Weighted Inverse Gaussian - a Versatile Lifetime Model

Ramesh C. Gupta[1] & Debasis Kundu[2]

Abstract

Jorgensen, Seshadri and Whitmore (1991, Scandinavian Journal of Statistics, 77 - 89) introduced a three-parameter generalized inverse Gaussian distribution, which is a mixture of the inverse Gaussian distribution and length biased inverse Gaussian distribution. Also Birnbaum-Saunders distribution is a special case for $p = 1/2$, where p is the mixing parameter. It is observed that the estimators of the unknown parameters can be obtained by solving a three-dimensional optimization process, which may not be a trivial issue. Most of the iterative algorithms are quite sensitive to the initial guesses. In this paper we propose to use the EM algorithm to estimate the unknown parameters for complete and censored samples. In the proposed EM algorithm, at the M-step the optimization problem can be solved analytically, and the observed Fisher information matrix can be obtained. These can be used to construct asymptotic confidence intervals of the unknown parameters. Some simulation experiments are conducted to examine the performance of the proposed EM algorithm, and it is observed that the performances are quite satisfactory. The methodology proposed here is illustrated by three data sets.

Key Words and Phrases: Inverse Gaussian distribution; length biased inverse Gaussian distribution; Birnbaum-Saunders distribution; EM algorithm, Fisher information; maximum likelihood estimators; censored samples.

[1] Department of Mathematics and Statistics, University of Maine, Orono, ME 04469 - 5752, USA.

[2] Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. E-mail: kundu@iitk.ac.in, Corresponding author.

1 Introduction

Two-parameter inverse Gaussian (IG) distribution is a right skewed distribution and it plays an important role in reliability analysis. Because of its flexibility and several other interesting properties, it has always been a popular alternative to the Weibull, log-normal, gamma and other similar skewed distributions. Due to these properties, considerable efforts have been made by different authors in developing different aspects of this distribution. A comprehensive review on the IG distribution and for its different applications the readers are referred to the excellent monograph by Chhikara and Folks [7] and the references cited therein.

Along with the IG distribution, its complementary reciprocal, also known as the length biased inverse Gaussian (LBIG) distribution has received considerable attention due to its various applications in different biomedical areas, such as family history of diseases, early detection of diseases, latency periods of AIDS etc, see for example Akman and Gupta [1] and Gupta and Akman [11] in this respect.

Jorgensen, Seshadri and Whitmore [15] introduced a new three-parameter generalized IG distribution as follows. Suppose $X_1 \sim \text{IG}(\mu, \lambda)$, *i.e.* X_1 has a IG distribution with the parameters $\mu > 0$ and $\lambda > 0$ and it has the probability density function (PDF)

$$f_{X_1}(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}}; \qquad x > 0. \tag{1}$$

Moreover, suppose X_2 is a random variable such that $X_2^{-1} \sim \text{IG}(\mu^{-1}, \lambda/\mu^2)$, and it is independent of X_1. Then consider the new random variable X such that, for $0 \le p \le 1$,

$$X = \begin{cases} X_1 & \text{with probability} \quad 1-p \\ & \\ X_2 & \text{with probability} \quad p. \end{cases} \tag{2}$$

Clearly, X is a mixture of X_1 and X_2 and the PDF of X is

$$f_X(x; \mu, \lambda, p) = (1-p) f_{X_1}(x; \mu, \lambda) + p f_{X_2}(x; \mu, \lambda), \tag{3}$$

where $f_{X_1}(x;\mu,\lambda)$ and $f_{X_2}(x;\mu,\lambda)$ are the PDFs of X_1 and X_2 respectively. It may be observed that $f_{X_2}(x)$ can be written in the following form;

$$f_{X_2}(x;\mu,\lambda) = x f_{X_1}(x;\mu,\lambda)/\mu. \tag{4}$$

From now on we call the distribution of X as defined in (3), as the JSW distribution with parameters μ, λ, p, and it will be denoted by JSW(μ, λ, p). The three-parameter JSW distribution has several interesting properties and different real life applications, see for example Jorgensen, Seshadri and Whitmore [15]. This model has been studied quite intensely by Gupta and Akman [11, 12] also, and a brief review of this mixture distribution will be presented in the next section.

Although, several interesting properties of JSW distribution have been discussed in the literature, but estimation problem of the unknown parameters still remains an open issue. It is observed that the maximum likelihood estimators (MLEs) can be obtained by solving a three dimensional optimization problem, but it has been mentioned by Jorgensen, Seshadri and Whitmore [15] and also by Gupta and Akman [11] that finding the efficient initial guesses and solving three non-linear equations simultaneously are non-trivial issues.

The main aim of this paper is to develop the expectation-maximization (EM) algorithm to find the MLEs of the unknown parameters in case of complete as well as censored samples. As the PDF of JSW distribution can be written in a mixture form, the EM algorithm seems to be a natural choice for obtaining MLEs of its parameters. It is observed that for each E-step, the M-step can be obtained explicitly. Therefore, the implementation of the proposed EM algorithm is very simple. Moreover, using the idea of Louis [17], from the last step of the EM algorithm we obtain the observed Fisher information matrix, which can be used to construct the asymptotic confidence intervals of the unknown parameters and perform certain testing of hypotheses. Extensive simulation experiments are conducted to investigate the behavior of the proposed EM algorithm and the performances are quite satisfactory. The methodology proposed here is illustrated by three data sets- one generated and the other two drawn from literature.

The rest of the paper is organized as follows. Section 2 briefly describes the JSW distribution. Section 3 gives the EM algorithm. Sections 4 and 5, respectively, contain simulation and data analysis results. Finally, Section 6 contains conclusions.

