

Retraction

Retracted: EM Algorithm for Estimating the Parameters of Quasi-Lindley Model with Application

Journal of Mathematics

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Journal of Mathematics. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] M. Kayid and N. S. Al-Maflehi, "EM Algorithm for Estimating the Parameters of Quasi-Lindley Model with Application," *Journal of Mathematics*, vol. 2022, Article ID 8467291, 9 pages, 2022.

Research Article

EM Algorithm for Estimating the Parameters of Quasi-Lindley Model with Application

M. Kayid ¹ and Nassr S. Al-Maflehi ²

¹Department of Statistics and Operations Research, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

²Department of Periodontics and Community Dentistry, College of Dentistry, King Saud University, Riyadh, Saudi Arabia

Correspondence should be addressed to M. Kayid; drkayid@ksu.edu.sa

Received 10 November 2021; Revised 12 December 2021; Accepted 21 December 2021; Published 25 January 2022

Academic Editor: Lazim Abdullah

Copyright © 2022 M. Kayid and Nassr S. Al-Maflehi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The quasi-Lindley distribution is a flexible model useful in reliability analysis, management science, and engineering analysis. In this paper, an expectation-maximization (EM) algorithm was applied to estimate the parameters of this model for uncensored and right-censored data. Simulation studies show that the estimates of EM perform better than maximum likelihood estimates (MLEs) for both uncensored and censored data. In an illustrative example, the waiting times of a bank's customers are analyzed and the estimator of the EM algorithm is compared with the MLE. The analysis of the data can be useful for the management of the bank.

1. Introduction

The quasi-Lindley distribution proposed by Shanker and Mishra [1] is a generalization of the Lindley distribution introduced by Lindley [2] and is quite useful in reliability theory and survival analysis. The probability density function (pdf) of the quasi-Lindley distribution is given by

$$f(x) = \frac{\lambda}{\alpha + 1} (\alpha + \lambda x) e^{-\lambda x}, x > 0, \alpha > 0, \lambda > 0, \quad (1)$$

and is a mixture of gamma distributions $G(1, \lambda)$ and $G(2, \lambda)$ with weights $\alpha/(\alpha + 1)$ and $1/(\alpha + 1)$, respectively. The hazard rate function of the quasi-Lindley model is

$$h(x) = \frac{\lambda(\alpha + \lambda x)}{1 + \alpha + \lambda x}, x > 0, \alpha > 0, \lambda > 0, \quad (2)$$

which is an increasing function.

An important feature of the quasi-Lindley model is that, unlike the Lindley model and its many other generalizations, it is scale invariant. Nevertheless, it is not complicated but sufficiently flexible. Shanker and Mishra [1] studied some of its basic properties and dynamic reliability measures. They also discussed the maximum likelihood estimator (MLE) for its parameters. The MLE is theoretically consistent and

efficient, but in practice, it strongly depends on the initial values and the computational approach, which can be achieved by directly maximizing the log-likelihood function or by solving the likelihood equations. Moreover, the simulation results for the MLE of the quasi-Lindley distribution (especially for α) show extremely large values for the mean square error (MSE) (see Tables 1 and 2). This motivates us to investigate the EM algorithm for estimating the parameters.

In statistics, when the data are collected from a mixture or competing risks model, the EM algorithm is an effective tool for estimating parameters of latent variable models. In their foundational work, Dempster et al. [3] introduced the EM algorithm. Many authors after them have used the idea of EM in their work to provide better estimation of the parameters of the models they are considering. For example, Elmahdy and Aboutahoun [4] and Almhana et al. [5] used the EM algorithm to estimate the parameters for a mixture of Weibull models and a mixture of gamma models, respectively. In addition, Bee et al. [6] applied the EM algorithm to estimate the parameters of Pareto mixture models, Ghosh et al. [7] used the EM algorithm for a mixture of Weibull and Pareto (IV) models, and Balakrishnan and Pal [8] used the EM-based likelihood inference in their work. For detailed discussions of the EM algorithm, we refer the readers to McLachlan and Krishnan [9]

TABLE 1: Simulation results for MLE and EM estimators of the QL distribution for uncensored data (in every cell, the first and second lines are related to α and λ , respectively).

Method	n α, λ	100				200			
		B	MSE	CP	CILM	B	MSE	CP	CILM
MLE	0.4, 0.02	0.6081	131.03	0.8918	29.48	0.2018	113.33	0.9092	3.20
		0.00005	0.000005	0.954	0.0099	0.00008	0.000002	0.9484	0.0065
	0.5, 0.5	2.9955	2514.5	0.8804	57.31	0.1250	26.72	0.91	4.849
		0.0026	0.0037	0.9472	0.2382	0.00125	0.00181	0.9554	0.1649
	0.7, 0.1	7.9912	5877.70	0.8742	215.57	0.6875	251.728	0.9018	19.197
		0.00041	0.00019	0.9454	0.0540	0.00045	0.00009	0.9574	0.0372
EM	0.4, 0.02	-0.00007	0.01200	0.9868	1.1915	0.00304	0.00599	0.9994	0.8047
		0.00017	0.000002	0.979	0.00917	0.00011	0.000001	0.9936	0.0062
	0.5, 0.5	0.00394	0.01673	0.9872	1.4474	0.00187	0.00863	0.9992	0.9888
		0.00357	0.00164	0.9802	0.2420	0.00155	0.00080	0.9952	0.1664
	0.7, 0.1	0.00767	0.02847	0.9762	2.0985	0.00021	0.01351	0.9976	1.4379
		0.00083	0.00006	0.974	0.0548	0.00051	0.00003	0.9944	0.0378

TABLE 2: Simulation results for MLE and EM estimators of the QL distribution when the censorship rate is $p = 0.2$ (in every cell, the first and second lines are related to α and λ , respectively).

