Order Statistics for Exponential-Poisson Distribution

Sadegh Rezaei, Rasool Tahmasbi and Fatemeh Hasantabar

Department of Statistics, Shahid Chamran University, Ahwaz, Iran
Department of Statistics, Amirkabir University of Technology, Tehran, Iran

Abstract: In this study, we introduced the order statistics of exponential-Poisson distribution arising from combination of exponential distribution with a truncated Poisson distribution. Various properties of the proposed distribution are discussed. The estimation of the parameters attained by the EM and nested EM algorithms and their asymptotic variances and covariance are obtained. In order to assess the accuracy of the approximation of variances and covariance of the maximum likelihood estimators, simulation studies are performed and experimental results are illustrated, based on real data sets.

Key words: DFR, exponential distribution, IFR, nested EM algorithm and truncated Poisson distribution

INTRODUCTION

The distribution of order statistics is useful in many field of science, such as reliability, earthquake engineering, queueing theory (Harel and Cheng, 1997) and networks (Yang and Petropulu, 2001). Kus (2007) and Tahmasbi and Rezaei (2008) used the first order statistics to model the minimum time interval between two successive earthquakes, using EP and EL distributions, respectively. Recently, the order statistics from different distributions and their properties are studied by many researchers. The distributions of order statistics from bivariate skew-normal and bivariate skew-t distributions are proposed by Jamalizadeh and Balakrishnan (2008). Balakrishnan and Stepanov (2008) presented the asymptotic results for the ratio of order statistics. Rukhin (2006) discussed the Gamma-distribution order statistics. Barakat and Abdelkader (2000) established a general recurrence relations satisfied by the product moments of bivariate order statistics from any arbitrary bivariate uniform distribution function. Lien and Balakrishnan (2003) considered the order statistics from a bivariate normal distribution and developed a conditional correlation analysis. They applied these results to evaluate the presence of inventory effects in futures market. Chen et al. (2004) estimated the means and covariance of inverse Gaussian order statistics. Order statistics from inverse Weibull distribution and discrete uniform distribution are considered in Mahmoud et al. (2003), Calik and Gungor (2004), respectively. Ahsanullah (2000) and Ragab (1998) studied on the order statistics from exponential distribution and Burr Type X Model, respectively. Barakat and Abdelkader (2000) computed the moments of order statistics from non-identically distributed Weibull variables.

In this study, we introduced order statistics from exponential-Poisson (EP) distribution. The first order statistics is studied in Kus (2007), which has decreasing failure rate (DFR) property. The distribution with DFR are studied in the work of Lomax (1954), Proschan (1963), Barlow et al. (1963), Barlow and Marshall (1964, 1965), Marshall and Proschan (1965), Cozzolino (1968), Dahiya and Gurland (1972), McNolty et al. (1980), Saunders and Myhre (1983), Nassar (1988), Gleser (1989), Gurland and Sethuraman (1994), Adamidis and Loukas (1998), Kus (2007) and Tahmasbi and Rezaei (2008). For orders greater than one, the failure rate function has two different shapes: It strictly increases (IFR) or increases at first and then decreases. Properties of the failure rate order and IFR is studied in Belzunce et al. (2004) and Rychlik (2001).

MATERIALS AND METHODS

The proposed distribution: Let \( Y_1, \ldots, Y_z \) be a random sample from an exponential distribution with probability density function (pdf)

\[
  f_Y(y; \beta) = \beta e^{-\beta y} I_{(0,\infty)}(y); \beta > 0
\]

and \( Z \) is a truncated at \((k-1)\) Poisson distribution with probability mass function as follows:

\[
  P_Z(z; \lambda) = \frac{e^{-\lambda} \lambda^z}{Z! \left(1 - e^{-\lambda} \sum_{i=0}^{k-2} \frac{\lambda^i}{i!}\right)}
\]
Since we want to model the $k^{th}$ order statistics of $Y_i$'s and we assumed the number of $Y_i$'s, $Z$, is a random variable, so it is possible that the realization of r.v $Z, z$, be less than $k$, which is not proper. So, it is necessary to follow a method that assures us that the realization of $Z$ be greater than or equal to $k$. This condition can be obtained by truncating $Z$ at $k-1$, simply.

