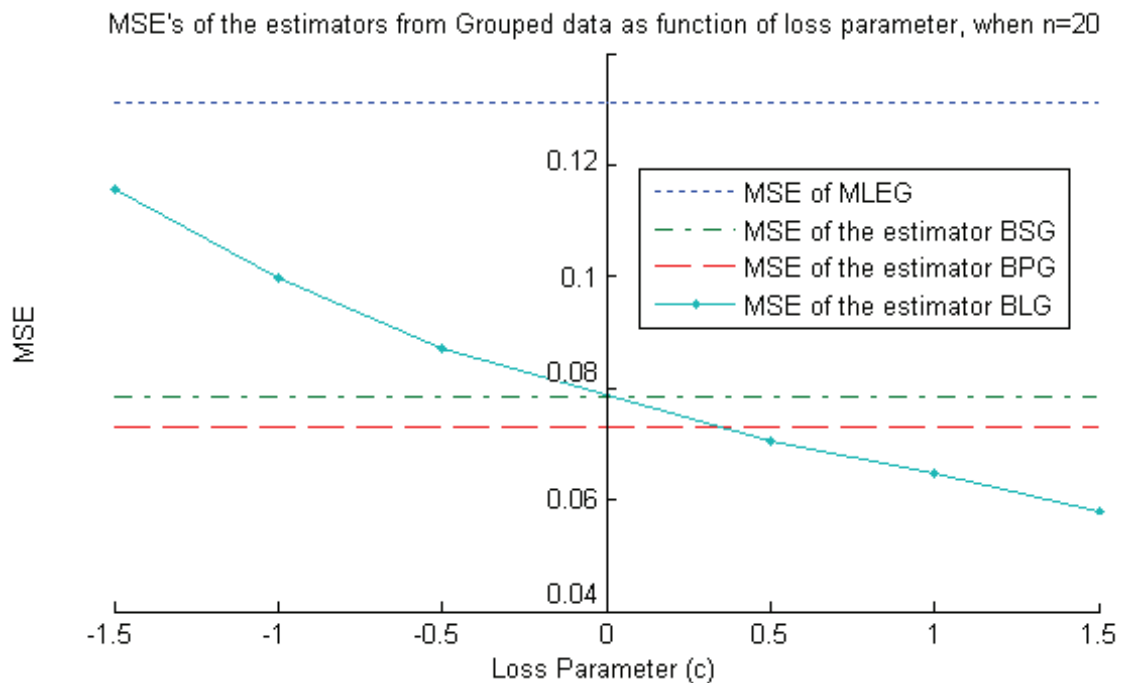Figure 4: MSE's of the Estimators as Function of Loss Parameter Based on the Un-grouped Data when $n = 20$

MSE's of the estimators from un-grouped data as function of loss parameter, when n=20

Figure 5: MSE's of the Estimators as Function of Loss Parameter Based on the Grouped Data when $n = 20$

MSE's of the estimators from Grouped data as function of loss parameter, when n=20

Conclusion

In this study, the Bayesian and non-Bayesian estimators for the shape parameter of the generalized exponential distribution based on grouped and un-grouped data were obtained. In addition, the interval estimator for θ was considered when the data are given in groups. The Bayes estimators under symmetric squared error loss function and asymmetric precautionary and LINEX loss functions were derived. Observations regarding the results are stated in the following points.

Results Based on Un-Grouped Data

Table 1 shows that all of the estimators are overestimations. Also, it is noted that derived estimators are consistent, because both bias and MSE decrease as the sample size increases. The Bayes estimates have the smallest estimated MSE's compared with the classical estimate. The Bayes estimates under the precautionary loss function ($\hat{\theta}_{BP}$) always have the smallest estimated MSE's as compared with the Bayes estimates under Squared error loss function, $\hat{\theta}_{BS}$ (see Figures 1 and 4). When the underestimation is more serious than overestimation (for $c \leq 0.5$), the performance of the Bayes estimates under precautionary loss function ($\hat{\theta}_{BP}$) are better than the rest: however, when the overestimation is more serious than underestimation (for $c > 0.5$), the performance of the Bayes estimates under LINEX loss function, $\hat{\theta}_{BL}$, are better than the rest (see Figures 1 and 4).