Method	n α, λ	100				200			
		B	MSE	CP	CILM	B	MSE	CP	CILM
MLE	0.4, 0.02	3.09912	996.321	0.8758	171.84	0.28132	58.7714	0.92	9.9925
		-0.00013	0.000008	0.9452	0.0107	-0.00009	0.000004	0.9588	0.0075
	0.5, 0.5	6.4487	2928.10	0.867	439.53	0.6962	126.6613	0.9054	61.3642
		-0.00363	0.0063	0.939	0.3021	-0.00235	0.00289	0.9604	0.2062
	0.7, 0.1	15.625	6471.68	0.853	958.73	2.6733	556.14	0.8947	215.307
		-0.00139	0.00031	0.9182	0.0699	-0.00099	0.00016	0.9512	0.0498
EM	0.4, 0.02	0.00538	0.01108	0.9812	1.3233	0.00231	0.00557	0.9976	0.8762
		0.00008	0.000003	0.978	0.01113	0.00009	0.0000015	0.994	0.0075
	0.5, 0.5	0.00293	0.01506	0.9698	1.6263	0.00202	0.00752	0.9966	1.1044
		0.00327	0.00193	0.967	0.2993	0.00101	0.00096	0.9944	0.2047
	0.7, 0.1	0.00785	0.02433	0.9452	2.3950	0.00380	0.01193	0.9872	1.6584
		0.00029	0.00007	0.9434	0.06918	0.00038	0.00004	0.9856	0.0482

and Mengersen et al. [10]. In addition, Wu [11] proved some results related to the convergence of the EM algorithm.

In this paper, a specific EM algorithm is developed to obtain a more reliable estimate for the parameters of the quasi-Lindley distribution for uncensored and right-censored data. The paper is organized as follows. Section 2 discusses the EM algorithm for the quasi-Lindley distribution when the data were uncensored. In Section 3, the EM algorithm is extended to right-censored data. Section 4 examines the behavior of the MLE and EM estimates and compares them through simulations. In Section 5, both MLE and EM estimates are computed for a real dataset. Finally, Section 6 draws the conclusions for the paper.

2. Uncensored Data

Assume that x_1, x_2, \dots, x_n be an independent and identically distributed (iid) random sample from quasi-Lindley distribution with parameters (α, λ) , briefly $QL(\alpha, \lambda)$. The log-likelihood function of the parameters is

$$l(\alpha, \lambda; \mathbf{x}) = n \ln \lambda - n \ln(\alpha + 1) + \sum_{i=1}^n \ln(\alpha + \lambda x_i) - \lambda \sum_{i=1}^n x_i. \quad (3)$$

The likelihood equations can be obtained by partial differentiation of this log-likelihood function with respect to α and λ as follows:

$$\begin{aligned} \frac{\partial}{\partial \alpha} l(\alpha, \lambda; \mathbf{x}) &= \sum_{i=1}^n \frac{1}{\alpha + \lambda x_i} - \frac{n}{\alpha + 1} = 0, \\ \frac{\partial}{\partial \lambda} l(\alpha, \lambda; \mathbf{x}) &= \frac{n}{\lambda} + \sum_{i=1}^n \frac{x_i}{\alpha + \lambda x_i} - \sum_{i=1}^n x_i = 0. \end{aligned} \quad (4)$$

The MLE can be calculated by directly maximizing the log-likelihood function (3) directly or by solving the likelihood equations. Let $l = \ln f(X)$, and the Fisher information matrix of the quasi-Lindley distribution is

$$K = \begin{bmatrix} E\left(-\frac{\partial^2 l}{\partial \alpha^2}\right) & E\left(-\frac{\partial^2 l}{\partial \alpha \partial \lambda}\right) \\ E\left(-\frac{\partial^2 l}{\partial \lambda \partial \alpha}\right) & E\left(-\frac{\partial^2 l}{\partial \lambda^2}\right) \end{bmatrix}. \quad (5)$$

When we have an iid random sample $X_i, i = 1, 2, \dots, n$ from $QL(\alpha_0, \lambda_0)$, the MLE, $(\hat{\alpha}, \hat{\lambda})$, weakly converges to bivariate normal $N((\alpha_0, \lambda_0), n^{-1}K^{-1})$ where K^{-1} is the inverse of the information matrix.

