



On the distribution of the sum of independent and non-identically extended exponential random variables

Masato Kitani¹ · Hidetoshi Murakami²

Received: 17 January 2019 / Accepted: 29 April 2019 / Published online: 18 May 2019
© Japanese Federation of Statistical Science Associations 2019

Abstract

Considering the sum of the independent and non-identically distributed random variables is a most important topic in many scientific fields. An extension of the exponential distribution based on mixtures of positive distributions is proposed by Gómez et al. (Rev Colomb Estad 37:25–34, 2014). Distribution of the sum of the independent and non-identically distributed random variables is obtained using inverse transformation of the moment generating function. A saddlepoint approximation is used to approximate the derived distribution. Simulations are used to investigate the accuracy of the saddlepoint approximation. Parameters are estimated by the maximum likelihood method. The method is illustrated by the analysis of real data.

Keywords Extended exponential distribution · Independent and non-identically distributed random variables · Saddlepoint approximation

1 Introduction

Determination of the distribution of the sum of independent random variables is one of the most important topics for real data analysis. For example, distribution of the sum of independent and non-identically distributed (i.n.i.d.) uniform random variables is well known. The first result was obtained by Olds (1952) with induction. In addition, Bradley and Gupta (2002) derived the explicit formulae for the distribution by inverting its characteristic function. As a different approach, Sadooghi-Alvandi

Research of the H. Murakami was supported by JSPS KAKENHI Grant Number 18K11199.

✉ Masato Kitani
1418502@ed.tus.ac.jp

¹ Department of Applied Mathematics, Graduate School of Science, Tokyo University of Science, Tokyo, Japan

² Department of Applied Mathematics, Tokyo University of Science, Tokyo, Japan

et al. (2009) gave the closed form of the distribution using the Laplace transform. For the sum of the exponential random variables, Khuong and Kong (2006) obtained the density function with distinct or equal parameters using the characteristic function. The most general case for non-identically exponential random variables is discussed in Amari and Misra (1997). Furthermore, distribution of the sum of the i.n.i.d. gamma random variables is obtained by Mathai (1982). Additionally, Moschopoulos (1985) gave the distribution of the sum of the i.n.i.d. gamma random variables, which is expressed as a single gamma series whose coefficients are computed by simple recursive relations. Alouini et al. (2001) considered applying Moschopoulos's approach for distribution of the sum of the correlated gamma random variables.

Many researchers discussed the extension of exponential distribution. For example, Gupta and Kundu (2001) introduced an extended exponential distribution such as

$$f_{\text{GK}}(x; \alpha, \lambda) = \alpha \lambda (1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}, \quad x > 0,$$

for $\alpha > 0, \lambda > 0$. In addition, Nadarajah and Haghghi (2011) introduced another extension of the exponential distribution. Its density function is given by

$$f_{\text{NH}}(x; \alpha, \lambda) = \alpha \lambda (1 + \lambda x)^{\alpha-1} \exp\{1 - (1 + \lambda x)^\alpha\}, \quad x > 0,$$

where $\alpha > 0$ and $\lambda > 0$. Recently, Lemonte et al. (2016) proposed a three-parameter extension of exponential distribution. More recently, Almarashi et al. (2019) extended the exponential distribution using the type I half-logistic family of distributions.

An extension of the exponential distribution based on mixtures of positive distributions is introduced by Gómez et al. (2014). A random variable X is distributed according to the extended exponential distribution with parameters α and β if the probability density function and cumulative distribution function are given by

$$f_X(x; \alpha, \beta) = \frac{\alpha^2(1 + \beta x)e^{-\alpha x}}{\alpha + \beta},$$

$$F_X(x; \alpha, \beta) = \frac{\alpha + \beta - (\beta + \alpha + \alpha\beta x)e^{-\alpha x}}{\alpha + \beta},$$

respectively, where $x > 0, \alpha > 0$, and $\beta \geq 0$. Then, the moment generating function is given by

$$M_X(t) = \frac{\alpha^2(\alpha + \beta - t)}{(\alpha + \beta)(t - \alpha)^2}, \quad t < \alpha. \quad (1)$$

Gómez et al. (2014) showed that the extended exponential distribution is more useful in fitting real data than are other extensions of the exponential distribution for the life of the fatigue fracture of Kevlar 49/epoxy.