2 A Brief Review JSW Model

The three-parameter JSW model, as defined in (3), has the cumulative distribution function (CDF);

$$F_X(x;\mu,\lambda) = \Phi(\alpha(x)) + (1-2p)e^{\frac{2\lambda}{\mu}}\Phi(\beta(x)), \tag{5}$$

where $\alpha(x) = \dfrac{\lambda^{\frac{1}{2}}(x-\mu)}{\mu x^{\frac{1}{2}}}$, $\beta(x) = -\dfrac{\lambda^{\frac{1}{2}}(x+\mu)}{\mu x^{\frac{1}{2}}}$ and $\Phi(\cdot)$ denotes the CDF of a standard normal distribution. Moreover, the PDF of X as provided in (3) can also be written as

$$f_X(x;\mu,\lambda,p) = f_{X_1;\mu,\lambda}(x)w(x), \tag{6}$$

where $w(x) = \left(\dfrac{\gamma+x}{\gamma+\mu}\right)$, and $\gamma \geq 0$ is a new parameter defined by the relation $p = \mu/(\mu+\gamma)$. Therefore, the PDF of X can be observed as a weighted IG distribution with the weight function as $w(x)$.

From (3) it is clear that the IG distribution and its complementary reciprocal can be obtained as special cases of (3). Interestingly, Jorgensen, Seshadri and Whitmore [15] observed that the Birnbaum-Saunders distribution, as proposed by Birnbaum and Saunders [3], can be obtained as a special case of JSW distribution when $p = 1/2$. It may be mentioned that the Birnbaum-Saunders distribution is a versatile distribution in its own right and has been used in fatigue crack growth and reliability, see Birnbaum and Saunders [3], Englehardt *et al.* [10], Chang and Tang [5, 6], Rieck [24] and Desmond and Yang [8]. More recently for the censored samples, special attention has been paid in estimating the unknown parameters of the Birnbaum-Saunders model, see for example Ng, Kundu and Balakrishnan [21, 22]. Therefore, the methodology developed here can be used for Birnbaum-Saunders distribution also.

It is observed that the shape of the PDF of JSW distribution is always unimodal. Gupta and Akman [11] showed that the shape of the hazard function of the JSW distribution is also unimodal. In this respect the JSW model is quite similar to the log-normal model. Gupta and Akman [12] also investigated the change point of the hazard function in great detail. Unimodality of the hazard function of the JSW distribution immediately implies that the hazard function of the Birnbaum-Saunders distribution is also unimodal. It may be mentioned here that Kundu, Kannan and Balakrishnan [16] also studied independently the shape of the hazard function of the Birnbaum-Saunders distribution and provided some approximation of the change point of its hazard function.

The JSW model is an exponential dispersion model in the sense of Jorgensen [14]. Therefore many key properties of the exponential dispersion model will be directly applicable here. For different reliability properties of the JSW model, the readers are referred to Gupta and Akman [11].

3 EM Algorithm

3.1 Complete Sample

In this subsection we discuss how to compute the MLEs of the unknown parameters of the JSW model using EM algorithm based on a complete sample say $\{x_1, \cdots, x_n\}$. The log-likelihood function based on the observed sample can be written as;

$$l(\mu, \lambda, p|x_1 \cdots, x_n) = \sum_{i=1}^{n} \ln \left\{ (1-p) f_{X_1}(x_i; \mu, \lambda) + p f_{X_2}(x_i; \mu, \lambda) \right\}. \tag{7}$$

It is observed in Jorgensen, Seshadri and Whitmore [15] that the normal equations obtained by differentiating (7) with respect to μ, λ, p do not have explicit solutions. They have to be solved iteratively.

We treat this problem as a missing value problem. It is assumed that the complete observation is as follows; $Y = (X, Z)$, where Z is an indicator variable taking value 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.

Based on the complete sample $\{y_1, \cdots, y_n\}$, where $y_i = (x_i, z_i)$ for $i = 1 \cdots, n$, the complete data log-likelihood function is

$$l_{complete}(\mu, \lambda, p|y_1, \cdots, y_n) = \sum_{i=1}^{n} (1-z_i) \ln \left\{ (1-p) f_{X_1}(x_i; \mu, \lambda) \right\} + \sum_{i=1}^{n} z_i \ln \left\{ p f_{X_2}(x_i; \mu, \lambda) \right\}. \tag{8}$$

First we compute the MLEs of μ, λ, p, based on complete data. Note that without the additive constant, (8) can be written as

$$l_{complete}(\mu, \lambda, p|y_1, \cdots, y_n) = n(1-\bar{z}) \ln(1-p) + n\bar{z} \ln p + \frac{n}{2} \ln \lambda - n\bar{z} \ln \mu - \frac{\lambda}{2\mu^2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{x_i}, \tag{9}$$

where $\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$. The MLEs of μ, λ and p based on the complete sample, will be denoted by $\widehat{\mu}, \widehat{\lambda}$ and $\widehat{p}$, respectively and can be obtained by maximizing (9) and they are as follows;

$$\widehat{p} = \bar{z} \qquad \text{and} \quad \widehat{\lambda} = \frac{\widehat{\mu}^2}{\frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \widehat{\mu})^2}{x_i}}. \tag{10}$$

Here $\widehat{\mu}$ can be obtained by maximizing $g(\mu)$, where

$$g(\mu) = -\frac{n}{2} \ln \left[\frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{x_i} \right] + n(1-\bar{z}) \ln \mu. \tag{11}$$

To compute $\widehat{\mu}$, we differentiate with respect μ and equate it to 0. It is a quadratic equation and it can be easily shown that it has one positive and one negative root. The positive root becomes the MLE of μ, and it is

$$\widehat{\mu} = \frac{(2\bar{z}-1) + \sqrt{(1-2\bar{z})^2 + 4\bar{x}(1-\bar{z})\bar{z}s_1}}{2\bar{z}s_1} \tag{12}$$

here $s_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Now we are in a position to provide the EM algorithm. Note that in the E-step of the EM algorithm, the 'pseudo' log-likelihood function is obtained be replacing the missing values by their expectation. In our case Z_i's are missing, therefore the 'pseudo' log-likelihood function can be obtained from (9) by replacing z_i with $E(Z_i)$.

