Inference for Compound Truncated Poisson Log-Normal Model with Application to Maximum Precipitation Data

M. Z. Raqab[a,*], Debasis Kundu[c], M. A. Meraou[d]

[a] *Department of Statistics & OR, Kuwait University, Al-Shadadiyah, KUWAIT*

[c] *Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, Pin 208016, INDIA*

[d] *Laboratory of Statistics and Stochastic Processes, University of Djillali Liabes, BP 89, Sidi Bel Abbes 22000, ALGERIA*

Abstract

The main objective of this paper is to propose a new general family of distributions, namely compound truncated Poisson log-normal distribution of which log-normal distribution is a special case. The proposed model has three unknown parameters, and it can take variety of shapes. It can be used effectively in analyzing maximum precipitation data during a particular period of time obtained from different stations. It is assumed that the number of stations operate follows a zero-truncated Poisson random variables, and the daily precipitation follows a log-normal random variable. The maximum likelihood estimators can be obtained quite conveniently using Expectation-Maximization (EM) algorithm. Approximate maximum likelihood estimators are also

* *Corresponding author. E-mail addresses: mraqab@ju.edu.jo; mraqab4@gmail.com.*

derived. The associated confidence intervals can also be obtained from the observed Fisher information matrix. Simulation results have been performed to check the performance of the EM algorithm, and it is observed that the EM algorithm works quite well in this case. When we analyze the precipitation data set using the proposed model it is observed that the proposed model provides better fit than some of the existing models.

Keywords: Compound Poisson log-normal distribution; EM algorithm; maximum likelihood estimation; approximate maximum likelihood estimation; Fisher information; skew distribution.

AMS Subject Classification (2010): 62E15; 62F10.

1 Introduction

Rainfall modeling plays a significant role in prediction and forecasting purposes for agriculture, in pricing of weather derivatives, in hydrology and in risk and disaster management. Several attempts have been made in rainfall modeling for different regions, see for example Gentilucci et al. (2019), Dzupire et al. (2018), Odening et al. (2007) and the references cited therein. In this paper we have analyzed the maximum precipitation data during three winter months (December, 2018 - February, 2019) of of the United States. Here precipitation means total amount of rain fall plus melted snow, and it is measured in cm. The data have been obtained from 117 stations in the Global Historical Climatological Network of the United States. The data are available in the National Centers for Environmental Information (NCEI) (https://www.ncdc.noaa.gov/cdo-web/datatools/records). The NCEI is responsible for hosting and providing access to one of the most significant archives on Earth, with comprehensive oceanic, atmospheric, and geophysical data. The data have been obtained from 117 different such stations. One important point to note that the stations are not operational on everyday. They are closed on certain days for some reason or the other, and this number is random. Hence, the maximum values are recorded based on the available days.

Let D denote the total number of days we are interested in a particular time interval, and $Y_1, \ldots, Y_D$ denote the precipitation in those D days for a particular station. Out of those D days, the station is operational only on N days, say on $\{i_1, \ldots, i_N\}$, where N is a random quantity. Hence, the maximum precipitation $T = \max\{Y_{i_1}, \ldots, Y_{i_N}\}$ is observed. It is assumed that Y_i's are independent identically distributed (i.i.d.) log-normal random variables, and N is a zero truncated Poisson random variable and it is independent of Y_i's. Equivalently, we can express T as follows: $T = \max\{X_1, \ldots, X_N\}$, where $X_1, X_2, \ldots, X_N$ are i.i.d. log-normal random variables, N is same as defined above and all the random variables are independently distributed. We call this distribution as compound truncated Poisson log-normal (CTP-LN) distribution. It may be mentioned Bulmer (1974) has also defined a compound Poisson log-normal distribution, which is a discrete distribution and clearly it is different from CTP-LN distribution. The choice of the log-normal distribution as a baseline during modeling processes is motivated by their characteristics that make it suitable for the maximum precipitation data. The log-normal allows the resulting compound model to capture the right-skewness and heavy-tailed nature the data. For general families of distributions, one may refer to Alzaatreh et al. (2021).

It may be mentioned that the idea of compounding is quite common in the statistical literature. Marshal and Olkin (1997) used i.i.d. exponential random variables compounding with the geometric random variables. Since then quite a bit of work has been done along that line. Adamidis and Loukas (1998) introduced an exponential-geometric distribution by compounding an exponential distribution with a geometric distribution. Kus (2007) and Karlis (2009) considered the exponential-Poisson distribution. Barreto-Souza and Cribari-Neto (2009) provided generalized exponential-Poisson distribution and Bakouch et al. (2012) introduced exponentiated exponential-binomial distribution.

It is observed that the proposed CTP-LN distribution has three parameters, and because of the presence of the three parameters, it is a fairly flexible distribution. It has an absolute continuous distribution function and the log-normal distribution can be obtained as a special case. The probability density function (PDF) of a CTP-LN distribution can take variety of shapes, and it is positively skewed. The hazard function can be increasing, decreasing or unimodal depending on the parameter values.

The estimation of the unknown parameters is essentially the main problem while analyzing a data set. It is worth mentioning that the maximum likelihood estimators (MLEs) of the unknown parameters require solving three non-linear equations simultaneously. The Expectation-Maximization (EM) algorithm was originally suggested by Dempster et al. (1977) and used by Ng et al. (2002) and Raqab and Madi (2011) to compute the MLEs of all model parameters. Numerical methods such as Newton-Raphson (NR) method or some of its variants may be used to compute these MLEs. For computing the MLEs, one is concerned about the non-convergence issue and even if the convergence does exist, it may converge to a local maximum rather than the global maximum. The EM algorithm can be computationally more efficient compared to the NR method, especially for large data sets. This alternating procedure may reduce the number of iterations compared to the iterative calculations involved in other numerical methods. Considering these factors and the specific structure of our problem, the EM algorithm is the most appropriate choice. The confidence intervals of the unknown parameters are also developed using the arguments given in Louis (1982).

The rest of the paper is organized as follows. In Section 2 we describe the CTP-LN distribution and derive its distributional properties. The MLEs and approximate MLEs (AMLEs) of the model parameters are discussed in Section 3. Numerical simulation is performed and concluding results are presented in Section 4 for comparison purposes. One real data set representing the monthly maximum precipitation is analyzed in Section 5. Finally, we summarized our findings in this paper in Section 6.

2 Model Description and Its Distribution Forms

Suppose $\{X_1, X_2, \ldots,\}$ is a sequence of i.i.d. lognormal random variables with location parameter μ and scale parameter σ (LN(μ,σ)). Let N be a zero-truncated Poisson (ZTP) random variable with parameter θ, i.e.

$$P(N=k) = \frac{e^{-\theta}\theta^k}{k!(1-e^{-\theta})}; \qquad k = 1, 2, \ldots,$$

and N is independent of $X_1, X_2, \ldots, X_N$. Let us define the random variable

$$T_N = \max\{X_1, \ldots, X_N\}.$$

Then T_N is said to have a CTP-LN distribution, and it will be denoted by CTP-LN(μ, σ, θ). Now, we would like to derive the probability distribution of T_N. Let us denote the cumulative distribution function (CDF) and the probability density function (PDF) of a $N(0,1)$ variate by $\Phi(.)$ and $\phi(.)$, we observe that

$$\begin{aligned} P(T_N \leq t, N = k) &= P(T_N \leq t | N = k) P(T_N = k) \\ &= [P(X_1 \leq t]^k \frac{\theta^k e^{-\theta}}{k!(1-e^{-\theta})} \\ &= \frac{\theta^k}{k!(e^{\theta}-1)} \left[\Phi\left(\frac{\ln t - \mu}{\sigma}\right)\right]^k. \end{aligned}$$

Note that

$$\lim_{\theta \to 0} P(T_N \leq t, N = k) = \begin{cases} \Phi\left(\frac{\ln\ t - \mu}{\sigma}\right) & \text{if} \quad k = 1, \\ 0 & \text{if} \quad k > 1. \end{cases}$$

The joint density of T_N and N becomes

$$f_{T_N,N}(t,k) = \frac{\theta^k}{\sigma(e^{\theta}-1)(k-1)!} \frac{1}{t} \left[\Phi\left(\frac{\ln\ t - \mu}{\sigma}\right)\right]^{k-1} \phi\left(\frac{\ln\ t - \mu}{\sigma}\right). \tag{1}$$

The marginal CDF of T_N for $0 < t < \infty$, is given by

$$\begin{aligned} F_{T_N}(t) = P(T_N \leq t) &= \sum_{k=1}^{\infty} P(T_N \leq t, N = k) = \frac{1}{e^{\theta}-1} \sum_{k=1}^{\infty} \left[\Phi\left(\frac{\ln\ t - \mu}{\sigma}\right)\right]^k \frac{\theta^k}{k!} \\ &= \frac{e^{\theta \Phi\left(\frac{\ln\ t - \mu}{\sigma}\right)} - 1}{e^{\theta}-1}. \end{aligned} \tag{2}$$

The corresponding marginal PDF can be obtained as

$$f_{T_N}(t) = \frac{\theta}{\sigma(e^{\theta}-1)} \frac{1}{t}\, e^{\theta \Phi\left(\frac{\ln\ t - \mu}{\sigma}\right)}\, \phi\left(\frac{\ln\ t - \mu}{\sigma}\right). \tag{3}$$

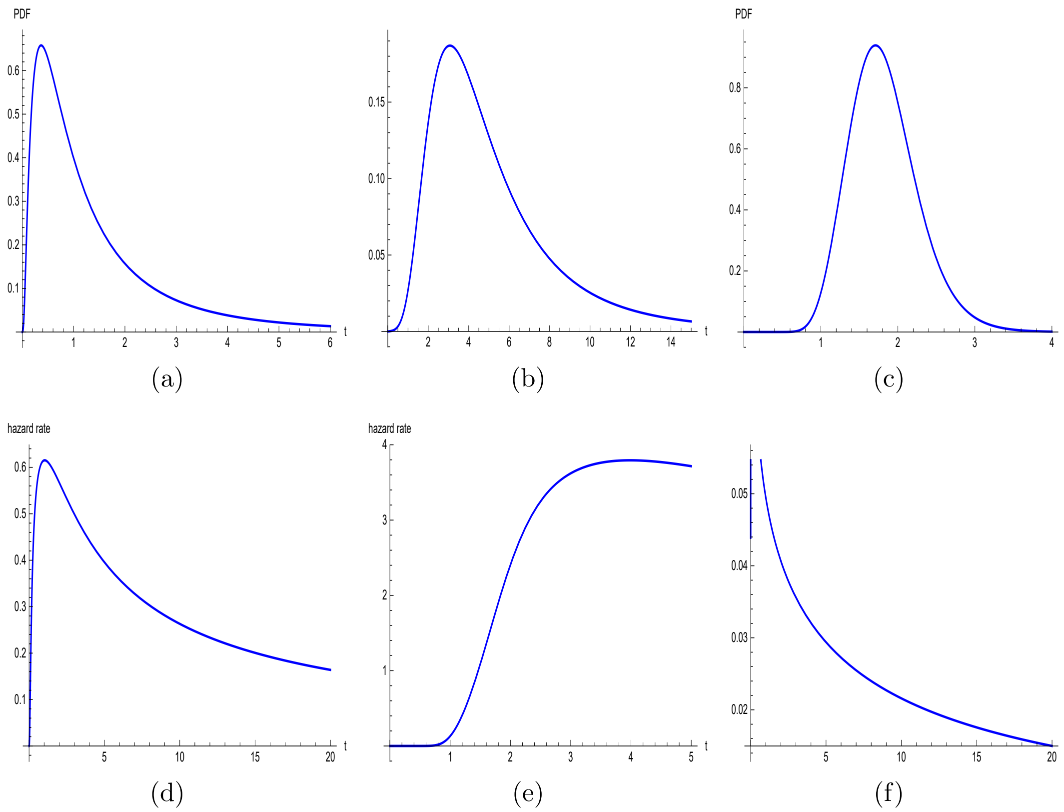

Figure 1: PDF and HR plots of CTP-LN(μ, σ, θ) distribution for different parameter values: (a) (0,1,0) (b) (0,1,10) (c) (0.5, 0.25, 1) (d) (0, 1, 1) (e) (0.5, 0.25, 1) (f) (0.25, 3, 5).

