Original research article

Point Estimation of Birnbuam-Saunders Distribution Using EM-Algorithm

Wikanda Phaphan*
Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bang Sue, Bangkok 10800, Thailand

Received 23 June 2016; Received in revised form 20 September 2016
Accepted 27 September 2016; Available online 24 March 2017

ABSTRACT

The Birnbaum-Sanders (BS) distribution was first introduced in 1969 by Birnbaum and Saunders as a combination of inverse Gaussian distributions with a length-biased inverse Gaussian distribution. Later, in 2008, Ahmed et al. introduced a new parametrization of the BS distribution based on Birnbaum-Sanders, and they also proposed a parameter estimation using the method of moments and regression-quantile estimation. In this paper, we emphasize the Birnbaum-Sanders distribution presented by Ahmed et al., and we develop an EM-algorithm to estimate two unknown parameters of this distribution. The EM-algorithm is a general method used to estimate the parameters when the probability density function is complicated and it is the best alternative for the estimation of a mixture distribution. We assumed that this problem has a missing value, and maximized complete data log-likelihood function instead log-likelihood function because it is analytically easier. Moreover, some simulation experiments were conducted in order to examine the performance of the proposed parameter estimation, and it was observed that the performances were quite satisfactory. Specifically, the MSE, variance and bias tend to decrease as n increases.

Keywords: Inverse Gaussian distribution; Length biased inverse Gaussian distribution; Maximum likelihood methods; Lifetime distribution; Parametrization

Introduction

The Birnbaum-Sanders distribution is a positively-skewed model, which was originally proposed by Birnbaum and Saunders [1] as a failure time distribution for fatigue failure caused under cyclic loading. The model was also established under the assumption that failure is due to the development and growth of a dominant crack. This distribution is the so- called two-parameter Birnbaum-Saunders distribution (herein after called the BS distribution). It is the mixture of the inverse Gaussian (IG) distribution and length-biased inverse Gaussian (LBIG) distribution with the weight parameter equal to 0.5. Birnbaum and Saunders [2] presented a theoretical and practical review of applying this distribution with fatigue data. Desmond [3, 4] proposed a more general derivation based on a biological model and strengthened the physical justification for the use of this distribution. Ahmed et al. [9] introduced a new parametrization of BS distribution based on that originally provided in Birnbaum and Saunders [1]. Essentially, this re-parametrization fits the physics of studying phenomena since the proposed parameters characterize the thickness of the sample and

***Corresponding author:** Wikanda.p@sci.kmutnb.ac.th

the nominal treatment loading on the sample, respectively. The usual shape and scale parameters of the distribution do not allow this physical interpretation. Ahmed et al. [9] also presented the relationship between the usual parameters and the proposed parameters. For addition details concerning the BS distribution refer to Cordeiro and Lemonte [14], Kundu et al. [13], Lisawadi [11], Balakrishnan et al. [10], Ng et al. [8].

The relevance of the probability density function (pdf.) of the distributions mentioned above is as follows. Suppose that X_1 and X_2 are the independent random variables such that $X_1 \sim IG(\lambda,\theta)$ i.e., X_1 has IG distribution with the parameter $\lambda,\theta > 0$ and it has pdf

$$f_{IG}(x_1;\lambda,\theta)=\begin{cases}\dfrac{\lambda}{\theta\sqrt{2\pi}}\left(\dfrac{\theta}{x_1}\right)^{\frac{3}{2}} e^{-\frac{1}{2}\left(\sqrt{\frac{x_1}{\theta}}-\lambda\sqrt{\frac{\theta}{x_1}}\right)^2} & ;x_1>0\\ 0 & ;\text{otherwise}\end{cases}$$

In addition, suppose $X_2 \sim LBIG(\lambda,\theta)$

The pdf of X_2 is given by:

$$f_{LBIG}(x_2;\lambda,\theta)=\begin{cases}\dfrac{1}{\theta\sqrt{2\pi}}\left(\dfrac{\theta}{x_2}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\left(\sqrt{\frac{x_2}{\theta}}-\lambda\sqrt{\frac{\theta}{x_2}}\right)^2} & ;x_2>0\\ 0 & ;\text{otherwise}\end{cases}$$