Therefore, if at the k-th stage the estimates of μ, λ and p are $\mu^{(k)}, \lambda^{(k)}$ and $p^{(k)}$ respectively, then the 'pseudo' log-likelihood function at the k-th stage becomes;

$$l^{(k)}_{pseudo}(\mu,\lambda,p|x_1,\cdots,x_n) = n(1-a^{(k)})\ln(1-p) + na^{(k)}\ln p + \frac{n}{2}\ln\lambda - na^{(k)}\ln\mu - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}, \tag{13}$$

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

$$a_i^{(k)} = E(Z_i|x_1,\cdots,x_n,\mu^{(k)},\lambda^{(k)}p^{(k)}) = \frac{p^{(k)} f_{X_2}(x_i;\mu^{(k)},\lambda^{(k)})}{(1-p^{(k)})f_{X_1}(x_i;\mu^{(k)},\lambda^{(k)}) + p^{(k)} f_{X_2}(x_i;\mu^{(k)},\lambda^{(k)})}, \tag{14}$$

see for example McLachlan and Krishnan [18]. Now at the M-step we maximize (13) with respect to μ, λ and p to obtain $\mu^{(k+1)}, \lambda^{(k+1)}$ and $p^{(k+1)}$. They will be as follows;

$$p^{(k+1)} = a^{(k)} \quad \text{and} \quad \lambda^{(k+1)} = \frac{(\mu^{(k+1)})^2}{\frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\mu^{(k+1)})^2}{x_i}}. \tag{15}$$

Here $\mu^{(k+1)}$ is obtained by maximizing $g^{(k+1)}(\mu)$ with respect to μ, where $g^{(k+1)}(\mu)$ is obtained from (11) by replacing $\bar{z}$ with $a^{(k)}$. Moreover, $\mu^{(k+1)}$ can be obtained explicitly from (12) by replacing $\bar{z}$ with $a^{(k)}$. The process should be continued until the convergence occurs.

Now we discuss how to choose the initial guesses mainly for μ and λ. Usually the initial guess of p is chosen to be $1/2$, unless some other information is available. The following two facts will be useful to find the initial guesses of μ and λ. Note that if $\{x_1, \cdots, x_n\}$ is a random sample from X_1, then the MLEs of μ and λ are

$$\bar{x} \quad \text{and} \quad \left(\frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{x_i \bar{x}^2} \right)^{-1} \tag{16}$$

respectively. Here $\bar{x}$ is the same as defined before. Similarly, if $\{x_1, \cdots, x_n\}$ is a random sample from X_2, then the MLEs of μ and λ are

$$\frac{1}{s_1} \quad \text{and} \quad \left(\frac{1}{n} \sum_{i=1}^{n} \left(x_i^{-1} - s_1 \right)^2 x_i \right)^{-1}, \tag{17}$$

respectively. Here s_1 is the same as defined before. Therefore, we suggest to use the simple averages of theses estimates as initial guesses of μ and λ, *i.e.*

$$\mu^{(0)} = \frac{1}{2} \left[\bar{x} + \frac{1}{s_1} \right] \text{ and } \lambda^{(0)} = \frac{1}{2} \left[\left(\frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{x_i \bar{x}^2} \right)^{-1} + \left(\frac{1}{n} \sum_{i=1}^{n} \left(x_i^{-1} - s_1 \right)^2 x_i \right)^{-1} \right]. \tag{18}$$

Therefore we can use the following algorithm to find the MLEs of the unknown parameters of the JSW model.

ALGORITHM:

- Step 1: From the given sample $\{x_1, \cdots, x_n\}$ find $\mu^{(0)}$ and $\lambda^{(0)}$ as given in (18), and assume $p^{(0)} = \frac{1}{2}$.
- Step 2: Compute $a_i^{(k)}$ for $i = 1 \cdots, n$ using (14), and compute $a^{(k)} = \frac{1}{n} \sum_{i=1}^{n} a_i^{(k)}$.
- Step 3: Obtain

$$\mu^{(1)} = \frac{(2a^{(1)} - 1) + \sqrt{(1 - 2a^{(1)})^2 + 4\bar{x}(1 - a^{(1)}) a^{(1)} s_1}}{2a^{(1)} s_1},$$

and $p^{(1)}$, $\lambda^{(1)}$ as given in (15).

- Step 4: Repeat Step 2 and Step 3, until convergence is met.

The above EM algorithm can be easily applied to find the MLEs of the unknown parameters of the two-parameter Birnbaum-Saunders distribution. Since $p = 1/2$, we do not need to estimate p at each stage. $\mu^{(k)}$ and $\lambda^{(k)}$ are the same as before, and $a_i^{(k)}$'s can be obtained from (14) by replacing $p^{(k)} = 1/2$.

3.2 Censored Sample

In this subsection we discuss the implementation of the EM algorithm based on a Type-II sample, namely $\{x_{(1)}, \cdots, x_{(r)}\}$. Here $r < n$, and $x_{(1)}, \cdots, x_{(r)}$, denote the first r failure times when n items are put on a test and the experiment stops at $x_{(r)}$.

