

Estimation of the PDF and the CDF of exponentiated moment exponential distribution

Yogesh Mani Tripathi¹ · Tanmay Kayal¹ · Sanku Dey²

Received: 17 August 2016/Revised: 4 February 2017/Published online: 18 March 2017

© The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2017

Abstract This article addresses the different methods of estimation of the probability density function and the cumulative distribution function for the exponentiated moment exponential distribution. Following estimation methods are considered: uniformly minimum variance unbiased estimators, maximum likelihood estimators, percentile estimators, least squares estimators, weighted least square estimators, maximum product of spacings estimators, Cramér–von-Mises estimators and Anderson–Darling estimators. Analytical expressions are derived for the bias and the mean squared error. Monte Carlo simulations are performed to compare the performances of the proposed methods of estimation for both small and large samples. Simulation studies and real data applications show that the ML estimator performs better than others. Finally, one real data set has been analyzed for illustrative purposes.

Keywords Exponentiated moment exponential distribution · Least squares estimator · Maximum likelihood estimator · Model selection criteria · Percentile estimator · Weighted least squares estimator

1 Introduction

Moment distributions have found their applications in many studies of practical interest including bio-assay, clinical, income and life testing experiments. Hasnain (2013) studied an exponentiated moment exponential (EME) distribution and developed various distributional properties of it. He showed that this distribution is more flexible than Dara (2012) moment exponential distribution and Gupta and Kundu (2001) exponentiated exponential distribution. He discussed unbiasedness and sufficiency of the parameters of EME distribution. He also discussed different properties of n th order statistic and obtained dual generalized order statistics for this distribution. The moment distributions have applicability in numerous situations. For example, in the field of physics, Krumbein and Pettijohn (1938) and Gy (1982) studied moment distributions to examine the relation between particle diameter and its frequency. Preston (1962) applied canonical distributions in ecology. Zelen (1974) analyzed cell cycle and pulse labeling on length biased data. Zelen (1976) also used length biased sample for the screening of disease and scheduling of examinations of patients. Brown (1972) studied various aspects of the traffic streams. Warren (1975) examined different case studies to observe the role of statistical distributions in forest and forestry products research. Taillie et al. (1979) modeled populations of fish stocks using weighted distributions. Temkin (1976) studied length biased survival models to analyze the cardiac transplantation situations and so on. In these and other application areas, it is pertinent that one has accurate estimation and prediction procedures based on records and moments of record values.

✉ Yogesh Mani Tripathi
yogesh@iitp.ac.in

¹ Department of Mathematics, Indian Institute of Technology Patna, Bihta 801106, India

² Department of Statistics, St. Anthony's College, Shillong, Meghalaya, India

The probability density function (PDF) of an exponentiated moment exponential (EME) distribution is given by

$$f(x; \alpha, \beta) = \frac{\alpha x}{\beta^2} \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]^{\alpha-1} e^{-\frac{x}{\beta}}, \quad (1.1)$$

$$x > 0, \alpha > 0, \beta > 0,$$

and the corresponding cumulative distribution function (CDF) is

$$F(x; \alpha, \beta) = \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]^{\alpha}, \quad x > 0, \alpha > 0, \beta > 0. \quad (1.2)$$

The survival function is of the form

$$S(x; \alpha, \beta) = 1 - \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]^{\alpha}, \quad x > 0, \alpha > 0, \beta > 0, \quad (1.3)$$

and the hazard function is

$$h(x; \alpha, \beta) = \frac{\alpha x \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^2 \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]^{\alpha}}. \quad (1.4)$$

Here the parameters α controls the shape of the distribution and β controls the scale of the distribution. We assume that β is known except for real data analysis. In this paper, we estimate the PDF and the CDF of the model under the assumption that the shape parameter α is unknown. For this purpose, we employ various methods of estimation which includes maximum likelihood estimation (MLE), uniformly minimum variance unbiased (UMVU) estimation, percentile (PC) estimation, least-square estimation (LSE), weighted least-square estimation (WLSE), maximum product spacings (MPS) method of estimation, Cramér–von-Mises method of estimation and Anderson–Darling method of estimation. Several researchers have carried out similar kind of studies for various lifetime models. For example, Dixit and Jabbari Nooghabi (2010) obtained efficient estimates for the PDF and the CDF of a Pareto distribution. They showed that MLEs are more efficient than UMVUEs of the PDF and the CDF. In a subsequent work, Dixit and Jabbari Nooghabi (2011) derived efficient estimators of the PDF and the CDF of a four-parameter Pareto distribution under the assumption that its shape parameter is unknown and other parameters are known quantities. Bagheri et al. (2013) considered efficient estimation of the PDF and the CDF of a three-parameter generalized exponential-Poisson distribution with unknown shape parameter. Recently, Bagheri et al. (2016) estimated the CDF and the PDF of an exponentiated Gumbel distribution using different methods of estimation. They compared the performance of the

proposed estimators using a simulation study and analyzed a real data set for illustrative purpose.

The main difference between our work with the existing work is that we have considered eight methods of estimation for estimating the pdf and the cdf whereas in existing literature only five methods of estimation is considered to the best of our knowledge.

We organize the contents of this paper as follows. In Sect. 2 maximum likelihood estimators of the PDF and the CDF are derived. Mean squared errors (MSEs) of both the estimators are presented as well. The uniformly minimum variance unbiased (UMVU) estimators and their MSEs are discussed in Sect. 3. We present least squares estimators (LSEs) and the weighted LSEs (WLSEs) in Sect. 4 and in Sect. 5 percentile estimators (PCEs) are obtained. The maximum product spacings estimators are obtained in Sect. 6. We derive Cramér–von-Mises estimators of PDF and CDF in Sect. 7 and in Sect. 8 Anderson–Darling estimators are discussed. In Sect. 9 performance of suggested estimators is studied using a Monte Carlo simulation study. We analyze a real data set in Sect. 10 for illustration purpose. Finally, the paper ends with a conclusion in Sect. 11.

2 Maximum likelihood estimation

Let X_1, X_2, \dots, X_n be a random sample of size n drawn from a $\text{EME}(\alpha, \beta)$ distribution with known scale β . The maximum likelihood estimator (MLE) α is then obtained as

$$\hat{\alpha} = - \frac{n}{\sum_{i=1}^n \log \left[1 - \left(1 + \frac{x_i}{\beta} \right) e^{-\frac{x_i}{\beta}} \right]} = \frac{n}{T_1} \quad (2.1)$$

where $T_1 = - \sum_{i=1}^n \log \left[1 - \left(1 + \frac{x_i}{\beta} \right) e^{-\frac{x_i}{\beta}} \right]$. Now if X follows $\text{EME}(\alpha, \beta)$ distribution then $T = - \log \left[1 - \left(1 + \frac{X}{\beta} \right) e^{-\frac{X}{\beta}} \right]$ has an exponential distribution with rate α . Thus T_1 being sum of n independent such exponential distributions, it follows that T_1 is distributed as gamma $G(n, \alpha)$ distribution with density function given by

$$f_{T_1}(t_1) = \frac{\alpha^n}{\Gamma n} t_1^{n-1} e^{-\alpha t_1}, \quad t_1 > 0, \alpha > 0. \quad (2.2)$$

Also observe that $\frac{1}{T_1}$ has inverse gamma distribution and then the distribution of MLE $\hat{\alpha} = \frac{n}{T_1} = S$ is given by

$$f_S(s) = \frac{n^n \alpha^n}{\Gamma n} s^{-n-1} e^{-\frac{n\alpha}{s}}, \quad s > 0.$$

Next we observe that MLEs of the PDF and the CDF are respectively given by

$$\hat{f}(x) = \frac{\hat{\alpha}x}{\hat{\beta}^2} \left[1 - \left(1 + \frac{x}{\hat{\beta}} \right) e^{-\frac{x}{\hat{\beta}}} \right]^{\hat{\alpha}-1} e^{-\frac{x}{\hat{\beta}}} \tag{2.3}$$

and

$$\hat{F}(x) = \left[1 - \left(1 + \frac{x}{\hat{\beta}} \right) e^{-\frac{x}{\hat{\beta}}} \right]^{\hat{\alpha}} \tag{2.4}$$

In the theorem presented below we show that the estimators $\hat{f}(x)$ and $\hat{F}(x)$ are biased for estimating $f(x)$ and $F(x)$ respectively. Before we proceed further we rewrite these estimators as $\hat{f}(x) = \hat{\alpha}b(x)(a(x))^{\hat{\alpha}-1}$ and $\hat{F}(x) = (a(x))^{\hat{\alpha}}$ where $b(x) = \frac{xe^{-\frac{x}{\beta}}}{\beta^2}$, $a(x) = \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]$.