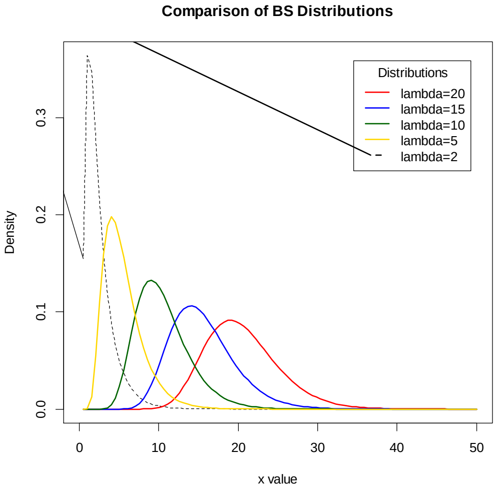

Fig. 1. The Birnbaum-Sanders density functions for $\theta = 1$ and different values of λ.

The variable X_2 is the so-called complementary reciprocal of X_1. For the BS distribution, we considered a new random variable X such that:

$$X=\begin{cases}X_1 \text{ with probability } 1/2\\ X_2 \text{ with probability } 1/2\end{cases}$$

Obviously, X is a mixture of X_1 and X_2 and the pdf of X is given by the following formula:

$$f_{BS}(x;\lambda,\theta)=\frac{1}{2}f_{IG}(x;\lambda,\theta)+\frac{1}{2}f_{LBIG}(x;\lambda,\theta)$$

The above equation can be expressed as

$$f_{BS}(x;\lambda,\theta)=\begin{cases}\dfrac{1}{2\theta\sqrt{2\pi}}\left[\lambda\left(\dfrac{\theta}{x}\right)^{\frac{3}{2}}+\left(\dfrac{\theta}{x}\right)^{\frac{1}{2}}\right]e^{-\frac{1}{2}\left(\sqrt{\frac{x}{\theta}}-\lambda\sqrt{\frac{\theta}{x}}\right)^2} & ;x>0\\ 0 & ;\text{otherwise}\end{cases}$$

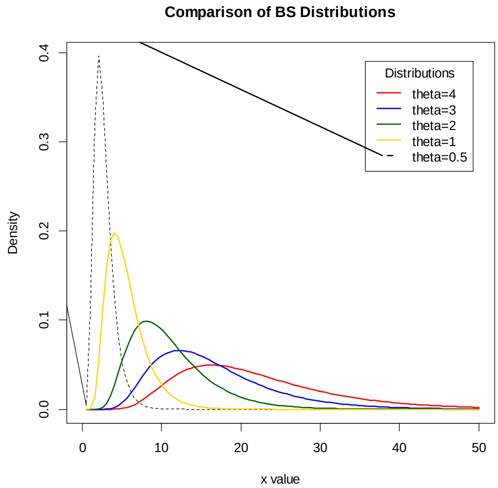

Fig. 2. The Birnbaum-Sanders density functions for $\lambda = 5$ and different values of θ.

It was observed that the pdf of X is complicated and intractable, so for this problem it is difficult to find the maximum-likelihood estimate of the unknown parameters. Thus, more elaborate techniques have to be considered. As the pdf. of the BS distribution can be written in a mixed form, the EM algorithm seems to be a natural choice for obtaining the maximum-likelihood estimate of its parameters. In this research, we proposed the use of an alternative technique to estimate the two unknown parameters; that is, the EM algorithm in the case of complete samples. Extensive simulation experiments were conducted to examine the performance of the proposed EM algorithm by using the R program.

EM Algorithm

Complete sample

Here we discuss how to investigate the MLEs of the two unknown parameters of the BS model using the EM algorithm based on a complete sample, i.e., $\{x_1, x_2, ..., x_n\}$. The log-likelihood function based on the observed sample can be expressed as:

$$l(\lambda,\theta \mid x_1,...,x_n) = \sum_{i=1}^{n} \ln\left\{\frac{1}{2} f_{IG}(x;\lambda,\theta) + \frac{1}{2} f_{LBIG}(x;\lambda,\theta)\right\}.$$