By assuming that the random variables $Y_i$'s and $Z$ are independent, we define $X = k^{th}$ order statistic of $\{Y_1, ..., Y_z\}$. Then,

$$f_X(z|x; \theta) = \beta k \frac{z^{k-1} e^{-\beta z}}{k!}$$

and marginal pdf of $X$ is

$$f_X(x, \lambda, \beta) = Ce^{-\beta x} \left(1 - e^{-\beta x}\right)^{k-1} e^{\lambda x - \beta x}$$

where

$$C = \frac{\beta k e^{-\lambda}}{(k-1) \left[1 - e^{-\lambda} \sum_{i=0}^{k-1} \frac{\lambda^i}{i!}\right]}$$

After here, the distribution of $X$ will be referred as the order statistics of exponential-Poisson distribution (OEP). The OEP probability density functions are displayed in Fig. 1 for selected parameter values. The pdf of OEP distribution for $k = 1$ is monotone decreasing with modal value $\frac{\beta \lambda}{1 - e^{-\lambda}} e^{-\lambda}$ at $x = 0$ (Kus, 2007). For $k > 1$, it increases from 0 to its maximum at

$$\left[\ln(2\lambda) - \ln(\sqrt{(k-\lambda)^2 + 4\lambda - k + \lambda})\right] / \beta.$$ 

For all values of parameters, the pdf tends to 0 as $x \to \infty$.

**Properties of the distribution:** In this section, the properties of the proposed distribution are presented, which are equal to those of Kus (2007), for $k = 1$.

**The distribution:** The distribution function is given by:

$$F_X(x; \lambda, \beta) = \frac{1 - e^{-\beta x} \sum_{i=0}^{k-1} \frac{(1-e^{-\beta x})^i}{i!}}{1 - e^{-\beta x} \sum_{i=0}^{k-1} \frac{\lambda^i}{i!}}$$

and hence, the median is obtained by solving the equation

$$x = \beta^{-1} \ln \lambda - \beta^{-1} \ln \left[ -\ln(2) + \ln \left( e^x + \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} \right) - \ln \left( \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} \right) \right].$$

The moment generating function of $X$ is determined from Eq. (1) by direct integration and is given by:

$$M_X(t) = \frac{C \sum_{j=0}^{k-1} \binom{k-1}{j}}{\beta} \left(1 - \frac{(-1)^j}{j+1-t/\beta} \right) F_{p,q} \left( j + 1 - \frac{t}{\beta}, j + 2 - \frac{t}{\beta}, \lambda \right),$$

where $C$ is defined in Eq. (2) and $F_{p,q}(.)$ is generalized hypergeometric function, defined by:

$$F_{p,q}(n,d,\lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k \prod_{i=1}^{p} \Gamma(n_i + k) \Gamma^{-1}(n_i)}{\prod_{i=1}^{q} \Gamma(d_i + k) \Gamma^{-1}(d_i)}$$

where $n = [n_1, n_2, ..., n_p]$ and $d = [d_1, d_2, ..., d_q]$. For $r \in \mathbb{R}$, the raw moments are given by:

$$E(X^r) = \frac{C r! \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j}{\beta^{r+1}} \left[ j + 1, ..., j + 1 \right] \left[ j + 2, ..., j + 2 \right] \lambda.$$ 

Hence the mean of the OEP distribution are given, respectively, by:
Fig. 1: Probability density functions of the OEP distribution for different values of $\beta$, $\lambda$, and $k$

The survival and hazard functions: Using Eq. (1) and (2), the survival function (also known as reliability function) and the hazard function (also known as failure rate function) of the OEP distribution are given, respectively, by:

$$S(x) = 1 - F_X(x) = e^{-\beta(1-e^{-\beta x})} \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} - e^{-\lambda} \sum_{i=0}^{k-1} \frac{\lambda^i}{i!}$$

$$h(x) = \beta e^{\beta x} \left( 1 - e^{-\beta x} \right)^{k-1} e^{\lambda} e^{-\beta x} - \frac{1}{(k-1)!} \left[ e^{-\beta(1-e^{-\beta x})} \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} - e^{-\lambda} \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} \right]$$

Both functions have simple forms. The hazard function is strictly decreasing for $k=1$ because the DFR.
Fig. 2: Hazard functions of the OEP distribution for different values of $\beta$, $\lambda$, and $k$

Property follows from the results of Barlow et al. (1963) on mixtures and is strictly increasing (IFR) for $\lambda \leq k - 1$. For $\lambda > k - 1$, it increases and then decreases. The OEP hazard functions are displayed in Fig. 2, for selected parameter values.