In Sect. 2, we derive the distribution of the sum of n i.n.i.d. extended exponential random variables using a simple gamma series and recurrence relation in a similar way as in Moschopoulos (1985). However, it is difficult to calculate the exact

probability when the number of random variables increases. Hence, we need a more accurate approximation of the distribution in terms of the amount of calculation. Under these circumstances, the saddlepoint approximation is commonly used. Saddlepoint approximations have been used with great success by many researchers. Their applications and usefulness are discussed by Butler (2007), Huzurbazar (1999), Jensen (1995), and Kolassa (2006). Eisinga et al. (2013) discussed the use of the saddlepoint approximation for the sum of the i.n.i.d. binomial random variables. In addition, Murakami (2014) and Nadarajah et al. (2015) applied the saddlepoint approximation for the sum of the i.n.i.d. uniform and beta random variables, respectively. Additionally, Murakami (2015) gave the approximation for the sum of the i.n.i.d. gamma random variables. In Sect. 3, we discuss the use of the saddlepoint approximation of the i.n.i.d. extended exponential random variables. We compare the accuracy of a saddlepoint approximation with the exact distribution function. Moreover, we discuss the parameter estimation using the maximum likelihood method for the case of $n = 2$ and the real data analysis. Finally, we conclude with our results in Sect. 4.

2 The exact density

In this section, we derive the exact distribution of the sum of the i.n.i.d. extended exponential random variables. Let X_1, \dots, X_n be independent, extended exponential random variables with parameters $\alpha_i > 0, \beta_i \geq 0$ for $i = 1, \dots, n$. By (1), the moment generating function of $Y = X_1 + X_2 + \dots + X_n$ is

$$\begin{aligned}
 M_Y(t) &= \prod_{i=1}^n \frac{\alpha_i^2(\alpha_i + \beta_i - t)}{(\alpha_i + \beta_i)(t - \alpha_i)^2} \\
 &= \left(\prod_{i=1}^n \frac{\alpha_i^2}{\alpha_i + \beta_i} \right) \sum_{\boldsymbol{\gamma} \in \mathbb{B}^n} \prod_{i=1}^n \frac{\beta_i^{(\boldsymbol{\gamma})_i}}{(\alpha_i - t)^{(\boldsymbol{\gamma})_i+1}} \\
 &= \left(\prod_{i=1}^n \frac{\alpha_i^2}{\alpha_i + \beta_i} \right) \sum_{\boldsymbol{\gamma} \in \mathbb{B}^n} \left(\prod_{i=1}^n \frac{\beta_i^{(\boldsymbol{\gamma})_i}}{\alpha_i^{(\boldsymbol{\gamma})_i+1}} \right) \prod_{i=1}^n \frac{1}{(1 - \frac{t}{\alpha_i})^{(\boldsymbol{\gamma})_i+1}},
 \end{aligned}
 \tag{2}$$

where $\mathbb{B} = \{0, 1\}$, and $(\boldsymbol{\gamma})_i$ is the i th component of $\boldsymbol{\gamma}$.

Herein, let

$$h(t) = \prod_{i=1}^n \frac{1}{(1 - \frac{t}{\alpha_i})^{(\boldsymbol{\gamma})_i+1}}.$$

The inverse transformation of the moment generating function is applicable to $h(t)$. Then we obtain Theorem 1.

Theorem 1 *The probability density function of Y is expressed as*

$$f_Y(y) = C \sum_{\gamma \in \mathbb{B}^n} D_\gamma \sum_{k=0}^{\infty} \frac{\delta_k}{\Gamma(\rho_\gamma + k) \left(\frac{1}{\alpha_1}\right)^{\rho_\gamma + k}} y^{\rho_\gamma + k - 1} \exp(-\alpha_1 y), \quad y > 0, \tag{3}$$

where

$$C = \prod_{i=1}^n \frac{\alpha_i^2}{\alpha_i + \beta_i}, \quad D_\gamma = \prod_{i=1}^n \frac{\beta_i^{(\gamma)_i}}{\alpha_i^{(\gamma)_i + 1}}, \quad \rho_\gamma = \sum_{i=1}^n \{(\gamma)_i + 1\},$$