Theorem 2.1 We have

$$E[\hat{f}(x)]^m = 2 \left[\frac{b(x)}{a(x)} \right]^m \frac{(n\alpha)^{\frac{m+n}{2}}}{\Gamma n} \left[-\frac{1}{m \ln a(x)} \right]^{\frac{m-n}{2}} K_{m-n} \left(2\sqrt{-n\alpha m \ln a(x)} \right)$$

and

$$E[\hat{F}(x)]^m = 2 \frac{(n\alpha)^{\frac{m}{2}}}{\Gamma n} \left[-\frac{1}{m \ln a(x)} \right]^{-\frac{m}{2}} K_{-n} \left(2\sqrt{-n\alpha m \ln a(x)} \right).$$

Proof We can write

$$\begin{aligned} E[\hat{f}(x)]^m &= \int_0^\infty \left[sb(x)(a(x))^{s-1} \right]^m \frac{(n\alpha)^n}{\Gamma n} \frac{1}{s^{n+1}} e^{-\frac{ns}{s}} ds \\ &= \frac{(n\alpha)^n}{\Gamma n} (b(x))^m \int_0^\infty s^m (a(x))^{sm-m} \frac{1}{s^{n+1}} e^{-\frac{ns}{s}} ds \\ &= \frac{(n\alpha)^n}{\Gamma n} (b(x))^m \int_0^\infty e^{(sm-m) \ln a(x)} s^{m-n-1} e^{-\frac{ns}{s}} ds \\ &= 2 \left[\frac{b(x)}{a(x)} \right]^m \frac{(n\alpha)^{\frac{m+n}{2}}}{\Gamma n} \left[-\frac{1}{m \ln a(x)} \right]^{\frac{m-n}{2}} K_{m-n} \left(2\sqrt{-n\alpha m \ln a(x)} \right), \end{aligned}$$

where K_ν denotes the modified Bessel function of second kind of order ν . Also

$$\begin{aligned} E[\hat{F}(x)]^m &= \int_0^\infty (a(x))^{sm} \frac{(n\alpha)^n}{\Gamma n} \frac{1}{s^{n+1}} e^{-\frac{ns}{s}} ds \\ &= \frac{(n\alpha)^n}{\Gamma n} \int_0^\infty e^{sm \ln a(x)} \frac{e^{-\frac{ns}{s}}}{s^{n+1}} ds \\ &= 2 \frac{(n\alpha)^{\frac{m}{2}}}{\Gamma n} \left[-\frac{1}{m \ln a(x)} \right]^{-\frac{m}{2}} K_{-n} \left(2\sqrt{-n\alpha m \ln a(x)} \right) \end{aligned}$$

□

Theorem 2.2 The mean squared error of estimates $\hat{f}(x)$ and $\hat{F}(x)$ are given by respectively

$$\begin{aligned} MSE(\hat{f}(x)) &= 2 \left(\frac{a(x)}{b(x)} \right)^2 \frac{(n\alpha)^{\frac{n+2}{2}}}{\Gamma n} \left[-\frac{1}{2 \ln a(x)} \right]^{\frac{2-n}{2}} \\ &\quad K_{2-n} \left(2\sqrt{-2\alpha n \ln a(x)} \right) \\ &\quad - 4f(x) \frac{a(x)}{b(x)} \frac{(n\alpha)^{\frac{n+1}{2}}}{\Gamma n} \left[-\frac{1}{\ln a(x)} \right]^{\frac{1-n}{2}} \\ &\quad K_{1-n} \left(2\sqrt{-n\alpha \ln a(x)} \right) + (f(x))^2 \end{aligned}$$

and

$$\begin{aligned} MSE(\hat{F}(x)) &= 2 \frac{(n\alpha)^{\frac{n}{2}}}{\Gamma n} \left[-\frac{1}{2 \ln a(x)} \right]^{-\frac{n}{2}} K_{-n} \left(2\sqrt{-2n\alpha \ln a(x)} \right) \\ &\quad - 4F(x) \frac{(n\alpha)^{\frac{n}{2}}}{\Gamma n} K_{-n} \left(2\sqrt{-n\alpha \ln a(x)} \right) + (F(x))^2 \end{aligned}$$

Proof The proof follows by proceeding in a manner similar to the proof of the previous theorem. □

3 Uniformly minimum variance unbiased estimation

This section deals with finding uniformly minimum variance unbiased (UMVU) estimators of the PDF and the CDF of an EME distribution. The mean square errors (MSE) of these estimators are also obtained. Suppose X_1, X_2, \dots, X_n denotes a random sample of size n drawn from the EME distribution as defined in (1.1). We consider estimation of $f(x)$ and $F(x)$ under the assumption that the scale parameter β is known. For this problem $T_1 = \sum_{i=1}^n \log \left[1 - \left(1 + \frac{x_i}{\beta} \right) e^{-\frac{x_i}{\beta}} \right]^{-1}$ is a complete and sufficient statistic for α and it has a gamma $G(n, \alpha)$ distribution. Following Lehmann–Scheffe theorem if $g_{X_1|T_1}(x_1 | T_1 = t) = h^*(t)$ is the conditional pdf of X_1 given $T_1 = t$ then we have

$$\begin{aligned} E[h^*(T_1)] &= \int g_{X_1|T_1}(x_1 | t) g_{T_1}(t) dt = \int g(x_1, t) dt \\ &= f(x_1) \end{aligned}$$

where $g(x_1, t)$ is the joint pdf of (X_1, T_1) . Thus $h^*(T_1)$ is the UMVU estimator of $f(x)$.

Lemma 3.1 The conditional distribution of Z_1 given $T_1 = t_1$ is obtained as

$$\begin{aligned} f_{Z_1|T_1}(z_1 | t_1) &= \frac{(n-1) \left[t_1 - \log \left(1 - \left(1 + \frac{x_1}{\beta} \right) e^{-\frac{x_1}{\beta}} \right) \right]^{n-2}}{t_1^{n-1}} \\ &\quad - \log \left[1 - \left(1 + \frac{x_1}{\beta} \right) e^{-\frac{x_1}{\beta}} \right] < t_1 < \infty \end{aligned}$$

where $z_1 = \log \left[1 - \left(1 + \frac{x_1}{\beta} \right) e^{-\frac{x_1}{\beta}} \right]^{-1}$.

Proof We can write the joint pdf of (Z_1, T_1) as

$$f_{Z_1, T_1}(z_1, t_1) = f(t_1)f(z_1 | t_1).$$

We now have

$$\begin{aligned} f_{Z_1|T_1}(z_1 | t_1) &= \frac{f(Z_1 = z_1, T_1 = t_1)}{f_{T_1}(t_1)} \\ &= \frac{f(Z_1 = z_1, \sum_{i=2}^n [1 - (1 + \frac{x_i}{\beta})e^{-\frac{x_i}{\beta}}]^{-1} = t_1 - z_1)}{f_{T_1}(t_1)} \\ &= \frac{\alpha e^{-\alpha z_1} \frac{\alpha^{n-1}}{\Gamma(n-1)} (t_1 - z_1)^{n-2} e^{-\alpha(t_1 - z_1)}}{\frac{\alpha^n t_1^{n-1} e^{-\alpha t_1}}{\Gamma n}} \\ &= (n-1) \frac{(t_1 - z_1)^{n-2}}{t_1^{n-1}} \\ &= (n-1) \frac{[t_1 + \log(1 - (1 + \frac{x_1}{\beta})e^{-\frac{x_1}{\beta}})]^{n-2}}{t_1^{n-1}}, \\ &\quad -\log\left(1 - \left(1 + \frac{x_1}{\beta}\right)e^{-\frac{x_1}{\beta}}\right) < t_1 < \infty. \end{aligned}$$

In the following theorem we provide UMVU estimators of $f(x)$ and $F(x)$. □

Theorem 3.1 For a given $T = t$, the estimator $\tilde{f}(x)$ with

$$\tilde{f}(x) = \frac{x e^{-\frac{x}{\beta}}(n-1)}{\beta^2 \left[1 - \left(1 + \frac{x}{\beta}\right)e^{-\frac{x}{\beta}}\right]} \frac{[t_1 + \log(1 - (1 + \frac{x}{\beta})e^{-\frac{x}{\beta}})]^{n-2}}{t_1^{n-1}}$$

is the UMVUE of $f(x)$ and the estimator $\tilde{F}(x)$ with

$$\tilde{F}(x) = \left[1 + \frac{\log(1 - (1 + \frac{x}{\beta})e^{-\frac{x}{\beta}})}{t_1}\right]^{n-1}$$

is the UMVUE of $F(x)$ where

$$-\log\left(1 - \left(1 + \frac{x}{\beta}\right)e^{-\frac{x}{\beta}}\right) < t_1 < \infty$$

Proof We observe that

$$\begin{aligned} E[\tilde{f}(x)] &= \left[\frac{x e^{-\frac{x}{\beta}}(n-1)}{\beta^2 \left(1 - \left(1 + \frac{x}{\beta}\right)e^{-\frac{x}{\beta}}\right)}\right] \int_{-\log a(x)}^{\infty} \frac{(t_1 + \log a(x))^{n-2} \alpha^n t_1^{n-1} e^{-\alpha t_1}}{\Gamma n} dt \\ &= \frac{x e^{-\frac{x}{\beta}}(n-1)}{\beta^2 \left(1 - \left(1 + \frac{x}{\beta}\right)e^{-\frac{x}{\beta}}\right)} \frac{\alpha e^{\alpha \log(1 - (1 + \frac{x}{\beta})e^{-\frac{x}{\beta}})}}{n-1} \\ &= \frac{\alpha x}{\beta^2} e^{-\frac{x}{\beta}} \left(1 - \left(1 + \frac{x}{\beta}\right)e^{-\frac{x}{\beta}}\right)^{\alpha-1} = f(x) \end{aligned}$$

Thus the estimator $\tilde{f}(x)$ is UMVU estimator of $f(x)$. In order to prove that $\tilde{F}(x)$ is UMVU estimator of $F(x)$ we show that $\frac{d}{dx} \tilde{F}(x) = \tilde{f}(x)$.