In this case we use the same procedure as the complete sample case by replacing each $x_{(r+1)}, \cdots, x_{(n)}$ with $b^{(k)} = b(\mu^{(k-1)}, \lambda^{(k-1)}, p^{(k-1)})$, at the k-th step, where

$$b(\mu, \lambda, p) = E(X|X > x_{(r)}), \tag{19}$$

and X has the JSW distribution with parameters μ, λ, p. Using Gupta and Akman [11], $b(\cdot)$ can be written as

$$\begin{aligned} b(\mu, \lambda, p) \;=\; & x_{(r)} + \left[\left(\mu - x_{(r)} + \frac{p\mu^2}{\lambda}\right)\Phi(-\alpha(x_{(r)})) + (1-2p)\left(\mu + x_{(r)} - \frac{p\mu^2}{\lambda}\right)e^{\frac{2\lambda}{\mu}}\Phi(\beta(x_{(r)}))\right. \\ & \left. + \frac{2p}{\sqrt{2\pi}}\sqrt{\frac{x_{(r)}\mu^2}{\lambda}}\; e^{-\frac{\lambda}{2x_{(r)}}\left(1-\frac{x_{(r)}}{\mu}\right)^2}\right]\left[\Phi(-\alpha(x_{(r)})) - (1-2p)e^{\frac{2\lambda}{\mu}}\Phi(\beta(x_{(r)}))\right]^{-1}. \end{aligned} \tag{20}$$

Here $\alpha(\cdot)$ and $\beta(\cdot)$ are the same as defined before.

Note that similar procedure will work for Type-I censored sample also. Suppose the experiment stops at the time point T and we observe m failures, say at $x_{(1)}, \cdots, x_{(m)}$, before time point T. In this $b(\mu, \lambda, p)$ can be obtained from (20) by replacing $x_{(r)}$ with T.

4 Monte Carlo Simulations

In this section we present some simulation results to verify how the proposed EM algorithm works in this case. All the experiments are performed at IIT Kanpur using Pentium IV dual core processor. We have used the random number generator RAN2 of Press *et al.* [23] and all the programs are written in FORTRAN-77.

We have taken different sample sizes; 10, 20, 40, 80, and different models; Model 1: $p = 0.5$, $\mu = 1.0$, $\lambda = 2.0$, Model 2: $p = 0.25$, $\mu = 1.0$, $\lambda = 2.0$ Model 3: $p = 0.75$, $\mu = 1.0$, $\lambda = 2.0$, Model 4: $p = 0.5$, $\mu = 2.0$, $\lambda = 2.0$. In each case we have first generated the inverse Gaussian random variable using the method proposed by Michael *et al.* [19], and then using the representation (2), a random sample from JSW(μ, λ, p) can be easily obtained. In each case we have computed the estimates of p, λ and μ using the EM algorithm proposed in the previous section and obtained the average estimates and square root of the mean squared errors over 1000 replications. For comparison purposes we have also provided the Cramer-Rao lower bound also. The results are reported in Tables 1 - 4. We further conduct a small simulation study for Type-I censored sample. In this case we have taken $n = 20$ and different censoring times T. Three different T values are chosen such that approximately 50%, 25% and 10% data are censored. We report the average estimates and the square root of the mean squared errors for Model 1. For other models the results are similar in nature, so they are not provided. The results are presented in Table 5.

It is observed that the proposed EM algorithm is working quite well. For each model as the sample size increases the average biases and the mean squared errors decrease. This verifies the consistency properties of the maximum likelihood estimates. Comparing Tables 1 - 3, it is observed that for fixed μ and λ as p changes, the average estimates and the mean squared errors of μ and λ do not change much at least for moderate or large sample sizes. It

$n \downarrow$	$p = 0.5$	$\mu = 1$	$\lambda = 2$
10	0.4907 (0.1987) (0.2124)	1.0284 (0.2424) (0.2528)	2.9802 (1.0112) (1.1034)
20	0.4954 (0.1698) (0.1736)	1.0109 (0.1659) (0.1788)	2.3690 (0.8201) (0.7802)
40	0.4969 (0.1198) (0.1227)	1.0026 (0.1207) (0.1264)	2.1630 (0.5464) (0.5517)
80	0.5011 (0.0889) (0.0868)	0.9993 (0.0836) (0.0894)	2.0763 (0.3897) (0.3901)

Table 1: The average estimates, square root of the mean squared errors and the associated Cramer-Rao lower bound of p, λ and μ for Model 1. In each box, corresponding to the sample size, the first figure represents the average estimate, the corresponding square root of the mean squares error is reported within bracket and the associated Cramer-Rao lower bound is reported within bracket below.

$n \downarrow$	$p = 0.25$	$\mu = 1$	$\lambda = 2$
10	0.2655 (0.1969) (0.2107)	1.0076 (0.2339) (0.2409)	2.8517 (0.9981) (1.0442)
20	0.2614 (0.1399) (0.1440)	1.0009 (0.1657) (0.1704)	2.3350 (0.7192) (0.7384)
40	0.2597 (0.0910) (0.1054)	0.9929 (0.1198) (0.1205)	2.1492 (0.5169) (0.5221)
80	0.2524 (0.0723) (0.0745)	0.9978 (0.0845) (0.0852)	2.0695 (0.3684) (0.3692)

Table 2: The average estimates, square root of the mean squared errors and the associated Cramer-Rao lower bound of p, λ and μ for Model 2. In each box, corresponding to the sample size, the first figure represents the average estimate, the corresponding square root of the mean squares error is reported within bracket and the associated Cramer-Rao lower bound is reported within bracket below.