$$\delta_{k+1} = \frac{1}{k+1} \sum_{j=1}^{k+1} j \eta_j \delta_{k+1-j}, \quad k = 0, 1, 2, \dots, \quad \delta_0 = 1,$$

$$\eta_j = \sum_{s=1}^n \{(\gamma)_s + 1\} \left(1 - \frac{\alpha_s}{\alpha_1}\right)^j / j, \quad \alpha_1 = \max_i(\alpha_i).$$

Proof We follow a similar procedure as in Moschopoulos (1985) to prove the theorem. Without loss of generality, we assume $\alpha_1 = \max_i(\alpha_i)$. We apply the identity

$$1 - \frac{t}{\alpha_i} = \left(1 - \frac{t}{\alpha_1}\right) \frac{\alpha_1}{\alpha_i} \left[1 - \frac{1 - \frac{\alpha_i}{\alpha_1}}{1 - \frac{t}{\alpha_1}}\right]$$

to $h(t)$, and then

$$\begin{aligned} \log h(t) &= \log \prod_{i=1}^n \left(1 - \frac{t}{\alpha_i}\right)^{-(\gamma)_i - 1} \\ &= \log \prod_{i=1}^n \left(1 - \frac{t}{\alpha_1}\right)^{-(\gamma)_i - 1} \left(\frac{\alpha_1}{\alpha_i}\right)^{-(\gamma)_i - 1} \left[1 - \frac{1 - \frac{\alpha_i}{\alpha_1}}{1 - \frac{t}{\alpha_1}}\right]^{-(\gamma)_i - 1} \\ &= \log \prod_{i=1}^n \left(1 - \frac{t}{\alpha_1}\right)^{-(\gamma)_i - 1} \left(\frac{\alpha_1}{\alpha_i}\right)^{-(\gamma)_i - 1} + \sum_{i=1}^n \log \left[1 - \frac{1 - \frac{\alpha_i}{\alpha_1}}{1 - \frac{t}{\alpha_1}}\right]^{-(\gamma)_i - 1}. \end{aligned}$$

Using the Maclaurin expansion $\log(1 - x) = -\sum_{j=1}^{\infty} x^j / j$, we have

$$\begin{aligned} \log h(t) &= \log \prod_{i=1}^n \left(1 - \frac{t}{\alpha_i}\right)^{-(\gamma)_i-1} \left(\frac{\alpha_i}{\alpha_1}\right)^{-(\gamma)_i-1} \\ &\quad + \sum_{i=1}^n (-(\gamma)_i - 1) \left[- \sum_{j=1}^{\infty} \left(\frac{1 - \frac{\alpha_i}{\alpha_1}}{1 - \frac{t}{\alpha_1}}\right)^j / j \right] \\ &= \log \prod_{i=1}^n \left(1 - \frac{t}{\alpha_i}\right)^{-(\gamma)_i-1} \left(\frac{\alpha_i}{\alpha_1}\right)^{-(\gamma)_i-1} \\ &\quad + \sum_{j=1}^{\infty} \left(\sum_{i=1}^n ((\gamma)_i + 1) \left(1 - \frac{\alpha_i}{\alpha_1}\right)^j / j \right) \left(1 - \frac{t}{\alpha_1}\right)^{-j} \\ &= \log \left[\prod_{i=1}^n \left(\frac{\alpha_i}{\alpha_1}\right)^{(\gamma)_i+1} \left(1 - \frac{t}{\alpha_1}\right)^{-\rho_\gamma} \right] + \sum_{j=1}^{\infty} \eta_j \left(1 - \frac{t}{\alpha_1}\right)^{-j}, \end{aligned}$$

where

$$\rho_\gamma = \sum_{i=1}^n \{(\gamma)_i + 1\}, \quad \eta_j = \sum_{i=1}^n \{(\gamma)_i + 1\} \left(1 - \frac{\alpha_i}{\alpha_1}\right)^j / j.$$