Now

$$\begin{aligned} \frac{d}{dx} \tilde{F}(x) &= \frac{d}{dx} \left[1 + \frac{\log\left(1 - \left(1 + \frac{x}{\beta}\right)e^{-\frac{x}{\beta}}\right)}{t_1}\right]^{n-1} \\ &= \frac{e^{-\frac{x}{\beta}}}{\beta^2} (n-1) \frac{[t_1 + \log(1 - (1 + \frac{x}{\beta})e^{-\frac{x}{\beta}})]^{n-2}}{t_1^{n-1} \left(1 - \left(1 + \frac{x}{\beta}\right)e^{-\frac{x}{\beta}}\right)} = \tilde{f}(x). \end{aligned}$$

This completes the proof of the theorem. □

Theorem 3.2 The mean squared errors of $\tilde{f}(x)$ and $\tilde{F}(x)$ are respectively given by

$$\begin{aligned} MSE(\tilde{f}(x)) &= ((n-1)a^*(x))^2 \sum_{i=0}^{(2(n-2))} \binom{2(n-2)}{i} \\ &\quad (a(x))^i \Gamma(n-2-i, a(x)\alpha) - f^2(x) \end{aligned}$$

and

$$\begin{aligned} MSE(\tilde{F}(x)) &= \frac{\alpha^n}{\Gamma n} \sum_{j=0}^{(n-1)2} \binom{(n-1)2}{j} (\log(a(x)))^j \\ &\quad \Gamma(n-j, -\log a(x)\alpha) - F^2(x). \end{aligned}$$

Proof Note that

$$\tilde{f}(x) = a^*(x)(n-1) \frac{(t_1 + \log a(x))^{n-2}}{t_1^{n-1}},$$

with $a^*(x) = \frac{x}{\beta^2} \frac{e^{-\frac{x}{\beta}}}{1 - (1 + \frac{x}{\beta})e^{-\frac{x}{\beta}}}$ and $\tilde{F}(x) = \left[1 + \frac{\log a(x)}{t_1}\right]^{n-1}$. We

calculate the following expectations

$$\begin{aligned} E[\tilde{f}(x)]^p &= (a^*(x))^p (n-1)^p \int_{-\log a(x)}^{\infty} \frac{(t_1 + \log a(x))^{(n-2)p} \alpha^n t_1^{n-1} e^{-\alpha t_1}}{\Gamma n} dt_1 \\ &= ((n-1)a^*(x))^p \frac{\alpha^n}{\Gamma n} \int_{-\log a(x)}^{\infty} \left(1 + \frac{\log a(x)}{t_1}\right)^{(n-2)p} t_1^{n-p-1} e^{-\alpha t_1} dt_1 \\ &= ((n-1)a^*(x))^p \frac{\alpha^n}{\Gamma n} \int_{-\log a(x)}^{\infty} \sum_{i=0}^{(n-2)p} \binom{(n-2)p}{i} (\log a(x))^i t_1^{n-p-i-1} e^{-\alpha t_1} dt_1 \\ &= ((n-1)a^*(x))^p \sum_{i=0}^{(n-2)p} \binom{(n-2)p}{i} (\log a(x))^i \Gamma(n-p-i, \log a(x)\alpha) \end{aligned}$$

and

$$\begin{aligned}
 E[\tilde{F}(x)]^p &= \int_{-\log a(x)}^{\infty} \left[\left(1 + \frac{\log a(x)}{t_1} \right)^{n-1} \right]^p \frac{\alpha^n}{\Gamma n} t_1^{n-1} e^{-\alpha t_1} dt_1 \\
 &= \frac{\alpha^n}{\Gamma n} \int_{-\log a(x)}^{\infty} \sum_{j=0}^{(n-1)p} \binom{(n-1)p}{j} \left(\frac{\log a(x)}{t_1} \right)^j t_1^{n-1} e^{-\alpha t_1} dt_1 \\
 &= \frac{\alpha^n}{\Gamma n} \sum_{j=0}^{(n-1)p} \binom{(n-1)p}{j} (\log a(x))^j \Gamma(n-j, -\log a(x)\alpha).
 \end{aligned}$$

We further observe that $MSE(\tilde{f}(x)) = E[\tilde{f}(x)]^2 - (f(x))^2$ and the corresponding expectation can be obtained using the above calculations. Similarly we can obtain mean squared error of the estimator $\tilde{F}(x)$. \square

4 Least squares and weighted least squares estimators

This section deals with regression based estimators for the unknown shape parameter α . This method has been introduced in literature by Swain et al. (1988) for estimating the unknown parameters of a beta distribution. Let $F(X_{(j:n)})$ denotes the distribution function of the order random variables $(X_{1:n}, X_{2:n}, \dots, X_{n:n})$ with X_1, X_2, \dots, X_n being a random sample drawn from the distribution as defined in (1). Observe that for a sample of size n we have $E(F(X_{(j:n)})) = \frac{1}{n+1}$, $V(F(X_{(j:n)})) = \frac{j(n-j+1)}{(n+1)^2(n+2)}$ and $cov(F(X_{(j:n)}), F(X_{(k:n)})) = \frac{j(n-k+1)}{(n+1)^2(n+2)}$, $j < k$ (see, Johnson et al. 1994) we provide two different least square estimators for the unknown parameter α which can be used to estimate the PDF and the CDF of the considered model.

4.1 Least square estimators (LSEs)

In this method we minimize the expression $\sum_{j=1}^n \left[F(X_{j:n}) - \frac{j}{n+1} \right]^2$ with respect to the unknown parameter. For the considered model the desired estimator α_{ls} of α (with known β) is the value minimizing $\sum_{j=1}^n \left[\left(1 - \left(1 + \frac{X_{(j:n)}}{\beta} \right) e^{-\frac{X_{(j:n)}}{\beta}} \right)^\alpha - \frac{j}{n+1} \right]^2$. Thus the least square estimators of the PDF and the CDF are given by

$$f_{ls}(x) = \frac{\tilde{\alpha}_{ls}}{\beta^2} \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]^{\tilde{\alpha}_{ls}-1} e^{-\frac{x}{\beta}} \tag{4.1}$$

and

$$F_{ls}(x) = \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]^{\tilde{\alpha}_{ls}}. \tag{4.2}$$

The expected values and mean square errors of these estimators are computed using simulations as it is difficult to evaluate these quantities analytically.

4.2 Weighted least square estimators

In this set up we minimize the expression $\sum_{j=1}^n w_j \left[F(X_{(j:n)}) - \frac{j}{n+1} \right]^2$ with respect to the unknown quantity α where the weight function is $w_j = \frac{(n+1)^2(n+2)}{j(n-j+1)}$. For the EME distribution the weighted LSE $\tilde{\alpha}_{wls}$ of α is obtained by minimizing the expression

$$\sum_{j=1}^n \frac{(n+1)^2(n+2)}{j(n-j+1)} \left[\left(1 - \left(1 + \frac{X_{(j:n)}}{\beta} \right) e^{-\frac{X_{(j:n)}}{\beta}} \right)^\alpha - \frac{j}{n+1} \right]^2.$$

Thus the weighted LS estimators of $f(x)$ and $F(x)$ are obtained as

$$\tilde{f}_{wls}(x) = \tilde{\alpha}_{wls} \frac{x}{\beta^2} \left(1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right)^{\tilde{\alpha}_{wls}-1} e^{-\frac{x}{\beta}} \tag{4.3}$$

and

$$\tilde{F}_{wls}(x) = \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]^{\tilde{\alpha}_{wls}} \tag{4.4}$$

The bias and MSE values of these estimators are calculated using simulation.

5 Estimators based on percentiles

In this section we consider estimation of the pdf and the cdf based percentile estimators. This method is originally discussed in Kao (1958), however, one may also refer to Mann et al. (1974) and Johnson et al. (1994) for some further results on this method of estimation. This method has found easy applications in estimation problems where distribution function has a closed form expression. Since the considered model has a closed form distribution function we can estimate its parameters using this technique.