2.1. The EM Algorithm for the Complete Data. Since $QL(\alpha, \lambda)$ is a mixture of two gamma distributions $G(1, \lambda)$ and $G(2, \lambda)$, the EM algorithm can be used to estimate its parameters. Let $X_i, i = 1, 2, \dots, n$, be an iid random sample of $QL(\alpha, \lambda)$. In the EM approach, for each X_i , we consider one latent random variable Z_i which determines that X_i belongs

to $G(1, \lambda)$ or $G(2, \lambda)$. In other words, $X_i|Z_i = j \sim IG(j, \theta)$, $j = 1, 2$, $P(Z_i = 1) = \alpha/(\alpha + 1)$, and $P(Z_i = 2) = 1/(\alpha + 1)$. For a brief representation, take $\theta = (\alpha, \lambda)$. The likelihood function can be written in the following form.

$$L(\theta; \mathbf{x}, \mathbf{z}) = \prod_{i=1}^n \prod_{j=1}^2 (g_j(x_i; \theta) P(Z_i = j))^{I(z_i=j)}, \quad (6)$$

where the indicator $I(z_i = j)$ equals 1 when $z_i = j$ and otherwise equals to 0. Also,

$$g_j(x_i; \lambda) = \lambda^j x_i^{j-1} \exp(-\lambda x_i), \quad (7)$$

is the pdf of the underlying gamma distribution and

$$P(Z_i = j) = \frac{1}{\alpha + 1} \alpha^{I(j=1)}. \quad (8)$$

Then, the log-likelihood function is

$$l(\theta; \mathbf{x}, \mathbf{z}) = \ln L(\theta; \mathbf{x}, \mathbf{z}) = \sum_{i=1}^n \sum_{j=1}^2 I(z_i = j) \ln \left(\lambda^j x_i^{j-1} \exp(-\lambda x_i) \frac{1}{\alpha + 1} \alpha^{I(j=1)} \right). \quad (9)$$

The EM algorithm goes through two steps: the expectation step (*E*) and the maximization step (*M*). In each iteration, the *E* step constructs the expected value of the log-likelihood with respect to the current estimate of the conditional latent variable. In the *M* step, the constructed in the *E* step is maximized to provide the estimates. The iterative process can be terminated when the improvement

in the expectation function falls below a predetermined small value.

2.1.1. The E Step. Given the estimate of the parameters at iteration t , θ_t , the conditional distribution of Z_i is obtained by Bayes theorem:

$$\begin{aligned} p_{ij,t} &= P(Z_i = j | X_i = x_i, \theta_t) = \frac{f(X_i = x_i | Z_i = j, \theta_t) P(Z_i = j | \theta_t)}{f(X_i = x_i | \theta_t)} \\ &= \frac{\lambda_t^j x_i^{j-1} e^{-\lambda_t x_i} (\alpha_t + 1)^{-1} \alpha_t^{I(j=1)}}{\sum_{j=1}^2 \lambda_t^j x_i^{j-1} e^{-\lambda_t x_i} (\alpha_t + 1)^{-1} \alpha_t^{I(j=1)}}, i = 1, 2, \dots, n, j = 1, 2, \end{aligned} \quad (10)$$

and after simplification, we have

$$p_{i1,t} = \frac{\alpha_t}{\alpha_t + \lambda_t x_i}, i = 1, 2, \dots, n, \quad (11)$$

and $p_{i2,t} = 1 - p_{i1,t}$. These probabilities are known as membership probabilities at iteration t and are used to construct the expectation function $Q(\theta | \theta_t)$ as follows:

$$\begin{aligned}
Q(\theta|\theta_t) &= E_{Z|X,\theta_t}(l(\theta; \mathbf{x}, \mathbf{Z})) \\
&= \sum_{i=1}^n E_{Z_i|X_i,\theta_t} \sum_{j=1}^2 I(z_i = j) \ln\left(\lambda^j x_i^{j-1} \exp(-\lambda x_i) \frac{1}{\alpha+1} \alpha^{I(j=1)}\right) \\
&= \sum_{i=1}^n P(Z_i = 1|X_i = x_i, \theta_t) \ln\left(\lambda e^{-\lambda x_i} \frac{\alpha}{\alpha+1}\right) \\
&\quad + P(Z_i = 2|X_i = x_i, \theta_t) \ln\left(\lambda^2 x_i e^{-\lambda x_i} \frac{1}{\alpha+1}\right) \\
&= \sum_{i=1}^n \left((1 + p_{i2,t}) \ln \lambda - \lambda x_i + p_{i2,t} \ln x_i \right) + \sum_{i=1}^n (p_{i1,t} \ln \alpha - \ln(\alpha + 1)).
\end{aligned} \tag{12}$$

The last expressions of (12) show that expectation can be expressed as the sum of two statements, one of which depends only on λ and the other only on α , i.e.,

$$Q(\theta|\theta_t) = Q_1(\lambda|\theta_t) + Q_2(\alpha|\theta_t), \tag{13}$$

where

$$\begin{aligned}
Q_1(\lambda|\theta_t) &= \sum_{i=1}^n \left((1 + p_{i2,t}) \ln \lambda - \lambda x_i + p_{i2,t} \ln x_i \right), \\
Q_2(\alpha|\theta_t) &= \sum_{i=1}^n (p_{i1,t} \ln \alpha - \ln(\alpha + 1)).
\end{aligned} \tag{14}$$