This expression is defined by t such that $\max_i |(1 - \alpha_i/\alpha_1)/(1 - t/\alpha_1)| < 1$. Therefore,

$$h(t) = \prod_{i=1}^n \left(\frac{\alpha_i}{\alpha_1}\right)^{(\gamma)_i+1} \left(1 - \frac{t}{\alpha_1}\right)^{-\rho_\gamma} \exp\left(\sum_{j=1}^{\infty} \eta_j \left(1 - \frac{t}{\alpha_1}\right)^{-j}\right).$$

Herein, we calculate together the terms of the same order in the Taylor series, and we obtain

$$\exp\left(\sum_{j=1}^{\infty} \eta_j \left(1 - \frac{t}{\alpha_1}\right)^{-j}\right) = \sum_{k=0}^{\infty} \delta_k \left(1 - \frac{t}{\alpha_1}\right)^{-k}.$$

The coefficient δ_k is obtained by the recursive formula:

$$\delta_{k+1} = \frac{1}{k+1} \sum_{j=1}^{k+1} j \eta_j \delta_{k+1-j}, \quad k = 0, 1, 2, \dots,$$

with $\delta_0 = 1$. Thus, the moment generating function of Y is

$$M_Y(t) = \left(\prod_{i=1}^n \frac{\alpha_i^2}{\alpha_i + \beta_i} \right) \sum_{\gamma \in \mathbb{B}^n} \prod_{i=1}^n \frac{\beta_i^{(\gamma)_i}}{\alpha_1^{(\gamma)_i+1}} \sum_{k=0}^{\infty} \delta_k \left(1 - \frac{t}{\alpha_1} \right)^{-(\rho_\gamma+k)}.$$

Remark that $\left(1 - \frac{t}{\alpha_1} \right)^{-(\rho_\gamma+k)}$ is the same as the moment generating function of the gamma distribution. Then, we apply the inverse transformation of the moment generating function term-by-term. Therefore, the theorem is completely proved. \square

Theorem 2 *The exact cumulative distribution function $F_Y(y) = P(Y \leq y)$ is derived by term-by-term integration of (3), that is,*

$$F_Y(y) = C \sum_{\gamma \in \mathbb{B}^n} D_\gamma \sum_{k=0}^{\infty} \delta_k \int_0^y \frac{1}{\Gamma(\rho_\gamma + k) \left(\frac{1}{\alpha_1} \right)^{\rho_\gamma+k}} w^{\rho_\gamma+k-1} \exp(-\alpha_1 w) dw. \tag{4}$$

In addition, the truncation error is obtained by

$$E_m(y) = C \sum_{\gamma \in \mathbb{B}^n} D_\gamma \frac{\alpha_1^{\rho_\gamma}}{\Gamma(\rho_\gamma)} \int_0^y w^{\rho_\gamma-1} \exp \{ -(1-a)\alpha_1 w \} dw - F_m(y),$$

where $F_m(y)$ is the sum of the first $m + 1$ terms of (4) for $k = 0, 1, \dots, m$.

Proof The interchange of the integration and summation in $F_Y(y)$ can be justified from the uniform convergence. For $i = 1, 2, \dots$ and $a = \max_{2 \leq \ell \leq n} (1 - \alpha_\ell / \alpha_1)$, we have

$$|\eta_j| = \sum_{s=1}^n \frac{((\gamma)_s + 1)(1 - \alpha_s / \alpha_1)^j}{j} \leq \frac{\rho_\gamma \alpha^j}{j}, \quad j = 1, 2, \dots, k + 1.$$

From the definition of δ , we obtain

$$|\delta_{k+1}| \leq \frac{\rho_\gamma}{k + 1} \sum_{j=1}^{k+1} \alpha^j |\delta_{k+1-j}|, \quad k = 0, 1, 2, \dots,$$

from the recursive equation as

$$|\delta_{k+1}| \leq \frac{\rho_\gamma(\rho_\gamma + 1) \cdots (\rho_\gamma + k)}{(k + 1)!} \alpha^{k+1}.$$

