

Bayesian Prediction under a Finite Mixture of Generalized Exponential Lifetime Model

M. Mahmoud
Department of Mathematics
Faculty of Science, Ain Shams University
Cairo, Egypt
mmahmoud46@hotmail.com

Elsayed H. Saleh
Department of Statistics
Faculty of Science, King Abdul-Aziz University
Jeddah, Saudi Arabia
noused7000@yahoo.com

Shaymaa M. Helmy
Department of Mathematics
Faculty of Science, Ain Shams University
Cairo, Egypt
shaymaa_moh_helmy@hotmail.com

Abstract

In this article a heterogeneous population is represented by a mixture of two generalized exponential distributions. Using the two-sample prediction technique, Bayesian prediction bounds for future order statistics are obtained based on type II censored and complete data. A numerical example is given to illustrate the procedures and the accuracy of the prediction intervals is investigated via extensive Monte Carlo simulation.

Keywords: Heterogeneous population; mixture of two Generalized exponential distribution; type II censored sample; complete sample; Bayesian prediction; Life testing model; Monte Carlo simulation.

AMS Classification: 62F15, 62F25, 62M20, 62N01, 62N05, 60E05.

1. Introduction

Recently generalized exponential distribution has received considerable attention. Two-parameter generalized exponential distribution $GE(\alpha, \lambda)$ was originally introduced by (Gupta & Kundu, 1999) as a skewed distribution, and as an alternative to Weibull, gamma and log-normal distributions, and studied its different properties. It is observed that $GE(\alpha, \lambda)$ distribution can be used quite effectively to analyze skewed data. Extensive work has been done by several authors on $GE(\alpha, \lambda)$ distribution. Some of the recent references on $GE(\alpha, \lambda)$ distribution are (Gupta & Kundu, 2001), (Gupta & Kundu, 2002), (Jaheen, 2004), (Raqab & Madi, 2005), (Kundu, et al., 2005), (Gupta & Kundu, 2007), (Kundu & Gupta, 2008), (Kundu & Pardhan, 2008) and (Kim & Song, 2010).

The analysis of finite mixture models has received an increasing attention during the last years. Finite mixture models have been much studied both theoretically and practically by several authors e.g.: (Everitt & Hand, 1981), (Tittergington, et al., 1985), (McLachlan & Peel, 2000). A finite mixture of some suitable probability distribution is recommended to study a population that is supposed to comprise a number of subpopulations mixed in an unknown proportion. Finite mixture models have been successfully applied to several fields of knowledge, such as economics, biology, medicine and engineering. (Ateya, 2012a) proved the identifiability of a finite mixture of generalized exponential distributions and obtained maximum likelihood estimates of the parameters, using EM algorithm, based on a general form of right censored failure times. (Ali, et al., 2012) provide Bayesian estimation of the mixture of generalized exponential distribution using censored sample under different loss functions.

Complete knowledge on the lifetimes or failure times of all the experimental units may not be available for many reasons. In many studies, experiments often must terminate before all units on test have failed. In such cases exact lifetimes are known for only a portion of the units under study and the remainder of the lifetimes is known only to exceed certain values. Such data are called censored. One of the most common censoring schemes is type II censoring. In type II censoring, a total of n units are put on a life test, but instead of continuing until all n units have failed, the life test stopped at the time of the m^{th} ($1 \leq m \leq n$) unit failure. See (Zheng, 2002), (Sarhan, 2007) and (Yarmohammadi, 2010).

In many practical problems of statistics, one wishes to infer the value of unknown observable that belongs to a future sample by using current available information known as the informative sample. One way to do this is to construct an interval, which will contain these results with a specified probability. This interval is called prediction interval. Prediction has been applied in medicine, engineering, business, and other areas as well. For details on the history of statistical prediction, analysis and applications, see for example, (Aitchison & Dunsmore, 1975), (Geisser, 1993).

(Jaheen, 2003) studied Bayesian prediction under a finite mixture of two component Gompertz lifetime model, based on type-I censored sample. (Ateya, 2011) discussed Bayesian prediction under generalized exponential distribution based on one and two sample schemes using MCMC algorithm. (Ateya, 2012b) obtained Bayesian prediction intervals of future nonadjacent generalized order statistics from generalized exponential distribution using Markov chain Monte Carlo method. (Ateya & Rizk, 2013) discussed Bayesian prediction intervals (BPI's) of future generalized order statistics under a finite mixture of two components of generalized exponential distributions based on a type II censored samples.

In this article, Bayesian prediction bounds for the s^{th} future observable from a heterogeneous population represented by a finite mixture of two components generalized exponential lifetime model, based on type-II censored and complete data, and using the two-sample prediction technique, are obtained. Prediction bounds are obtained when the parameters λ_1 and λ_2 are assumed to be known. The accuracy of prediction intervals is investigated via Monte Carlo simulation. In section 2, the generalized exponential

distribution GED and the mixture of two generalized component are defined. The Bayesian prediction bounds of future observables based on type-II censored and complete data are studied in section 3. In section 4, examples for samples generated from the mixture model are given and a simulation study is performed to illustrate the computations of the obtained results. Finally, some comments and concluding remarks are given in section 5.

2. The Generalized Exponential Distribution and Its Mixture Model

The probability density function (pdf), cumulative distribution function (cdf) and reliability function of $GE(\alpha, \lambda)$ are given, respectively, by:

$$f(t) = \alpha\lambda(1 - e^{-\lambda t})^{\alpha-1} e^{-\lambda t}, \quad t > 0, \alpha, \lambda > 0 \quad (2.1)$$

$$F(t) = (1 - e^{-\lambda t})^\alpha, \quad t > 0, \alpha, \lambda > 0 \quad (2.2)$$

$$R(t) = 1 - (1 - e^{-\lambda t})^\alpha, \quad t > 0, \alpha, \lambda > 0 \quad (2.3)$$

Where α is a shape parameter and λ is the reciprocal of the scale parameter.