2.1.2. The M Step. To estimate the parameters at $t+1$ iteration, we maximize the $Q(\theta|\theta_t)$ in terms of θ . So, we have

$$\theta_{t+1} = \operatorname{argmax}_{\theta} Q(\theta|\theta_t), \tag{15}$$

which, by (13), reduces to the following separate maximization problems.

$$\begin{aligned}
\lambda_{t+1} &= \operatorname{argmax}_{\lambda} Q_1(\lambda|\theta_t), \\
\alpha_{t+1} &= \operatorname{argmax}_{\alpha} Q_2(\alpha|\theta_t),
\end{aligned} \tag{16}$$

where $Q_1(\lambda|\theta_t)$ and $Q_2(\alpha|\theta_t)$ are determined by (14) and (15), respectively. By solving the equation $\partial/\partial\lambda Q_1(\lambda|\theta_t) = 0$, the estimation of λ at $t+1$ iteration is obtained.

$$\lambda_{t+1} = \frac{n + \sum_{i=1}^n p_{i2,t}}{\sum_{i=1}^n X_i}. \tag{17}$$

On the other hand, solving the equation $\partial/\partial\alpha Q_2(\alpha|\theta_t) = 0$, we have

$$\alpha_{t+1} = \frac{\sum_{i=1}^n p_{i1,t}}{\sum_{i=1}^n p_{i2,t}}. \tag{18}$$

The sequence θ_t will converge to θ , and the iterative process can be concluded when for some predefined small $\epsilon > 0$, $Q(\theta_{t+1}|\theta_{t+1}) < Q(\theta_t|\theta_t) + \epsilon$. This means that further iterations do not improve the objective function considerably. For detailed information about convergence of the EM algorithm, see Wu [11].

3. Right-Censored Data

Consider an iid random sample X_i , $i = 1, 2, \dots, n$, from $QL(\alpha, \lambda)$ which is exposed to right censorship. We say that X_i is censored from the right by a censoring random variable C_i , if $X_i > C_i$, and in this case, the only information about event time is that it is greater than censoring time C_i . The observations consist of $T_i = \min(X_i, C_i)$ and d_i , where $d_i = 1$, when the event has not been censored, $X_i \leq C_i$, and $d_i = 0$, when the event has been censored, $X_i > C_i$. Given a right-censored sample (t_i, d_i) , $i = 1, 2, \dots, n$, the log-likelihood function is

$$l(\alpha, \lambda; \mathbf{t}, \mathbf{d}) = \sum_{i=1}^n d_i \ln f(t_i) + \sum_{i=1}^n (1 - d_i) \ln R(t_i), \tag{19}$$

where f and R show the density and the reliability functions of the quasi-Lindley distribution, respectively. The log-likelihood function simplifies to

$$l(\theta; \mathbf{t}, \mathbf{d}) = n_1 \ln\left(\frac{\lambda}{\alpha+1}\right) + \sum_{i=1}^n d_i \ln(\alpha + \lambda t_i) + \sum_{i=1}^n (1 - d_i) \ln\left(1 + \frac{\lambda t_i}{\alpha+1}\right) - \sum_{i=1}^n \lambda t_i, \tag{20}$$

where $n_1 = \sum_{i=1}^n d_i$ and $\theta = (\alpha, \lambda)$.

3.1. *The EM Algorithm for Right-Censored Data.* To implement the EM algorithm, we should include the latent

variable $Z_i, i = 1, 2, \dots, n$, defined in the previous section. Then, the likelihood function for the censored data is

$$L(\theta; \mathbf{t}, \mathbf{d}, \mathbf{z}) = \prod_{i=1}^n \left(\prod_{j=1}^2 (g_j(t_i; \theta) P(Z_i = j))^{I(z_i=j)} \right)^{d_i} \prod_{i=1}^n \left(\prod_{j=1}^2 (\bar{G}_j(t_i; \theta) P(Z_i = j))^{I(z_i=j)} \right)^{1-d_i}, \quad (21)$$

where g_j shows the gamma pdf considered in the previous section and \bar{G}_j is its corresponding reliability function. By

taking logarithm from (21), the log-likelihood function has the following form:

$$l(\theta; \mathbf{t}, \mathbf{d}, \mathbf{z}) = \sum_{i=1}^n d_i \left(I(z_i = 1) \ln \left(\lambda e^{-\lambda t_i} \frac{\alpha}{\alpha + 1} \right) + I(z_i = 2) \ln \left(\lambda^2 t_i e^{-\lambda t_i} \frac{1}{\alpha + 1} \right) \right) + \sum_{i=1}^n (1 - d_i) \left(I(z_i = 1) \ln \left(e^{-\lambda t_i} \frac{\alpha}{\alpha + 1} \right) + I(z_i = 2) \ln \left((\lambda t_i + 1) e^{-\lambda t_i} \frac{1}{\alpha + 1} \right) \right). \quad (22)$$

Similar to the uncensored data, we should iterate two E and M steps to find an improved estimation.