Therefore,

$$\begin{aligned}
 f_Y(y) &= C \sum_{\gamma \in \mathbb{B}^n} D_\gamma \frac{\alpha_1^{\rho_\gamma}}{\Gamma(\rho_\gamma)} y^{\rho_\gamma-1} \exp(-\alpha_1 y) \sum_{k=0}^{\infty} \frac{\delta_k}{\rho_\gamma(\rho_\gamma + 1) \cdots (\rho_\gamma + k - 1)} (\alpha_1 y)^k \\
 &\leq C \sum_{\gamma \in \mathbb{B}^n} D_\gamma \frac{\alpha_1^{\rho_\gamma}}{\Gamma(\rho_\gamma)} y^{\rho_\gamma-1} \exp(-\alpha_1 y) \sum_{k=0}^{\infty} \frac{(\alpha_1 a y)^k}{k!} \\
 &= C \sum_{\gamma \in \mathbb{B}^n} D_\gamma \frac{\alpha_1^{\rho_\gamma}}{\Gamma(\rho_\gamma)} y^{\rho_\gamma-1} \exp\{-(1-a)\alpha_1 y\}.
 \end{aligned}
 \tag{5}$$

Here, (5) shows the uniform convergence of (3), and then we have (4). □

3 Numerical results

In this section, we discuss the evaluation of the tail probability using the saddlepoint approximation. Furthermore, we estimate parameters by the maximum likelihood method and apply it for the real data analysis.

3.1 Saddlepoint approximation

We consider the use of the saddlepoint approximation for the distribution of the sum of the i.n.i.d. extended exponential random variables. The saddlepoint approximation was first proposed by Daniels (1954), and a higher order approximation was also given by Daniels (1987).

Herein, we consider an approximation of the distribution of Y . The cumulant generating function of Y is given by

$$k(t) = \sum_{i=1}^n \log \left(\frac{\alpha_i^2(\alpha_i + \beta_i - t)}{(\alpha_i + \beta_i)(t - \alpha_i)^2} \right).$$

Lugannani and Rice (1980) gave the formula to approximate the distribution function as follows:

$$F_S(x) = \Phi(\hat{w}) + \phi(\hat{w}) \left(\frac{1}{\hat{w}} - \frac{1}{\hat{u}} \right) + O(n^{-\frac{3}{2}}),
 \tag{6}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal probability density function and its corresponding cumulative distribution function, respectively. In addition, we denote

$$\hat{w} = \text{sgn}(\hat{s}) \sqrt{2\{\hat{s}x - k(\hat{s})\}}, \quad \hat{u} = \hat{s} \sqrt{k''(\hat{s})},$$

where \hat{s} is the root of $k'(s) = x$, which is solved numerically by the Newton–Raphson algorithm; $\text{sgn}(\hat{s}) = \pm 1, 0$ if \hat{s} is positive, negative, or zero; and

$$k'(t) = \sum_{i=1}^n \frac{t - \alpha_i - 2\beta_1}{(t - \alpha_i)(\alpha_i + \beta_i - t)},$$

$$k''(t) = \sum_{i=1}^n \left(\frac{2}{(t - \alpha_i)^2} - \frac{1}{(\alpha_i + \beta_i - t)^2} \right).$$

Herein, we compare the accuracy of approximation with the following distributions by calculating the probability. \hat{y} is a percentile derived from 100, 000, 000 random numbers generated by Y , and p is its exact probability. Note that we generate X as a mixture distribution of two random variables (see Gómez et al. 2014). More precisely, they are exponential and gamma distribution. Then, a random number Y is obtained by the sum of n random numbers X .

- F_m : the approximate cumulative distribution function, which is truncated in the infinite series in (4) after $m + 1$ terms.
- F_N : the normal approximation.
- F_S : the saddlepoint cumulative distribution function from (6).

In the tables, Conv. represents the exact probability calculated by convolution, r.e. is the relative error between the approximation and p , and MCT is the mean calculating time. In our study, we use Mathematica version 11 (CPU 2.80 GHz and 32.0 GB RAM). The parameters $\alpha = (\alpha_1, \dots, \alpha_n)$, and $\beta = (\beta_1, \dots, \beta_n)$ are generated from the uniform distribution $U(0, 3)$, as in Murakami (2014) and Nadarajah et al. (2015):

Case 1: $n = 2$

Case 1-A: $\alpha = (2.30699, 1.43842)$, $\beta = (2.13769, 2.69432)$.