A random variable T is said to follow a finite mixture distribution with k components, if the density function of T can be written in the form:

$$f(t) = \sum_{j=1}^k p_j f_j(t), \quad (2.4)$$

Where p_j is a non-negative real number (known as the j^{th} mixing proportion) such that $\sum_{j=1}^k p_j = 1$ and f_j is a density function of the j^{th} component; $j = 1, 2, \dots, k$.

The corresponding cdf is given by:

$$F(t) = \sum_{j=1}^k p_j F_j(t), \quad (2.5)$$

Where $F_j(t)$ is the j^{th} cdf component.

The reliability function of the mixture is given by:

$$R(t) = \sum_{j=1}^k p_j R_j(t), \quad (2.6)$$

Where $R_j(t)$ is the j^{th} reliability component.

A finite mixture of two-component generalized exponential lifetime model may describe a heterogeneous population. The pdf of a finite mixture of two $GE(\alpha_j, \lambda_j)$, $j = 1, 2$ components is given by:

$$f_T(t; p, \alpha_1, \alpha_2, \lambda_1, \lambda_2) = p f_1(t; \alpha_1, \lambda_1) + (1 - p) f_2(t; \alpha_2, \lambda_2) \quad (2.7)$$

Where, $p \in [0,1]$ and for $j = 1,2$, $f_j(t)$ are given by (2.1) after indexing α and λ by j .

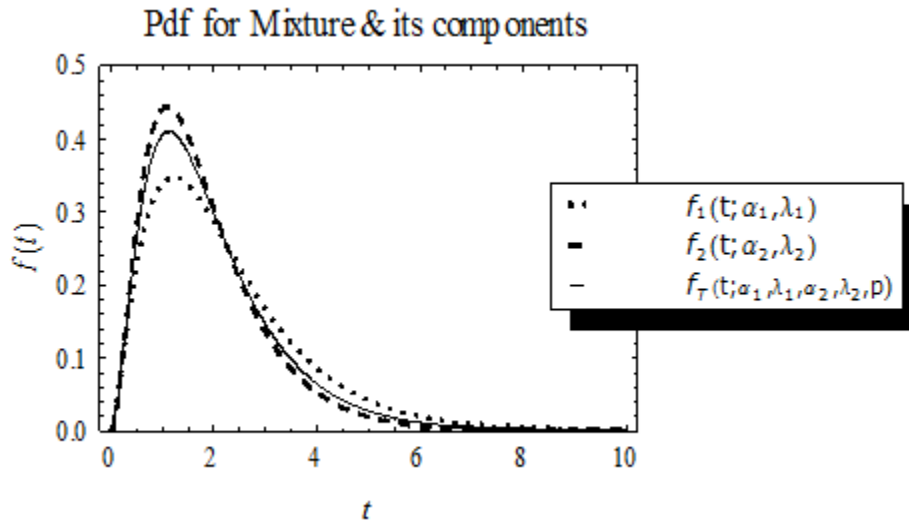


Figure 1: The pdf $f_T(t; p, \alpha_1, \alpha_2, \lambda_1, \lambda_2)$ of a finite mixture of two $GE(\alpha_j, \lambda_j)$, $j = 1, 2$ and its components $f_1(t; \alpha_1, \lambda_1)$ and $f_2(t; \alpha_2, \lambda_2)$ with $p = 0.35$, $\alpha_1 = 2.5$, $\alpha_2 = 3$, $\lambda_1 = 0.75$, $\lambda_2 = 1$.

3. Posterior Distributions and Bayesian Two-sample Prediction

In this section, we present the posterior densities of the parameters α_1, α_2 and p as well as the Bayesian prediction distribution for the future failure times, based on the observed type-II censored and complete data, when λ_1 and λ_2 are known.

When the parameters λ_1 and λ_2 are assumed known, we suppose that α_1, α_2 and p are independent random variables. The joint prior pdf of the vector (α_1, α_2, p) is thus given by:

$$g(\alpha_1, \alpha_2, p) = g_1(\alpha_1)g_2(\alpha_2)g_3(p) \tag{3.1}$$

Where $g_i(\alpha_i)$ is the prior pdf of α_i ; $i = 1,2$ and $g_3(p)$ is a prior pdf of p . Let $g_3(p)$ be *uniform*[0, 1] and the random variables α_i ; $i = 1,2$ are assumed to follow the gamma density with parameters (a_i, b_i) .

Then:

$$\left. \begin{aligned} g_i(\alpha_i) &= \frac{b_i^{a_i} \alpha_i^{a_i-1}}{\Gamma(a_i)} e^{-b_i \alpha_i} \quad ; \quad \alpha_i > 0; a_i, b_i > 0; i = 1,2 \\ \text{and } g_3(p) &= 1 \end{aligned} \right\} \tag{3.2}$$

3.1 Bayesian Prediction Based on Type-II Censored Data

Suppose that n units from a population with pdf of a finite mixture of two $GE(\alpha_j, \lambda_j)$; $j = 1,2$ components given by (2.7) are subjected to a life testing experiment, and that the test is terminated after some fixed predetermined number of failures r . It is assumed that

an item can be attributed to the appropriate subpopulation after it had failed. Then r units have failed during the interval $(0, t_{(r)})$: r_1 from the first and r_2 from the second subpopulation, such that $r = r_1 + r_2$ and $n - r$ units which cannot be identified as to subpopulation are still functioning. Let t_{ij} denotes the failure time of the j^{th} unit that belongs to the i^{th} subpopulation and $t_{ij} \leq t_{(r)}$; $j = 1, 2, \dots, r_i$; $i = 1, 2$. See (Jaheen, 2003) based on such scheme of sampling, the likelihood function as described by (Mendenhall & Hader, 1958), is given by:

$$\begin{aligned}
 L(p, \alpha_1, \alpha_2, \lambda_1, \lambda_2; \underline{t}) &= \frac{n!}{(n-r)!} \times \left[\prod_{j=1}^{r_1} p f_1(t_{1j}; \alpha_1, \lambda_1) \right] \times \left[\prod_{j=1}^{r_2} (1-p) f_2(t_{2j}; \alpha_2, \lambda_2) \right] \\
 &\times [R(t_{(r)})]^{n-r} \\
 &= \frac{n!}{(n-r)!} \sum_1 \Omega_1 p^{\delta_1} (1-p)^{\delta_2} \times \prod_{i=1}^2 [B_i(\alpha_i, \lambda_i)] \tag{3.3}
 \end{aligned}$$