3.1.1. *The E Step.* Given the estimate of the parameters at iteration t , θ_t , applying the Bayes theorem, we can compute the conditional distribution of Z_i as follows:

$$\begin{aligned} p_{ij,t} &= d_i P(Z_i = j | X_i = t_i, \theta_t) + (1 - d_i) P(Z_i = j | X_i > t_i, \theta_t) \\ &= d_i \frac{f(X_i = t_i | Z_i = j, \theta_t) P(Z_i = j | \theta_t)}{f(X_i = t_i | \theta_t)} + (1 - d_i) \frac{P(X_i > t_i | Z_i = j, \theta_t) P(Z_i = j | \theta_t)}{f(X_i > t_i | \theta_t)} \\ &= d_i \frac{\lambda^j t_i^{j-1} e^{-\lambda t_i} (\alpha_t + 1)^{-1} \alpha_t^{I(j=1)}}{\sum_{j=1}^2 \lambda^j t_i^{j-1} e^{-\lambda t_i} (\alpha_t + 1)^{-1} \alpha_t^{I(j=1)}} \\ &\quad + (1 - d_i) \frac{I(j=1) e^{-\lambda t_i} (\alpha_t / (\alpha_t + 1)) + I(j=2) (\lambda t_i + 1) e^{-\lambda t_i} (1 / (\alpha_t + 1))}{e^{-\lambda t_i} (\alpha_t / (\alpha_t + 1)) + (\lambda t_i + 1) e^{-\lambda t_i} (1 / (\alpha_t + 1))} \\ &= d_i \frac{I(j=1) \alpha_t + I(j=2) \lambda t_i}{\alpha_t + \lambda t_i} + (1 - d_i) \frac{I(j=1) \alpha_t + I(j=2) \lambda t_i + 1}{\alpha_t + \lambda t_i + 1}, i = 1, 2, \dots, n, j = 1, 2. \end{aligned} \quad (23)$$

Specifically, for $j = 1$,

$$p_{i1,t} = d_i \frac{\alpha_t}{\alpha_t + \lambda t_i} + (1 - d_i) \frac{\alpha_t}{\alpha_t + \lambda t_i + 1}, i = 1, 2, \dots, n, \quad (24)$$

and $p_{i2,t} = 1 - p_{i1,t}$. Then, using (22), the expectation function at iteration t can be written in the following form.

$$\begin{aligned} Q(\theta | \theta_t) &= E_{Z|t,d,\theta_t} (l(\theta; \mathbf{t}, \mathbf{d}, \mathbf{z})) \\ &= \sum_{i=1}^n d_i \left(p_{i1,t} \left(\ln \left(\lambda e^{-\lambda t_i} \frac{\alpha}{\alpha + 1} \right) \right) + p_{i2,t} \left(\ln \left(\lambda^2 t_i e^{-\lambda t_i} \frac{1}{\alpha + 1} \right) \right) \right) \\ &\quad + \sum_{i=1}^n (1 - d_i) \left(p_{i1,t} \left(\ln \left(e^{-\lambda t_i} \frac{\alpha}{\alpha + 1} \right) \right) + p_{i2,t} \left(\ln \left((\lambda t_i + 1) e^{-\lambda t_i} \frac{1}{\alpha + 1} \right) \right) \right). \end{aligned} \quad (25)$$

Similar to uncensored case, it is straightforward to check that $Q(\theta|\theta_t)$ can be written as two statements in which one of them just depends on α and the other depends on λ . More precisely,

$$Q(\theta|\theta_t) = Q_1(\lambda|\theta_t) + Q_2(\alpha|\theta_t), \quad (26)$$

where

$$Q_1(\lambda|\theta_t) = \sum_{i=1}^n d_i(1 + p_{i2,t}) \ln \lambda + \sum_{i=1}^n d_i p_{i2,t} \ln t_i + \sum_{i=1}^n (1 - d_i) p_{i2,t} \ln(\lambda t_i + 1) - \lambda \sum_{i=1}^n t_i, \quad (27)$$

$$Q_2(\alpha|\theta_t) = \sum_{i=1}^n (p_{i1,t} \ln \alpha - \ln(\alpha + 1)) \quad (28)$$

3.1.2. The M Step. In this step, we should maximize the $Q(\theta|\theta_t)$ function to compute the estimations at the $t + 1$ iteration.

$$\theta_{t+1} = \arg \max_{\theta} Q(\theta|\theta_t), \quad (29)$$

which, by (26), reduces to the following separate maximization problems.