Results Based on Grouped Data

The results (based on the un-grouped data) are true for the grouped data. Table 2 shows that almost all of the estimators are overestimations. Also, it is clear that derived estimators are consistent, because both, bias and MSE decrease as the sample size increases. The Bayes estimates have the smallest estimated MSE's compared with the classical estimate. The Bayes estimates under the precautionary loss function, $\hat{\theta}_{BPG}$, always have the smallest estimated MSE's compared with the Bayes estimates under squared error loss function,

$\hat{\theta}_{BSG}$, (see Figures 2 and 5). When the underestimation is more serious than overestimation (for $c < 0.5$), the performance of the Bayes estimates under Precautionary loss function, $\hat{\theta}_{BPG}$, are better than the rest: however, when the overestimation is more serious than underestimation (for $c \geq 0.5$), the performance of the Bayes estimates under LINEX loss function, $\hat{\theta}_{BLG}$, are better than the rest (see Figures 2 and 5). Otherwise, the confidence intervals work quite well.

General Conclusions

In general, when the data are given in groups, the proposed Bayes estimators ($\hat{\theta}_{B.G}$) are more efficient than the corresponding Bayes estimators based on un-grouped data ($\hat{\theta}_{B.}$) for very small sample sizes, thus these estimators work very well (see Figure 3). Also, whereas the proposed Bayes estimators are better than the proposed estimators by Gupta and Kundu (1999, 2001a, 2008), it is suggested that the Bayes estimators be used for estimating the shape parameter of GE distribution when the data are given in groups, for example, in life testing experiments.

References

Alodat, M. T., & Al-Saleh, M. F. (2000). Bayesian estimation using grouped data with application to the exponential distribution. *Soochow Journal of Mathematics*, *26*, 342-357.

Alodat, M. T., Aludaat, K. M., & Alodat, T. T. (2007). Bayesian prediction intervals from grouped data: exponential distribution. *Abhath Al-yarmouk*, (*Accepted*).

Aludaat, K. M., Alodat, M. T., & Alodat, T. T. (2008). Parameter estimation of Burr type X distribution for grouped data. *Journal of Applied Mathematical Sciences*, *2*(*9*), 415-423.

Amjad, A., & Ayman, B. (2006). Interval estimation for the scale parameter of Burr type X distribution based on grouped data. *Journal of Modern Applied Statistical Methods*, *3*, 386-398.

Basu, A. P., & Ebrahimi, N. (1991). Bayesian approach to life testing and reliability estimation using asymmetric loss function. *Journal of Statistical Planning and Inference*, *29*, 21-31.

Calabria, R., & Pulcini, G. (1996). Point estimation under asymmetric loss functions for lift-truncated exponential samples. *Communications in Statistical Theory and Methods*, *25*(*3*), 585-600.

Gupta, R. D., & Kundu, D. (1999). Generalized-Exponential Distribution. *Australia and New Zealand Journal of Statistics*, *41*, 173-188.

Gupta, R. D., & Kundu, D. (2001a). Generalized-Exponential Distribution, Different method of estimations. *Journal of Statistical Computation and Simulation*, 69, 315-338.

Gupta, R. D., & Kundu, D. (2001b). Generalized-Exponential Distribution: An alternative to gamma or Weibull distribution. *Biometrical Journal*, *43*, 117-130.

Gupta, R. D., & Kundu, D. (2002). Generalized-Exponential Distribution. *Journal of Applied Statistical Society*, *1*, 101-118.

Gupta, R. D., & Kundu, D. (2003a). Closeness between the gamma and generalized exponential distributions. *Communications in Statistics-Theory and Methods*, *32*, 705-722.