The initial hazards are, for $k = 1$ and $h(0) = 0$, for $k > 1$. The long-term is $h(\infty) = \beta$. They are both finite in contrast to those of Weibull distribution with $h(0) = \infty$ and $h(\infty) = 0$.

**Random number generation:** Let $U$ be a random variable (r.v.) from a uniform distribution with the parameters \( \left\{ \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} e^{\lambda x} \right\} \). The root of the following equation will generate r.v. from the OEP distribution with parameters $\lambda$ and $\beta$:

\[
U = e^{\lambda x} - \beta x^{k-1} \left( \sum_{i=0}^{k-1} \frac{\lambda^i}{i!} e^{\lambda x} \right)^i
\]
Estimation of the parameters:

**Estimation by maximum likelihood:** The log-likelihood function based on the observed sample size of \( n, y_{\text{obs}} = (x_i; i = 1, \ldots, n) \), from the OEP distribution is given by

\[
\begin{align*}
(\beta, \lambda, y_{\text{obs}}) &= n \ln \beta + n \lambda \ln \lambda - n \ln (k - 1) \\
&- \beta \sum_{i=1}^{n} x_i + \lambda \sum_{i=1}^{n} e^{-\beta x_i} \\
&- n \ln \left(1 - e^{-\lambda \sum_{i=0}^{k-1} x_i} + (k-1) \ln(1 - e^{-\beta x_i})\right)
\end{align*}
\]

and subsequently the associated gradients are found to be

\[
\begin{align*}
\frac{\partial}{\partial \beta} &= \frac{n}{\beta} + (k - 1) \sum_{i=1}^{n} x_i e^{-\beta x_i} \\
&- \sum_{i=1}^{n} x_i - \lambda \sum_{i=1}^{n} x_i e^{-\beta x_i}
\end{align*}
\]

and

\[
\begin{align*}
\frac{\partial}{\partial \lambda} &= \frac{k}{\lambda} - n \left(1 - \sum_{i=0}^{k-1} e^{-\lambda x_i} \right) \\
&+ \sum_{i=1}^{n} e^{-\beta x_i} - n
\end{align*}
\]

To achieve estimations via ML method, it is not easy to solve the equations \( \partial / \partial \beta = 0 \) and \( \partial / \partial \lambda = 0 \), directly. In the following, Theorems 1 and 2 express the conditions for existence of the MLE, when the other parameter is given or known.

**Theorem 1:** Let \( g(\beta; \lambda, y_{\text{obs}}) \) denote the function on the right hand side of the expression in (ref{r_beta}) and

\[
\bar{x} = n^{-1} \sum_{i=1}^{n} x_i,
\]

where \( \lambda \) is the true value of the parameter. Then, for a given \( \lambda > 0 \) the root of \( g(\beta; \lambda, y_{\text{obs}}) \) lies in the interval \( \left[ (\bar{x}(\lambda + 1))^{-1}, +\infty \right] \).

**Theorem 2:** Let \( h(\lambda; \beta, y_{\text{obs}}) \) denote the function on the right hand side of the expression in (ref{r_lambda}), where \( \beta \) is the true value of the parameter. Then for \( \lambda < (k + 1) \sum_{i=1}^{n} e^{-\beta x_i} \), the equation \( h(\lambda; \beta, y_{\text{obs}}) = 0 \) has at least one root.

**EM algorithm:** The MLEs of \( \lambda \) and \( \beta \) must be derived numerically. Newton-Raphson (NR) algorithm is one of the standard methods to determine the MLEs of the parameters. To employ the algorithm, second derivatives of the log-likelihood are required for all iterations. Another method is using the EM algorithm. This algorithm is a very powerful tool in handling the incomplete data problem (Dempster et al., 1977; McLachlan and Krishnan, 1997). It is an iterative method by repeatedly replacing the missing data with estimated values and updating the parameters. It is especially useful if the complete data set is easy to analyze. Recently, an EM algorithm has been used by several authors such as Adamidis and Loukas (1998), Adamidis (1999), Ng et al. (2002), Kus (2007) and Tahmasbi and Rezaei (2008).