Case 1-B: $\alpha = (2.76659, 0.30096)$, $\beta = (0.65938, 1.07876)$.

Case 2: $n = 5$

Case 2-A: $\alpha = (1.79204, 1.52231, 0.827571, 1.69002, 1.18927)$,
 $\beta = (1.23271, 0.343942, 0.840907, 1.7232, 0.585717)$.

Case 2-B: $\alpha = (0.62841, 2.73007, 0.795827, 1.55644, 1.22646)$,
 $\beta = (0.942398, 0.866528, 1.43792, 0.631225, 2.01407)$.

Case 3: $n = 10$

Case 3-A: $\alpha = (1.82047, 1.53669, 0.887251, 1.70666,$
 $0.866679, 1.85662, 1.3083, 0.16474, 1.40535, 1.19952)$,
 $\beta = (1.17206, 0.331381, 0.811715, 1.10704, 0.575096,$
 $1.63428, 0.412453, 0.151229, 1.83359, 1.37744)$.

Case 3-B: $\alpha = (1.36622, 2.63063, 2.94352, 0.105937, 1.76873,$
 $2.20458, 2.22706, 2.34693, 1.91121, 1.76058)$,

$$\beta = (2.62005, 0.21023, 1.00028, 0.472836, 2.63518, 1.26672, 1.25457, 0.743258, 0.103399, 1.609467).$$

To determine the appropriate value of m is difficult. Therefore, we compared percentiles of F_m for various m with the exact probability calculated by convolution. Then, we determine $m = 5000$ for simulating F_m . There is no difference between the Conv. and p , as shown in Table 1. Hence, we use the simulated p as the exact probability for Case 2 ($n = 5$) and Case 3 ($n = 10$).

Tables 1, 2 and 3 show that F_m is closer to p than any other approximation. However, it is difficult to apply the real data analysis because it takes many times to calculate the probability. On the other hand, F_N and F_S overcome the calculating time; in particular, F_S gives better accuracy than F_N .

3.2 Parameter estimation

In this section, we discuss the parameter estimation for the sum of extended exponential distribution for the case of $n = 2$.

We estimate the parameter from the random number generated from the distribution of the sum of the extended exponential random variables.

Table 1 Numerical results for $n = 2$

\hat{y}	Conv.	p	F_m	F_N	F_S	r.e. F_N	r.e. F_S
Case 1-A							
1.8912	0.6000	0.6000	0.6000	0.5226	0.6005	0.1290	0.0009
2.2256	0.7000	0.7000	0.7000	0.6373	0.7006	0.0895	0.0008
2.6609	0.8000	0.8000	0.8000	0.7688	0.8005	0.0390	0.0006
3.3471	0.9000	0.9000	0.9000	0.9098	0.9003	0.0109	0.0003
3.9896	0.9500	0.9500	0.9500	0.9717	0.9502	0.0228	0.0002
4.6048	0.9750	0.9750	0.9750	0.9928	0.9751	0.0183	0.0001
5.3904	0.9900	0.9900	0.9900	0.9992	0.9900	0.0092	0.0000
MCT (s)		84	1039	0.016	0.031		
Case 1-B							
6.4006	0.6000	0.6000	0.6001	0.5042	0.5999	0.1597	0.0002
7.7789	0.7000	0.7000	0.7000	0.6203	0.6998	0.1139	0.0003
9.6176	0.8000	0.8000	0.8000	0.7582	0.7998	0.0522	0.0002
12.586	0.9000	0.9000	0.9000	0.9095	0.8998	0.0105	0.0002
15.421	0.9500	0.9500	0.9500	0.9742	0.9499	0.0254	0.0001
18.170	0.9750	0.9750	0.9750	0.9944	0.9749	0.0199	0.0001
21.713	0.9900	0.9900	0.9900	0.9995	0.9900	0.0096	0.0000
MCT (s)		70	731	0.000	0.016		