We assumed that this problem had a missing value $\{z_1, z_2, ..., z_n\}$, and the complete data set was as follows: $Y = (X, Z)$, where Z is an indicator variable with a value of 0 or 1. The random variable Z takes the value 0 or 1 depending on whether the observation X comes from X_1 or X_2, respectively. Now, we suppose X is the incomplete data. The complete data log-likelihood function of $\{y_1, y_2, ..., y_n\}$, where $y_i = (x_i, z_i)$ for $i = 1, 2, ..., n$, is

$$l_{complete}(\lambda,\theta \mid y_1,...,y_n) = \sum_{i=1}^{n}(1-z_i)\ln\left\{\frac{1}{2} f_{IG}(x;\lambda,\theta)\right\} + \sum_{i=1}^{n}(z_i)\ln\left\{\frac{1}{2} f_{LBIG}(x;\lambda,\theta)\right\}.$$

We simplified the above formula. Therefore, the complete data log-likelihood function without the additive constant can be written as:

$$l_{complete} = -\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{\frac{x_i}{\theta}} - \lambda\sqrt{\frac{\theta}{x_i}}\right)^2 + (1-\bar{z})n\ln\lambda + \left(\frac{1}{2}-\bar{z}\right)n\ln\theta,$$

where $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$. We can set the derivative of $l_{complete}$ to zero, and solve directly for λ so the MLEs of λ based on the complete sample, denoted by $\hat{\lambda}$, and it is as follows:

$$\hat{\lambda} = \frac{1+\sqrt{1+4\theta s_1(1-\bar{z})}}{2\theta s_1},$$

where $s_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}$. Here the MLEs of θ, denoted as $\hat{\theta}$, can be obtained by maximizing $g(\theta)$, where

$$g(\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{\frac{x_i}{\theta}} - \left(\frac{1+\sqrt{1+4\theta S_1(1-\bar{Z})}}{2S_1\sqrt{\theta}\sqrt{x_i}}\right)\right)^2 + \left(\frac{1}{2}-\bar{Z}\right)n\ln\theta.$$

At the E-step of the EM algorithm, the "pseudo" log-likelihood function was obtained be replacing the missing values by their expectation $E(Z_i)$; then the "pseudo" log-likelihood function at the k[th] replication becomes:

$$l_{pseudo}^{(k)} = -\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{\frac{x_i}{\theta}} - \lambda\sqrt{\frac{\theta}{x_i}}\right)^2 + \left(1-a^{(k)}\right)n\ln\lambda + \left(\frac{1}{2}-a^{(k)}\right)n\ln\theta,$$

here $a^{(k)} = \frac{1}{n}\sum_{i=1}^{n} a^{(k)}$ and $a_i^{(k)}$ is given by

$$a_i^{(k)} = E\left(Z_i \mid x_1,...,x_n,\lambda^{(k)},\theta^{(k)}\right) = \frac{\frac{1}{2} f_{LBIG}\left(x;\lambda^{(k)},\theta^{(k)}\right)}{\frac{1}{2} f_{IG}\left(x;\lambda^{(k)},\theta^{(k)}\right) + \frac{1}{2} f_{LBIG}\left(x;\lambda^{(k)},\theta^{(k)}\right)}.$$

In the M-step of the EM algorithm, we maximized the "pseudo" log-likelihood function with respect to λ and θ to obtain $\lambda^{(k+1)}$ and $\theta^{(k+1)}$, where

$$\lambda^{(k+1)} = \frac{1+\sqrt{1+4\theta^{(k+1)}s_1\left(1-a^{(k)}\right)}}{2\theta^{(k+1)}s_1},$$

and $\theta^{(k+1)}$ obtained by maximizing $g^{(k+1)}(\theta)$, where

$$g^{(k+1)}(\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{\frac{x_i}{\theta}} - \left(\frac{1+\sqrt{1+4\theta S_1\left(1-a^{(k)}\right)}}{2S_1\sqrt{\theta}\sqrt{x_i}}\right)\right)^2 + \left(\frac{1}{2}-a^{(k)}\right)n\ln\theta,$$

by using a numerical procedure, here we use the Newton-Raphson method in the R program. Each iteration is guaranteed to increase the value of $\hat{\lambda}$ and $\hat{\theta}$, and the process stops when convergence occurs. The algorithm is guaranteed to converge to a local maximum of the likelihood function.