To start the algorithm, hypothetical complete-data distribution is defined with density function

\[
f(x, z, \beta, \lambda) = \frac{k \beta^x e^{-\lambda z} \left(\frac{z}{k}\right)^{k-1} e^{-\beta x(z-k+1)} \left(1 - e^{-\beta x}\right)^{k-1}}{z \left(1 - e^{-\lambda \sum_{i=0}^{k-1} \lambda x_i} \right)}
\]

Thus, it is straightforward to verify that the E-step of an EM cycle requires the computation of the conditional expectation of \( \left(Z|X, \beta(h), \lambda(h)\right) \) where \( (\beta(h), \lambda(h)) \) is the current estimate \( (\beta, \lambda) \). Using that

\[
P(Z|X; \beta, \lambda) = \frac{e^{-\beta x(z-k)} e^{-\lambda x}}{(z-k)!} x^{-k} e^{-\beta x} e^{-\lambda x}
\]

this is found to be \( \mathbb{E}(Z|X, \beta, \lambda) = \lambda x + \beta x - k \). The EM cycle completes with M-step, which is complete data maximum likelihood over \((\beta, \lambda)\), with the missing Z's replaced by their conditional expectations \( \mathbb{E}(Z|X, \beta(h), \lambda(h)) \). Thus, EM iteration is given by
**E-step:** With the given values of parameters $\beta^{(h)}$, $\lambda^{(h)}$ calculate

$$E_i = \mathbb{E}(Z_i | \beta^{(h)}, \lambda^{(h)}) = k + \lambda^{(h)} - e^{-\lambda^{(h)} x},$$

$$i = 1, \ldots, n$$

**M-step:** Update the parameters by

$$\beta^{(h+1)} = n \left\{ \frac{\sum_{i=1}^{n} x_i (s_i - k + 1)}{\left( k - 1 \right) \sum_{i=1}^{n} x_i e^{-\lambda^{(h+1)} x_i}} \right\}^{-1}$$

$$\lambda^{(h+1)} = \frac{1}{n} \sum_{i=1}^{n} s_i - \frac{e^{-\lambda^{(h+1)} x}}{\left( k - 1 \right) \sum_{i=1}^{k-1} \lambda^{(h+1)}}$$

From the above discussion, a nested EM algorithm is proposed for the OEP distribution where the M-step of the algorithm is solved by applying another EM algorithm. So, the new algorithm takes the form:

**E-step:** With the given values of parameters $\beta^{(h)}$, $\lambda^{(h)}$ calculate

**E1-step:**

$$E_i = \mathbb{E}(Z_i | \beta^{(h)}, \lambda^{(h)}) =$$

$$k + \lambda^{(h)} - e^{-\lambda^{(h)} x}, \quad i = 1, \ldots, n$$

**E2-step:**

$$E_i^{(h+1)} = \left( n + \sum_{j=0}^{k-1} n_j \right) \frac{e^{-\lambda^{(h)} x}}{i!},$$

$$i = 0, \ldots, k - 1$$

**M-step:** Update the parameters by

$$\beta^{(h+1)} = \left\{ \frac{\sum_{i=1}^{n} x_i (s_i - k + 1)}{\left( k - 1 \right) \sum_{i=1}^{k-1} \lambda^{(h+1)}} \right\}^{-1}$$

$$\lambda^{(h+1)} = \frac{1}{n} \sum_{i=1}^{n} x_i e^{-\lambda^{(h+1)} x_i}$$

Thus, with observed data $Z_1, \ldots, Z_n$, the EM algorithm for a simple truncated at $k - 1$ Poisson distribution is described as:

Use current estimates for

**E-step:** For each $\eta_i^{(h)}, \Lambda^{(h)}$, $i \in \{0, \ldots, k - 1\}$ obtain

$$\eta_i^{(h+1)} = \left( \frac{n + \sum_{j=0}^{k-1} n_j}{i!} \right)^{-1}$$

**M-step:** Update Poisson parameter

$$\Lambda^{(h+1)} = \frac{\sum_{i=1}^{n} Z_i + \sum_{j=0}^{k-1} n_j^{(h+1)}}{n + \sum_{j=0}^{k-1} n_j^{(h+1)}}$$
M2-step:

\[
\hat{\lambda}(t+1) = \frac{\sum_{i=1}^{n} \lambda_i + \sum_{j=0}^{k-1} j \lambda_j}{n + \sum_{j=0}^{k-1} j \lambda_j} \tag{15}
\]

Note that, there are two strategies that the EM algorithms will be used. The first one needs to run several steps of the inside EM algorithm (steps E2 and M2) so as to ensure the convergence of the estimates for the truncated Poisson part. On the contrary the second strategy needs just one iteration of the inside EM at each outer EM (steps E1 and M1) iteration. In both cases the convergence of the algorithm can be based on the results of Van Dyk (2000), which follows from making use of the monotonic property of the algorithm, i.e. the log-likelihood increases at each step.