$n \downarrow$	$p = 0.75$	$\mu = 1$	$\lambda = 2$
10	0.7250 (0.2134) (0.2327)	1.0407 (0.2474) (0.2520)	3.0567 (1.2101) (1.2447)
20	0.7290 (0.1593) (0.1646)	1.0271 (0.1696) (0.1732)	2.4608 (0.9012) (0.8802)
40	0.7337 (0.1021) (0.1163)	1.0140 (0.1198) (0.1260)	2.2186 (0.4919) (0.5082)
80	0.7447 (0.0803) (0.0823)	1.0088 (0.0847) (0.0891)	2.1093 (0.4388) (0.4401)

Table 3: The average estimates, square root of the mean squared errors and the associated Cramer-Rao lower bound of p, λ and μ for Model 3. In each box, corresponding to the sample size, the first figure represents the average estimate, the corresponding square root of the mean squares error is reported within bracket and the associated Cramer-Rao lower bound is reported within bracket below.

$n \downarrow$	$p = 0.5$	$\mu = 2$	$\lambda = 2$
10	0.4900 (0.2224) (0.2327)	2.1101 (0.6314)	3.0854 (1.1129) (1.2343)
20	0.4969 (0.1587) (0.1646)	2.0451 (0.4441) (0.4341)	2.3998 (0.8594) (0.8728)
40	0.4950 (0.0989) (0.1163))	2.0160 (0.3180) (0.3310)	2.1766 (0.6098) (0.6172)
80	0.5004 (0.0879) (0.0884)	2.0039 (0.2352) (0.2398)	2.0801 (0.4327) (0.4364)

Table 4: The average estimates, square root of the mean squared errors and the associated Cramer-Rao lower bound of p, λ and μ for Model 4. In each box, corresponding to the sample size, the first figure represents the average estimate, the corresponding square root of the mean squares error is reported within bracket and the associated Cramer-Rao lower bound is reported within bracket below.

$T \downarrow$	$p = 0.5$	$\mu = 1$	$\lambda = 2$
1	0.4016 (0.2518)	1.2258 (0.2247)	2.9802 (1.3989)
1.5	0.4123 (0.2112)	1.1965 (0.1992)	2.8196 (1.1614)
2.0	0.4823 (0.1795)	1.1123 (0.1711)	2.3981 (0.9549)

Table 5: The average estimates and square root of the mean squared errors of p, λ and μ for Model 1, when the data are Type-I censored. In each box, corresponding to the sample size, the first figures represents the average estimates and the corresponding square root of the mean squares errors are reported in bracket below.

may be due to the fact that both X_1 and X_2 contribute in estimating μ and λ. It seems that the average biases and the MSEs of the MLEs of p is maximum when $p = \frac{1}{2}$. Comparing Table 1 and Table 4, it is clear that if μ changes the average estimates and the mean squared errors of the MLEs of λ do not change. Moreover, the average biases of the MLEs of p and μ also do not change, but the square root of the mean squared errors of both p and μ increase two fold as μ changes from 1 to 2. It is also observed (not reported here) that for fixed p and μ if λ is doubled, then the average biases of the MLEs of all the parameters and the mean squared errors of the MLEs of μ do not change. The square root of the mean squared errors of p become half and the square root of the mean squared errors of λ become doubled. In Table 5 it is observed that when $T = 2$ (95% data are observed) the results are quite similar with the corresponding complete sample results. As the censoring percentage increases the biases and the mean squared errors increase.

Another important point which has been observed that in the convergence of the EM algorithm, the choice of the initial estimate of p plays an important role. In all our simulation experiments, we use different initial guesses of p and choose that estimate which provides the maximum log-likelihood value. Since the range of p is bounded, it is not a problem. The details will be explained in the next section.

Table 6: Simulated Data set.

0.87	1.11	1.22	1.27	1.29
1.71	2.06	2.10	2.24	2.59
2.80	3.68	4.63	4.82	4.89
5.34	5.80	7.39	9.35	9.61
12.44	12.74	15.51	15.71	23.64

5 Illustrative Examples

In this section we present three illustrative data analysis. (i) Simulated Data Set, (ii) Guinea Pigs Data, (iii) Guinea Pigs Data: Censored. The first two are complete data sets whereas the third one is Type-II censored data.

5.1 Simulated Data Set:

We have generated the data from the model (1) with $p = \frac{1}{2}$, $\mu = 5.0$ and $\lambda = 5.0$. The generated data are presented in Table 6.

Based on (18), we obtain the initial guesses of μ and λ as $\mu^{(0)} = 4.4693$ and $\lambda^{(0)} = 3.5614$ respectively. We use different initial guesses of $p \in (0, 1)$ and then use the proposed EM algorithm to compute the MLEs of the unknown parameters. Finally, we choose those estimates that provide the maximum log-likelihood value. In Figure 1 we provide the plot of the maximum log-likelihood value obtained, as a function of the different initial estimate of p. We obtain the MLEs of p, μ and λ as $\widehat{p} = 0.3847$, $\widehat{\mu} = 4.4568$ and $\widehat{\lambda} = 4.4024$ respectively, and the corresponding 95% confidence intervals obtained using the observed Fisher information matrix are (0.0000, 0.7819), (3.6797, 5.2339) and (1.8914, 6.9134) respectively.