$$\begin{aligned} \lambda_{t+1} &= \arg \max_{\lambda} Q_1(\lambda|\theta_t), \\ \alpha_{t+1} &= \arg \max_{\alpha} Q_2(\alpha|\theta_t), \end{aligned} \quad (30)$$

in which $Q_1(\lambda|\theta_t)$ and $Q_2(\alpha|\theta_t)$ are determined by (27) and (28), respectively. The likelihood equation $\partial/\partial\lambda Q_1(\lambda|\theta_t) = 0$ which after some algebra can be simplified to

$$\lambda = \frac{\sum_{i=1}^n d_i(1 + p_{i2,t})}{\sum_{i=1}^n t_i} + \frac{\sum_{i=1}^n (1 - d_i)(\lambda t_i/(1 + \lambda t_i)) p_{i2,t}}{\sum_{i=1}^n t_i} \quad (31)$$

which does not yield to an analytical solution for λ , so the solution can be computed by numerical methods. But, clearly (31) implies that the solution of this equation, namely, λ_{t+1} , satisfies the inequality

$$\lambda_{t+1} > \frac{\sum_{i=1}^n d_i(1 + p_{i2,t})}{\sum_{i=1}^n t_i}. \quad (32)$$

On the other hand, since $1 + p_{i2,t} > (\lambda t_i/(1 + \lambda t_i)) p_{i2,t}$, one upper bound for the solution is

$$\lambda_{t+1} < \frac{\sum_{i=1}^n (1 + p_{i2,t})}{\sum_{i=1}^n t_i}, \quad (33)$$

and in turn, by (32) and (33), we have

$$\frac{\sum_{i=1}^n d_i(1 + p_{i2,t})}{\sum_{i=1}^n t_i} < \lambda_{t+1} < \frac{\sum_{i=1}^n (1 + p_{i2,t})}{\sum_{i=1}^n t_i}. \quad (34)$$

These bounds can be applied in numerical processes to find optimized answer. The solution for α can be obtained by solving the equation $\partial/\partial\alpha Q_2(\alpha|\theta_t) = 0$ as follows:

$$\alpha_{t+1} = \frac{\sum_{i=1}^n p_{i1,t}}{\sum_{i=1}^n p_{i2,t}}. \quad (35)$$

Similar to the uncensored case, the iterative process can be concluded when for some predefined small $\epsilon > 0$, $Q(\theta_{t+1}|\theta_{t+1}) < Q(\theta_t|\theta_t) + \epsilon$.

Let $\hat{\theta}$ and θ_0 be the EM estimator and the real parameter, respectively. Then, $\hat{\theta} - \theta_0$ converges asymptotically to a bivariate normal distribution $N(0, V)$, where V can be approximated by the inverse of the observed information matrix with respect to the observed data (see Meng and Rubin [12]). It is computed by evaluating the Hessian matrix of the log-likelihood function with respect to the observed data at the point $\hat{\theta}$, and then calculating the inverse of the obtained Hessian matrix, briefly $V = I_o^{-1}(\hat{\theta}|\mathbf{x})$. Fortunately, in the case of this study, the log-likelihood function of the observed data is not complicated and can be used to calculate the Hessian matrix and finally the variance approximation. For this purpose, the function "hessian" of the library "pracma" in R is used. Since the asymptotic distribution of the EM estimator is normal, the standard normal quantiles are used to obtain approximate confidence intervals of the parameters.

4. Simulations

In a simulation study, we investigate the behavior of the MLE and EM estimators and compare them. The fact that the quasi-Lindley model is a mixture of gamma distributions is applied to generate random samples. To generate right-censored sample y_1, y_2, \dots, y_n , we assume that the censoring random variable C_i follows the degenerate distribution with mean M . Thus, if p is the censoring rate, we can calculate M by solving the equation $M = F^{-1}(1 - p)$ where F^{-1} is the inverse of the distribution function of the quasi-Lindley model. Now, an uncensored sample x_1, x_2, \dots, x_n is taken from the quasi-Lindley model. Then, the i th instance of the desired right-censored sample is $y_i = \min\{x_i, M\}$.

Each cell of Tables 1 and 2 shows the results of one run. In every run, $r = 5000$ replicates of samples of size $n = 100$ or 200 were generated by the quasi-Lindley model with selected parameters, and in each run, the MLE and EM estimators were calculated. To calculate the MLE, the log-likelihood function was maximized by using the "optim" function built into R with the standard "Nelder-Mead" optimization method. In both the maximum likelihood method and EM, the initial values are generated from a uniform distribution.

TABLE 3: Waiting times (in minutes) of customers to receive service in a bank.

0.8	0.8	1.3	1.5	1.8	1.9	1.9	2.1	2.6	2.7
2.9	3.1	3.2	3.3	3.5	3.6	4.0	4.1	4.2	4.2
4.3	4.3	4.4	4.4	4.6	4.7	4.7	4.8	4.9	4.9
5.0	5.3	5.5	5.7	5.7	6.1	6.2	6.2	6.2	6.3
6.7	6.9	7.1	7.1	7.1	7.1	7.4	7.6	7.7	8.0
8.2	8.6	8.6	8.6	8.8	8.8	8.9	8.9	9.5	9.6
9.7	9.8	10.7	10.9	11.0	11.0	11.1	11.2	11.2	11.5
11.9	12.4	12.5	12.9	13.0	13.1	13.3	13.6	13.7	13.9
14.1	15.4	15.4	17.3	17.3	18.1	18.2	18.4	18.9	19.0
19.9	20.6	21.3	21.4	21.9	23.0	27.0	31.6	33.1	38.5

TABLE 4: Results of fitting quasi-Lindley model to dataset of Table 3.