Hence, from (2), we immediately obtain that T_N tends to be $LN(\mu, \sigma)$ variate as $\theta \to 0$. It is clear that the CDF of a CTP-LN distribution is an infinite mixture of the CDFs of sample maximum from log-normal with Poisson probability weights. The plots of the PDFs and hazard rates (HRs) of the CTP-LN distribution for different parameter values are displayed in Figure 1. It can take variety of shapes and it can have heavy tails.

It is immediate that the hazard function of CTP-LN distribution is an increasing, decreasing, and increasing-decreasing function depending the values of μ, σ and θ. Now we derive the moments of the proposed CTP-LN distribution. The k-th moment of T can be expressed as

$$\begin{aligned} E(T_N^k) &= \frac{\theta}{\sigma(e^\theta - 1)} \int_0^\infty t^{k-1}\, e^{\theta\, \Phi\left(\frac{\ln\, t - \mu}{\sigma}\right)}\, \phi\left(\frac{\ln\, t - \mu}{\sigma}\right)\, dt \\ &= \frac{\theta e^{k\mu}}{e^\theta - 1}\, \eta_k(\sigma, \theta), \end{aligned}$$

where

$$\eta_k(\sigma,\theta) = \int_0^1 e^{\theta y + k\sigma\Phi^{-1}(y)}\, dy.$$

This in turns out that the mean and variance of T_N can be obtained as

$$E(T_N) = \frac{\theta e^{\mu}}{e^{\theta}-1}\eta_1(\sigma,\theta), \text{ and } Var(T_N) = \frac{\theta e^{2\mu}}{e^{\theta}-1}\left(\eta_2(\sigma,\theta) - \frac{\theta}{e^{\theta}-1}\eta_1^2(\sigma,\theta)\right).$$

The skewness (SKW) and kurtosis (KURT) measures of T_N are also derived to be

$$\begin{aligned} SK(\sigma,\theta) &= \frac{\theta^{-1/2}}{(e^{\theta}-1)^{-1/2}}\left[\eta_2(\sigma,\theta) - \frac{\theta}{e^{\theta}-1}\eta_1^2(\sigma,\theta)\right]^{-3/2} \\ &\quad \left\{\eta_3(\sigma,\theta) - 3\frac{\theta}{e^{\theta}-1}\eta_1(\sigma,\theta)\eta_2(\sigma,\theta) + 2\frac{\theta^2}{(e^{\theta}-1)^2}\eta_1^3(\sigma,\theta)\right\}, \end{aligned}$$

and

$$\begin{aligned} KURT(\sigma,\theta) &= \frac{(e^{\theta}-1)}{\theta}\left[\eta_2(\sigma,\theta) - \frac{\theta}{e^{\theta}-1}\eta_1^2(\sigma,\theta)\right]^{-2} \\ &\quad \left\{\eta_4(\sigma,\theta) - 4\frac{\theta\eta_1(\sigma,\theta)\eta_3(\sigma,\theta)}{e^{\theta}-1} - 6\frac{\theta^3\eta_1^2(\sigma,\theta)}{(e^{\theta}-1)^2}\left(\eta_2(\sigma,\theta) - \frac{\theta}{e^{\theta}-1}\eta_1^2(\sigma,\theta)\right)\right. \\ &+ \left. 3\frac{\theta^3}{(e^{\theta}-1)^3}\eta_1^4(\sigma,\theta)\right\}, \end{aligned}$$

respectively. Furthermore, from (1) and (3), it can be easily checked that the probability mass function (PMF) of the conditional distribution of N given $T_N = t$ is

$$P(N = k|T_N = t) = \frac{\left[\theta\ \Phi\left(\frac{\ln\ t-\mu}{\sigma}\right)\right]^{k-1}}{(k-1)!}e^{-\theta\ \Phi\left(\frac{\ln\ t-\mu}{\sigma}\right)}, k = 1, 2, \ldots. \tag{4}$$

Therefore, $N-1|T_N = t$ is a Poisson random variate with mean $\theta\ \Phi((\ln\ t-\mu)/\sigma)$ and then

$$E(N|T_N = t) = 1 + \theta\ \Phi\left(\frac{\ln\ t-\mu}{\sigma}\right). \tag{5}$$

To assess the model's skewness and kurtosis, we provide Table 1 below for further illustration. Table 1 shows the skewness (SKW) and kurtosis (KURT) for different values of μ, σ and θ. It is observed from Table (1) that the distribution of T_N is positively skewed and it has heavier tails. The highest kurtosis value, the more extreme the tails compared to a normal distribution.

Table 1: The SKW and KURT of the CTP-LN model for different parameters μ, σ and θ.

μ	σ	θ	Skw	Kurt
1.5	1	0.5	4.8726	52.0263
		1.0	4.6393	47.304
		1.5	4.4491	43.6495
		2.0	4.2956	40.8193
1.75	1.5	0.5	2.3837	10.6896
		1.0	2.2724	9.8395
		1.5	2.1831	9.2082
		2.0	2.1142	8.7531
2.0	1.75	0.5	1.9126	6.7163
		1.0	1.8184	6.1769
		1.5	1.7431	5.7869
		2.0	1.6859	5.5187
2.5	2.0	0.5	1.6007	4.6423
		1.0	1.5163	4.2582
		1.5	1.4488	3.9883
		2.0	1.3982	3.8124

3 Parameter Estimation Methods

Here in this section, we develop the MLEs of the unknown model parameters based on a observed sample $\boldsymbol{t} = \{t_1, t_2, \ldots, t_m)$ from CTP-LN(μ, σ, θ). In addition, we develop approximate MLEs of the unknown parameters which might be quite useful in obtaining quick estimates for these parameters for practical purposes. The approximate estimates can also be used as initial values for finding the MLEs.

3.1 EM Algorithm

Initially, the MLEs of $\mu, \sigma,$ and θ can be investigated by maximizing the log-likelihood function

$$l(\mu, \sigma, \theta|\boldsymbol{t}) = \sum_{i=1}^{m} \ln f_{T_N}(t_i; \mu, \sigma, \theta), \tag{6}$$

where $f_{T_N}(t; \mu, \sigma, \theta)$ is the right hand side of (3). Clearly, the MLEs can be computed by solving numerically three non-linear equations simultaneously. To avoid the convergence problems for the solutions of these equations, it is quite useful to adopt the EM algorithm for obtaining the MLEs of the model parameters. For this, we consider the underlying problem as a missing value problem. The combination $(\boldsymbol{t}, \boldsymbol{n})$, forms the complete data set for the variates $(\boldsymbol{T}, \boldsymbol{N})$, where $\boldsymbol{t}$ and $\boldsymbol{n}$ are observed and missing data, respectively. For simplicity, let us write $\boldsymbol{\vartheta} = (\mu, \sigma, \theta)$. Based on the complete data, $(\boldsymbol{t}, \boldsymbol{n})$ and (1), the likelihood function is given by

$$L_c(\boldsymbol{\vartheta}) \propto \frac{\prod_{i=1}^{m} \theta^{n_i}}{\sigma^m (e^\theta - 1)^m} \prod_{i=1}^{m} [\Phi(\xi_i)]^{n_i - 1} \prod_{i=1}^{m} \phi(\xi_i). \tag{7}$$

It follows from (7), the log-likelihood function based on $(\boldsymbol{t}, \boldsymbol{n})$ is

$$l_c(\boldsymbol{\vartheta}) \propto \left(\sum_{i=1}^{m} n_i \right) \ln\ \theta - m \ln(e^\theta - 1) - m \ln\ \sigma + \sum_{i=1}^{m} (n_i - 1) \ln\ \Phi(\xi_i) + \sum_{i=1}^{m} \ln\ \phi(\xi_i), \tag{8}$$

where $\xi = (\ln\ t - \mu)/\sigma$. Firstly, in the E-step, we compute the 'pseudo' log-likelihood function by replacing n_i by its expectation given in (5) (say, E_i) . Therefore the pseudo log-likelihood function is

$$\begin{aligned} l_{ps}(\boldsymbol{\vartheta}) \ &\propto \ \sum_{i=1}^{m} E_i(\boldsymbol{\vartheta}) \ln\ \theta - m \ln(e^\theta - 1) - m \ln\ \sigma \\ &+ \ \sum_{i=1}^{m} (E_i(\boldsymbol{\vartheta}) - 1) \ln\ \Phi(\xi_i) + \sum_{i=1}^{m} \ln\ \phi(\xi_i). \end{aligned} \tag{9}$$

Secondly, the M-step involves the maximization of the pseudo-likelihood function (9). Let us assume that at the k-th stage, the estimate of $\boldsymbol{\vartheta} = (\mu, \sigma, \theta)$ is $\boldsymbol{\vartheta}^{(k)} = (\mu^{(k)}, \sigma^{(k)}, \theta^{(k)})$. Then

$\boldsymbol{\vartheta}^{(\boldsymbol{k+1})}$ can be obtained by maximizing

$$\begin{aligned} l^*_{ps}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^{(\boldsymbol{k})}) &\propto \left(\sum_{i=1}^{m} E_i(\boldsymbol{\vartheta}^{(k)})\right) \ln\ \theta - m \ln(e^{\theta} - 1) - m \ln\ \sigma \\ &+ \sum_{i=1}^{m} (E_i(\boldsymbol{\vartheta}^{(k)}) - 1) \ln\ \Phi(\xi_i) + \sum_{i=1}^{m} \ln\ \phi(\xi_i), \end{aligned} \tag{10}$$

with respect to μ, σ and θ. Here $E_i(\boldsymbol{\vartheta}^{(k)})$ is the missing n_i value at the k-th stage and in the EM algorithm, the choice of $E_i(\boldsymbol{\vartheta}^{(k)})$ is obtained using (5) by replacing $\boldsymbol{\vartheta}$ by $\boldsymbol{\vartheta}^{(k)}$ as follows:

$$E_i(\boldsymbol{\vartheta}^{(k)}) = 1 + \theta^{(k)}\ \Phi(\xi_i^{(k)}),$$

where $\xi_i^{(k)} = (\ln\ t_i - \mu^{(k)})/\sigma^{(k)}$.