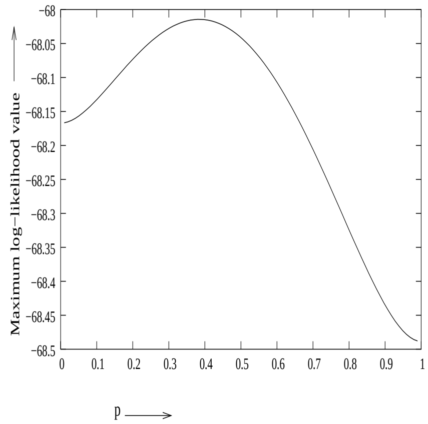


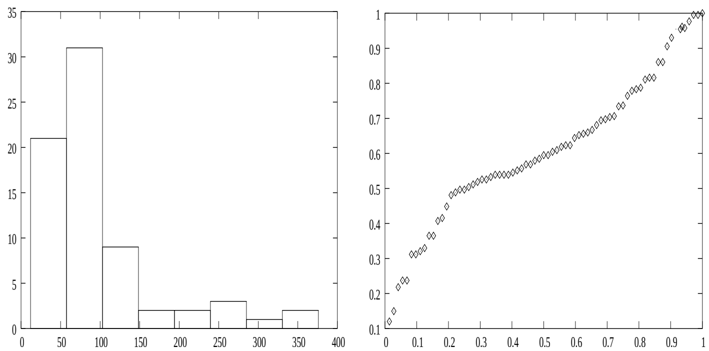
Figure 1: Maximum log-likelihood value as a function of different initial estimates of p.

Table 7: Guinea Pigs Data.

12	15	22	24	24	32	32	33	34
38	38	43	44	48	52	53	54	54
55	56	57	58	58	59	60	60	60
60	61	62	63	65	65	67	68	70
70	72	73	75	76	76	81	83	84
85	87	91	95	96	98	99	109	110
121	127	129	131	143	146	146	175	175
211	233	258	258	263	297	341	341	376

5.2 Guinea Pigs Data

This data set represents the survival times of guinea pigs injected with different doses of tubercle bacilli. The data set was first analyzed by Bjerkedal [4], later by Gupta *et al.* [13], and also by Kundu *et al.* [16]. Typically guinea pigs are chosen for tuberculosis experiments because of their high susceptibility. The 72 observations are presented in Table 7. The mean, standard deviation and the coefficient of skewness are calculated as 99.82, 80.55 and 1.80, respectively. The skewness measure indicates that the data are positively skewed. We have plotted the histogram and the scaled TTT transform of Aarset [2]. The histogram clearly indicates that the data are right skewed. The scaled TTT transform is first concave and then



(a) (b)

Figure 2: Guinea Pig Data: (a) Histogram (b) Scaled-TTT Plot.

convex, therefore it signifies that the hazard function should be unimodal. This justifies the use of the JSW model for analyzing this data set. It may be mentioned that Gupta *et al.* [13] used the log-normal distribution and Kundu *et al.* [16] used Birnbaum-Saunders distribution to analyze this data set and both these distributions have unimodal hazard functions.

In this case we fit the JSW model. We start with the initial values as $\mu^{(0)} = 79.9584$ and $\lambda^{(0)} = 120.9735$, and we use different $p^{(0)}$ and we obtain the MLEs of p, μ and λ as $\widehat{p} = 0.0331$, $\widehat{\mu} = 97.7250$, $\widehat{\lambda} = 150.9174$, and the corresponding log-likelihood value is -390.7221. The associated 95% confidence intervals of p, μ and λ become (0.0000, 0.0759), (88.6165, 106.8333) and (101.6089, 200.2259) respectively. The reliability function and the associated 95% confidence bounds are provided in Figure 3.

When we use the Birnbaum-Saunders distribution, we obtain the estimates of μ and λ as $\widehat{\mu} = 77.5348$ and $\widehat{\lambda} = 134.2443$ respectively and the associated log-likelihood value becomes -390.9173. Therefore, if we want to test whether the data are coming from Birnbaum-Saunders distribution or not, *i.e.*

$$H_0 : \text{Birnbaum-Saunders distribution} \quad vs. \quad H_1 : \text{JSW distribution},$$

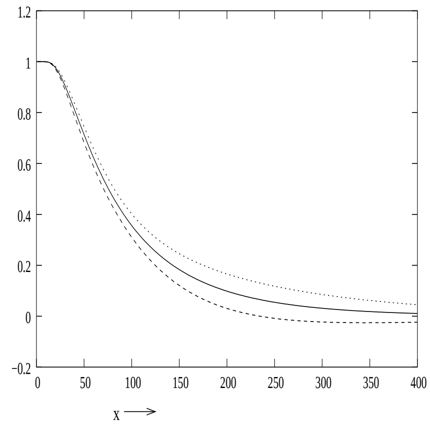


Figure 3: Reliability function and the associated 95% confidence bands for Guinea Pigs data.

then we cannot reject the null hypothesis based on the likelihood ratio test. Interestingly, if we consider the 95% confidence interval for p, we reject the null hypothesis with 5% level of significance. To check the goodness-of-fit we plot the empirical survival function, and the best fitted survival functions based on JSW and Birnbaum-Saunders distributions and they are provided in Figure 4. It is immediate that the two fitted survival functions are almost identical, and that is reflected in their log-likelihood values also. Furthermore, the Kolmogorov- Smirnov (KS) distance between the empirical survival function and the fitted survival function is 0.104 with the associated p-value, being 0.413, indicates that JSW model and in particular Birnbaum-Saunders distribution can be used to analyze this data set.

5.3 Guinea Pigs Data: Censored

To see how the proposed EM algorithm works for censored sample, we use the same guinea pigs data used in the previous subsection, but it is assumed that we observe only first 67 observations and the last 5 observations are censored at time point 258. Therefore, it is a Type-II censored sample with $n = 72$, and $r = 67$.