Now we discuss how to choose the current parameter estimates for λ and θ. First we considered the MLEs of λ and θ, when $\{x_1, x_2, ..., x_n\}$ is a random sample of X_1 and they will be as follows:

$$\tilde{\lambda} = \frac{1+\sqrt{1+4\tilde{\theta}s_1}}{2\tilde{\theta}s_1}.$$

Here $\tilde{\theta}$ can be obtained by maximizing $A(\theta)$ with respect to θ by using the Newton-Raphson method in the R program, where:

$$A(\theta) = n\ln(1+\sqrt{1+4\theta s_1}) - n\ln(2\theta s_1) - n\ln\theta + \frac{3n}{2}\ln\theta - \frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{\frac{x_i}{\theta}} - \left[\frac{1+\sqrt{1+4\theta s_1}}{2s_1\sqrt{x_1\theta}}\right]\right)^2.$$

Similarly, if $\{x_1, x_2, ..., x_n\}$ is a random sample of X_2 the MLE of λ is

$$\tilde{\tilde{\lambda}} = \frac{1}{\tilde{\tilde{\theta}}s_1},$$

and the MLE of θ (denoted by $\tilde{\tilde{\theta}}$) can be obtained by maximizing $B(\theta)$ with respect to θ by using the Newton-Raphson method, where:

$$B(\theta) = -n\ln\theta + \frac{n}{2}\ln\theta - \frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{\frac{x_i}{\theta}} - \frac{1}{s_1\sqrt{\theta x_i}}\right)^2.$$

We suggest using the simple averages of the MLEs of IG and LBIG distribution as the current parameter estimates of λ and θ, i.e.:

$$\lambda^{(0)} = \frac{\tilde{\lambda}+\tilde{\tilde{\lambda}}}{2} \text{ and } \theta^{(0)} = \frac{\tilde{\theta}+\tilde{\tilde{\theta}}}{2}.$$

The following algorithm was used to find the MLEs of the unknown parameters of the BS distribution.

Algorithm:

Step 1: Generate a random sample $\{x_1, x_2, ..., x_n\}$ following the BS distribution.

Step 2: Compute $\lambda^{(0)}$ and $\theta^{(0)}$.

Step 3: Compute $a_i^{(0)}$ for $i = 1, 2, ..., n$ and $a^{(0)} = \frac{1}{n}\sum_{i=1}^{n} a_i^{(0)}$, where

$$a_i^{(0)} = \frac{\frac{1}{2}f_{LBIG}\left(x \mid \lambda^{(0)}, \theta^{(0)}\right)}{\frac{1}{2}f_{IG}\left(x \mid \lambda^{(0)}, \theta^{(0)}\right) + \frac{1}{2}f_{LBIG}\left(x \mid \lambda^{(0)}, \theta^{(0)}\right)},$$

n is the number of samples, and k is the number of iterations.

Step 4: Obtain $\theta^{(1)}$ by maximizing

$$g^{(1)}(\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{\frac{x_i}{\theta}} - \left(\frac{1+\sqrt{1+4\theta S_1\left(1-a^{(0)}\right)}}{2S_1\sqrt{\theta}\sqrt{x_i}}\right)\right)^2 + \left(\frac{1}{2} - a^{(0)}\right)n\ln\theta,$$

and

$$\lambda^{(1)} = \frac{1+\sqrt{1+4\theta^{(1)}s_1\left(1-a^{(0)}\right)}}{2\theta^{(1)}s_1}.$$

Step 5: Repeat Step 3 and Step 4, until convergence is achieved.

Monte Carlo Simulations

In this section, a Monte Carlo simulation study was conducted in order to appraise the performance of the proposed strategies for given sample sizes. All of the computational parts were run on R program version 3.2.0.

For the computer simulations we considered different sample sizes n; 10, 30, 50, 80, and different models; Model 1: $\lambda = 1$, $\theta = 3$, Model 2: $\lambda = 1$, $\theta = 4$, Model 3: $\lambda = 2$, $\theta = 1$, Model 4: $\lambda = 2$,

$\theta = 4$, Model 5: $\lambda = 3$, $\theta = 4$, and Model 6: $\lambda = 4$, $\theta = 4$. The number of iterations is fixed at 1,000 for each model.