In the M-step, $\mu^{(k+1)}$ and $\sigma^{(k+1)}$ are obtained by solving the following equations:

$$\sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} (E_i(\boldsymbol{\vartheta}^{(k)}) - 1) \frac{\phi(\xi_i)}{\Phi(\xi_i)} = 0, \tag{11}$$

and

$$\sum_{i=1}^{m} \xi_i^2 - \sum_{i=1}^{m} (E_i(\boldsymbol{\vartheta}^{(k)}) - 1) \xi_i \frac{\phi(\xi_i)}{\Phi(\xi_i)} - m = 0, \tag{12}$$

while $\theta^{(k+1)}$ is computed by maximizing the profile pseudo log-likelihood function with respect to θ, namely,

$$\eta(\theta) = \sum_{i=1}^{m} E_i(\boldsymbol{\vartheta}^{(k)}) \ln\ \theta - m \ln(e^{\theta} - 1). \tag{13}$$

For $\sum_{i=1}^{m} E_i(\boldsymbol{\vartheta}^{(k)}) > m$, it can be shown that Eq.(13) has a unique maximum and it can be obtained numerically by solving the non-linear equation:

$$\frac{\sum_{i=1}^{n} E_i(\boldsymbol{\vartheta}^{(k)})}{\theta} - \frac{m\ e^{\theta}}{e^{\theta} - 1} = 0. \tag{14}$$

The steps of the EM algorithm for computing the unknown parameters of CTP-LN model are presented in Appendix A.

As a consequence of that, we continue the iteration process until the convergence is achieved. In the last iteration, the EM-algorithm produces the MLEs $\widehat{\mu}$, $\widehat{\sigma}$ and $\widehat{\theta}$ of μ, σ and θ, respectively. It can be shown that the CTP-LN is a regular family. Hence, the consistency and the asymptoic normality properties of the MLEs automatically hold. The asymptotic normal distribution of the MLEs of the parameters can be obtained as follows:

$$\sqrt{n}\left(\hat{\mu}-\mu, \hat{\sigma}-\sigma, \hat{\theta}-\theta\right) \to N_3(\mathbf{0}, I^{-1}),$$

where I is the expected Fisher information matrix. For more convenience, we use the observed Fisher information based on the EM-algorithm (see, for example, Louis (1982)). Let $C = ((c_{ij}))$ and $b = ((b_{ij}))$ be the Hessian matrix of the negatives of the second derivatives and the gradient vector of the associated pseudo log-likelihood $l_{pc}(\vartheta)$. That is, $I_{obs} = C - \boldsymbol{b}\,\boldsymbol{b}^T$

$$\mathbf{C} = \begin{bmatrix} \frac{\partial^2 l_{ps}}{\partial \mu^2} & \frac{\partial^2 l_{ps}}{\partial \mu \partial \sigma} & \frac{\partial^2 l_{ps}}{\partial \mu \partial \theta} \\ \frac{\partial^2 l_{ps}}{\partial \sigma \partial \mu} & \frac{\partial^2 l_{ps}}{\partial \sigma^2} & \frac{\partial^2 l_{ps}}{\partial \sigma \partial \theta} \\ \frac{\partial^2 l_{ps}}{\partial \theta \partial \mu} & \frac{\partial^2 l_{ps}}{\partial \theta \partial \sigma} & \frac{\partial^2 l_{ps}}{\partial \theta^2} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \frac{\partial l_{ps}}{\partial \mu} & \frac{\partial l_{ps}}{\partial \sigma} & \frac{\partial l_{ps}}{\partial \theta} \end{bmatrix}^T,$$

with

$$\begin{aligned}
\frac{\partial^2 l_{ps}}{\partial \mu^2} &= -\frac{1}{\sigma^2}\left[m + \sum_{i=1}^{m}(E_i(\boldsymbol{\vartheta}^{(k)})-1)\frac{\phi(\xi_i)}{\Phi^2(\xi_i)}\left(\xi_i\Phi(\xi_i)+\phi(\xi_i)\right)\right], \\
\frac{\partial^2 l_{pc}}{\partial \mu \partial \sigma} &= -\frac{2}{\sigma^2}\sum_{i=1}^{m}\xi_i - \frac{1}{\sigma^2}\sum_{i=1}^{m}(E_i(\boldsymbol{\vartheta}^{(k)})-1)\frac{\phi(\xi_i)}{\Phi^2(\xi_i)}\left(\xi_i^2\Phi(\xi_i)+\xi_i\phi(\xi_i)-\Phi(\xi_i)\right), \frac{\partial^2 l_{ps}}{\partial \mu \partial \theta} = 0, \\
\frac{\partial^2 l_{ps}}{\partial \sigma^2} &= \frac{m}{\sigma^2} - \frac{3}{\sigma^2}\sum_{i=1}^{m}\xi_i^2 - \frac{1}{\sigma^2}\sum_{i=1}^{m}(E_i(\boldsymbol{\vartheta}^{(k)})-1)\frac{\phi(\xi_i)}{\Phi(\xi_i)}\left(\xi_i\phi(\xi_i)-\Phi(\xi_i)(\xi_i^2+\xi_i+1)\right), \\
\frac{\partial^2 l_{ps}}{\partial \theta^2} &= -\frac{\sum_{i=1}^{m}E_i(\boldsymbol{\vartheta}^{(k)})}{\theta^2} + \frac{me^{-\theta}}{(1-e^{-\theta})^2}, \frac{\partial^2 l_{ps}}{\partial \sigma \partial \theta} = 0, \\
\frac{\partial l_{ps}}{\partial \mu} &= -\frac{1}{\sigma}\left[\sum_{i=1}^{m}(E_i(\boldsymbol{\vartheta}^{(k)})-1)\frac{\phi(\xi_i)}{\Phi(\xi_i)} - \sum_{i=1}^{m}\xi_i\right], \\
\frac{\partial l_{ps}}{\partial \sigma} &= -\frac{1}{\sigma}\left[m + \sum_{i=1}^{m}(E_i(\boldsymbol{\vartheta}^{(k)})-1)\xi_i\frac{\phi(\xi_i)}{\Phi(\xi_i)} - \sum_{i=1}^{m}\xi_i^2\right], \\
\frac{\partial l_{ps}}{\partial \theta} &= \frac{\sum_{i=1}^{m}E_i(\boldsymbol{\vartheta}^{(k)})}{\theta} - \frac{me^{\theta}}{e^{\theta}-1}.
\end{aligned}$$

Furthermore, $(1-\alpha)100\%$ confidence intervals (CIs) for μ, σ and θ can be established using the observed Fisher information matrix and asymptotic normal distribution of the MLEs.

3.2 Approximation-Based Methods in the EM Algorithm

Let $Z_{1:m} < Z_{2:m} < \ldots < Z_{m:m}$, denote the order statistics of a random sample from standard normal distribution with CDF $\Phi(.)$ and PDF $\phi(.)$. Let $\Phi_{k:m}$ and $\phi_{k:m}$ denote the CDF and PDF of $Z_{k:m}$, respectively, for $k = 1$ to m. The joint likelihood of $(\boldsymbol{T}, \boldsymbol{N})$ can be written as a product of two functions, L_1 and L_2, where

$$L_1 \propto \frac{1}{\sigma^m} \prod_{i=1}^{m} [\Phi(\xi_{i:m})]^{n^{(i)}-1} \prod_{i=1}^{m} \phi(\xi_{i:m}), \quad \text{and} \quad L_2 \propto \prod_{i=1}^{m} \frac{\theta^{n^{(i)}}}{(e^\theta - 1)},$$

where $\xi_{i:m} = (\ln t_{i:m} - \mu)/\sigma$ and $n^{(i)}$ is the corresponding variate to $t_{i:m}$ which is a ZTP distribution. Arguments similar to those in subsection (3.1), we apply the EM-algorithm for obtaining the estimate of θ by solving

$$\frac{\sum_{i=1}^{m} E_i(\boldsymbol{\vartheta}^{(k)})}{\theta} - \frac{me^\theta}{e^\theta - 1} = 0,$$

where $E_i(\boldsymbol{\vartheta}^{(k)})$ is computed based on (5). Once the estimate of θ is obtained, we approximate the estimates of μ and σ by modifying the likelihood equations induced based on the likelihood function, L_1. To this end, we propose two methods of approximation. The first method is based on approximating the reversed hazard rate $Q(\xi_{i:m}) = \phi(\xi_{i:m})/\Phi(\xi_{i:m})$ in (11) and (12) by its expansion using the Taylor series around the quantile functions at η_i, where

$$\eta_i = E(Z_{i:m}) \approx \Phi^{-1}(\nu_i),$$

where $\nu_i = E(U_{i:m}) = i/(m+1)$, the expected value of the i-th order statistic from a sample of size m from the uniform $U(0,1)$ distribution. This type of approximation is used in the estimation and prediction problems, see for example, Raqab (1997) and Asgharzadeh et al. (2013). Let $Q'(.)$ be the derivative of $Q(.)$. Now, by keeping only the first two terms, $Q(\xi_{i:m})$

can be approximated as follows:

$$\begin{aligned} Q(\xi_{i:m}) &\approx Q(\eta_i) + Q'(\eta_i)(\xi_{i:m} - \eta_i). \\ &= a_i + b_i \xi_{i:m}. \end{aligned} \tag{15}$$

The constants can be found to be

$$a_i = Q(\eta_i) - \eta_i Q'(\eta_i), \quad \text{and} \quad b_i = Q'(\eta_i).$$

By applying (15), the likelihood equations, (11) and (12) are modified to be

$$\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)})\xi_{i:m} - \sum_{i=1}^{m} \rho_i(\boldsymbol{\vartheta}^{(k)}) = 0, \tag{16}$$

and

$$\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)})\xi_{i:m}^2 - \sum_{i=1}^{m} \rho_i(\boldsymbol{\vartheta}^{(k)})\xi_{i:m} - m = 0, \tag{17}$$

where

$$\Delta_i(\boldsymbol{\vartheta}^{(k)}) = 1 - (E_i(\boldsymbol{\vartheta}^{(k)}) - 1)b_i,$$

and

$$\rho_i(\boldsymbol{\vartheta}^{(k)}) = (E_i(\boldsymbol{\vartheta}^{(k)}) - 1)a_i.$$