Where: $\underline{t} = (t_{11}, \dots, t_{1r_1}, t_{21}, \dots, t_{2r_2})$ and for $i = 1, 2$,

$$\begin{aligned}
 B_i(\alpha_i, \lambda_i) &= (\alpha_i \lambda_i)^{r_i} e^{-\lambda_i T_i} C_i^{\alpha_i - 1} \times \phi(\lambda_i, t_{(r)})^{\alpha_i s_i} \\
 T_i &= \sum_{j=1}^{r_i} t_{ij} \\
 C_i &= \prod_{j=1}^{r_i} (1 - e^{-\lambda_i t_{ij}}), \quad \phi(\lambda_i, t) = (1 - e^{-\lambda_i t}) \\
 \sum_1 &= \sum_{k=0}^{n-r} \sum_{s_1=0}^{n-r-k} \sum_{s_2=0}^k, \\
 \Omega_1 &= \binom{n-r}{k} \binom{n-r-k}{s_1} \binom{k}{s_2} (-)^{s_1+s_2}, \\
 \delta_1 &= r_1 + n - r - k, \delta_2 = r_2 + k
 \end{aligned}$$

From Eqs. (3.2) and (3.3) the posterior pdf of α_1, α_2 and p given \underline{t} based on type-II censored data is given by:

$$\begin{aligned}
 P(p, \alpha_1, \alpha_2 | \underline{t}) &= A_1 \left[\sum_1 \Omega_1 p^{\delta_1} (1-p)^{\delta_2} \right. \\
 &\quad \left. \times \prod_{i=1}^2 [(\alpha_i)^{r_i + a_1 - 1} e^{-b_i \alpha_i} C_i^{\alpha_i} \phi(\lambda_i, t_{(r)})^{\alpha_i s_i}] \right] \tag{3.4}
 \end{aligned}$$

A_1 is a normalizing constant given by:

$$A_1^{-1} = \int_0^1 \int_0^\infty \int_0^\infty P(p, \alpha_1, \alpha_2 | \underline{t}) d\alpha_1 d\alpha_2 dp$$

$$\begin{aligned}
 &= \sum_1 \Omega_1 \int_0^1 p^{\delta_1} (1-p)^{\delta_2} dp \\
 &\quad \times \prod_{i=1}^2 \left[\int_0^\infty (\alpha_i)^{r_i+a_i-1} e^{-b_i \alpha_i} C_i^{\alpha_i} \phi(\lambda_i, t_{(r)})^{\alpha_i s_i} d\alpha_i \right] \\
 &= \sum_1 \Omega_1 \times \beta(\delta_1 + 1, \delta_2 + 1) \times \prod_{i=1}^2 I_1[i] \tag{3.5}
 \end{aligned}$$

And:

$$\begin{aligned}
 I_1[i] &= \int_0^\infty (\alpha_i)^{r_i+a_i-1} e^{-b_i \alpha_i} C_i^{\alpha_i} \phi(\lambda_i, t_{(r)})^{\alpha_i s_i} d\alpha_i \\
 &= \int_0^\infty (\alpha_i)^{r_i+a_i-1} \times e^{-\alpha_i w_i} d\alpha_i \\
 &= \frac{\Gamma(r_i + a_i)}{w_i^{(r_i+a_i)}}
 \end{aligned}$$

Where:

$$w_i = b_i - \sum_{j=1}^{r_i} \ln(1 - e^{-\lambda_i t_{ij}}) - s_i \ln(1 - e^{-\lambda_i t_{(r)}})$$

A future sample of size m is assumed to be independent of the informative sample of size n and is obtained from the same population with pdf (2.7). Predictions are made for the s^{th} order statistic in the future sample based on the informative sample. This is the two-sample prediction technique. Let y_s be the ordered lifetime of the s^{th} components to fail in a future sample of size m , $1 \leq s \leq m$. The s^{th} order statistic in a sample of size m represents the life length of a $(m - s + 1)$ out of m system. The distribution function of y_s is given by: (see (Arnold, et al., 2008), (Jaheen, 2003))

$$\begin{aligned}
 F_{Y_s}(y_s|p, \alpha_1, \alpha_2) &= \sum_{l=s}^m \binom{m}{l} [F_T(y_s|p, \alpha_1, \alpha_2)]^l \times [1 - F_T(y_s|p, \alpha_1, \alpha_2)]^{m-l} \\
 &= \sum_{l=s}^m \left[\sum_{i=0}^l \binom{l}{i} \binom{m}{l} (-)^i (R(y_s))^{i+m-l} \right] \tag{3.6}
 \end{aligned}$$

Where $F_T(y_s|p, \alpha_1, \alpha_2) = 1 - R(y_s)$ is the distribution function of the mixture model and $R(y_s)$ is the reliability function of the mixture model after replacing t by y_s .

$$F_{y_s}(y_s|p, \alpha_1, \alpha_2) = \sum_{l=s}^m \sum_{i=0}^l \binom{m}{l} \binom{l}{i} (-)^i [p \times R_1(y_s) + (1 - p) \times R_2(y_s)]^{m-l+i}$$

Applying the binomial expansion, we get:

$$\begin{aligned}
 F_{y_s}(y_s|p, \alpha_1, \alpha_2) &= \sum_{l=s}^m \sum_{i=0}^l \sum_{k=0}^{m-l+i} (-)^i \binom{m}{l} \binom{l}{i} \binom{m-l+i}{k} p^{m-l+i-k} \times (1-p)^k \\
 &\times [1 - (1 - e^{-\lambda_1 y_s}) \alpha_1]^{m-l+i-k} \times [1 - (1 - e^{-\lambda_2 y_s}) \alpha_2]^k \\
 &= \sum_2 \Omega_1^* p^{\delta_1^*} \times (1-p)^k \prod_{i=1}^2 \phi(\lambda_i, y_s)^{\alpha_2 u_2} \tag{3.7}
 \end{aligned}$$