Regarding the simulation results, it was observed that the proposed EM algorithm exhibited quite fast convergence. The results are summarized numerically in Tables 1-6. It is clear from these tables, that the performance of estimator of λ and θ was more accurate as n increased from $n = 10$ to $n = 80$ except for $\lambda = 2$ and $\theta = 1$ in which case the biases are all close to zero for $n \geq 30$. For example, in Table 1, for $\lambda = 1$, $\theta = 3$ and $n = 10$, the simulated biases of $\hat{\lambda}$ and $\hat{\theta}$ were 0.41878 and -0.56919 respectively, while for $n = 80$ the simulated biases of $\hat{\lambda}$ and $\hat{\theta}$ are 0.14964 and -0.43316 respectively. Similarly, the performance of the estimator of λ and θ was less accurate for $n = 30$ than for $n = 50$. For instance, in Table 6, for $\lambda = 4$, $\theta = 4$ and $n = 30$ the simulated biases of $\hat{\lambda}$ and $\hat{\theta}$ were 0.03547 and -0.00166 respectively, while for $n = 50$ the simulated biases of $\hat{\lambda}$ and $\hat{\theta}$ were 0.02019 and -0.00006 respectively.

It was observed that for almost all models, as the sample size increased the biases decreased and tend to zero. This examines the consistency properties of the maximum likelihood estimates. Similarly, the MSE is decreasing function of n and it approaches to zero as $n \to \infty$.

Table 1. The average estimates, bias, the mean squared errors, and variance of $\lambda = 1$ and $\theta = 3$.

n	$\lambda = 1$	$\theta = 3$
10	1.41878 (0.41878) (0.45150)(0.27612)	2.43081 (-0.56919) (1.19713)(0.87316)
30	1.20655 (0.20655) (0.10558)(0.06292)	2.52831 (-0.47169) (0.59358)(0.37109)
50	1.16171 (0.16171) (0.06443)(0.03827)	2.56980 (-0.43020) (0.41988)(0.23480)
80	1.14964 (0.14964) (0.04496)(0.02256)	2.56684 (-0.43316) (0.33635)(0.14873)

Table 2. The average estimates, bias, the mean squared errors, and variance of $\lambda = 1$ and $\theta = 4$.

n	$\lambda = 1$	$\theta = 4$
10	1.45804 (0.45804) (0.48383)(0.27403)	2.94080 (-1.05920) (1.96811)(0.84620)
30	1.23433 (0.23433) (0.11309)(0.05818)	3.23806 (-0.76194) (0.97893)(0.39837)
50	1.18158 (0.18158) (0.06625) (0.03328)	3.31567 (-0.68433) (0.72573) (0.25742)
80	1.15155 (0.15155) (0.04180)(0.01884)	3.38430 (-0.61570) (0.57314)(0.19405)

Table 3. The average estimates, bias, the mean squared errors, and variance of $\lambda = 2$ and $\theta = 1$.

n	$\lambda = 2$	$\theta = 1$
10	2.20920 (0.20920) (0.48144)(0.43768)	1.03321 (0.03321) (0.14091)(0.13981)
30	1.95373 (-0.04627) (0.11423)(0.11209)	1.07941 (0.07941) (0.05320)(0.04690)
50	1.91105 (-0.08895) (0.06825)(0.06033)	1.08555 (0.08555) (0.03388)(0.02656)
80	1.91058 (-0.08942) (0.05014)(0.04214)	1.07755 (0.07755) (0.02301)(0.01700)

Table 4. The average estimates, bias, the mean squared errors, and variance of $\lambda = 2$ and $\theta = 4$.

n	$\lambda = 2$	$\theta = 4$
10	2.42616 (0.42616) (0.52562)(0.34400)	3.47954 (-0.52046) (0.69914)(0.42827)
30	2.13451 (0.13451) (0.08978)(0.07169)	3.79258 (-0.20742) (0.16103)(0.11801)
50	2.06988 (0.06988) (0.04344)(0.03855)	3.89238 (-0.10762) (0.05728)(0.04569)
80	2.04368 (0.04368) (0.02384)(0.02193)	3.91308 (-0.08692) (0.04282)(0.03527)