Upon simplifying (16), we readily obtain $\mu^{(k+1)}$ expressed in terms of $\sigma^{(k+1)}$. That is,

$$\mu^{(k+1)} = \frac{\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)}) \ln\ t_{i:m}}{\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)})} - \sigma^{(k+1)} \frac{\sum_{i=1}^{m} \rho_i(\boldsymbol{\vartheta}^{(k)})}{\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)})}, \tag{18}$$

By substituting (18) into (17), we deduce the following quadratic equation:

$$m\sigma^2 + D_1\, \sigma + D_2 = 0, \tag{19}$$

where

$$D_1 = \sum_{i=1}^{m} \rho_i(\boldsymbol{\vartheta}^{(k)}) \ln\ t_{i:m} - \frac{\left(\sum_{i=1}^{m} \rho_i(\boldsymbol{\vartheta}^{(k)})\right)\left(\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)}) \ln\ t_{i:m}\right)}{\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)})},$$

and

$$D_2 = \frac{\left(\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)})\ \ln\ t_{i:m}\right)^2}{\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)})} - \sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)})\ \ln^2\ t_{i:m}.$$

Therefore, solving Eq. (19) yields

$$\sigma^{(k+1)} = \frac{-D_1 \pm \sqrt{D_1^2 - 4mD_2}}{2m}. \tag{20}$$

On differentiating $Q(\xi) = \frac{d}{d\xi}\ \ln\ \Phi(\xi)$, we obtain

$$Q'(\xi) = \frac{-\phi(\xi)}{[\Phi(\xi)]^2} \left[\xi \Phi(\xi) + \phi(\xi)\right].$$

This derivative is negative since

$$\begin{aligned} \xi\ \Phi(\xi) &= \xi \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2}}\ dx \\ &\geq \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi}}\ x\ e^{-\frac{x^2}{2}}\ dx \\ &= -\int_{\infty}^{\xi} Q'(x)\ dx \\ &= -\phi(\xi). \end{aligned}$$

This implies that $Q'(\xi) < 0$ and consequently, $a_i > 0$ and $b_i < 0$, for $i = 1, 2, \ldots, m$. On using the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned} \left(\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)}) \ln^2 t_{i:m}\right)\left(\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)})\right) &\geq \left(\sum_{i=1}^{m} \Delta_i^2(\boldsymbol{\vartheta}^{(k)}) \ln^2 t_{i:m}\right)^2 \\ &\geq \left(\sum_{i=1}^{m} \Delta_i(\boldsymbol{\vartheta}^{(k)}) \ln t_{i:m}\right)^2, \end{aligned}$$

which implies that $D_2 < 0$ and then one of the two roots in (20) drops out. Therefore, the approximate MLE (AMLE-I) of σ can be found using the following iterative identity:

$$\sigma^{(k+1)} = \frac{-D_1 + \sqrt{D_1^2 - 4mD_2}}{2m}. \tag{21}$$

Now, we introduce the following lemma which is useful in approximating $\phi(\xi_{i:m})/\Phi(\xi_{i:m})$ appearing in the likelihood equations in (11) and (12).

Lemma 1: For $i \leq m$,

$$E\left(\frac{\phi(Z_{i:m})}{\Phi(Z_{i:m})}\right) = \frac{1}{i-1} \sum_{j=i}^{m} E(Z_{i:m}), \tag{22}$$

where $E(Z_{i:m})$ is the expected value of the i-th order statistic from a sample of size m having standard normal distribution.

Proof: The left hand side of (22) can be rewritten as

$$\begin{aligned} E\left(\frac{\phi(Z_{i:m})}{\Phi(Z_{i:m})}\right) &= \frac{m!}{(i-1)!(m-i)!} \int_{-\infty}^{\infty} \phi(\xi)[\Phi(\xi)]^{i-2} \left[1-\Phi(\xi)\right]^{m-i} \phi(\xi)\, d\xi \\ &= \frac{m}{i-1} \int_{-\infty}^{\infty} \phi(\xi)\ \phi_{i-1:m-1}(\xi) d\xi \\ &= \frac{m}{i-1} E\left(\phi(Z_{i-1:m-1})\right). \end{aligned} \tag{23}$$

Upon setting $\psi(\xi) = \phi'(\xi)/\phi(\xi)$, we rewrite (23) as

$$E\left(\phi(Z_{i-1:m-1})\right) = \int_{-\infty}^{\infty} \phi(\xi)\ d\Phi_{i-1:m-1}(\xi),$$

where dg is derivative of g. Upon integrating by parts, we immediately obtain

$$\begin{aligned} E\left(\phi(Z_{i-1:m-1})\right) &= -\int_{-\infty}^{\infty} \psi(\xi)\phi(\xi)\Phi_{i-1:m-1}(\xi)\, d\xi \\ &= -\sum_{k=i-1}^{m-1} \frac{(m-1)!}{k!(m-1-k)!} \int_{-\infty}^{\infty} \psi(\xi)[\Phi(\xi)]^{k}[1-\Phi(\xi)]^{m-1-k}\phi(\xi)\, d\xi \\ &= -\frac{1}{m}\sum_{j=i}^{m} \frac{m!}{(j-1)!(m-j)!} \int_{-\infty}^{\infty} \psi(\xi)[\Phi(\xi)]^{j-1}[1-\Phi(\xi)]^{m-j}\phi(\xi)\, d\xi \\ &= -\frac{1}{m}\sum_{j=i}^{m} E\psi(Z_{j:m}) \\ &= \frac{1}{m}\sum_{j=i}^{m} E(Z_{j:m}). \end{aligned} \tag{24}$$

The expected values of normal order statistics can be computed easily and some selected values are available in Arnold et al. (2008). The main point in this approach is that the likelihood equations (11) and (12) involve a messy term, $\phi(\xi_{i:m})/\Phi(\xi_{i:m})$ and it is impossible to obtain an explicit form for the MLE. In this technique, this term is replaced by its respective expected value. This allows us to solve the intended likelihood equations and obtain simple estimates of μ and σ.

By using (23) and (24), we may modify the likelihood equations in (11) and (12) as

$$\sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} (E^{(i)}(\boldsymbol{\vartheta}^{(k)}) - 1)c_{i,m} = 0, \tag{25}$$

and

$$\sum_{i=1}^{m} \xi_i^2 - \sum_{i=1}^{m} (E^{(i)}(\boldsymbol{\vartheta}^{(k)}) - 1)\xi_i c_{i,m} - m = 0, \tag{26}$$

where

$$c_{i,m} = \frac{1}{i-1}\sum_{j=i}^{m} E(Z_{j:m}).$$

It follows from (25), the approximate MLE of μ can be obtained based on the following identity:

$$\mu^{(k+1)} = \frac{\sum_{i=1}^{m} \ln t_{i:m} - \sigma^{(k+1)} \sum_{i=1}^{m} \beta_i(\boldsymbol{\vartheta}^{(k)})}{m}. \tag{27}$$

On using (25), Eq. (26) can be expressed as

$$m\sigma^2 + D_3\sigma - \frac{m-1}{m} D_4 = 0. \tag{28}$$

where

$$D_3 = \sum_{i=1}^{m} \beta_i(\boldsymbol{\vartheta}^{(k)}) \ln\ t_{i:m} - \frac{1}{m}\left(\sum_{i=1}^{m} \beta_i(\boldsymbol{\vartheta}^{(k)})\right)\left(\sum_{i=1}^{m} \ln\ t_{i:m}\right), D_4 = (\sum_{i=1}^{m} \ln\ t_{i:m})^2,$$

and

$$\beta_i(\boldsymbol{\vartheta}^{(k)}) = (E_i(\boldsymbol{\vartheta}^{(k)}) - 1)\ c_{i,m}.$$

On solving Eq. (28) in σ, we readily have the approximate MLE (AMLE-II) of σ

$$\sigma^{(k+1)} = \frac{-D_3 + \sqrt{D_3^2 + 4(m-1)D_4}}{2m}, \tag{29}$$

The procedures for approximating the unknown parameters μ and σ of the CTP-LN model using AMLE-I and AMLE-II can be found in Appendices B and C, respectively.

4 Simulation results

Here in this section, we perform a Monte Carlo simulation study to examine the performance of the MLE of $\boldsymbol{\vartheta} = (\mu, \sigma, \theta)$ obtained based on the EM algorithm and their related approximations for different sample sizes, and for different parameter values. We assess the three different methods of estimation, namely, the maximum likelihood estimation based on EM-algorithm, modifications to the likelihood Eq.'s (11)-(12) by approximating the term involving the reversed hazard rate function either by Taylor series expansion (Type-I approximation) or replacing this term by its respective expected value (Type II-approximation). All numerical computations are conducted using the R package.

In this simulation, the values of CTP-LN parameters are considered as ($\mu = 1.0, \sigma = 1.0, \theta = 1.0$), ($\mu = 0.5, \sigma = 1.0, \theta = 1.5$) and ($\mu = 2.5, \sigma = 10.0, \theta = 1.0$). Here we kept σ fixed, and we have changed μ and θ and consider different sample sizes, $n = 20(10)100$. For a given set of parameter values for μ, σ and θ. We compute the values of the average estimates (AEs) and mean square errors (MSEs) of MLE, AMLE-I, AMLE-II. In each case of estimation, we stop the EM-algorithm when the absolute difference between two successive estimates is less than 10^{-5}. Based on 1000 replications, we compute all the estimates and 95% simulated CIs as well as their average lengths (ALs) and coverage probabilities (CPs). The results of AEs, MSEs, ALs and CPs are presented in Tables 2-7.

It is clear from the numerical experiments that the MSEs presented in Tables 2-4 tend to decrease as the sample size increases. Also it is noticed that as μ and θ increase, the MSEs increase. By considering the MSE as an optimality criterion, the AMLE-I performs better than AMLE-II when both compared with the MLE. For more clarification, Figures 2 and 3 ensures the superiority of AMLE-I to AMLE-II. For comparing the CIs of the CTP-LN parameters, it appears from Tables 5-7 that although the CIs derived based on the AMLE-II is more efficient than the ones based on MLEs and AMLEs-I in the sense of the AL, both CIs based on MLEs and AMLE-I's have highest CPs. Another aspect can be seen, the CIs established based on the MLEs and AMLEs-I are very close in terms of AL criterion.