Let $F(X_{(j:n)})$ be the CDF of the ordered random sample $(X_{1:n}, X_{2:n}, \dots, X_{(n:n)})$ with (X_1, X_2, \dots, X_n) taken from an EME distribution. For further consideration let $p_i = \frac{i}{n+1}$ then percentile estimator $\tilde{\alpha}_{pe}$ of α minimizes the expression

$$\sum_{i=1}^n \left[\ln p_i - \alpha \ln \left(1 - \left(1 + \frac{X_{i:n}}{\beta} \right) e^{-\frac{X_{i:n}}{\beta}} \right) \right]^2. \tag{5.1}$$

The desired estimator $\tilde{\alpha}_{pe}$ is the solution to equation

$$\sum_{i=1}^n \left[\ln p_i - \alpha \ln \left(1 - \left(1 + \frac{x_{i:n}}{\beta} \right) e^{-\frac{x_{i:n}}{\beta}} \right) \right] \ln \left(1 - \left(1 + \frac{x_{i:n}}{\beta} \right) e^{-\frac{x_{i:n}}{\beta}} \right) = 0 \tag{5.2}$$

This implies that

$$\tilde{\alpha}_{pe} = \frac{\sum_{i=1}^n \left[(\ln p_i) \ln \left(1 - \left(1 + \frac{x_{i:n}}{\beta} \right) e^{-\frac{x_{i:n}}{\beta}} \right) \right]}{\sum_{i=1}^n \left[\ln \left(1 - \left(1 + \frac{x_{i:n}}{\beta} \right) e^{-\frac{x_{i:n}}{\beta}} \right) \right]^2}. \tag{5.3}$$

So the percentile estimator of the PDF and the CDF are now given by

$$\tilde{f}_{pe}(x) = \frac{\tilde{\alpha}_{pe} x}{\beta^2} \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]^{\tilde{\alpha}_{pe}-1} e^{-\frac{x}{\beta}} \tag{5.4}$$

and

$$\tilde{F}_{pe}(x) = \left[1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right]^{\tilde{\alpha}_{pe}}. \tag{5.5}$$

6 Method of maximum product spacing

This method was introduced by Cheng and Amin (1979) as an alternative method to the maximum likelihood estimation method. For a given sample of size of n taken from an EME distribution, the corresponding uniform spacing is given by

$$\Delta_i(\alpha, \beta) = F(x_{i:n} | \alpha, \beta) - F(x_{i-1:n} | \alpha, \beta) \tag{6.1}$$

for $i = 1, 2, \dots, n$ with $F(x_{0:n} | \alpha, \beta) = 0, F(x_{n+1:n} | \alpha, \beta) = 1$ and $\sum_{i=1}^n \Delta_i(\alpha, \beta) = 1$.

The maximum product spacing estimator $\tilde{\alpha}_{mps}$ of α is computed by maximizing the expression

$$\Delta(\alpha, \beta) = \left[\prod_{i=1}^{n+1} \Delta_i(\alpha, \beta) \right]^{\frac{1}{n+1}} \tag{6.2}$$

with respect to α . Equivalently we maximize the function

$$\Delta^*(\alpha, \beta) = \frac{1}{n+1} \sum_{i=1}^{n+1} \ln \Delta_i(\alpha, \beta),$$

and observe that $\tilde{\alpha}_{mps}$ can be computed from the equation

$$\frac{\partial \Delta^*(\alpha, \beta)}{\partial \alpha} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{1}{\Delta_i(\alpha, \beta)} \left[\Delta_0(x_{i:n} | \alpha, \beta) - \Delta_0(x_{i-1:n} | \alpha, \beta) \right] \tag{6.3}$$

where $\Delta_0(x_{i:n} | \alpha, \beta) = \left[1 - \left(1 + \frac{x_{i:n}}{\beta} \right) e^{-\frac{x_{i:n}}{\beta}} \right]^\alpha \ln \left[1 - \left(1 + \frac{x_{i:n}}{\beta} \right) e^{-\frac{x_{i:n}}{\beta}} \right]$. Cheng and Amin (1983) observed that

maximizing $\Delta^*(\alpha, \beta)$ as a method of parameter estimation is as efficient as the MLEs. The MPS estimator of the PDF and the CDF are, respectively,

$$\tilde{f}_{mps}(x) = \left(1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right)^{\tilde{\alpha}_{mps}-1} \frac{\tilde{\alpha}_{mps} x}{\beta} e^{-\frac{x}{\beta}} \tag{6.4}$$

and

$$\tilde{F}_{mps}(x) = \left(1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right)^{\tilde{\alpha}_{mps}}. \tag{6.5}$$

7 Method of Cramér–von-Mises

Macdonald (1971) gave empirical evidence that bias values of the Cramér–von-Mises type estimators are smaller than the other minimum distance estimators. The Cramér–von-Mises estimates $\tilde{\alpha}_{cvm}$ of the parameter α is computed by minimizing, with respect to α , the function

$$V(\alpha, \beta) = \frac{1}{12n} + \sum_{i=1}^n \left(F(x_{i:n} | \alpha, \beta) - \frac{2i-1}{2n} \right)^2. \tag{7.1}$$

The estimator $\tilde{\alpha}_{cvm}$ can also be derived from the following non-linear equation

$$\sum_{i=1}^n \left(F(x_{i:n} | \alpha, \beta) - \frac{2i-1}{2n} \right) \Delta_0(x_{i:n} | \alpha, \beta) = 0 \tag{7.2}$$

where $\Delta_0(x_{i:n} | \alpha, \beta) = 0$ is defined in the previous section. Now the CVM estimators of the PDF and the CDF are given by

$$\tilde{f}_{cvm}(x) = \left(1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right)^{\tilde{\alpha}_{cvm}-1} e^{-\frac{x}{\beta}} \tag{7.3}$$

and

$$\tilde{F}_{cvm}(x) = \left(1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right)^{\tilde{\alpha}_{cvm}}. \tag{7.4}$$

The bias and MSE values of these estimators are computed using simulations.

8 Method of Anderson–Darling

Anderson and Darling (1952) developed a test to detect normality distribution of various data. The Anderson–Darling estimator $\tilde{\alpha}_{ad}$ of α is derived by minimizing the function

$$AD(\alpha, \beta) = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \left[\ln F(x_{i:n} | \alpha, \beta) + \ln \bar{F}(x_{n+1-i:n} | \alpha, \beta) \right]. \tag{8.1}$$

Fig. 1 $\alpha = 1.5, \beta = 0.5$

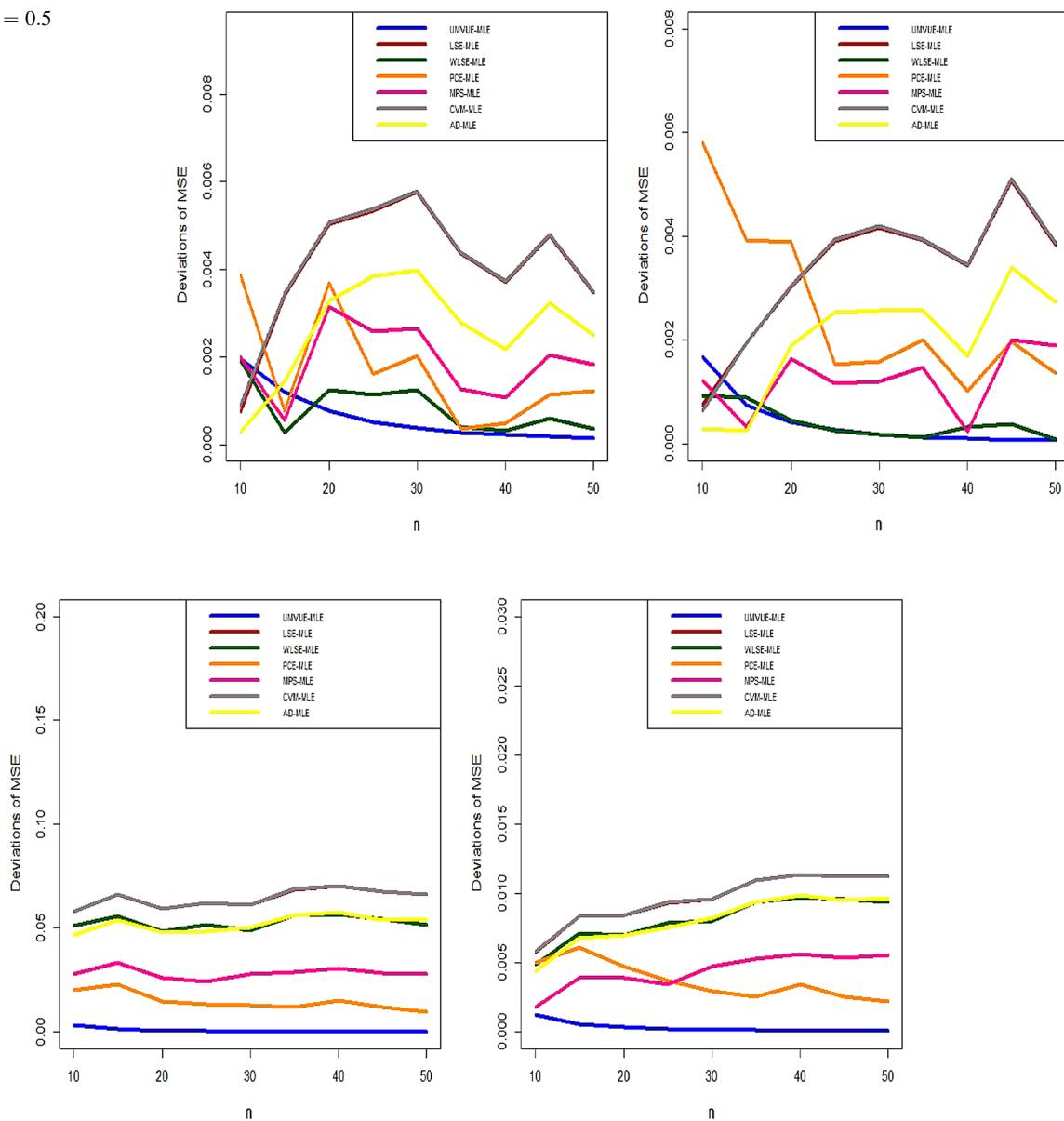


Fig. 2 $\alpha = 1.5, \beta = 0.75$

The estimator $\tilde{\alpha}_{ad}$ can also be obtained by solving the following non-linear equations

$$\sum_{i=1}^n (2i - 1) \left[\frac{\Delta_0(x_{(i:n)} | \alpha, \beta)}{F(x_{(i:n)} | \alpha, \beta)} - \frac{\Delta(x_{(i:n)} | \alpha, \beta)}{\bar{F}(x_{(i:n)} | \alpha, \beta)} \right] = 0 \quad (8.2)$$

where $\Delta_0(\cdot | \alpha, \beta)$ is defined previously.