The nested EM scheme is more advantageous rather than the NR scheme. First of all if the parameters are in the admissible range the estimates remain always in the admissible range, which is not the case for the NR iterations during the M-step. The closed form expressions are much simpler than calculating derivatives as needed for the NR scheme even in such a simple case as the one here with the truncated Poisson. Simulations based evidence is provided to show this concept.

Asymptotic variance and covariance of the MLEs: Applying the usual large sample approximation, the MLE of \( \theta = (\beta, \lambda) \) can be treated as being approximately bivariate normal with mean \( \bar{\theta} \) and variance-covariance matrix, which is the inverse of the expected information matrix \( J(\theta) = E(I(\theta), \theta) \), where \( I = I(\theta; \gamma_{\text{obs}}) \) is the observed information matrix with elements \( I_{ij} = \partial^2 \ell / \partial \theta_i \partial \theta_j \) with \( i, j = 1, 2 \) and the expectation is to be taken with respect to the distribution of \( X \). Differentiating Eq. (5) and (6), the elements of the symmetric, second-order observed information matrix are found to be:

\[
I_{11} = n \beta^2 + (k-1) \sum_{i=0}^{n} \frac{x_i \beta e^{-\beta x}}{(1-e^{-\beta x})^2} - 2 \sum_{i=0}^{n} x_i e^{-\beta x}
\]

\[
I_{12} = I_{21} = \sum_{i=0}^{n} x_i e^{-\beta x}
\]

\[
I_{22} = \frac{nk \lambda}{k^2} + \frac{ne^{-\lambda} \lambda^{2k-2} (k-1)}{(k-1)!} \left[ 1 - e^{-\lambda} \sum_{i=0}^{k-1} \frac{1}{i!} \right] \]

\[
- n \left[ \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!} \left[ 1 - e^{-\lambda} \sum_{i=0}^{k-1} \frac{1}{i!} \right] \right]^2
\]

The elements of the expected information matrix, \( J(\theta) \), are calculated by taking the expectations of \( I_{ij}, i, j = 1, 2 \) with respect to the distribution of \( X \). The following expectations are required:

\[
E[X e^{-\beta X}] = C \beta^{2} \sum_{j=0}^{k-1} \frac{(k-1)}{j} \left( -1 \right)^j \left( j+2 \right)^2
\]

\[
F_{2,2} \left[ j+2, j+2, \left[ j+3, j+3 \right], \lambda \right]
\]

\[
E[X^2 e^{-\beta X}] = 2C \beta^{3} \sum_{j=0}^{k-1} \frac{(k-1)}{j} \left( -1 \right)^j \left( j+2 \right)^3
\]

\[
F_{3,3} \left[ j+2, j+2, j+2, \left[ j+3, j+3, j+3 \right], \lambda \right]
\]

\[
E \left[ \frac{X^2 e^{-\beta X}}{\left( 1-\beta X \right)^2} \right] =
\]

\[
2C \beta^{3} \sum_{j=0}^{\infty} \frac{3-k+j+1}{j} \left( -1 \right)^j \left( j+2 \right)^3
\]

\[
F_{3,3} \left[ j+2, j+2, j+2, \left[ j+3, j+3, j+3 \right], \lambda \right] \text{ if } k < 3
\]

\[
2C \beta^{3} \sum_{j=0}^{k-3} \frac{(k-3)}{j} \left( -1 \right)^j \left( j+2 \right)^3
\]

\[
F_{3,3} \left[ j+2, j+2, j+2, \left[ j+3, j+3, j+3 \right], \lambda \right] \text{ if } k \geq 3
\]
Table 1: Variances and covariances of the MLEs. The simulated values of $[\text{Var} \hat{\beta}, \text{Cov} \{ \hat{\beta}, \hat{\alpha} \}, \text{Var} \hat{\alpha}]$ denoted by $\text{Cov} \{ \hat{\theta} \}$, the approximate values of variances and covariances from the expected information matrix ($\text{Cov} (E)$) and observed information matrix ($\text{Cov} (O)$).