Gupta, R. D., & Kundu, D. (2003b). Discriminating between the Weibull and generalized exponential distributions. *Computational Statistics and Data Analysis*, *43*, 179-196.

Gupta, R. D., & Kundu, D. (2004). Discriminating between the gamma and generalized exponential distributions. *Journal of statistical Computation and Simulation*, *74*, 107-122.

Heitjan, D. (1989). Inference from grouped data: a review (with discussion). *Statistical Sciences*, *4*, 164-183.

Kuldorff. (1961). *Estimation from grouped and partially grouped samples*. New York: John Wiley, Inc.

Kundu, D., Gupta, R. D. & Manglick, A. (2004). Discriminating between the log-normal and generalized exponential distributions. *Journal of Statistical Planning and Inference*, (*in press*).

Kundu, D., & Gupta, R. D. (2008) Generalized-exponential distribution, Bayesian estimations. *Computational Statistics and Data Analysis*, *52*, 1873-1883.

Mudholkar, G. S., & Hutson, A. D. (1996). The exponentiated Weibull family, some properties and a flood data applications. *Communications in Statistics-Theory and Methods*, *25*, 3059-3083.

Mudholkar, G. S., & Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, *42*, 299-302.

Mudholkar, G. S., Srivastava, D. K., & Freimer, M. (1995). The exponentiated Weibull family: A reanalysis of the bus-motor failure data. *Technometrics*, *37*, 436-445.

Nasiri, P., & Pazira, H. (2010). Bayesian approach on the Generalized Exponential distribution in the presence of outliers, *Journal of Statistical Theory and Practice, 4(3),* 453-475.

Nassar, M. M., & Eissa, F. H. (2004). Bayesian estimation for the exponentiated Weibull model. *Communications in Statistics-Theory and Methods*, *33*(*10*), 2343-2362.

Norstrom, J. G. (1996). The use of precautionary loss function in risk analysis. *IEEE Transactions on Reliability*, *45*(*3*), 400-403.

Pandey, B. N. (1997). Testimator of the scale parameter of the exponential distribution using LINEX loss function. *Communications in Statistics-Theory and Methods*, *26*(*9*), 2191-2202.

Pipper, C. B., & Ritz, C. (2006). Cheking the grouped data version of Cox model for interval-grouped survival data. *Scandinavian Journal of Statistics*, *10*, 1467-9469.

Raqab, M. Z. (2002). Inference for generalized exponential distributions based on record statistics. *Journal of Statistical Planning and Inference*, *104*, 339-350.

Raqab, M. Z., & Ahsanullah, M. (2001). Estimation of the location and scale parameters of the generalized exponential distributions based on order statistics. *Journal of Statistical Computation and Simulation*, *69*, 109-124.

Schervish, M. J. (1995). *Theory of statistics*. New York: Springer-Verlag, Inc.

Singh, R., Singh, S. K., Singh, U., & Singh, G. P. (2008). Bayes estimator of generalized-exponential parameters under LINEX loss function using Lindley's approximation. *Data Science Journal*, *7*, 65-75.

Soliman, A. A. (2000). Inference for reliability and stress-length for a scaled Burr type X distribution. *Communications in Statistics-Theory and Methods*, *29*, 95-107.

Surles, J. G., & Padgett, W. J. (2001). Inference for reliability and stress-length for a scaled Burr type X distribution. *Lifetime Data analysis*, *7*, 187-202.

Varian, H. R. (1975). *A Bayesian approach to real estate assessment*. Amsterdam: North Holland.

Wu, X., & Perloff, J. M. (2005). China's income distribution: 1985-2001. *Review of Econometrics and Statistics*, *87*, 763-775.

Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, *81*, 446-451.

Zheng, G. (2002). On the Fisher information matrix in type-II censored data from the exponentiated exponential family. *Biometrical Journal*, *44*, 353-357.