The Anderson–Darling estimators of the PDF and the CDF are now given by

$$\tilde{f}_{ad}(x) = \left(1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right)^{\tilde{\alpha}_{ad}} \frac{\tilde{\alpha}_{ad} x}{\beta^2} e^{-\frac{x}{\beta}} \quad (8.3)$$

and

$$\tilde{F}_{ad}(x) = \left(1 - \left(1 + \frac{x}{\beta} \right) e^{-\frac{x}{\beta}} \right)^{\tilde{\alpha}_{ad}} \quad (8.4)$$

9 Simulation study

In this section, we study the behavior of the suggested methods of estimation using a simulation study in terms of their mean squared error values. Here, we compute the deviation of MSEs which represent difference between MSE of an estimator from the MSE of

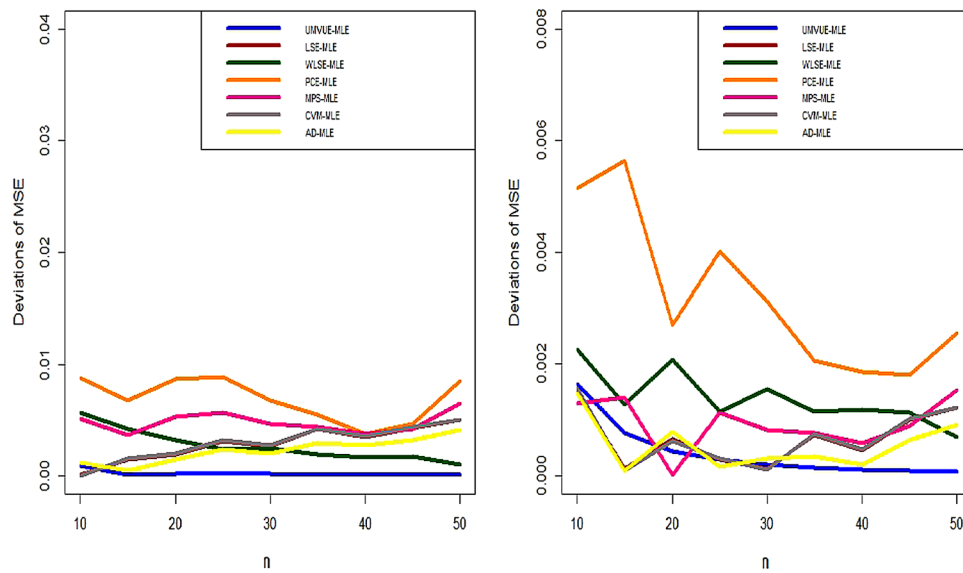


Fig. 3 $\alpha = 2, \beta = 0.5$

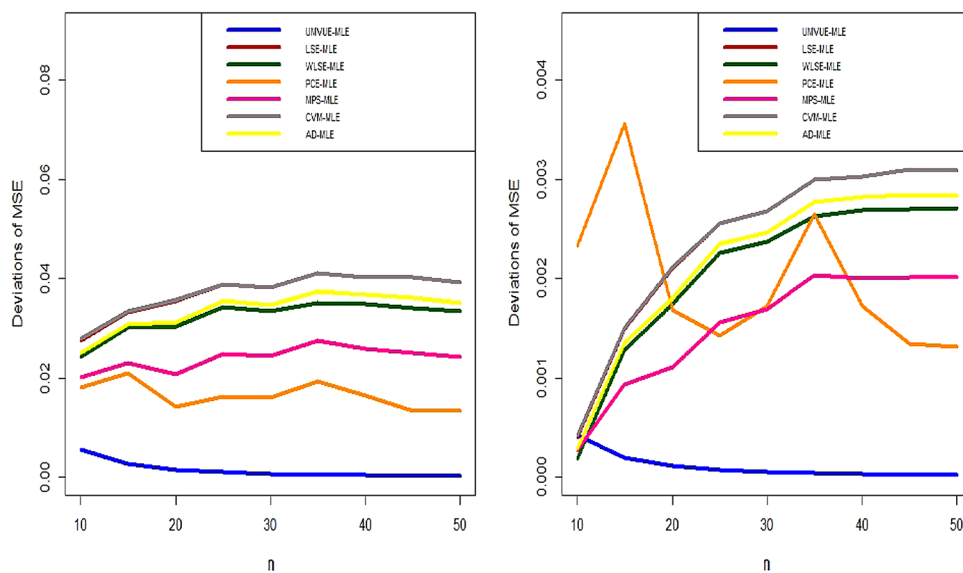


Fig. 4 $\alpha = 2, \beta = 0.75$

maximum likelihood estimator. These values are computed based on 1000 simulated samples from the EME distribution with different sample sizes $n = 10, 15, 20, 25, 30, 35, 40, 45, 50$ and for arbitrarily assigned values of parameters α and β . We compute MSEs for different values of x as $x = 0.75, 1$ and 1.5 . Figures 1, 2, 3 and 4 represent deviation of MSEs at $x = 0.75$. Similarly, Figs. 5, 6, 7, 8, 9, 10, 11 and 12 represent deviation of MSEs at $x = 1$ and $x = 1.5$ respectively. Note that, left

hand side figures are for the PDF estimation and right hand side figures are for the CDF estimation. It is observed from these figures that maximum likelihood estimators of the PDF and the CDF perform better than other proposed estimators and their efficiency improve with the increase in sample sizes. In general, the performance of UMVU estimators is good as far as its MSE is concerned. With moderate increase in α , we get better estimates of the PDF and the CDF for a given β .

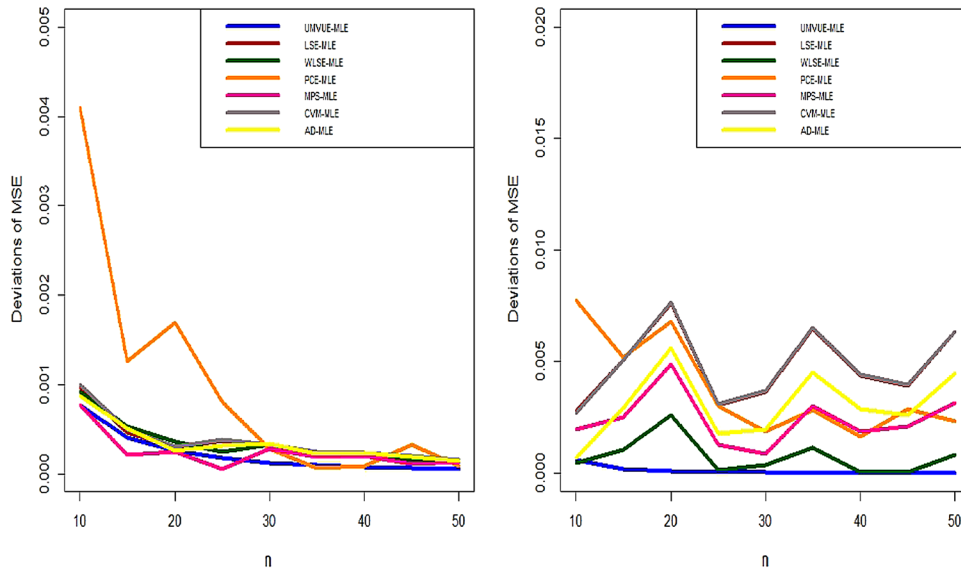


Fig. 5 $\alpha = 1.5, \beta = 0.5$

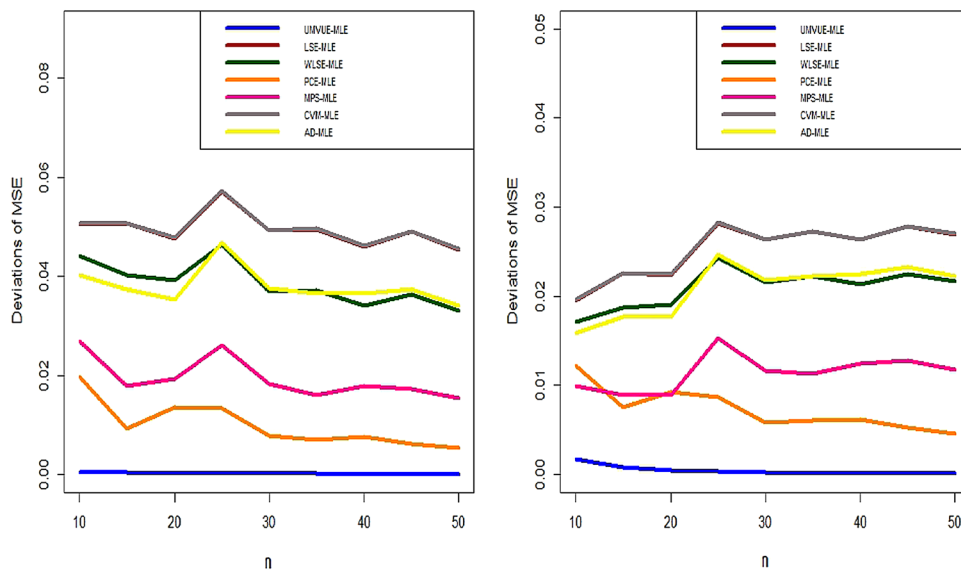


Fig. 6 $\alpha = 1.5, \beta = 0.75$

10 Data analysis

Here we analyze a real data set to illustrate the use of proposed estimators in practice. This data set as reported in Lawless (1982) represents the survival times for the control group laboratory mice that were exposed to a fixed dose of radiation for a duration of 5–6 weeks and the mice died due to tymic lymphoma. The data set is listed below as