Method	$(\hat{\alpha}, \hat{\lambda})$	AIC	KS <i>p</i> value	AD <i>p</i> value	CVM <i>p</i> value
MLE	(0.000002, 0.2025)	638.6014	0.0442 0.9842	0.1914 0.9743	0.0301 0.9621
EM	(0.0118, 0.2013)	---	0.04198 0.9946	0.1858 0.9938	0.0279 0.9828

Note that checking the termination conditions of the EM process in each EM iteration results in very slow and time-consuming runs. Therefore, the EM algorithm has been tested many times to find a suitable constant for the number of iterations. In this way, we find that 5 iterations are sufficient.

Four measures bias (B), mean squared error (MSE), coverage probability (CP), and confidence interval length mean (CILM) for α and λ have been computed. The B and MSE for α are defined to be

$$\begin{aligned}
 B &= \frac{1}{r} \sum_{i=1}^r (\hat{\alpha}_i - \alpha), \\
 \text{MSE} &= \frac{1}{r} \sum_{i=1}^r (\hat{\alpha}_i - \alpha)^2, \\
 \text{CP} &= \frac{1}{r} \sum_{i=1}^r I(\alpha \in CI_{\alpha}(i)), \\
 \text{ILM} &= \frac{1}{r} \sum_{i=1}^r \text{length of } CI_{\alpha}(i),
 \end{aligned}
 \tag{36}$$

where $\hat{\alpha}_i$ shows the MLE/EM estimator in the run i and $CI_{\alpha}(i)$ shows an approximate asymptotic 95 percent confidence interval for α in the i th iteration (see the last paragraph of Section 3). Also, the indicator function I in (30) equals 1 when the real parameter falls inside the confidence interval and otherwise equals zero. These measures are defined for λ similarly. Tables 1 and 2 present the simulation results for uncensored data and censored data with censorship 0.2, respectively. The main observations from these tables are listed in the following:

- (i) The MSE decreases as sample size increases, for both MLE and EM estimators and both uncensored data and censored data which indicates that the MLE and EM estimators are consistent.

- (ii) The EM estimator outperforms the MLE in terms of the MSE.

- (iii) The results show higher CPs and lower CILMs for EM than MLE. Moreover, the CP increases and CILM decreases as sample size increases.

5. Application

Table 3 shows 100 waiting times of customers of a bank analyzed by Shanker [13]. The quasi-Lindley distribution was fitted to this dataset, and the parameters were estimated using the maximum likelihood method and EM. The “optim” function in the R language was used to calculate the MLE. Table 4 shows the results of the fitting. In terms of the KS, Anderson–Darling (AD), and Cramer–von Mises (CVM) statistics, both methods provide a good fit, but EM outperforms MLE in a close competition. The empirical and the fitted CDFs are shown in Figure 1(a) and also confirm a good fit. The histogram and estimated probability density function are also shown in Figure 1(b). Using the Hessian matrix calculated with the optim function, the variances of the MLE are estimated for the parameters, $\hat{v}(\hat{\alpha}) = 4 \times 10^{-12}$ and $\hat{v}(\hat{\lambda}) = 0.00020$. Using these variance estimates and standard normal quantiles, the 95% confidence intervals for α and λ are (0, 0.000006) and (0.1744, 0.2305), respectively. The left bound of the α confidence interval was a negative value; by the fact that $\alpha > 0$, it was set to 0.

To find the variances of the EM estimator of the parameters, the bootstrap method is used. In this way, $r = 1000$ samples are derived by the function “sample” of R. Then, for each sample, the EM estimates of the parameters are computed. The estimate of the variance of the EM estimators is approximated by the variance of these estimates which are $\hat{v}(\hat{\alpha}) = 0.0015$ and $\hat{v}(\hat{\lambda}) = 0.00023$. For each of the parameters, the 2.5% and 97.5% quantiles of the EM estimator can be considered as upper and lower bounds of the 95% confidence intervals. Then, the 95%