Table 2: The AE and the associated MSEs of the MLE, AMLE-I and AMLE-II for the CTP-LN distribution when $\mu = 1.0$, $\sigma = 1.0$ and θ= 1.0.

n	Method	μ		σ		θ	
		AE	MSE	AE	MSE	AE	MSE
20	MLE	1.0174	0.0401	0.9682	0.0263	0.9951	0.0445
	AMLE-I	1.0441	0.0481	0.9664	0.0242	0.9246	0.0823
	AMLE-II	1.0337	0.0562	0.8770	0.0355	0.9833	0.1356
30	MLE	1.0125	0.0276	0.9755	0.0165	1.0016	0.0300
	AMLE-I	1.0209	0.0319	0.9737	0.0168	0.9421	0.0592
	AMLE-II	1.0297	0.0352	0.8982	0.0259	0.9805	0.0901
40	MLE	1.0120	0.0199	0.9836	0.0130	1.0001	0.0221
	AMLE-I	1.0241	0.0245	0.9838	0.0125	0.9555	0.0475
	AMLE-II	1.0233	0.0256	0.9020	0.0214	0.9997	0.0725
50	MLE	1.0114	0.0165	0.9768	0.0113	1.0071	0.0184
	AMLE-I	1.0306	0.0205	0.9847	0.0108	0.9550	0.0363
	AMLE-II	1.0254	0.0229	0.927	0.0191	0.9754	0.0571
60	MLE	1.0052	0.0144	0.9898	0.0088	1.0032	0.0148
	AMLE-I	1.0272	0.0154	0.9819	0.0087	0.9509	0.0295
	AMLE-II	1.0189	0.0173	0.9082	0.0162	0.9572	0.0436
70	MLE	0.9993	0.0116	0.9858	0.0076	1.0027	0.0145
	AMLE-I	1.0184	0.0129	0.9911	0.0075	0.9587	0.0280
	AMLE-II	1.0397	0.0153	0.9132	0.0135	0.9538	0.0360
80	MLE	1.0051	0.0112	0.9916	0.0068	1.0069	0.0129
	AMLE-I	1.0183	0.0124	0.9900	0.0061	0.9624	0.0247
	AMLE-II	1.0332	0.0128	0.9118	0.0128	0.9551	0.0358
90	MLE	1.0003	0.0093	0.9927	0.0061	0.9981	0.0101
	AMLE-I	1.0228	0.0107	0.9967	0.0054	0.9467	0.0212
	AMLE-II	1.0369	0.0116	0.9108	0.0132	0.9535	0.0313
100	MLE	0.9992	0.0088	0.9938	0.0053	1.0014	0.0094
	AMLE-I	1.0214	0.0093	0.9919	0.0050	0.9571	0.0207
	AMLE-II	1.0363	0.0112	0.9138	0.0119	0.9575	0.0274

Table 3: The AE and the associated MSEs of the MLEs, AMLE-I's and AMLE-II's for the CTP-LN distribution when $\mu = 0.5$, $\sigma = 1.0$ and $\theta = 1.5$.

n	Method	μ		σ		θ	
		AE	MSE	AE	MSE	AE	MSE
20	MLE	0.5242	0.0377	0.9603	0.0277	1.4937	0.0632
	AMLE-I	0.5842	0.0528	0.9599	0.0268	1.3223	0.1227
	AMLE-II	0.5174	0.0579	0.8364	0.0489	1.5948	0.3891
30	MLE	0.5120	0.0264	0.9728	0.0175	1.4957	0.0420
	AMLE-I	0.5679	0.0352	0.9772	0.0183	1.3583	0.0836
	AMLE-II	0.5154	0.0427	0.8374	0.0419	1.6186	0.2985
40	MLE	0.5059	0.0200	0.9797	0.0129	1.4923	0.0326
	AMLE-I	0.5529	0.0261	0.9797	0.0127	1.3510	0.0694
	AMLE-II	0.4931	0.0339	0.8495	0.0338	1.6480	0.2360
50	MLE	0.5083	0.0161	0.9810	0.0101	1.4893	0.0250
	AMLE-I	0.5583	0.0228	0.9755	0.0105	1.3634	0.0581
	AMLE-II	0.4878	0.0266	0.8540	0.0297	1.6567	0.2084
60	MLE	0.5056	0.0129	0.9884	0.0080	1.4987	0.0226
	AMLE-I	0.5534	0.0174	0.9872	0.0093	1.3699	0.0477
	AMLE-II	0.5017	0.0223	0.8537	0.0282	1.6258	0.1581
70	MLE	0.5104	0.0112	0.9899	0.0072	1.4952	0.0187
	AMLE-I	0.5528	0.0163	0.9841	0.0072	1.3762	0.0427
	AMLE-II	0.4848	0.0192	0.8500	0.0284	1.6392	0.1497
80	MLE	0.5072	0.0100	0.9871	0.0061	1.5024	0.0170
	AMLE-I	0.5467	0.0140	0.9916	0.0061	1.3759	0.0389
	AMLE-II	0.4814	0.0172	0.8535	0.0267	1.6511	0.1346
90	MLE	0.5042	0.0085	0.9865	0.0061	1.4948	0.0152
	AMLE-I	0.5467	0.0124	0.9855	0.0059	1.3732	0.0383
	AMLE-II	0.4764	0.0153	0.8537	0.0260	1.6543	0.1212
100	MLE	0.5053	0.0080	0.9974	0.0051	1.4926	0.0129
	AMLE-I	0.5483	0.0117	0.9860	0.0056	1.3782	0.0343
	AMLE-II	0.4736	0.0133	0.8577	0.0246	1.6454	0.0972

Table 4: The AE and the associated MSEs of the MLEs, AMLE-I's and AMLE-II's for the CTP-LN distribution when $\mu = 2.5$, $\sigma = 10.0$ and $\theta = 1.0$.

n	Method	μ		σ		θ	
		AE	MSE	AE	MSE	AE	MSE
20	MLE	2.9611	4.5408	9.1576	1.7426	0.9711	0.0562
	AMLE-I	2.9349	4.8906	9.1249	2.7116	0.9412	0.0799
	AMLE-II	2.9395	5.2386	9.1117	3.3857	0.9479	0.1241
30	MLE	2.9298	2.7560	9.3030	1.2129	0.9750	0.0331
	AMLE-I	2.9040	3.3187	9.2728	1.8289	0.9303	0.0578
	AMLE-II	2.9091	3.5734	9.2567	2.2584	0.9327	0.0921
40	MLE	2.8306	2.1533	9.4099	0.8741	0.9819	0.0317
	AMLE-I	2.8052	2.2814	9.3687	1.2444	0.9606	0.0466
	AMLE-II	2.8085	2.7110	9.3671	1.9921	0.9629	0.0683
50	MLE	2.7328	1.6319	9.5023	0.6712	0.9896	0.0229
	AMLE-I	2.7103	1.9905	9.4867	1.0335	0.9550	0.0376
	AMLE-II	2.7000	2.0512	9.4857	1.5903	0.9527	0.0534
60	MLE	2.6955	1.4563	9.5810	0.4853	0.9807	0.0183
	AMLE-I	2.6629	1.6538	9.5503	0.8606	0.9574	0.0295
	AMLE-II	2.6699	1.7960	9.5533	0.9812	0.9517	0.0442
70	MLE	2.6652	1.1111	9.5853	0.4468	0.9852	0.0154
	AMLE-I	2.6337	1.3815	9.5466	0.6815	0.9607	0.0262
	AMLE-II	2.6336	1.4528	9.5501	1.4188	0.9648	0.0366
80	MLE	2.6466	1.0391	9.6541	0.3612	0.9925	0.0152
	AMLE-I	2.6083	1.1861	9.5975	0.6179	0.9730	0.0244
	AMLE-II	2.6156	1.4127	9.6016	1.2883	0.9684	0.0339
90	MLE	2.6246	0.8935	9.6754	0.2976	0.9971	0.0126
	AMLE-I	2.5868	1.0742	9.6257	0.5720	0.9643	0.0209
	AMLE-II	2.5637	1.2165	9.6175	1.1648	0.9636	0.0334
100	MLE	2.5819	0.8593	9.6760	0.2735	0.9959	0.0112
	AMLE-I	2.5616	0.9461	9.6110	0.5429	0.9667	0.0198
	AMLE-II	2.5521	1.1842	9.6020	1.1263	0.9697	0.0287

Table 5: The ALs and the CPs of the MLEs, AMLE-I's and AMLE-II's for the CTP-LN distribution when μ = 1.0, σ = 1.0 and θ= 1.0.

n	Method	μ		σ		θ	
		AL	CP	AL	CP	AL	CP
20	MLE	0.7269	0.896	0.6101	0.895	1.0739	0.981
	AMLE-I	0.7278	0.886	0.6129	0.886	1.0407	0.917
	AMLE-II	0.6863	0.831	0.5064	0.752	1.0424	0.851
30	MLE	0.6044	0.921	0.5075	0.913	0.8772	0.987
	AMLE-I	0.6061	0.911	0.5106	0.911	0.8598	0.916
	AMLE-II	0.5621	0.850	0.4156	0.743	0.8507	0.844
40	MLE	0.5296	0.926	0.4449	0.924	0.7593	0.989
	AMLE-I	0.5286	0.905	0.4451	0.921	0.7437	0.910
	AMLE-II	0.4887	0.866	0.3605	0.715	0.7440	0.861
50	MLE	0.4745	0.921	0.3986	0.938	0.6787	0.985
	AMLE-I	0.4751	0.901	0.4001	0.928	0.6671	0.912
	AMLE-II	0.4384	0.866	0.3245	0.724	0.6616	0.834
60	MLE	0.4337	0.933	0.3643	0.928	0.602	0.989
	AMLE-I	0.4344	0.923	0.3658	0.936	0.6123	0.909
	AMLE-II	0.4025	0.872	0.2972	0.706	0.6087	0.842
70	MLE	0.4038	0.941	0.3392	0.934	0.5730	0.991
	AMLE-I	0.4028	0.913	0.3392	0.937	0.5653	0.900
	AMLE-II	0.3730	0.858	0.2756	0.701	0.5634	0.830
80	MLE	0.3772	0.938	0.3169	0.946	0.5373	0.984
	AMLE-I	0.3771	0.901	0.3375	0.932	0.5297	0.910
	AMLE-II	0.3499	0.858	0.2582	0.662	0.5293	0.834
90	MLE	0.3558	0.926	0.2990	0.944	0.5080	0.988
	AMLE-I	0.3569	0.915	0.3005	0.943	0.5006	0.912
	AMLE-II	0.3308	0.859	0.2446	0.680	0.4971	0.846
100	MLE	0.3379	0.939	0.2840	0.944	0.4811	0.987
	AMLE-I	0.3373	0.894	0.2840	0.941	0.4739	0.901
	AMLE-II	0.3145	0.869	0.2328	0.668	0.4705	0.808

Table 6: The ALs and the CPs of the MLEs, AMLE-I's and AMLE-II's for the CTP-LN distribution when $\mu = 0.5$, $\sigma = 1.0$ and $\theta = 1.5$.