<table>
<thead>
<tr>
<th>$k$</th>
<th>$n$</th>
<th>$\theta$</th>
<th>$\text{Cov} { \hat{\theta} }$</th>
<th>$\text{Cov} (O)$</th>
<th>$\text{Cov} (E)$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>50</td>
<td>$1.05$</td>
<td>$0.060$</td>
<td>$0.028$</td>
<td>$0.048$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.5$</td>
<td>$0.008$</td>
<td>$0.049$</td>
<td>$0.066$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$1.0$</td>
<td>$0.003$</td>
<td>$0.049$</td>
<td>$0.075$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$2.0$</td>
<td>$0.002$</td>
<td>$0.049$</td>
<td>$0.085$</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>$1.05$</td>
<td>$0.016$</td>
<td>$0.048$</td>
<td>$0.075$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.5$</td>
<td>$0.001$</td>
<td>$0.048$</td>
<td>$0.085$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$1.0$</td>
<td>$0.001$</td>
<td>$0.048$</td>
<td>$0.085$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$2.0$</td>
<td>$0.001$</td>
<td>$0.048$</td>
<td>$0.085$</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>50</td>
<td>$1.05$</td>
<td>$0.016$</td>
<td>$0.048$</td>
<td>$0.075$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.5$</td>
<td>$0.001$</td>
<td>$0.048$</td>
<td>$0.085$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$1.0$</td>
<td>$0.001$</td>
<td>$0.048$</td>
<td>$0.085$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$2.0$</td>
<td>$0.001$</td>
<td>$0.048$</td>
<td>$0.085$</td>
<td></td>
</tr>
</tbody>
</table>

where $F_{\text{res}}(.)$ is generalized hypergeometric function and $C$ is defined in (2). Thus $J(\hat{\theta})$ are derived in the form:

$$J_{11} = n \beta^{-2} - 2nC \lambda \beta^{-3} \sum_{j=0}^{k-1} \binom{k-1}{j} \binom{-1}{j+2}$$

$$F_{3,3}[j+2, j+2, j+2] \{j+3, j+3, j+3, j+3, \lambda \} + n(k-1)4$$

$$J_{12} = nC \lambda \beta^{-2} \sum_{j=0}^{k-1} \frac{(-1)^{j}}{(j+2)^{3}}$$

$$F_{3,2}[j+2, j+2] \{j+3, j+3, j+3, \lambda \}$$

$J_{22} = L_{22}$

where

$$A = B \left[ \frac{X^2 e^{-\beta X}}{(1-e^{-\beta X})^2} \right]$$

The inverse of $J(\hat{\theta})$, evaluated at $\tilde{\theta}$ provides the asymptotic variance-covariance matrix of the MLEs. Alternative estimates can be obtained from the inverse of the observed information matrix since, it is a consistent estimator of $J(\hat{\theta})$.

**Simulation study:** Simulation results are obtained based on the assumption of Theorem 1 and 2, to guarantee the existence of the solution in each iteration of the parameters in EM algorithm. No restriction has been imposed on the maximum number of iterations and Convergence is assumed when the absolute differences between two successive estimates are less than $10^{-5}$.

For each value of $\theta$, the parameter estimates have been found by employing three different schemes: Solving Eq. (8) and (9) using NR method, using the nested EM algorithm with one iteration at inside EM (NEM1), and using the nested EM algorithm with five iteration at inside EM (NEM5). In all cases, different initial values are considered.

In order to assess the accuracy of the approximation of the variances and covariance of the MLEs determined from the information matrix, a simulation study (based on 1000 simulations) has been carried out. The simulated
The close to the that of expected convergence of the prop

One thousand samples of size 100 and 500

also calculated the empirical type I error probability, \( \alpha \), as the proportion of rejecting the null hypothesis \( H_0: \theta = \hat{\theta} \), if twice the log likelihood ratio is greater than \( \chi^2(2) \), so that the nominal type I error probability is 0.05. From Table 1, it is observed that the approximate values determined from expected and observed information matrix are quite close to the simulated values for large values of \( n \). Furthermore, it is noted that the approximation becomes quite accurate as \( n \) increases. In addition, variances and covariances of the MLEs obtained from the observed information matrix is quite close to the that of expected information matrix for large values of \( n \).