Treating all the observations as complete, we obtain the initial estimates of μ and λ as

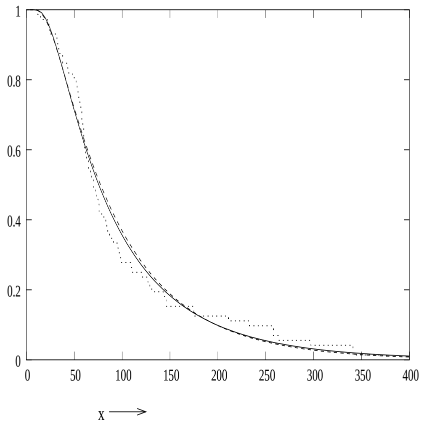


Figure 4: Empirical survival function, the best fitted JSW model and the best fitted Birnbaum-Saunders model for Guinea Pigs data.

$\widehat{\mu} = 77.5886$ and $\widehat{\lambda} = 131.4988$ respectively. As before we use different initial estimates of $p \in (0, 1)$ and perform the proposed EM algorithm and finally choose those estimates which provide the maximum log-likelihood value. We obtain the MLEs of the unknown parameters as $\widehat{p} = 0.4142$, $\widehat{\mu} = 78.2891$ and $\widehat{\lambda} = 135.6624$. The associated 95% confidence intervals of p, μ and λ are (0.2890, 0.5156), (68,4766, 88.1016) and (80.4245, 190.9003) respectively. The reliability function and the associated 95% confidence intervals are provided in Figure 5. It is interesting to see that although, due to censoring, the estimates of μ and λ do not change that much but, the estimate of p changes quite drastically.

Now to verify the goodness-of-fit and also the effect of censoring, we plot the Kaplan-Meier estimators and the best fitted JSW for the censored data and the best fitted JSW model for the complete data (obtained in the previous sub-section) in Figure 6. It is clear that JSW model fits quite well to the censored data also, and since the best fitted JSW models for the complete data and censored data are almost identical, it can be concluded that in this case the effect of censoring is not very significant.

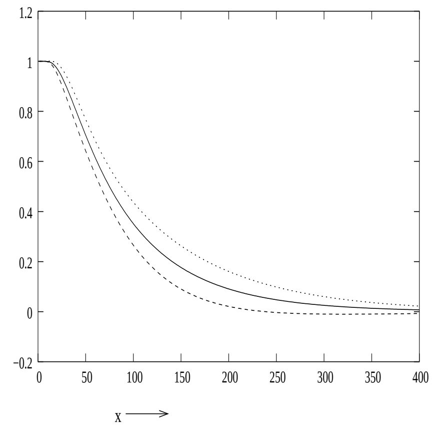


Figure 5: Reliability function and the associated 95% confidence bands for censored Guinea Pigs data.

6 Conclusions

In this paper we have considered the estimation procedure of a lifetime model proposed by Jorgensen, Seshadri and Whitmore [15]. This three parameter model has some interesting features. It is well known that this model has unimodal hazard function and the well known Birnbaum-Saunders distribution can be obtained as a special case of this model. We have proposed to use the EM algorithm to estimate the unknown parameters. If we want to find the MLEs by solving the normal equations, we need to solve three non-linear equations simultaneously, which is quite involved as explained earlier. However, in the proposed EM algorithm, at each E-step, the maximization (M-step) can be performed explicitly. Therefore, the implementation of the EM algorithm is quite simple in this case. From results of the simulation study it is observed that the performance of the proposed EM algorithm is quite satisfactory. Moreover, it is observed that the proposed EM algorithm can be easily used when the data are censored. Since Birnbaum-Saunders distribution is a special case of the JSW model, our algorithm is well suited for finding the MLEs of Birnbaum-Saunders model as well.

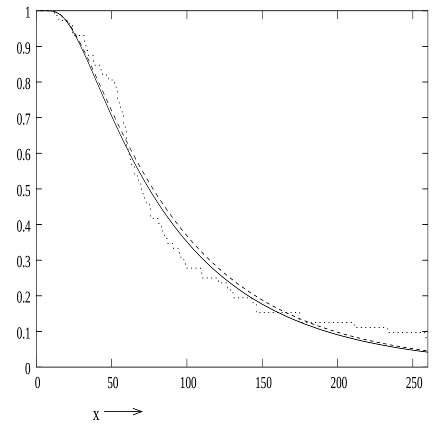


Figure 6: Kaplan-Meier estimator, the best fitted JSW model for censored Guinea Pigs data and the best fitted JSW model for complete Guinea Pigs data.

Acknowledgements:

The authors would like to thank the unknown referees and the editor for several important comments which had helped us to improve the paper significantly.

Appendix: Observed Fisher Information Matrix

For convenience we just present the observed Fisher information matrix obtained from the EM algorithm using the idea of Louis [17]. Using the same notation as in Louis [17], the observed Fisher information matrix can be written

$$\boldsymbol{F}_{obs} = \boldsymbol{B} - \boldsymbol{S}\boldsymbol{S}^T,$$

here $\boldsymbol{B}$ is the negative of the second derivative of the log-likelihood function and $\boldsymbol{S}$ is the derivative vector. We just provide the elements of the matrix $\boldsymbol{B} = ((B(i,j)))$ and the vector $\boldsymbol{S} = (S(j))$.