Where:

$$\begin{aligned}
 \sum_2 &= \sum_{l=s}^m \sum_{i=0}^l \sum_{k=0}^{m-l+i} \sum_{u_1=0}^{m-l+i-k} \sum_{u_2=0}^k, \\
 \Omega_1^* &= (-)^{i+u_1+u_2} \binom{m}{l} \binom{l}{i} \binom{m-l+i}{k} \binom{m-l+i-k}{u_1} \binom{k}{u_2}, \\
 \delta_1^* &= m - l + i - k
 \end{aligned}$$

The Bayes predictive pdf of y_s given \underline{t} is defined by:

$$f^*(y_s|\underline{t}) = \int_0^1 \int_0^\infty \int_0^\infty f(y_s|p, \alpha_1, \alpha_2) P(p, \alpha_1, \alpha_2|\underline{t}) d\alpha_1 d\alpha_2 dp \tag{3.8}$$

Where $P(p, \alpha_1, \alpha_2|\underline{t})$ is the posterior pdf and $f(y_s|p, \alpha_1, \alpha_2)$ is the pdf of the s^{th} component in a future sample which can be obtained from (2.7).

Bayesian prediction bounds for y_s can be obtained, for a given value of ε , by computing:

$$\begin{aligned}
 \Pr[y_s \geq \varepsilon|\underline{t}] &= \int_\varepsilon^\infty f^*(y_s|\underline{t}) dy_s \\
 &= 1 - \int_0^1 \int_0^\infty \int_0^\infty F_{y_s}(\varepsilon|p, \alpha_1, \alpha_2) P(p, \alpha_1, \alpha_2|\underline{t}) d\alpha_1 d\alpha_2 dp \tag{3.9}
 \end{aligned}$$

Where $F_{y_s}(\cdot|p, \alpha_1, \alpha_2)$ is the cumulative distribution function of s^{th} component in the future sample given by (2.7).

Therefore:

$$\begin{aligned}
 \Pr[y_s \geq \varepsilon|\underline{t}] &= 1 \\
 &- A_1 \sum_3 \Omega_1^{**} \int_0^1 p^{\delta_1^{**}} \times (1-p)^{\delta_2^{**}} dp \\
 &\times \prod_{i=1}^2 \left[\int_0^\infty (\alpha_i)^{r_i+a_i-1} e^{-b_i \alpha_i} \left[\prod_{j=1}^{r_i} (1 - e^{-\lambda_i t_{ij}})^{\alpha_i} \right] \right. \\
 &\left. \times (1 - e^{-\lambda_i t(r)})^{\alpha_i s_i} \times (1 - e^{-\lambda_i \varepsilon})^{\alpha_i u_i} d\alpha_i \right]
 \end{aligned}$$

$$\Pr[y_s \geq \varepsilon|t] = 1 - A_1 \sum_3 \Omega_1^{**} \beta(\delta_1^{**} + 1, \delta_2^{**} + 1) \times \prod_{i=1}^2 I_1^*[i] \tag{3.10}$$

Where:

$$\sum_3 = \sum_{l=s}^m \sum_{i=0}^l \sum_{k=0}^{m-l+i} \sum_{u_1=0}^{m-l+i-k} \sum_{u_2=0}^k \sum_{k=0}^{n-r} \sum_{s_1=0}^{n-r-k} \sum_{s_2=0}^k,$$

$$\Omega_1^{**} = (-)^{i+u_1+u_2+s_1+s_2} \binom{m}{l} \binom{l}{i} \binom{m-l+i}{k} \binom{m-l+i-k}{u_1} \binom{k}{u_2} \binom{n-r}{k} \times \binom{n-r-k}{s_1} \binom{k}{s_2},$$

$$\delta_1^{**} = r_1 + n - r - k + m - l + i - k,$$

$$\delta_2^{**} = r_2 + k + v$$

And

$$\begin{aligned} I_1^*[i] &= \int_0^\infty (\alpha_i)^{r_i+a_i-1} \times (1 - e^{-\lambda_i \varepsilon})^{\alpha_i u_i} \times \left[\prod_{j=1}^{r_i} ((1 - e^{-\lambda_i t_{ij}})^{\alpha_i}) \right] \times (1 - e^{-\lambda_i t(r)})^{\alpha_i s_i} \\ &\quad \times e^{-b_i \alpha_i} d\alpha_i \\ &= \int_0^\infty (\alpha_i)^{r_i+a_i-1} \times e^{-\alpha_i w_i^*} d\alpha_i \\ &= \frac{\Gamma(r_i + a_i)}{w_i^{*(r_i+a_i)}} \end{aligned}$$

Where:

$$w_i^* = b_i - \sum_{j=1}^{r_i} \ln(1 - e^{-\lambda_i t_{ij}}) - s_i \ln(1 - e^{-\lambda_i t(r)}) - u_i \ln(1 - e^{-\lambda_i \varepsilon})$$

3.2 Bayesian Prediction Based on Complete Data

Suppose that n -units from a population with pdf of a finite mixture of two $GE(\alpha_j, \lambda_j)$; $j = 1, 2$ components (1.7) are subjected to a life testing experiment and that the test is terminated after the failure of all n items. It is assumed that an item can be attributed to the appropriate subpopulation after it had failed. Then the n -units have failed during the interval $(0, t_{(n)})$: r from the first and $n - r$ from the second subpopulation.

Let t_{ij} denote the failure of the j^{th} unit that belongs to the i^{th} subpopulation and $t_{ij} \leq t_{(n)}$; $j = 1, 2, \dots, r_i$; $i = 1, 2$; $r_1 = r$ and $r_2 = n - r$. Where $t_{(n)}$ denotes the failure time of the n^{th} unit.