n	Method	μ		σ		θ	
		AL	CP	AL	CP	AL	CP
20	MLE	0.7202	0.916	0.6151	0.908	1.2495	0.983
	AMLE-I	0.7235	0.866	0.6237	0.896	1.1962	0.901
	AMLE-II	0.6320	0.784	0.4242	0.580	1.2668	0.752
30	MLE	0.5887	0.915	0.5031	0.905	1.0242	0.971
	AMLE-I	0.5957	0.854	0.5129	0.915	0.9841	0.903
	AMLE-II	0.5161	0.796	0.3419	0.514	1.0530	0.721
40	MLE	0.5182	0.934	0.4430	0.927	0.8853	0.987
	AMLE-I	0.5188	0.885	0.4466	0.932	0.8557	0.883
	AMLE-II	0.4550	0.771	0.3045	0.499	0.9044	0.719
50	MLE	0.4649	0.925	0.3972	0.928	0.7938	0.983
	AMLE-I	0.4671	0.881	0.4020	0.936	0.7661	0.854
	AMLE-II	0.4021	0.800	0.2701	0.433	0.8126	0.693
60	MLE	0.4256	0.924	0.3638	0.931	0.7226	0.977
	AMLE-I	0.4283	0.912	0.3684	0.930	0.7006	0.866
	AMLE-II	0.3721	0.808	0.2481	0.402	0.7440	0.712
70	MLE	0.3976	0.943	0.3398	0.928	0.6689	0.979
	AMLE-I	0.3933	0.851	0.3384	0.916	0.6494	0.833
	AMLE-II	0.3456	0.791	0.2307	0.388	0.6882	0.679
80	MLE	0.3722	0.941	0.3181	0.942	0.6262	0.986
	AMLE-I	0.3712	0.871	0.3194	0.922	0.6089	0.863
	AMLE-II	0.3230	0.793	0.2156	0.327	0.6444	0.684
90	MLE	0.3492	0.937	0.2985	0.942	0.5914	0.981
	AMLE-I	0.3493	0.861	0.3005	0.943	0.5732	0.838
	AMLE-II	0.3048	0.781	0.2034	0.282	0.6076	0.667
100	MLE	0.3334	0.944	0.2849	0.940	0.5596	0.982
	AMLE-I	0.3323	0.889	0.2858	0.951	0.5453	0.843
	AMLE-II	0.2885	0.803	0.1933	0.245	0.5749	0.697

5 Application to Maximum Precipitation

Here, we present the analysis of a real data set representing the highest precipitation during winter season (December 2018 - February 2019) from 117 stations of the Global Historical Climatological Network. The data are available in the National Centers for Environmental

Table 7: The ALs and the CPs of the MLEs, AMLEs1 and AMLEs2 for the CTP-LN distribution when $\mu = 2.5$, $\sigma = 10.0$ and $\theta = 1.0$.

n	Method	μ		σ		θ	
		AL	CP	AL	CP	AL	CP
20	MLE	5.2444	0.793	5.7199	0.912	1.0594	0.962
	AMLE-I	5.6043	0.776	6.4733	0.904	1.0704	0.907
	AMLE-II	4.8327	0.706	4.8949	0.854	1.0354	0.840
30	MLE	4.3555	0.802	4.7681	0.925	0.8669	0.967
	AMLE-I	4.5978	0.789	5.3055	0.912	0.8880	0.910
	AMLE-II	4.0628	0.754	4.1145	0.866	0.8553	0.853
40	MLE	3.8200	0.805	4.1981	0.934	0.7531	0.972
	AMLE-I	4.0331	0.798	4.6455	0.924	0.7790	0.936
	AMLE-II	3.5569	0.767	3.6218	0.875	0.7431	0.884
50	MLE	3.4558	0.812	3.8124	0.935	0.6765	0.974
	AMLE-I	3.6003	0.806	4.1462	0.932	0.6957	0.939
	AMLE-II	3.2214	0.785	3.2966	0.892	0.6634	0.903
60	MLE	3.1875	0.819	3.5203	0.953	0.6134	0.975
	AMLE-I	3.3151	0.809	3.8154	0.940	0.6391	0.954
	AMLE-II	2.9338	0.797	3.0055	0.916	0.6087	0.915
70	MLE	2.9515	0.839	3.2706	0.954	0.5714	0.975
	AMLE-I	3.0802	0.816	3.5447	0.947	0.5950	0.952
	AMLE-II	2.7389	0.808	2.8081	0.918	0.5627	0.927
80	MLE	2.7854	0.845	3.0871	0.957	0.5319	0.963
	AMLE-I	2.8744	0.814	3.3049	0.954	0.5512	0.956
	AMLE-II	2.5693	0.807	2.6434	0.922	0.5252	0.926
90	MLE	2.6323	0.869	2.9254	0.965	0.5026	0.971
	AMLE-I	2.7082	0.825	3.1152	0.951	0.5194	0.962
	AMLE-II	2.4161	0.816	2.4841	0.927	0.4962	0.940
100	MLE	2.4939	0.885	2.7723	0.962	0.4785	0.981
	AMLE-I	2.5780	0.851	2.9644	0.954	0.4943	0.966
	AMLE-II	2.3079	0.832	2.3760	0.931	0.4704	0.948

Information (NCEI) (https://www.ncdc.noaa.gov/cdo-web/datatools/records). The NCEI is responsible for hosting and providing access to one of the most significant archives on Earth, with comprehensive oceanic, atmospheric, and geophysical data. The stations are not operational in some days for some reason or the other, and hence, the maximum values are reported based on the available days only. The data points comprise $(t_i, n_i), i = 1, 2, \ldots, 117$ where t_i is the observed maximum precipitation and n_i is the number of days which the ith station is operational. The maximum monthly precipitation (rain, melted snow, etc.)

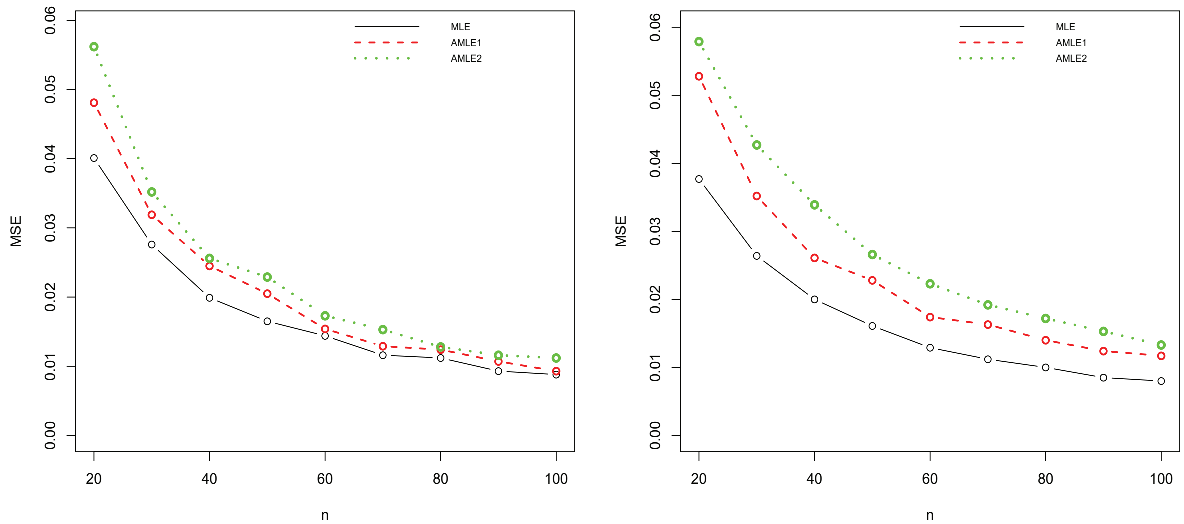

Figure 2: MSE of $\widehat{\mu}$ for μ=1, σ=1 and θ=1 (left) and MSE of $\widehat{\mu}$ for μ=0.5, σ=1 and θ=1.5 (right).

is recorded in inches. Before progressing further, let us provide some basic statistics of the observed value of T_N, it is presented in Table 5.

Table 8: Basic statistics of monthly highest precipitation data set

Size (m)	Mean	Median	Std.Dev.	Q_1	Q_3
117	3.294	2.800	1.974	2.000	4.450

Based on the rationale given above, we have opted to fit the proposed CTP-LN model to analyze the given data set. First, we display the plot of the scaled total time on test (TTT) transform in Figure 4. It is well known that the scaled TTT transform plot provides an indication about the shape of the hazard function. For example, if the plot is a concave (convex) function then it indicates that the hazard function is an increasing (decreasing) function. Therefore, the CTP-LN distribution can initially be an appropriate model to fit this data set.

Upon completing 11 iterations of the NR iterative method, we obtained the MLE of $\boldsymbol{\vartheta}$ =

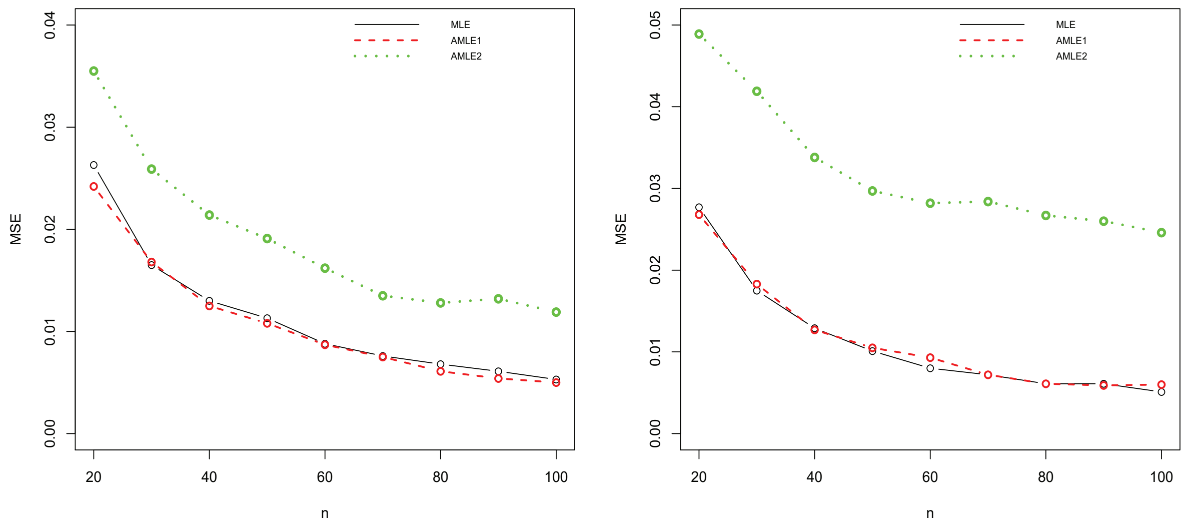

Figure 3: MSE of $\widehat{\sigma}$ for μ=1, σ=1 and θ=1 (left) and MSE of $\widehat{\sigma}$ for μ=0.5, σ=1 and θ=1.5 (right).