159 189 191 198 200 207 220 235 245 250 256 261 265 266 280
 343 350 383 403 414 428 432

We first examine whether an EME distribution can be used to draw inference from this data set. For this purpose we fitted this data set using EME and Weibull distributions. Since for real data the unknown parameters α and β of EME distribution are treated as unknown and so they need to be estimated using some estimation method for further analysis. Here we use maximum likelihood estimation method to estimate α and β . To proceed further we assume

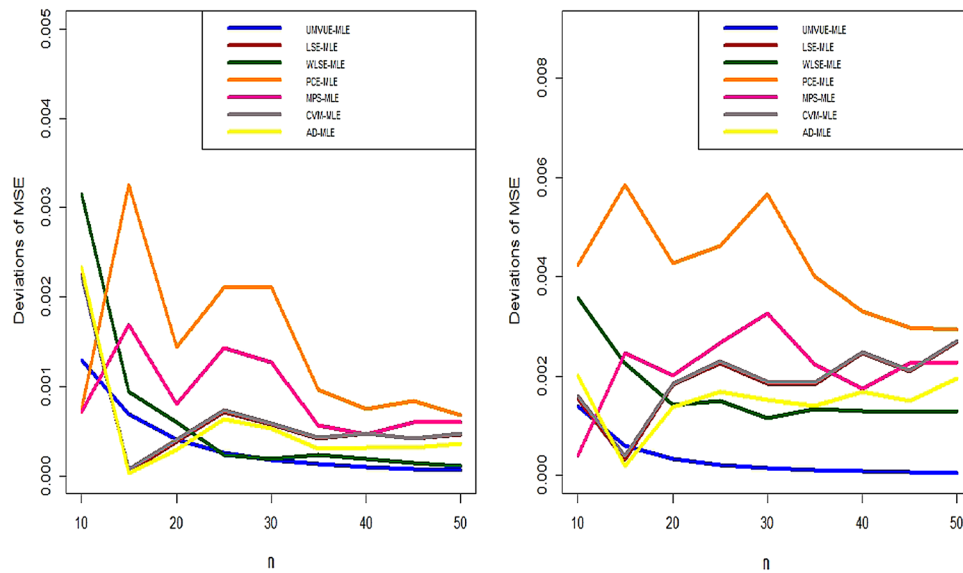


Fig. 7 $\alpha = 2, \beta = 0.5$

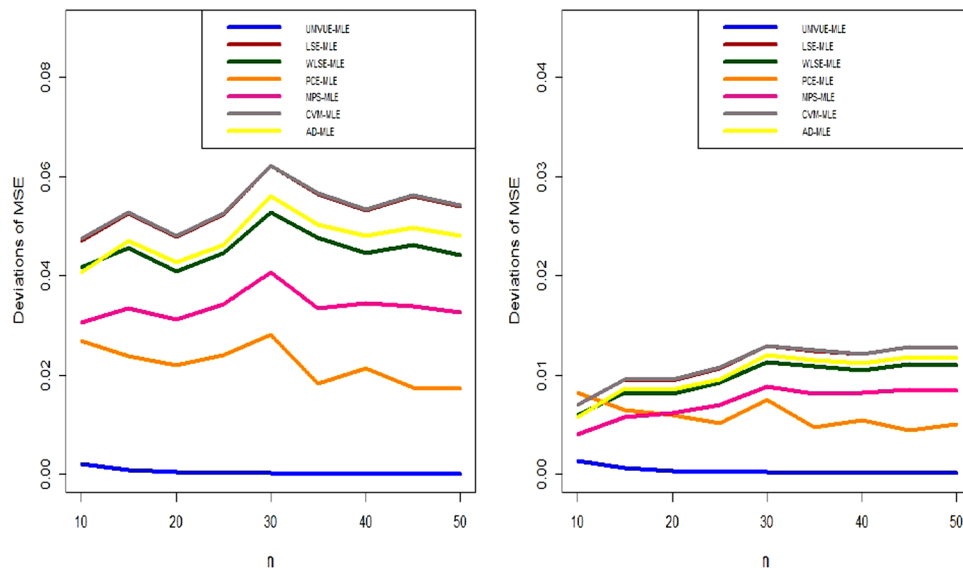


Fig. 8 $\alpha = 2, \beta = 0.75$

that X_1, X_2, \dots, X_n denote a random sample from a EME distribution. Below we derive MLE of α, β based on the given data set. The likelihood equation of (α, β) is

$$L(\alpha, \beta) = \prod_{i=1}^n \left[1 - \left(1 + \frac{x_i}{\beta} \right) e^{-\frac{x_i}{\beta}} \right] \frac{\alpha x_i}{\beta^2} e^{-\frac{x_i}{\beta}}$$

and the corresponding log likelihood function is obtained

$$l = \ln L = \alpha \sum_{i=1}^n \ln \left(1 - \left(1 + \frac{x_i}{\beta} \right) e^{-\frac{x_i}{\beta}} \right) + n \ln \alpha + \sum_{i=1}^n x_i - \frac{\sum_{i=1}^n x_i}{\beta} - 2n \ln \beta.$$

Now the likelihood equations of α and β are

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^n \ln \left(1 - \left(1 + \frac{x_i}{\beta} \right) e^{-\frac{x_i}{\beta}} \right) + \frac{n}{\alpha} = 0$$

$$\frac{\partial l}{\partial \beta} = \alpha \sum_{i=1}^n \frac{x_i e^{-\frac{x_i}{\beta}}}{\left(1 - \left(1 + \frac{x_i}{\beta} \right) e^{-\frac{x_i}{\beta}} \right)} + \sum_{i=1}^n x_i - 2n\beta = 0.$$

The desired estimates of α and β can now be obtained by simultaneously solving the above system of equations. Similarly unknown parameters of the Weibull distribution can be estimated. We report MLEs of unknown parameters of these two distributions along with the

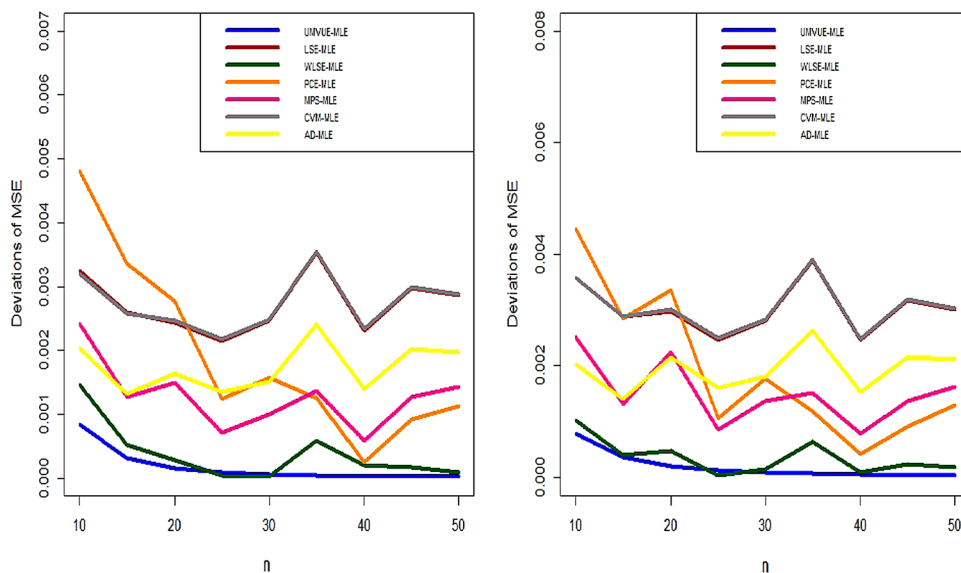


Fig. 9 $\alpha = 1.5, \beta = 0.5$

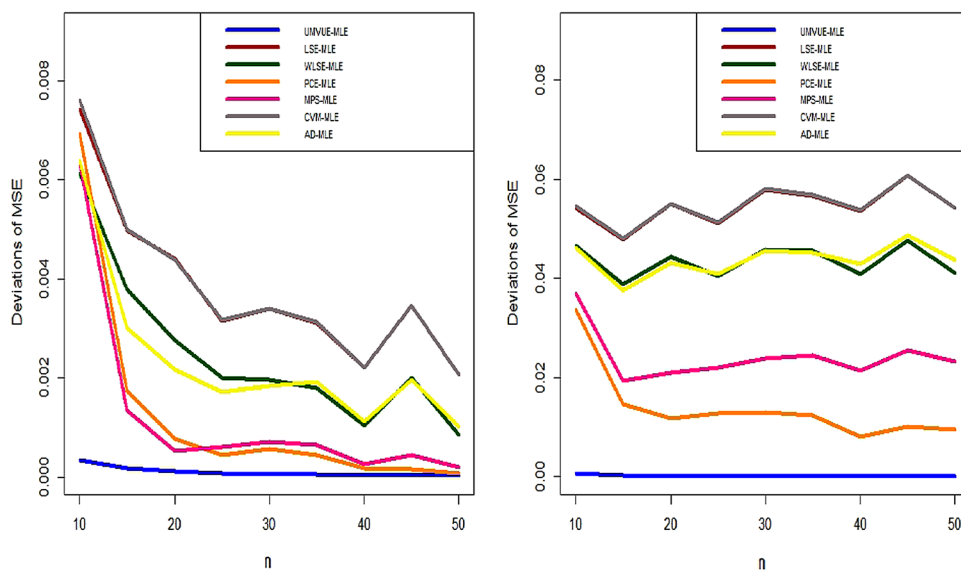


Fig. 10 $\alpha = 1.5, \beta = 0.75$

values of K-S test and the corresponding p values in Table 1. For computational simplification, we divided each data points by 100. The reported results suggest that the EME distribution fits the data set reasonably good. Further, we also fit the EME distribution to the given data set using the suggested procedures for model selection purpose. In Table 2 we have reported different estimates of α and β obtained using MLE, LSE, WLSE, PCE, CVE, MPS and AD methods of estimation. The