Based on such scheme of sampling, the likelihood function is given by:

$$L(p, \alpha_1, \alpha_2, \lambda_1, \lambda_2; \underline{t}) = n! \left[\prod_{j=1}^{r_1} p f_1(t_{1j}; \alpha_1, \lambda_1) \right] \times \left[\prod_{j=1}^{r_2} (1-p) f_2(t_{2j}; \alpha_2, \lambda_2) \right] \quad (3.11)$$

$$= n! \times p^{r_1} \times (1-p)^{r_2} \times \prod_{i=1}^2 [D_i(\alpha_i, \lambda_i)] \quad (3.12)$$

Where: $\underline{t} = (t_{11}, \dots, t_{1r_1}, t_{21}, \dots, t_{2r_2})$ and for $i = 1, 2$,

$$D_i(\alpha_i, \lambda_i) = (\alpha_i \lambda_i)^{r_i} e^{-\lambda_i T_i} C_i^{\alpha_i - 1}$$

$$T_i = \sum_{j=1}^{r_i} t_{ij}$$

It follows, from Eqs. (3.2) and (3.12), that the posterior density of α_1, α_2 and p given \underline{t} based on complete data is given by:

$$P(p, \alpha_1, \alpha_2 | \underline{t}) = A_2 \left(p^{r_1} \times (1-p)^{r_2} \times \prod_{i=1}^2 [(\alpha_i)^{r_i + a_i - 1} \times C_i^{\alpha_i} \times e^{-b_i \alpha_i}] \right) \quad (3.13)$$

A_2 is the normalizing constant given by:

$$A_2^{-1} = \int_0^1 p^{r_1} (1-p)^{r_2} dp \times \prod_{i=1}^2 \left(\int_0^\infty (\alpha_i)^{r_i + a_i - 1} C_i^{\alpha_i} e^{-b_i \alpha_i} d\alpha_i \right) = \beta(r_1 + 1, r_2 + 1) \times \prod_{i=1}^2 I_2[i] \quad (3.14)$$

Where:

$$I_2[i] = \int_0^\infty (\alpha_i)^{r_i + a_i - 1} \left[\prod_{j=1}^{r_i} (1 - e^{-\lambda_i t_{ij}})^{\alpha_i} \right] e^{-b_i \alpha_i} d\alpha_i; i = 1, 2 = \Gamma(r_i + a_i) \times v_i^{-(r_i + a_i)} \quad (3.15)$$

And

$$v_i = b_i - \sum_{j=1}^{r_i} \ln(1 - e^{-\lambda_i t_{ij}}) \quad ; i = 1, 2$$

Bayesian prediction bounds for y_s are obtained by evaluating $\Pr[y_s \geq \varepsilon|t]$, for a given value of ε , it follows from eq. (3.9) that:

$$\begin{aligned} \Pr[y_s \geq \varepsilon|t] &= 1 \\ &\quad - \left(A_2 \sum_4 \Omega_2 \int_0^1 p^{\delta_1} \times (1-p)^{\delta_2} dp \right. \\ &\quad \times \prod_{i=1}^2 \left(\int_0^\infty (\alpha_i)^{r_i+a_i-1} (1-e^{-\lambda_i \varepsilon})^{\alpha_i u_i} \right. \\ &\quad \times \left. \left[\prod_{j=1}^{r_i} (1-e^{-\lambda_i t_{ij}})^{\alpha_i} \right] e^{-b_i \alpha_i} d\alpha_i \right) \left. \right) \\ &= 1 - \left(A_2 \times \sum_4 \Omega_2 \times \beta(\delta_1 + 1, \delta_2 + 1) \times \prod_{i=1}^2 I_2^*[i] \right) \end{aligned} \tag{3.16}$$

Where:

$$\begin{aligned} \sum_4 &= \sum_{l=s}^m \sum_{i=0}^l \sum_{k=0}^{m-l+i} \sum_{u_1=0}^{m-l+i-k} \sum_{u_2=0}^k, \\ \Omega_2 &= \binom{m}{l} \binom{l}{i} \binom{m-l+i}{k} \binom{m-l+i-k}{u_1} \binom{k}{u_2} (-)^{i+u_1+u_2}, \\ \delta_1 &= m - l + i - k + r_1, \\ \delta_2 &= k + r_2 \end{aligned}$$

And

$$\begin{aligned} I_2^*[i] &= \int_0^\infty (\alpha_i)^{r_i+a_i-1} (1-e^{-\lambda_i \varepsilon})^{\alpha_i u_i} \left[\prod_{j=1}^{r_i} (1-e^{-\lambda_i t_{ij}})^{\alpha_i} \right] e^{-b_i \alpha_i} d\alpha_i \\ &= [(r_i + a_i) \times v_i^{*-(r_i+a_i)}] \quad ; i = 1,2 \end{aligned} \tag{3.17}$$

Where:

$$v_i^* = b_i - \sum_{j=1}^{r_i} \ln(1 - e^{-\lambda_i t_{ij}}) - u_i \ln(1 - e^{-\lambda_i \varepsilon}) \quad ; i = 1,2$$

Using (3.10) and (3.16), a $100\gamma\%$ prediction interval for y_s is then given by:

$$P[L(t) < y_s < U(t)] = \gamma$$

Where $L(\underline{t})$ and $U(\underline{t})$ are obtained, respectively, by solving the following two equations:

$$P[L(\underline{t}) > y_s] = \frac{1 + \gamma}{2}, \quad [y_s > U(\underline{t})] = \frac{1 - \gamma}{2} \quad (3.18)$$

4. Numerical Computation

In this section, algorithms used to generate type-II censored and complete samples are given, a numerical example is given to illustrate the results and the accuracy of the prediction intervals is investigated via Monte Carlo simulations.