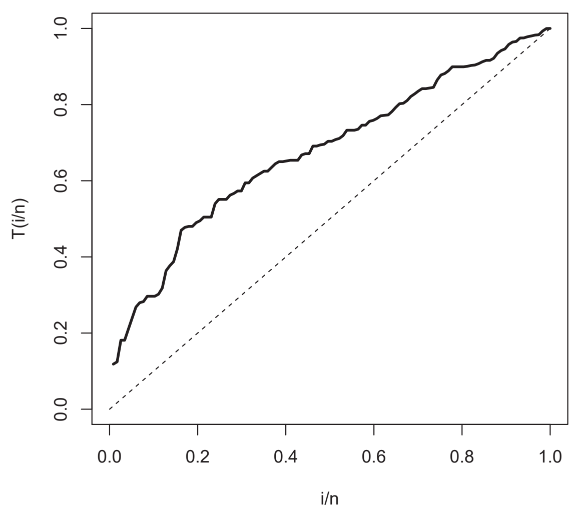

Figure 4: Scaled TTT plot of the monthly highest precipitation.

(μ, σ, θ) with corresponding log-likelihood (ll) values as follows:

$$\hat{\boldsymbol{\vartheta}}_{MLE} = (0.5422, 0.7021, 2.6193), \quad \text{and} \quad \text{ll}_{MLE} = -230.479.$$

Using the identical initial values and maintaining the same level of accuracy, the EM algorithm converged in 8 iterations yielding the values:

$$\hat{\boldsymbol{\vartheta}}_{ML} = (0.5428, 0.7020, 2.6157), \text{ and } \text{ll}_{ML} = -230.478.$$

The AMLE-I, AMLE-II of $\boldsymbol{\vartheta}$ and their corresponding ll are computed to be

$$\hat{\boldsymbol{\vartheta}}_{AM_1} = (0.5496, 0.7081, 2.6209), \ \hat{\boldsymbol{\vartheta}}_{AM_2} = (0.5534, 0.7119, 2.6324),$$

and

$$\text{ll}_{AM_1} = -230.509, \ \text{ll}_{AM_2} = -230.562.$$

The associated 95% CIs of μ, σ, and θ based on MLE, AMLE-Type I and AMLE-Type II are also established, respectively, as follows

$$(0.4403, 0.6453), \ (0.6129, 0.7911), \ (2.2990, 2.9324),$$

$$(0.4470, 0.6521), \ (0.6167, 0.7994), \ (2.3036, 2.9381),$$

and

$$(0.4508, 0.6559), \ (0.6190, 0.8047), \ (2.3139, 2.9508).$$

To verify whether the proposed EM algorithm actually converges to the MLEs, we have tried the EM algorithm with some other initial estimates also, and it converges to the same set of estimates. Moreover, we have conducted an extensive grid search method with a grid size 0.001, and it is noticed that they coincide, which guarantees that the estimates obtained via the proposed EM algorithm converges to the MLEs.

The extended exponential-geometric $EEG(\beta, \nu)$ distribution is fitted to the data set (see, Adamidis et al. (2005)). The EM algorithm is used to compute the MLEs of the unknown parameters. The MLEs of β and ν, the associated 95% CIs and the corresponding log-likelihood (ll) values are provided as follows

$$\hat{\beta} = 9.5234(\mp 0.7774), \ \hat{\nu} = 0.7815(\mp 0.0388), \ \text{ll} = -233.052.$$

Further, We fit the exponentiated Weibull-Poisson $EWP(\alpha, \beta, \gamma, \theta)$ distribution introduced by Mahmoudi and Sepahdar (2013) to the data set. The EM algorithm is used to obtain the MLEs of α, β, γ, and θ, associated CIs and log-likelihood (ll). These values are

$$\begin{aligned} \hat{\alpha} &= 133.911(\mp 24.264), \quad \hat{\beta} = 180.300(\mp 18.932), \quad \hat{\gamma} = 0.3026(\mp 0.0101), \\ \hat{\theta} &= 3.8517(\mp 0.852), \quad \text{ll} = -231.359. \end{aligned}$$

The LN distribution is a special case of the CTP-LN when θ tends to be 0. Let us fit the LN distribution to the data set. The MLEs of μ and σ based on the current data set and their 95% CIs as well as the corresponding log-likelihood (ll) are obtained to be

$$\hat{\mu} = 1.0042(\mp 0.1037), \quad \hat{\sigma} = 0.6497(\mp 0.0832), \quad \text{ll} = -233.068.$$

We fit normal distribution to the data set. Using the EM algorithm we obtain the following estimates of μ, and σ, and the associated 95% CIs. The corresponding log-likelihood (ll) value also has been reported below

$$\hat{\mu} = 3.2932(\mp 0.3563), \quad \hat{\sigma} = 1.9662(\mp 0.2519), \quad \text{ll} = -245.115.$$

For checking the model validity, we compute the Akaike Information Criterion (AIC), Akaike Information Criterion correction (AICc), Hannan–Quinn Information Criterion (HQIC), Bayesian Information Criterion (BIC) and Kolmogorov-Smirnov (K-S) distances of the empirical distribution function and the fitted distribution functions with associated p-values. Table 6 represents the results of all these measures. Therefore, these values indicate that the three-parameter CTP-LN distribution fits the data set well comparing with other distributions. Clearly, the CTP-LN distribution is a good alternative model comparing with other fitted models. The empirical and fitted distribution functions are presented in Figure 5. In Figure 5, we have plotted the empirical survival function (ESF) and fitted survival functions for the different models.

Another aspect of inference is the testing problem about the parameter θ. In this context,

Table 9: The goodness of fit tests for monthly highest precipitation data set.

Model	AIC	AICc	HQIC	BIC	K-S	p-value
CTP-LN	466.956	467.168	470.320	475.242	0.0467	0.9602
EEG	470.104	470.209	472.346	475.628	0.0669	0.6699
EWP	470.718	471.075	475.203	481.766	0.0644	0.7153
LN	470.136	470.241	472.378	475.660	0.0766	0.4971
Normal	494.230	494.335	496.472	499.754	0.1270	0.0457

the generalized likelihood ratio test can be used to test whether θ is 0 (LN) or not for this data set. Specifically, we wish test the hypotheses

$$H_0 : \theta = 0, \quad vs. \quad H_1 : \theta \neq 0.$$

Under the null hypothesis, the MLEs of the unknown parameters, and the corresponding log-likelihood value are provided as

$$\widetilde{\mu} = 1.0042, \quad \widetilde{\sigma} = 0.6497, \quad l_{H_0} = -233.068.$$

Therefore, the test statistic in this case becomes $2(l_{H_0 \cup H_1} - l_{H_0}) = 5.18$. Under the null hypothesis, θ is in the boundary, the standard results do not work and it follows by Theorem 3 of Self and Liang (1987) that

$$2(l_{H_0 \cup H_1} - l_{H_0}) \to \frac{1}{2} + \frac{1}{2}\chi_1^2.$$

Since p-value $= 0.0114$ and less than the significance level of 5 %, which leads us to reject the null hypothesis and the LN distribution cannot be used to analyze this data set.

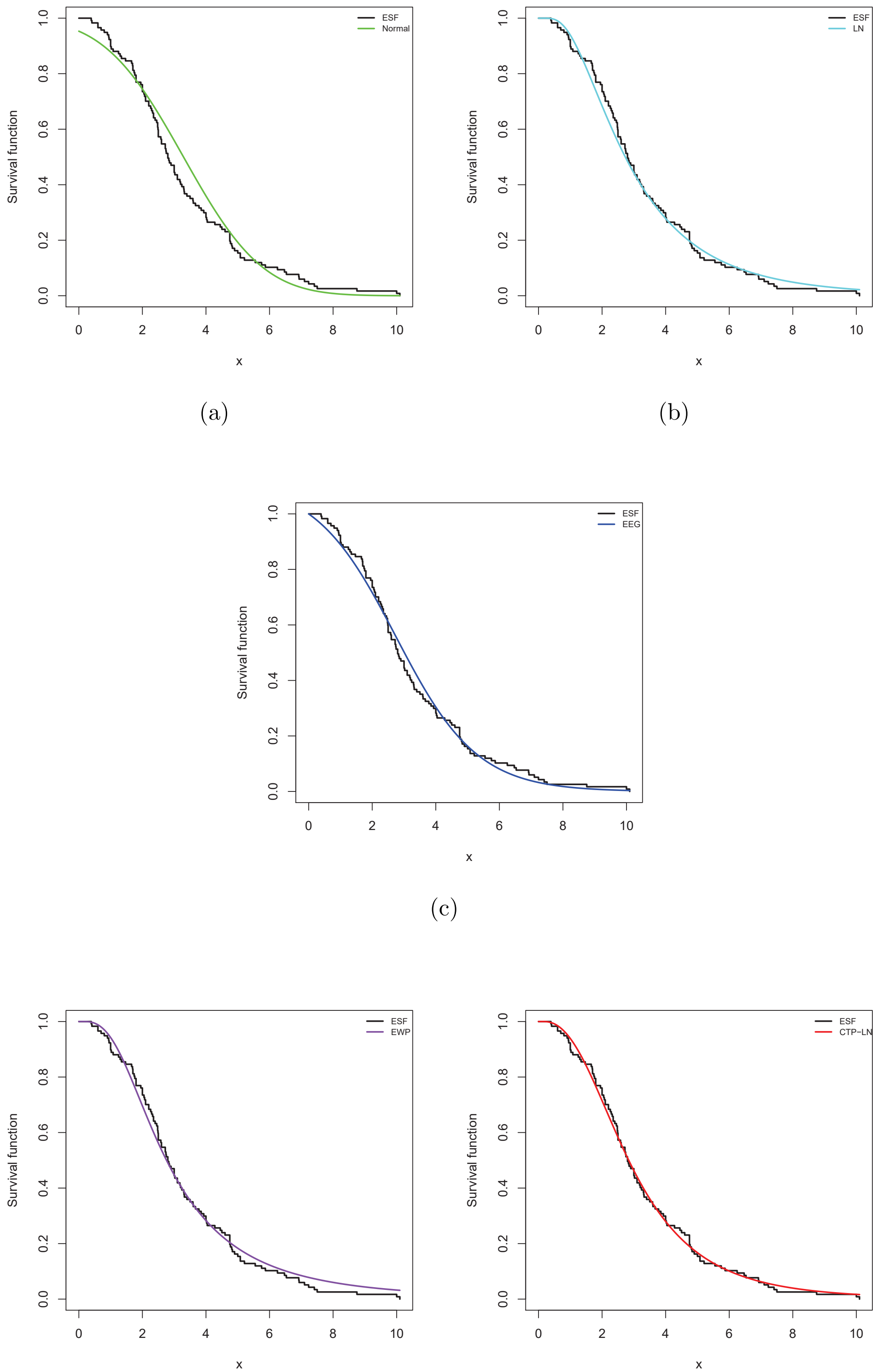

Figure 5: The ESF and fitted survival functions for different distributions: (a) Normal (b) LN (c) EEG (d) EWP (e) CTP-LN models for the monthly highest precipitation.