corresponding values of different model selection criteria such as maximum likelihood (ML) criterion, Akaike information criterion (AIC), corrected Akaike information criterion (AICc), Bayes information criterion (BIC), Hannan-Quinn criterion (HQC) are also displayed in this table. Results suggest that, the percentile estimators provide better fit among different model selection criteria which is reasonably in line with the simulation study as well. In Fig. 13, we provide $Q-Q$ plots

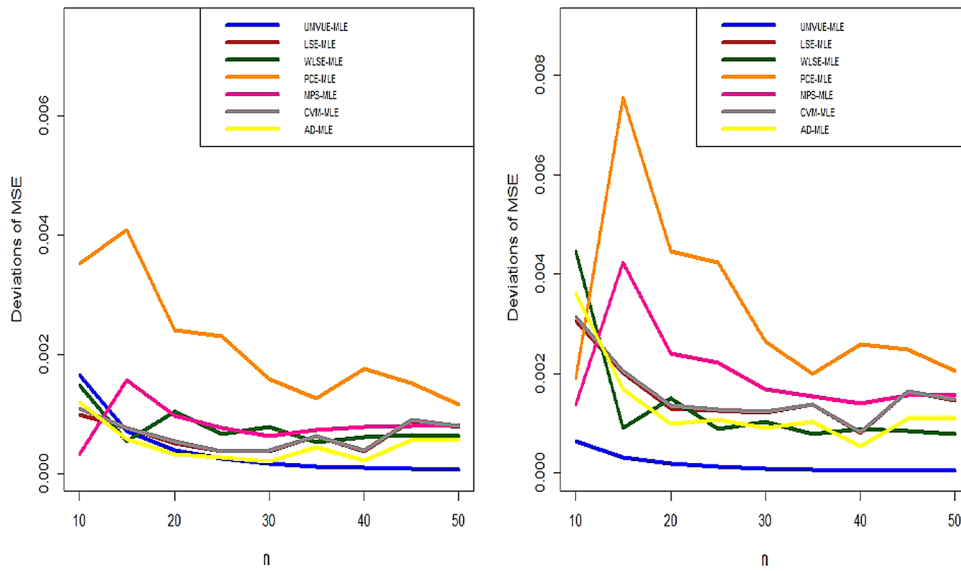


Fig. 11 $\alpha = 2, \beta = 0.5$

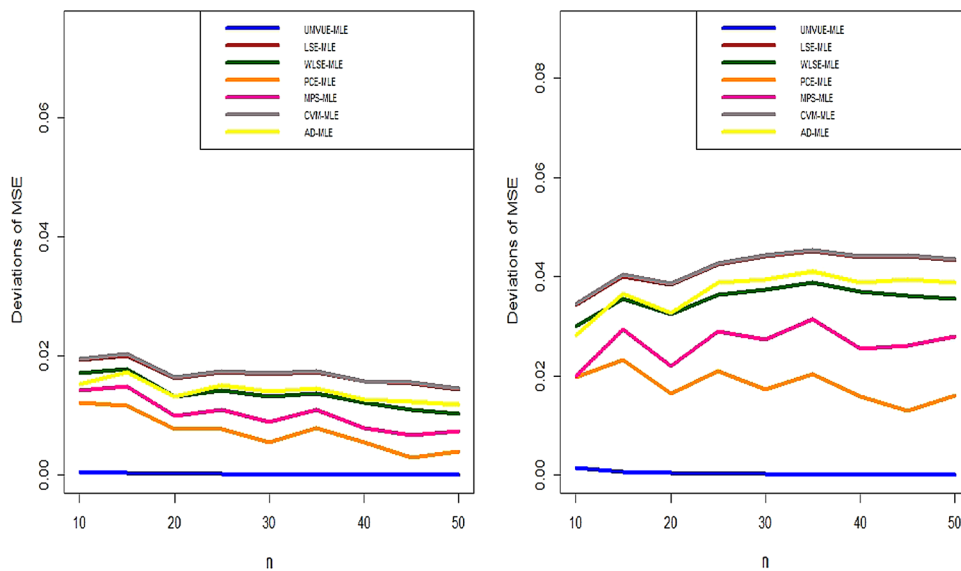


Fig. 12 $\alpha = 2, \beta = 0.75$

(observed versus expected quantiles) for all the suggested estimation methods. Displayed figures indicate that the percentile estimator (fourth in list) provides the better fit. Figure 14 displays the P–P plots (observed versus expected probabilities) for all the suggested methods of estimation. Figure 15 shows the density plots

Table 1 Goodness of fit tests for the real data set

Distribution	$\hat{\alpha}$	$\hat{\beta}$	K–S	<i>p</i> value
EME	13.54	0.5614	0.1388	0.7
Weibull	3.591	0.0168	0.2052	0.3

Table 2 Different estimates of the parameters of EME distribution and values of model selection criteria for the real data set

Estimator	$\hat{\alpha}$	$\hat{\beta}$	ML	AIC	AICc	BIC	HQC
MLE	13.54	0.5614	153.4	157.4	158.0	159.6	157.9
LSE	6.837	0.6718	123.3	127.3	127.9	129.5	127.8
WLSE	8.216	0.6400	131.1	135.1	135.7	137.3	135.6
PCE	9.167	0.8343	102.0	106.0	106.7	108.2	106.5
MPS	8.843	0.6314	133.3	137.3	137.9	139.5	137.8
CVM	8.301	0.6327	133.1	137.1	137.8	139.3	137.6
AD	8.819	0.6288	134.0	138.0	138.6	140.2	138.5