4.1 Generation Algorithms

Failure-censored samples:

The algorithm to generate a pseudorandom failure-censored sample (type-II censored) with n units and r failures is as follows (see (William & Luis, 1998)):

1. Generate U_1, \dots, U_r pseudorandom observations from UNIF(0,1)
2. Compute the uniform pseudorandom order statistics:

$$U_{(1)} = 1 - [1 - U_{(0)}] \times (1 - U_1)^{1/n} \quad ; U_{(0)} = 0$$

$$U_{(2)} = 1 - [1 - U_{(1)}] \times (1 - U_2)^{1/(n-1)}$$

$$U_{(r)} = 1 - [1 - U_{(r-1)}] \times (1 - U_r)^{1/(n-r+1)}$$

3. The pseudorandom sample from $F(t; \theta)$ is:

$$T_{(i)} = F^{-1}[U_{(i)}; \theta] \quad ; i = 1, \dots, r$$

For generalized exponential distribution with cdf $F(t; \alpha, \lambda) = (1 - e^{-\lambda t})^\alpha$

$$T_{(i)} = -\frac{1}{\lambda} \ln \left(1 - U_{(i)}^{1/\alpha} \right) \quad ; i = 1, \dots, r$$

In case of mixture of two components:

If $U_i \leq p$

$$T_{(i)} = -\frac{1}{\lambda_1} \ln \left(1 - U_{(i)}^{1/\alpha_1} \right)$$

Otherwise:

$$T_{(i)} = -\frac{1}{\lambda_2} \ln \left(1 - U_{(i)}^{1/\alpha_2} \right)$$

For the case of complete sample we are using the same algorithm with $r = n$

4.2 Example

The following complete sample ($n = 20$) has been generated from the mixture model of two generalized exponential distributions with $p = 0.35, \alpha_1 = 2.5, \alpha_2 = 3, \lambda_1 = 0.75, \lambda_2 = 1$ (See Figure 1).

{1.08522, 1.43312, 2.48984, 2.65207, 3.18065, 3.26202, 3.31147} ($r_1 = 7$)

{0.39141, 0.922536, 1.02482, 1.21247, 1.28588, 1.45316, 1.52692, 2.03428, 2.10718, 2.4974, 3.39261, 4.3734, 5.00216} ($r_2 = 13$)

Also the following type-II censored sample ($n = 20, r = 10$) has been generated from the mixture model of two generalized exponential distributions with the same parameters.

{0.176573, 0.319349, 0.560964} ($r_1 = 3$)

{0.445624, 0.676239, 0.904488, 0.987053, 1.13092, 1.21349, 1.36661} ($r_2 = 7$)

4.3 Monte Carlo Simulation

The behavior of the Bayes prediction bounds derived in section 3 is investigated via Monte Carlo simulations according to the following steps:

1. Making use of the vector of actual parameters $p = 0.35, \alpha_1 = 2.5, \alpha_2 = 3, \lambda_1 = 0.75, \lambda_2 = 1$ (see Figure 1) and for a given values of the prior parameters $a_1 = 0.9, a_2 = 0.75, b_1 = 0.5$ and $b_2 = 0.35$, 1000 samples $(t_{11}, \dots, t_{1r_1}, t_{21}, \dots, t_{2r_2})$ of different sizes n were generated from the mixture model of two generalized exponential with pdf (2.7) in case of complete data and also were generated for different values of r in case of type-II censored data using algorithm (4.1).
2. For each sample, the 100 $\gamma\%$ Bayesian prediction interval for the unobserved value Y_s was computed by solving numerically the system of equations (3.18) using Mathematica 7 via the routine Find Root. The lengths of the intervals were obtained.
3. The simulated coverage probabilities (CP), the average lower limits, the average upper limits (L,U) and the average interval lengths (AL) from the 1000 samples were computed.
4. Steps 1-3 were performed for $\gamma = 0.90, 0.95$ and 0.99 in case of complete sample (table 1, 2 & 3) and for $\gamma = 0.95$ in case of type-II censoring sample (table 4).

Table 1: Two sample 90% Bayesian prediction interval (BPI), simulated coverage probability (CP), average interval length (AL), average lower limit (L), and average upper limit (U), for $y_s, s = 1, (\frac{m}{2} \text{ or } \frac{m+1}{2})$ & m from 1000 complete samples for different sample sizes n and different future sample sizes m

n	m	y_1			y_3			y_m		
		AL	CP	(L,U)	AL	CP	(L,U)	AL	CP	(L,U)
10	5	1.52357	0.894977	(0.0827333, 1.6063)	2.3264	0.898441	(0.685326, 3.01172)	4.70999	0.896364	(1.69254, 6.40253)
15		1.47063	0.900596	(0.119915, 1.59054)	2.23691	0.895451	(0.744749, 2.98165)	4.68701	0.901387	(1.73733, 6.42434)
20		1.43898	0.896125	(0.143177, 1.58216)	2.19089	0.90083	(0.780129, 2.97102)	4.64995	0.897544	(1.75422, 6.40416)
n	m	y_1			y_5			y_m		
10	10	1.08825	0.892965	(0.0497455, 1.138)	1.72682	0.900654	(0.740225, 2.46704)	4.8347	0.894736	(2.40211, 7.23681)
15		1.03719	0.896857	(0.0751358, 1.11232)	1.62273	0.894326	(0.807276, 2.43)	4.79201	0.897378	(2.45084, 7.24285)
20		1.00899	0.897963	(0.0957232, 1.10472)	1.56545	0.899717	(0.846676, 2.41213)	4.77593	0.895302	(2.47602, 7.25195)
25		0.992403	0.901625	(0.109501, 1.1019)	1.52632	0.894637	(0.865671, 2.39199)	4.76084	0.898876	(2.48954, 7.25037)
n	m	y_1			y_7			y_m		
15	15	0.859737	0.900848	(0.0589288, 0.918666)	1.38938	0.900952	(0.844551, 2.23393)	4.83324	0.898049	(2.86772, 7.70096)
20		0.829055	0.899935	(0.0736785, 0.902733)	1.32348	0.894961	(0.88491, 2.20839)	4.82498	0.896674	(2.90317, 7.72815)
25		0.814188	0.893147	(0.0853863, 0.899575)	1.28099	0.899637	(0.905843, 2.18684)	4.81101	0.897679	(2.91158, 7.72259)