Also, simulations have been performed to investigate the convergence of the proposed EM and nested EM schemes. One thousand samples of size 100 and 500 of which are randomly sampled from the OEP for each of three values of \( \theta \) are generated.

The results from simulated data sets are reported in Table 2, which gives the averages of the 1000

<table>
<thead>
<tr>
<th>Data set</th>
<th>n</th>
<th>k</th>
<th>( \hat{\theta} )</th>
<th>KS</th>
<th>p-value</th>
<th>NR</th>
<th>Iterations</th>
<th>NEM1</th>
<th>NEM5</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDEI</td>
<td>13</td>
<td>2</td>
<td>(0.222, 2.979)</td>
<td>0.169</td>
<td>0.816</td>
<td>184</td>
<td>298</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(0.264, 4.628)</td>
<td>0.202</td>
<td>0.612</td>
<td>124</td>
<td>210</td>
<td>91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENA</td>
<td>24</td>
<td>2</td>
<td>(0.056, 4.568)</td>
<td>0.148</td>
<td>0.632</td>
<td>141</td>
<td>177</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(0.075, 6.156)</td>
<td>0.183</td>
<td>0.362</td>
<td>100</td>
<td>138</td>
<td>67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data set</th>
<th>n</th>
<th>k</th>
<th>( \hat{\theta} )</th>
<th>( R_{11}^{-1} )</th>
<th>( R_{12}^{-1} )</th>
<th>( R_{22}^{-1} )</th>
<th>( Ch(\hat{\theta}) )</th>
<th>( Ch(\hat{\lambda}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDEI</td>
<td>13</td>
<td>2</td>
<td>(0.2215, 2.9785)</td>
<td>0.0133</td>
<td>-0.2266</td>
<td>5.0795</td>
<td>(0.0000, 0.4475)</td>
<td>(0.0000, 7.3959)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(0.2643, 4.6276)</td>
<td>0.0105</td>
<td>-0.1746</td>
<td>4.3456</td>
<td>(0.0639, 0.4648)</td>
<td>(0.5417, 7.1314)</td>
<td></td>
</tr>
<tr>
<td>ENA</td>
<td>24</td>
<td>2</td>
<td>(0.0555, 4.5681)</td>
<td>0.0004</td>
<td>-0.0241</td>
<td>2.4177</td>
<td>(0.0187, 0.9224)</td>
<td>(1.5205, 7.6157)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(0.0730, 6.1562)</td>
<td>0.0003</td>
<td>-0.0206</td>
<td>2.0873</td>
<td>(0.0367, 0.1094)</td>
<td>(3.3245, 8.9879)</td>
<td></td>
</tr>
</tbody>
</table>

values of \( Var(\hat{\theta}), Var(\hat{\lambda}) \), and \( Cov(\hat{\theta}, \hat{\lambda}) \) (denoted by \( Cov(\hat{\theta}) \)) as well as the approximate values determined by averaging the corresponding values obtained from the expected and observed information matrix are presented in Table 1.
MLEs, $\hat{\theta}$, and averages numbers of iterations to convergence for NR method, \(\text{av}(NR)\), nested EM with one iteration inside EM, \(\text{av}(NEM1)\), and nested EM with five iteration inside EM method, \(\text{av}(NEM5)\), together with standard errors, \(\text{se}(\hat{\theta})\). Table 2 indicates the following results: convergence has been achieved in all cases, even when the starting values are poor and this emphasizes the numerical stability of the EM and nested EM algorithms. The values of $\text{av}(\tilde{\theta})$ and $\text{se}(\tilde{\theta})$ suggest the EM estimates performed, consistently. Standard errors of the MLEs decrease when sample size increases.

RESULTS AND DISCUSSION

In this section, the OEP parameters are estimated using three different schemes with \(k = 2\) and \(k = 3\).