Table 5. The average estimates, bias, the mean squared errors, and variance of $\lambda = 3$ and $\theta = 4$.

n	$\lambda = 3$	$\theta = 4$
10	3.24674 (0.24674) (0.36350)(0.30262)	3.81591 (-0.18409) (0.17415)(0.14026)
30	3.04985 (0.04985) (0.08161)(0.07912)	3.97267 (-0.02733) (0.01234)(0.01159)
50	3.02376 (0.02376) (0.05094)(0.05038)	3.99315 (-0.00685) (0.00271)(0.00266)
80	3.00537 (0.00537) (0.02924)(0.02922)	3.99887 (-0.00113) (0.00024)(0.00024)

Table 6. The average estimates, bias, the mean squared errors, and variance of $\lambda = 4$ and $\theta = 4$.

n	$\lambda = 4$	$\theta = 4$
10	4.16851 (0.16851) (0.36431)(0.33592)	3.95025(-0.04975) (0.04014)(0.03767)
30	4.03547 (0.03547) (0.11091)(0.10965)	3.99834(-0.00166) (0.00056)(0.00056)
50	4.02019 (0.02019) (0.06485)(0.06445)	3.99994(-0.00006) (0.00000)(0.00000)
80	4.01286 (0.01286) (0.04367)(0.04351)	3.99994(-0.00006) (0.00000)(0.00000)

Conclusion

In this paper we have considered the estimation procedure of a BS distribution given by Ahmed et al. [9]. We have proposed the use of the EM algorithm to estimate the two unknown BS parameters. If we want to find the MLEs of the BS distribution by solving the normal equations, we need to solve two non-linear equations simultaneously. However, in the proposed EM algorithm, the proposed estimators were analytically easier to compute estimates. Therefore, in this case, the process of the EM algorithm was quite simple. The study found that the performance of the presented EM algorithm was very satisfying. Specifically, the MSE, variance and bias tend to zero as $n \to \infty$.

Acknowledgements

I would like to thank our “army” of undergraduate research assistants, Kusuma Sengsui and Kitterut Thongaiem.

References

[1] Birnbaum ZW, Saunders SC. A new family of lifetime distribution. Appl Probab 1969;6:319-27.

[2] Birnbaum ZW, Saunders SC. Estimation for a family of life distributions with applications to fatigue. Appl Probab 1969;6:328-47.

[3] Desmond AF. Stochastic models of failure in random environments. Canad J Statist 1985;13:171-83.

[4] Desmond AF. On the relationship between two fatigue-life models. IEEE Trans Rel 1986;35:167-9.

[5] Gupta RC, Akman HO. Bayes estimation in a mixture inverse Gaussian model. Ann Inst Statist Math 1995;47(3):493-503.

[6] Bilmes JA. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. International Computer Science Institute, Berkeley; 1998.

[7] Jorgensen B, Seshadri V, Whitmore GA. On the mixture of the inverse Gaussian distribution with its complementary reciprocal. Scand J Statist 1991;18(1):77-89.

[8] Ng HKT, Kundu D, Balakrishnan N. Point and interval estimation for the two-parameter Birnbaum-Saunders distribution based on type-II censored samples. Comput Statist Data Anal 2007;51(8).

[9] Ahmed SE, Budsaba K, Lisawadi S, Volodin A. Parametric estimation for the Birnbaum-Saunders lifetime distribution based on a new parametric zation. Thailand Statistician 2008;6(2):213-40.

[10] Balakrishnan N, Leiva V, Sanhueza A, Vilca F. Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. SORT 2009;33(2):171-192.

[11] Lisawadi S. Parameter estimation for the two-sided BS and IG distribution: and laws of large numbers for arrays under a condition of uniform integrability. 1st ed. VDM Verlag, 2009.

[12] Robert CP, Casella G. Introducing Monte Carlo methods with R. London: Springer; 2010.

[13] Kundu D, Balakrishnan N, Jamalizadeh A. Bivariate Birnbaum-Saunders distribution and associated inference. J Multivar Anal 2010;101: 113-125.

[14] Cordeiro G, Lemonte A. The β Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling. Comput Statist Data Anal 2011;55:1445-61.