Table 2: Two sample 95% Bayesian prediction interval (BPI), simulated coverage probability (CP), average interval length (AL), average lower limit (L) and average upper limit (U), for $y_s, s = 1, (\frac{m}{2} \text{ or } \frac{m+1}{2})$ & m from 1000 complete samples for different sample sizes n and different future sample sizes m

n	m	y_1			y_3			y_m		
		AL	CP	(L,U)	AL	CP	(L,U)	AL	CP	(L,U)
10	5	1.79637	0.950026	(0.044897, 1.84127)	2.79574	0.943224	(0.543099, 3.33884)	5.79815	0.950846	(1.45314, 7.25129)
15		1.73395	0.949209	(0.0728518, 1.8068)	2.69362	0.943447	(0.605762, 3.29938)	5.75177	0.947511	(1.50159, 7.25336)
20		1.70154	0.948106	(0.0886755, 1.79021)	2.6393	0.948225	(0.648906, 3.2882)	5.73059	0.950217	(1.53134, 7.26193)
n	m	y_1			y_5			y_m		
10	10	1.25921	0.948961	(0.0235629, 1.28277)	2.06192	0.944242	(0.599317, 2.66124)	5.93527	0.947495	(2.15383, 8.08909)
15		1.20907	0.950941	(0.0411074, 1.25018)	1.94064	0.944264	(0.686045, 2.62668)	5.89988	0.949964	(2.20464, 8.10452)
20		1.18533	0.947614	(0.0566421, 1.24197)	1.87145	0.945398	(0.726748, 2.5982)	5.86573	0.943478	(2.22703, 8.09276)
25		1.16303	0.948546	(0.0696206, 1.23265)	1.83012	0.949189	(0.760101, 2.59023)	5.85423	0.953462	(2.24877, 8.103)
n	m	y_1			y_7			y_m		
15	15	1.00495	0.952518	(0.0316057, 1.0365657)	1.65906	0.942709	(0.747692, 2.40676)	5.95059	0.950825	(2.62565, 8.57624)
20		0.976104	0.949168	(0.0443871, 1.02049)	1.57802	0.952383	(0.776226, 2.35425)	5.94215	0.948965	(2.6625, 8.60465)
25		0.957671	0.948115	(0.0537799, 1.01145)	1.53014	0.949727	(0.813485, 2.34362)	5.9265	0.953689	(2.68065, 8.60715)

Table 3: Two sample 99% Bayesian prediction interval (BPI), simulated coverage probability (CP), average interval length (AL), average lower limit (L) and average upper limit(U), for $y_s, s = 1, (\frac{m}{2} \text{ or } \frac{m+1}{2})$ & m from 1000 complete samples for different sample sizes n and different future sample sizes m

n	m	y_1			y_3			y_m		
		AL	CP	(L,U)	AL	CP	(L,U)	AL	CP	(L,U)
10	5	2.29251	0.987181	(0.00681839, 2.29933)	3.74338	0.987483	(0.28241, 4.02579)	8.19861	0.990723	(1.01986, 9.21847)
15		2.24326	0.990664	(0.0172311, 2.26049)	3.62974	0.987625	(0.370428, 4.00016)	8.12556	0.987958	(1.10785, 9.23341)
20		2.22371	0.982906	(0.029191, 2.2529)	3.56137	0.986475	(0.41751, 3.97888)	8.10997	0.988774	(1.15297, 9.26294)
n	m	y_1			y_5			y_m		
10	10	1.60624	0.983668	(0.0036538, 1.6099)	2.73843	0.987565	(0.369212, 3.10764)	8.36944	0.991596	(1.70693, 10.0764)
15		1.55865	0.988872	(0.0113983, 1.57005)	2.57675	0.988676	(0.468379, 3.04513)	8.33296	0.990577	(1.78132, 10.1143)
20		1.51379	0.989423	(0.0167858, 1.53057)	2.48987	0.98664	(0.527971, 3.01784)	8.31547	0.980595	(1.81646, 10.1319)
25		1.49131	0.987447	(0.0229587, 1.51427)	2.43492	0.991578	(0.567722, 3.00264)	8.27781	0.986975	(1.85077, 10.1286)
n	m	y_1			y_7			y_m		
15	15	1.28187	0.984936	(0.00809931, 1.28997)	2.19564	0.990847	(0.526362, 2.722)	8.4301	0.980764	(2.19756, 10.6277)
20		1.24128	0.983625	(0.0132456, 1.25452)	2.09341	0.984503	(0.59066, 2.68407)	8.39823	0.991406	(2.2353, 10.6335)
25		1.21778	0.984581	(0.0180862, 1.23587)	2.02873	0.989669	(0.637201, 2.66593)	8.35343	0.991378	(2.26205, 10.6155)

From tables 1, 2 & 3, we can see that for each value of y_s , and given values of n and m , the relation between the Bayesian prediction intervals (BPI) corresponding to 90%, 95% and 99% is clear. The first is a subset from the second which itself a subset of the third. Also the average interval width tends to decrease as n increases. The limits of the intervals (L,U) for the future observation y_s are quite relative to the apparent graph of the mixture in Figure 1.

Table 4: Two sample 95% Bayesian prediction intervals (BPI), simulated coverage probability (CP), average interval length (AL), average lower limit (L) and average upper limit (U), for $y_s, s = 1, 3$ & 5 from 1000 Type-II censored samples for different sample sizes n , and different censoring value r and future sample size $m = 5$

r	n	y_1			y_3			y_m		
		AL	CP	(L,U)	AL	CP	(L,U)	AL	CP	(L,U)
7	10	1.81565	0.947822	(0.0352285, 1.85088)	2.83322	0.947604	(0.534211, 3.36743)	5.86858	0.950215	(1.45861, 7.32719)
	15	1.79981	0.944414	(0.0529655, 1.85277)	2.77659	0.943945	(0.606657, 3.38325)	5.8669	0.943727	(1.51174, 7.37864)
	20	1.79671	0.944121	(0.0513559, 1.84807)	2.74232	0.950496	(0.649935, 3.39226)	5.86291	0.950492	(1.55548, 7.41839)
10	10	1.78046	0.949724	(0.0411987, 1.82166)	2.79726	0.94555	(0.53949, 3.33675)	5.79843	0.947854	(1.42924, 7.22768)
	15	1.76608	0.94275	(0.0611684, 1.82725)	2.76059	0.950074	(0.617372, 3.37796)	5.84322	0.947587	(1.52699, 7.37021)
	20	1.76286	0.949625	(0.0737992, 1.83666)	2.72886	0.952973	(0.652338, 3.3812)	5.8683	0.945899	(1.55272, 7.42102)

From table 4, we can see that for fixed sample size n , by increasing the censoring value r , the average interval width tends to decrease. The limits of the intervals (L,U) for the future observation y_s are quite relative to the apparent graph of the mixture in Figure 1.