$$B(1,1) \;=\; \frac{n\lambda}{\mu^2}\left[\frac{3\bar{x}}{\mu^2} - \frac{10}{\mu} + 8s_1\right] - \frac{n\bar{z}}{\mu}, \quad B(2,2) = \frac{n}{2\lambda^2}, \quad B(3,3) = \frac{n\bar{z}}{p^2} - \frac{n(1-\bar{z})}{(1-p)^2}$$

$$
\begin{aligned}
B(1,2) &= B(2,1) = -\frac{1}{\mu^3}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i} - \frac{1}{\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)}{x_i} \\
B(1,3) &= B(3,1) = B(2,3) = B(3,2) = 0.
\end{aligned}
$$

$$
\begin{aligned}
S(1) &= -\frac{n\bar{z}}{\mu} + \left[\frac{\lambda}{\mu^3}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i} + \frac{\lambda}{\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)}{x_i}\right] \\
S(2) &= \frac{n}{2\lambda} - \frac{1}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i} \\
S(3) &= \frac{n\bar{z}}{p} - \frac{n(1-\bar{z})}{(1-p)}
\end{aligned}
$$

Note that in case of censored sample the observed Fisher information matrix can be obtained by replacing the censored observation with its expected value where the true parameters are replaced by their estimated values.

References

[1] Akman, O. and Gupta, R.C. (1992), "A comparison of various estimators of an inverse Gaussian distribution", *Journal of Statistical Computation and Simulation*, vol. 40, 71 - 81.

[2] Aarset, M.V. (1987), "How to identify a bathtub shaped hazard rate?", *IEEE Transactions on Reliability*, vol. 36, 106–108.

[3] Birnbaum, Z.W., Saunders, S.C. (1969), "A new family of life distributions", *Journal of Applied Probability*, vol. 6, 319–327.

[4] Bjerkedal, T. (1960). "Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli", *American Journal of Hygiene*, vol. 72, 130-148.

[5] Chang, D.S. and Tang, L.C. (1993), "Reliability bounds and critical time for the Birnbaum-Saunders distribution", *IEEE Transactions on Reliability*, vol. 42, 464 - 469.

[6] Chang, D.S. and Tang, L.C. (1994), "Parametric bounds and tolerance limits for the Birnbaum-Saunders distribution", *Communications in Statistics - Theory and Methods*, vol. 23, 2853 - 2863.

[7] Chhikara, R.S. and Folks, J.L. (1989), *The Inverse Gaussian Distribution*, Marcel and Dekker, New York.

[8] Desmond, A.F. and Yang, Z. (1995), "Shortest prediction intervals for Birnbaum-Saunders distribution", *Communications in Statistics - Theory and Methods*, vol. 24, 1383 - 1401..

[9] Efron, B. (1988), "Logistic regression, survival analysis and the Kaplan Meier curve", *Journal of the American Statistical Association*, vol. 83, 414 - 425.

[10] Englehardt, M., Bain, L.J. and Wright, F.T. (1981), "Inferences based on the parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation", *Technometrics*, vol. 23, 3, 251 - 256.

[11] Gupta, R.C. and Akman, O. (1995), "On the reliability studies of a weighted inverse Gaussian model", *Journal of Statistical Planning and Inference*, vol. 48, 69 - 83.

[12] Gupta, R.C. and Akman, O. (1997), "Estimation of critical points in the mixture inverse Gaussian model", *Statistical Papers*, vol. 38, 445 452.

[13] Gupta, R. C., Kannan, N. and RayChoudhuri, A. (1997), "Analysis of log-normal survival data", *Mathematical Biosciences*, vol. 139, 103-115.

[14] Jorgensen, B (1987), "Some properties of the exponential dispersion model", *Journal of the Royal Statistical Society*, Ser. B., vol. 49. 127 - 162.

[15] Jorgensen, B., Seshadri, V. and Whitmore, G.A. (1991), "On the mixture of the inverse Gaussian distribution with its complementary reciprocal", *Scandinavian Journal of Statistics*, vol. 18, 77 - 89.

[16] Kundu, D., Kannan, N., Balakrishnan, N. (2008), "On the hazard function of Birnbaum-Saunders distribution and associated inference", *Computational Statistics & Data Analysis*, vol. 52, 2692–2702.

[17] Louis, T. A. (1982), "Finding the observed information matrix when using the EM algorithm", *Journal of the Royal Statistical Society*, Series B 44, 2, 226233.

[18] McLachlan, G.J. and Krishnan, T. (1997), *The EM algorithm and extensions*, Wiley and Sons, New York.

[19] Michael, J.R., Schucany, W.R. and Haas, R.W. (1976), "Generating random variables using transformation with multiple roots", *American Statistician*, vol. 30, 88 - 90.

[20] Mudholkar, G.S., Srivastava, D.K. and Kollia, G.D. (1996), "A generalization of the Weibull distribution with application to the analysis of survival data", *Journal of the American Statistical Association*, vol. 91, 1575 - 1583.

[21] Ng, H.K.T., Kundu, D., Balakrishnan, N. (2003), "Modified moment estimation for the two-parameter Birnbaum-Saunders distributions", *Computational Statistics & Data Analysis*, vol. 43, 283–298.

[22] Ng, H.K.T., Kundu, D., Balakrishnan, N. (2006), "Point and interval estimation for the two-parameter Birnbaum-Saunders distribution based on Type-II censored samples", *Computational Statistics & Data Analysis*, vol. 50, 3222–3242.

[23] Press, W. H., Teukolsky, S. A., Vellerling, W. T. and Flannery, B. P. (1994), *Numerical Recipes in FORTRAN*, 2nd. edition, Cambridge University Press, Cambridge.

[24] Rieck, J.R. (1995), "Parametric estimation for the Birnbaum-Saunders distribution based on symmetrically censored samples", *Communications in Statistics - Theory and Methods*, vol. 24, 1721 - 1736.