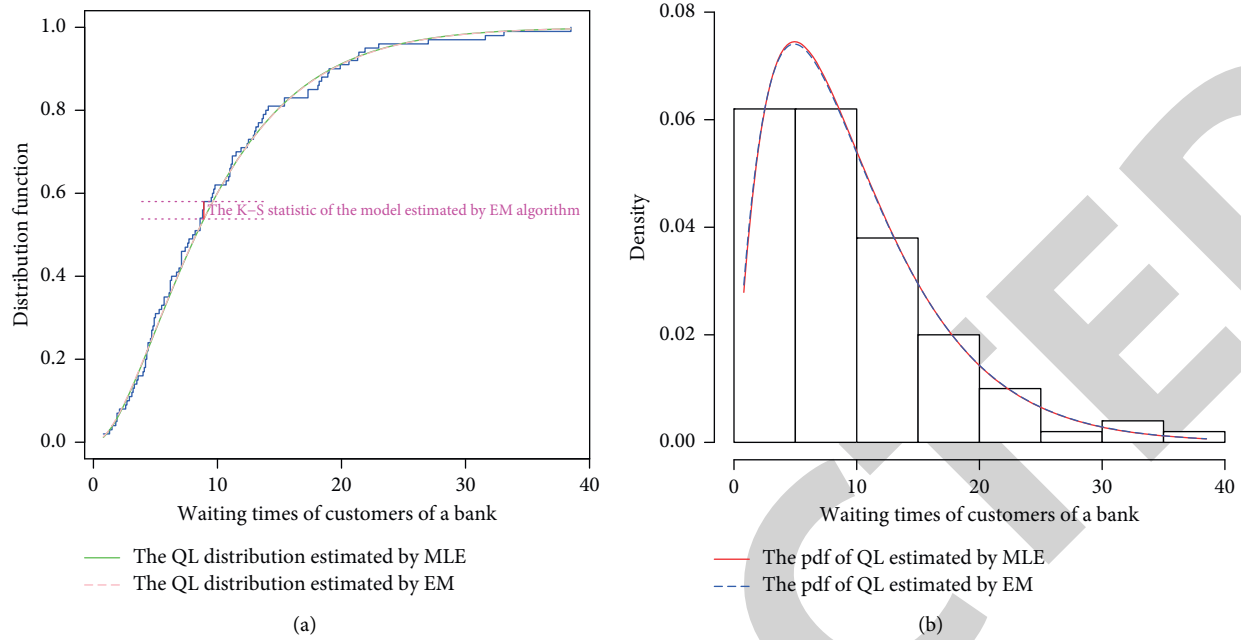


FIGURE 1: (a) The empirical distribution and fitted quasi-Lindley distribution for dataset of Table 3. (b) The histogram and the estimated probability density function for this dataset.

confidence intervals for α and λ are (0.00024, 0.12845) and (0.1750, 0.2325), respectively.

6. Conclusion

The quasi-Lindley distribution is a scale-invariant version of the Lindley distribution with a shape parameter α and a scale parameter λ and is a simple yet flexible model in reliability theory, survival analysis, management science, and many other fields. The MLE and EM approaches were investigated to estimate the parameters of this model. The simulation results show that the EM algorithm is better than the MLE for estimating the parameters for both uncensored and censored data.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by Researchers Supporting Project (RSP-2021/392), King Saud University, Riyadh, Saudi Arabia.

Supplementary Materials

Censored-MLE and EM for quasi Lindley. (*Supplementary Materials*)

References

- [1] R. Shanker and A. Mishra, "A quasi Lindley distribution," *African Journal of Mathematics and Computer Science Research*, vol. 6, no. 4, pp. 64–71, 2013.
- [2] D. V. Lindley, "Fiducial distributions and Bayes' theorem," *Journal of the Royal Statistical Society: Series B*, vol. 20, no. 1, pp. 102–107, 1958.
- [3] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B*, vol. 39, no. 1, pp. 1–22, 1977.
- [4] E. E. Elmahdy and A. W. Aboutahoun, "A new approach for parameter estimation of finite Weibull mixture distributions for reliability modeling," *Applied Mathematical Modelling*, vol. 37, no. 4, pp. 1800–1810, 2013.
- [5] J. Almhana, Z. Liu, V. Choulakian, and R. McGorman, "A recursive algorithm for gamma mixture models," in *Proceedings of the IEEE International Conference on Communications ICC 06*, vol. 1, pp. 197–202, Istanbul, Turkey, June 2006.
- [6] M. Bee, R. Benedetti, and G. Espa, "On maximum likelihood estimation of a Pareto mixture," *Computational Statistics*, vol. 28, no. 1, pp. 161–178, 2013.
- [7] I. Ghosh, G. G. Hamedani, N. Bansal, and M. Maadooliat, "On the mixtures of Weibull and Pareto (IV) distribution: an alternative to Pareto distribution," *Communications in Statistics-Theory and Methods*, vol. 47, no. 9, pp. 2073–2084, 2018.
- [8] N. Balakrishnan and S. Pal, "Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes," *Statistical Methods in Medical Research*, vol. 25, no. 4, pp. 1535–1563, 2016.
- [9] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley, New York, NY, USA, 1997.
- [10] K. L. Mengersen, C. P. Robert, and D. M. Titterton, *Mixtures: Estimation and Applications*, Wiley, New York, NY, USA, 2011.

- [11] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Annals of Statistics*, vol. 11, no. 1, pp. 95–103, 1983.
- [12] X.-L. Meng and D. B. Rubin, "Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm," *Journal of the American Statistical Association*, vol. 86, no. 416, pp. 899–909, 1991.
- [13] R. Shanker, "On generalized Lindley distribution and its applications to model lifetime data from biomedical science and engineering," *Insights in Biomedicine*, vol. 1, no. 2, 2016.

RETRACTED