5. Comments and Conclusion

In this paper, Bayesian prediction bounds for a future order statistic from the finite mixture of two-generalized exponential distribution are derived based on complete and type-II censored data. The two-sample prediction technique is used. It has been noticed, from Tables 1, 2, 3 and 4 that the prediction intervals are affected (getting shorter) by increasing n . Also the coverage probabilities are quite close to the pre-assigned confidence levels (90%, 95% and 99%), and therefore the intervals tend to perform very well in terms of simulated coverage probabilities. The average interval width tends to decrease as n increases. Also in case of type-II censored sample (as shown in Table 4) for fixed sample size n , by increasing the censoring value r , the average interval width tends to decrease and then the prediction intervals become better.

Acknowledgements

The authors would like to thank the managing editor and referee for their valuable remarks and comments, which improved the original manuscript.

References

1. Aitchison, J. & Dunsmore, I. R., 1975. *Statistical Prediction Analysis*. Cambridge: Cambridge University Press.
2. Ali, S., Aslam, M., Kundu, D. & Kazmi, S., 2012. Bayesian estimation of the mixture of generalized exponential distribution: a versatile lifetime model in industrial processes. *Journal of the Chinese Institute of Industrial Engineers*, 29(4), pp. 246-269.
3. Arnold, B. C., Balakrishnan, N. & Nagaraja, H. N., 2008. *A First Course in Order Statistics*. Siam publisher, Philadelphia.
4. Ateya, S. F., 2011. Prediction under generalized exponential distribution using MCMC. *international mathematical forum*, 6(63), pp. 3111-3119.
5. Ateya, S. F., 2012a. Maximum likelihood estimation under a finite mixture of generalized exponential distributions based on censored data. *Statistical Papers*.
6. Ateya, S. F., 2012b. Bayesian prediction intervals of future nonadjacent generalized exponential distribution using Markov Chain Monte Carlo. *Applied mathematical sciences*, 6(27), pp. 1335-1345.
7. Ateya, S. F. & Rizk, M. M., 2013. Bayesian prediction intervals of future Generalized order statistics under a finite mixture of generalized exponential distributions. *Applied mathematical sciences*, 7(32), pp. 1575-1592.
8. Everitt, B. S. & Hand, D. J., 1981. *Finite Mixture Distributions*. Cambridge: University Press.
9. Geisser, S., 1993. *Predictive Inference: An Introduction*. London: Chapman and Hall.
10. Gupta, R. & Kundu, D., 1999. Generalized exponential distribution. *Australian and New Zealand Journal of Statistics*, 41(2), pp. 173-188.

11. Gupta, R. & Kundu, D., 2001. Exponentiated exponential distribution, an alternative to Gamma and Weibull distributions. *Biometrical Journal*, 43(1), pp. 117-130.
12. Gupta, R. & Kundu, D., 2002. Generalized exponential distribution. *Journal of Statistical theory and applications*, Volume 1, pp. 101-118.
13. Gupta, R. & Kundu, D., 2007. Generalized exponential distribution: existing results and some recent developments. *Journal of Statistical Planning and Inference*, 137(11), pp. 3537-3547.
14. Jaheen, Z. F., 2003. Bayesian prediction under a mixture of two-component Gompertz lifetime model. *Sociedad de Estadística e Investigación Operativa*, 12(2), pp. 413-426.
15. Jaheen, Z. F., 2004. Empirical Bayes inference for generalized exponential distribution based on records. *Communications in Statistics Theory and Methods*, 33(8), pp. 1851-1861.
16. Kim, C. & Song, S., 2010. Bayesian estimation of the parameters of the generalized exponential distribution from doubly censored sample. *Statistical Papers*, 51(3), pp. 583-597.
17. Kundu, D. & Gupta, R., 2008. Generalized exponential distribution: Bayesian estimations. *Computational Statistics and Data Analysis*, 52(4), pp. 1873-1883.
18. Kundu, D., Gupta, R. & Manglick, A., 2005. Discriminating between the log-normal and generalized exponential distributions. *Journal of Statistical Planning and Inference*, Volume 127, pp. 213-227.
19. Kundu, D. & Pardhan, B., 2008. Estimating the parameters of the generalized exponential distribution in presence of hybrid censoring. *Communications in Statistics Theory and Methods*.
20. McLachlan, G. J. & Peel, D., 2000. *Finite Mixture Models*. New York: John Wiley and Sons.
21. Mendenhall, W. & Hader, R. J., 1958. Estimation of parameters of mixed exponentially distribution failure time distributions from censored life test data. *Biometrika*, Volume 45, pp. 504-520.
22. Raqab, M., 2002. Inferences for generalized exponential distribution based on record statistics. *Journal of Statistical Planning and Inference*, 104(2), pp. 339-350.
23. Raqab, M. & Madi, M., 2005. Bayesian inference for the generalized exponential distribution. *Journal of Statistical Computation and Simulation*, 75(10), pp. 841-852.
24. Sarhan, A., 2007. Analysis of incomplete, censored data in competing risks models with generalized exponential distributions. *IEEE Transactions on Reliability*, Volume 56, pp. 132-138.
25. Titterginton, D. M., Smith, A. F. & Markov, U. E., 1985. *Statistical Analysis of Finite Mixture Distributions*. New York: John Wiley and Sons.
26. William, M. Q. & Luis, E. A., 1998. *Statistical Methods for Reliability Data*. s.l.:John Wiley and sons.

27. Yarmohammadi, M., 2010. Classical and Bayesian estimations on the generalized exponential distribution using censored data. *Int. Journal of Math Analysis*, 4(29), pp. 1417-1431.
28. Zheng, G., 2002. On the Fisher information matrix in type-II censored data from the exponentiated exponential family. *Biometrical Journal*, Volume 44, pp. 353-357.