To compare the results, two data sets are considered. The first set consists of the time intervals (in day) of the most desolat ed earthquakes in Iran (MDEI). The data are presented in Tahmasbi and Rezaei (2008) containing the dates of successive earthquakes with their exact locations, magnitudes, depths and the number of killed persons from 1990. In the second set, the data are 24 observations on the period between successive earthquakes in the last century in North Anatolia fault zone (ENA). These data are analyzed by Kus (2007), Tahmasbi and Rezaei (2008) and can be found in Kus (2007) (Table 3 and 4).

To make the numbers smaller and avoid overflow errors, the original data are divided by 100. Figure 3 shows the history of the log-likelihood using the NR approach (NR), the nested EM with only one iteration at each M-step (NEM1) and the nested EM with five iterations at each M-step (NEM5). One can see that the NEM5 goes much faster towards the maximum. Note also that the NEM5 approach provides close convergence to
For all cases, the initial values used for the parameters are estimated via moment matching method.

For the OEP model with \( k = 2 \) and \( k = 3 \), the PP-plots are given in Fig. 4. As the figures show, in all cases, the models provide good fits to the data sets. The MLEs of \( \hat{\lambda} \) together with Kolmogorov-Smirnov (KS) statistics and \( p \)-values for these two data sets are presented in Table 3. The KS test statistics takes the smallest value and the largest \( p \)-value for MDEI data with \( k = 2 \).

Estimation of asymptotic variance-covariance matrix of the MLEs obtained from the inverse of the observed information matrix together with a 95 percent coefficient intervals for \( \hat{\beta} \) and \( \hat{\lambda} \) are reported in Table 4.

**Appendix**

**Proof of Theorem 1**: Let

\[ \omega_1(\beta, \lambda, y_{obs}) = (k - 1) \sum_{i=0}^{n} \frac{x_i e^{-\beta x_i}}{1 - e^{-\beta x_i}} \]

and

\[ \omega_2(\beta, \lambda, y_{obs}) = \lambda \sum_{i=0}^{n} x_i e^{-\beta x_i} \]

Clearly, \( \omega_1 \) is strictly decreasing in \( \beta \) and \( \lim_{\beta \to \infty} \omega_1 = 0 \) and \( \omega_2 \) is strictly increasing in \( \beta \) and
\[ \lim_{\beta \to 0} \omega_2 = -\lambda \sum_{i=0}^{n} x_i \]. Therefore,
\[ g(\beta, \lambda, y_{obs}) > \frac{n}{\beta} \sum_{i=1}^{n} x_i (\lambda + 1) \]
and hence \( g(\beta, \lambda, y_{obs}) < 0 \) when \( \beta > \left( \frac{\lambda + 1}{\lambda} \right)^{-1} \).

On the other hand, \( \lim_{\beta \to 0} \omega_2 = +\infty \), and so \( g(\beta; \lambda, y_{obs}) > 0 \). Therefore, there is at least one root of \( g(\beta; \lambda, y_{obs}) = 0 \) in the interval \( \left[ \left( \frac{\lambda + 1}{\lambda} \right)^{-1}, +\infty \right] \).

**Proof of Theorem 2:** Let
\[ \omega(\lambda; \beta, y_{obs}) = \frac{n \beta^k}{\lambda} - \frac{n \beta^k - 1}{(k-1)!} \left( 1 - e^{-\beta \sum_{i=0}^{k-1} x_i} \right) \]
It is clear that \( \omega \) is strictly decreasing in \( \lambda \) and
\[ \lim_{\lambda \to 0} \omega = n \frac{k}{k+1} \quad \text{and} \quad \lim_{\lambda \to \infty} \omega = 0 \]. Therefore,
\[ h(\lambda; \beta, y_{obs}) < n \frac{k}{k+1} e^{-\beta x_1} - n \]
and hence \( h(\lambda; \beta, y_{obs}) > 0 \)
when \( n < (k+1) \sum_{i=1}^{n} e^{-\beta x_i} \).

On the other hand, \( h(\lambda; \beta, y_{obs}) > \sum_{i=1}^{n} e^{-\beta x_i} - n \)
and therefore \( h < 0 \) when \( \sum_{i=1}^{n} e^{-\beta x_i} - n < 0 \) or
\[ \sum_{i=1}^{n} e^{-\beta x_i} < n \]. Therefore, the equation \( h(\lambda; \beta, y_{obs}) = 0 \)
has at least one root if \( (k+1) \sum_{i=1}^{n} e^{-\beta x_i} > n \).

**REFERENCES**