6 Conclusion

In this article, we have analyzed the maximum precipitation data during the Winter months of (Dec. 2018, Jan 2019 and Feb 2019) obtained from 117 different stations in US based on a new probability model named as CTP-LN. It is observed that the proposed model is a very flexible probability model and its PDF can take variety of shapes. The MLEs of the unknown parameters cannot be obtained in closed form and we have proposed to use EM algorithm to compute the MLEs. Extensive simulations have been performed and it is observed that the proposed EM algorithm and two of its approximations work quite well. The proposed model fits the monthly maximum precipitation data very well and it fits better than some of the well known models like Weibull, Extended Exponential, Exponentiated Weibull, log-normal and normal. It is expected that the proposed model can be used in analyzing other data which are obtained from maximization. It may be mentioned that in the present paper we have not used the information of the location of the stations. It might be of interest to incorporate this information in the model itself through covariates. More work is needed along that direction.

Declarations

- Authors' Contributions: MR performed the methodology, formal analysis, writing-original draft, DK managed the conceptualization, supervision, writing-review and editing, MM carried out the data curation, investigation, coding.

- Competing interests: The authors declare that they do not have any relevant financial or non-financial competing interests.

- Availability of Data and Materials: The data set analyzed in this study is available at https://www.ncdc.noaa.gov/cdo-web/datatools/records and will also be available upon a request from the corresponding author.

- Funding: No funding was obtained for this study.

Appendices:

In this section, we provide detailed descriptions of the pseudo codes for obtaining MLEs using an EM-type algorithm. Additionally, we include the corresponding approximations.

Appendix A: Pseudo code of EM algorithm for CTP-LN

- Given M an integer, denoting the maximum number of iterations allowed.
- Given ϵ for the stopping the iteration. If the absolute difference between the two successive log-likelihood values is less than ϵ, then we stop the iteration. We take $\epsilon = 10^{-6}$.
- Given m denote the sample size.
- Set the data vector $\{t_1, ..., t_m\}$.
- Set the initial values $\mu^{(0)}$, $\sigma^{(0)}$ and $\theta^{(0)}$.

  for (k in 1 to M){

  for (i in 1 to m) {

  Find the missing value $E_i^{(k)}$ as follows:

  $E_i^{(k)} = 1+\theta^k\Phi(\xi_i)$, where $\xi_i = (\ln t_i - \mu^k)/\sigma^k$ and $\Phi(.)$ denotes the CDF of standar normal distribution N(0,1).

  }

- Compute $\mu^{(k+1)}$, $\sigma^{(k+1)}$ and $\theta^{(k+1)}$ as a solution of (11), (12) and (14) respectively, by using Newton-Raphson or fixed point method.
- Compute the log-likelihood function $l^*_{ps}(\mu^{(k+1)}, \sigma^{(k+1)}, \theta^{(k+1)})$.
- If $|l^*_{ps}(\mu^{(k+1)}, \sigma^{(k+1)}, \theta^{(k+1)}) - l^*_{ps}(\mu^{(k)}, \sigma^{(k)}, \theta^{(k)})| < \epsilon$ we stop the iteration, else we continue until convergence.

  }

Appendix B: Pseudo code of EM algorithm for CTP-LN based on type I approximation

- Given M an integer, denoting the maximum number of iterations allowed.
- Given ϵ for the stopping the iteration. If the absolute difference between the two successive log-likelihood values is less than ϵ, then we stop the iteration. We take $\epsilon = 10^{-6}$.
- Given m denote the sample size.
- Set the data vector $\{t_1, \ldots, t_m\}$.
- Set the initial values $\mu^{(0)}$, $\sigma^{(0)}$ and $\theta^{(0)}$.

  for (k in 1 to M){

    for (i in 1 to m) {

    Find the missing value $E_i^{(k)}$ as follows:

    $E_i^{(k)} = 1 + \theta^k \Phi(\xi_i)$, where $\xi_i = (\ln t_i - \mu^k)/\sigma^k$ and $\Phi(.)$ denotes the CDF of standar normal distribution N(0,1).

    Find $\Delta_i(\boldsymbol{\vartheta}^{(k)})$ and $\rho_i(\boldsymbol{\vartheta}^{(k)})$ as: $\Delta_i(\boldsymbol{\vartheta}^{(k)}) = 1 - (E_i(\boldsymbol{\vartheta}^{(k)}) - 1)b_i$ and $\rho_i(\boldsymbol{\vartheta}^{(k)}) = (E_i(\boldsymbol{\vartheta}^{(k)}) - 1)a_i$.

    }

- Compute D_1 and D_2.
- Compute $\mu^{(k+1)}$ and $\sigma^{(k+1)}$ as given in Eq.'s (18) and (21) and $\theta^{(k+1)}$ as a solution of equation (12) using Newton-Raphson or fixed point method.
- Compute the log-likelihood function $l^*_{ps}(\mu^{(k+1)}, \sigma^{(k+1)}, \theta^{(k+1)})$.
- If $|l^*_{ps}(\mu^{(k+1)}, \sigma^{(k+1)}, \theta^{(k+1)}) - l^*_{ps}(\mu^{(k)}, \sigma^{(k)}, \theta^{(k)})| < \epsilon$ we stop the iteration, else we continue until convergence.

  }

Appendix C: Pseudo code of EM algorithm for CTP-LN based on type II approximation

- Given M an integer, denoting the maximum number of iterations allowed.
- Given ϵ for the stopping the iteration. If the absolute difference between the two successive log-likelihood values is less than ϵ, then we stop the iteration. We take $\epsilon = 10^{-6}$.
- Given m denote the sample size.
- Set the data vector $\{t_1, ..., t_m\}$.
- Set the initial values $\mu^{(0)}$, $\sigma^{(0)}$ and $\theta^{(0)}$.

for (k in 1 to M){

for (i in 1 to m) {

Find the missing value $E_i^{(k)}$ as follows:

$E_i^{(k)} = 1+\theta^k \Phi(\xi_i)$, where $\xi_i = (\ln t_i - \mu^k)/\sigma^k$ and $\Phi(.)$ denotes the CDF of standar normal distribution N(0,1).

Find $c_{i,m}$ and $\beta_i(\boldsymbol{\vartheta}^{(k)})$ as: $c_{i,m} = \dfrac{1}{i-1}\displaystyle\sum_{i=1}^{m} E(Z_{i:m})$ and $\beta_i(\boldsymbol{\vartheta}^{(k)}) = (E_i^{(k)} - 1)c_{i,m}$.

}

- Compute D_3 and D_4.
- Compute $\mu^{(k+1)}$ and $\sigma^{(k+1)}$ as given in Eq.'s (27) and (29) and $\theta^{(k+1)}$ as a solution of equation (12) using Newton-Raphson or fixed point method.
- Compute the log-likelihood function $l^*_{ps}(\mu^{(k+1)}, \sigma^{(k+1)}, \theta^{(k+1)})$.
- If $|l^*_{ps}(\mu^{(k+1)}, \sigma^{(k+1)}, \theta^{(k+1)}) - l^*_{ps}(\mu^{(k)}, \sigma^{(k)}, \theta^{(k)})| < \epsilon$ we stop the iteration, else we continue until convergence.

}

References

[1] Adamidis, K., Dimitrakopoulou, T., and Loukas, S., 2005. On an extension of the exponential-geometric distribution, *Statistics & Probability Letters* 73 (3), 259–269.

[2] Adamidis, K. and Loukas, S., 1998. A lifetime distribution with decreasing failure rate, *Statistics & Probability Letters*, 39, pp. 35–42.

[3] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N., 2008. *A First Course in Order Statistics*, 2nd Edition, SIAM, USA.

[4] Asgharzadeh, A., Esmaily, L. and Nadarajah, S., 2013. Approximate MLEs for the location and scale parameters of the skew logistic distribution, *Statistical Papers*, 54, 391-411.

[5] Alzaatreh, A., Aljarrah, M. A., Smithson, M. et al. (2021). Truncated family of distributions with applications to time and cost to start a business, *Methodol Comput Appl Probab*, 23, 5-27. https://doi.org/10.1007/s11009-020-09801-1.

[6] Bakouch, H. S., Ristić, M. M., Asgharzadeh, A., Esmaily, L., and Al-Zahrani, B. M., 2012. An exponentiated exponential binomial distribution with application, *Statistics & Probability Letters*, 82, 1067-1081.

[7] Barreto-Souza, W. and Cribari-Neto, F., 2009. A generalization of the exponential-Poisson distribution, *Statistics & Probability Letters*, 79, 2493–2500.

[8] Bulmer. M.G., 1974. On fitting the Poisson lognormal distribution to species-abundance data, *Biometrics*, 30, 101-110.

[9] Dempster, A. P., Laird, N. M., and Rubin, D. D., 1977. Maximum likelihood from incomplete data via EM algorithm, *Journal of the Royal Statistical Society*, Ser. B, Vol. 39, 1-38.

[10] Dzupire, N.C., Ngare, P. and Odongo, L., 2018. A Poisson-gamma model for zero inflated rainfall data, *Journal of Probability and Statistics*, Article ID 1012647.

[11] Gentilucci, M., Barbieri, M., Lee, H.S. and Zardi, D., 2019. Analysis of Rainfall Trends and Extreme Precipitation in the Middle Adriatic Side, Marche Region (Central Italy), *Water*, vol. 11, Article no. 1948.

[12] Karlis, D., 2009. A note on the exponential Poisson distribution: A nested EM algorithm, *Computational Statistics & Data Analysis*, 53(4), 894-899.

[13] Kus, C., 2007. A new lifetime distribution, *Computational Statistics & Data Analysis*, 51(9), 4497—4509.

[14] Louis, T. A., 1982. Finding the observed information matrix when uisng the EM algorithm, *Journal of the Royal Statistical Society*, Ser. B 44, 226–233.

[15] Mahmoudi, E. and Sepahdar, A., 2013. Exponentiated Weibull–Poisson distribution: model, properties and applications, *Mathematics and Computers in Simulation*, 92, 76-97.

[16] Marshall, A., and Olkin, I., 1997. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, *Biometrika*, 84(3), 641–652.

[17] Ng, H.K. T., Chan, C. S. and Balakrishnan, N., 2002. Estimation of parameters from progressively ordered censored data using EM algorithm, *Computational Statistics and Data Analysis*, 39, 371-386.

[18] Odening, M., Musshoff, O. and Xu, W., 2007. Analysis of rainfall derivatives using daily precipitation models: Opportunities and pitfalls, *Agricultural Finance Review*, vol. 67, no. 1, pp. 135–156.

[19] Raqab, M. Z., 1997. Modified maximum likelihood predictors of future order statistics from normal samples, *Computational Statistics and Data Analysis*, 25, 91-106.

[20] Raqab, M. Z. and Madi, M. T., 2011. Inference for the Generalized Rayleigh Distribution Based on Progressively Censored Data, *Journal of Statistical Planning & Inference*, 141, 3313-3322.

[21] Self, S.G. and Liang, K-L., 1987. Asymptotic properties of the maximum likelihood estimators and likelihood ratio test under non-standard conditions, *Journal of the American Statistical Association*, 82, 605 - 610.