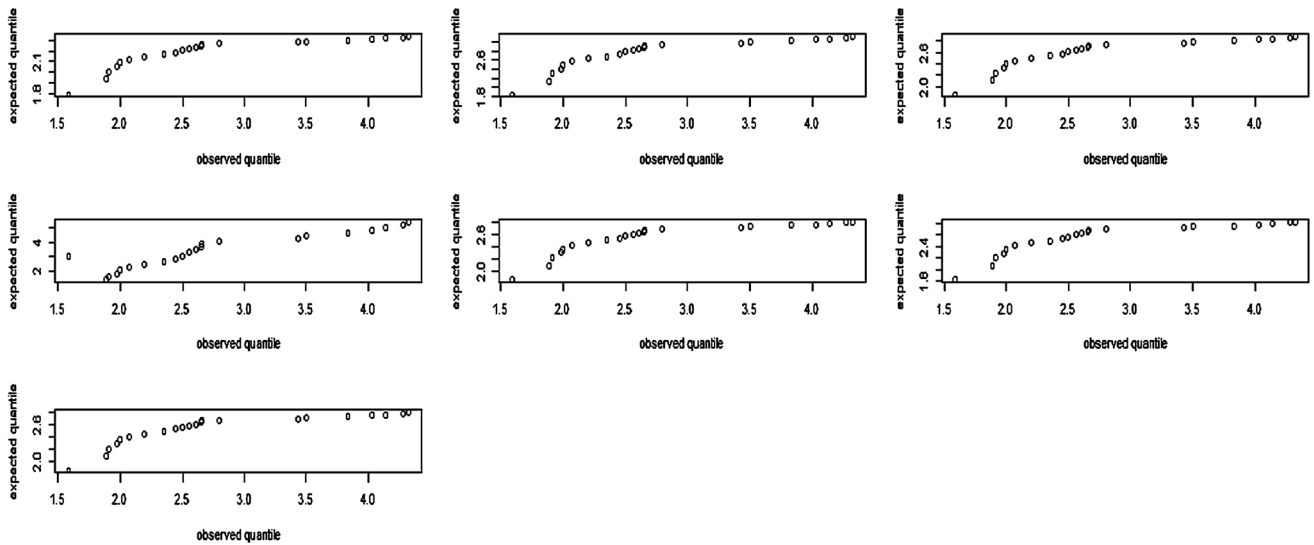


Fig. 13 Q–Q plots for the different estimation methods

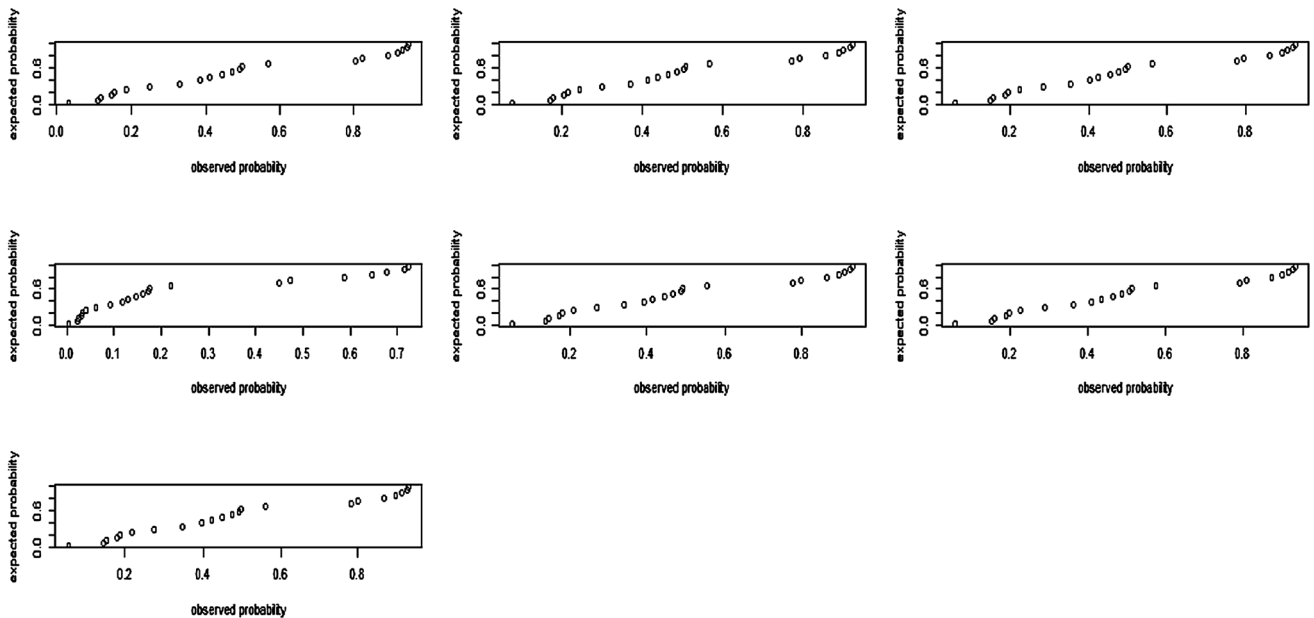


Fig. 14 P–P plots for the different estimation methods

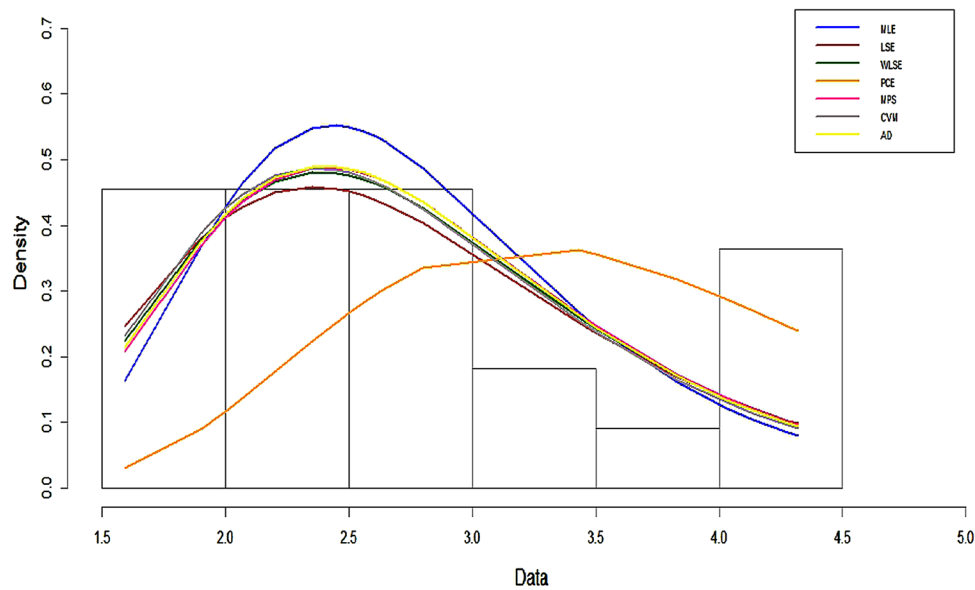


Fig. 15 Fitted PDFs and the histogram for the seven different estimation methods

(fitted density functions versus empirical histogram) for all the estimation methods. We may conclude from the figures that percentile estimator provides reasonably good fit.

11 Conclusion

In this article, we have considered eight methods of estimation of the probability density function and the cumulative distribution function for the EME distribution and comparisons are performed. Such comparisons can be useful to find the best estimators for the PDF and the CDF which can be used to estimate functionals like differential entropy, Rényi entropy, Kullback–Leibler divergence, Fisher information, cumulative residual entropy, the quantile function, Bonferroni curve, Lorenz curve, probability weighted moments, hazard rate function, mean deviation about mean etc. From both simulation study and real data analysis, we observe that both ML and UMVU estimators along with percentiles estimator perform better than their counter part. We hope our results and methods of estimation might attract wider sets of applications in the above mentioned functionals. It will be interesting to study the methods of estimation of the PDF and the CDF based on censoring data like progressive censoring and group censoring data. The work is in progress and it will be reported later.

Acknowledgements The authors would like to express thanks to the Editor, Associate Editor and anonymous referees for useful suggestions and comments which have improved the presentation of the

manuscript. The research work of Yogesh Mani Tripathi is partially supported by a Grant SR/S4/MS: 785/12 from the Department of Science and Technology, India.

References

- Anderson TW, Darling DA (1952) Asymptotic theory of certain ‘goodness-of-fit’ criteria based on stochastic processes. *Ann Math Stat* 23:193–212
- Bagheri SF, Alizadeh M, Baloui Jamkhaneh E, Nadarajah S (2013) Evaluation and comparison of estimations in the generalized exponential-Poisson distribution. *J Stat Comput Simul*. doi:10.1080/00949655.793342
- Bagheri SF, Alizadeh M, Nadarajah S (2016) Efficient estimation of the PDF and the CDF of the exponentiated Gumbel distribution. *Commun Stat Simul Comput* 45:339–361
- Brown M (1972) Low density traffic streams. *Adv Appl Probab* 4:177–192
- Cheng RCH, Amin NAK (1979) Maximum product of spacings estimation with application to the lognormal distribution. University of Wales Institute of Science and Technology, Cardiff, Math Report 79-1
- Cheng RCH, Amin NAK (1983) Estimating parameters in continuous univariate distributions with a shifted origin. *J R Stat Soc Ser B* 45:394–403
- Dara ST (2012) Recent advances in moment distributions and their hazard rate. Ph.D. thesis, National College of Business Administration and Economics, Lahore, Pakistan
- Dixit UJ, Jabbari Nooghabi M (2010) Efficient estimation in the Pareto distribution. *Stat Methodol* 7:687–691
- Dixit UJ, Jabbari Nooghabi M (2011) Efficient estimation in the Pareto distribution with the presence of outliers. *Stat Methodol* 8:340–355
- Gupta RD, Kundu D (2001) Generalized exponential distributions. An alternative to gamma or Weibull distribution. *Biom J* 43:117–130
- Gy P (1982) Sampling of particulate material: theory and practice. Elsevier, Amsterdam

- Hasnain SA (2013) Exponentiated moment exponential distribution. Ph.D. thesis, National College of Business Administration and Economics, Lahore, Pakistan
- Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distribution, vol 1, 2nd edn. Wiley, New York
- Kao JHK (1958) Computer methods for estimating Weibull parameters in reliability studies. *Trans IRE Reliab Qual Control* 13:15–22
- Krumbein WC, Pettijohn FJ (1938) Manual of sedimentary petrography. Appleton Century-Crofts Inc, New York
- Lawless JF (1982) Statistical models and methods for lifetime data. Wiley, New York
- Macdonald PDM (1971) Comment on a paper by Choi and Bulgren. *J R Stat Soc B* 33:326–329
- Mann NR, Schafer RE, Singpurwalla ND (1974) Methods for statistical analysis of reliability and life data. Wiley, New York
- Preston FW (1962) The canonical distribution of commonness and rarity. *Ecology* 43:186–215
- Swain J, Venkatraman S, Wilson J (1988) Least squares estimation of distribution function in Johnsons translation system. *J Stat Comput Simul* 29:71–297
- Taillie C, Patil GP, Hennemuth R (1979) Modelling and analysis of recruitment distributions. *Ecol Environ Stat* 2(4):315–329
- Temkin N (1976) Interactive information and distributional length biased survival models. Unpublished Ph.D. dissertation, University of New York at Buffalo
- Warren W (1975) Statistical distributions in forestry and forest products research. In: Patil GP, Kotz S, Ord JK (eds) *Statistical distributions in scientific work*, vol 2. D. Reidel, Dordrecht, pp 360–384
- Zelen M (1974) Problems in cell kinetics and the early detection of disease. In: Proschan F, Serfling RJ (eds) *Reliability and biometry*. SIAM, Philadelphia, pp 701–706
- Zelen M (1976) Theory of early detection of breast cancer in the general population. In: Heuson JC, Mattheiem WH, Rozenweig M (eds) *Breast cancer: trends in research and treatment*. Raven Press, New York, pp 287–301