Table 2 Numerical results for $n = 5$

\hat{y}	p	F_m	F_N	F_S	r.e. F_N	r.e. F_S
Case 2-A						
5.6105	0.6000	0.6000	0.5425	0.6008	0.0958	0.0014
6.2899	0.7000	0.7000	0.6544	0.7008	0.0651	0.0012
7.1533	0.8000	0.8000	0.7784	0.8007	0.0271	0.0009
8.4794	0.9000	0.9000	0.9089	0.9005	0.0099	0.0006
9.6923	0.9500	0.9500	0.9680	0.9503	0.0190	0.0003
10.8360	0.9750	0.9750	0.9904	0.9752	0.0158	0.0002
12.2754	0.9900	0.9900	0.9985	0.9900	0.0085	0.0000
MCT (s)	132	19385	0.016	0.031		
Case 2-B						
7.5544	0.6000	0.6001	0.5434	0.6006	0.0944	0.0009
8.4581	0.7000	0.7000	0.6553	0.7005	0.0639	0.0007
9.6059	0.8000	0.8000	0.7790	0.8004	0.0262	0.0006
11.3643	0.9000	0.9000	0.9090	0.9003	0.0100	0.0003
12.9702	0.9500	0.9500	0.9680	0.9502	0.0189	0.0002
14.4816	0.9750	0.9750	0.9903	0.9751	0.0157	0.0001
16.3811	0.9900	0.9900	0.9984	0.9901	0.0085	0.0001
MCT (s)	103	13621	0.000	0.016		

Table 3 Numerical results for $n = 10$

\hat{y}	p	F_m	F_N	F_S	r.e. F_N	r.e. F_S
Case 3-A						
17.5029	0.6000	More than 100 h	0.5177	0.6011	0.1372	0.0019
19.5850	0.7000	More than 100 h	0.6294	0.7002	0.1008	0.0004
22.3481	0.8000	More than 100 h	0.7611	0.7997	0.0486	0.0003
26.8066	0.9000	More than 100 h	0.9069	0.8997	0.0077	0.0004
31.0632	0.9500	More than 100 h	0.9717	0.9498	0.0229	0.0002
35.1904	0.9750	More than 100 h	0.9933	0.9749	0.0188	0.0001
40.5149	0.9900	More than 100 h	0.9993	0.9899	0.0094	0.0001
MCT (s)	229	–	0.016	0.031		
Case 3-B						
23.4805	0.6000	More than 100 h	0.5069	0.5999	0.1552	0.0002
27.4155	0.7000	More than 100 h	0.6223	0.6998	0.1111	0.0003
32.6564	0.8000	More than 100 h	0.7590	0.7998	0.0512	0.0003
41.1107	0.9000	More than 100 h	0.9091	0.8998	0.0101	0.0002
49.1727	0.9500	More than 100 h	0.9737	0.9499	0.0249	0.0001
56.9841	0.9750	More than 100 h	0.9942	0.9749	0.0197	0.0001
67.0520	0.9900	More than 100 h	0.9995	0.9900	0.0096	0.0000
MCT (s)	165	–	0.000	0.031		

Table 4 The mean (the variance) of numerical simulation based on 1000 iterations and the first four moments

	True	r			True	r			
		100	500	2000		100	500	2000	
$\hat{\alpha}_1$	2.2	1.99 (2.21)	2.07 (0.76)	2.12 (0.64)	$E[Y]$	1.67	1.74	1.70	1.68
$\hat{\alpha}_2$	1.4	1.45 (0.07)	1.39 (0.04)	1.38 (0.03)	$E[Y^2]$	3.99	4.21	4.06	4.01
$\hat{\beta}_1$	2.6	4.28 (18.4)	4.28 (13.9)	3.60 (5.34)	$E[Y^3]$	12.27	13.04	12.48	12.34
$\hat{\beta}_2$	0.8	0.60 (3.07)	0.43 (1.62)	0.49 (0.89)	$E[Y^4]$	46.26	49.08	46.98	46.57

Table 4 reveals that the estimated parameters are close to the true parameters as sample size r increases. $\hat{\beta}_1$ and $\hat{\beta}_2$ are different from true parameters β_1 and β_2 ; however, there is almost no difference in the moments. The simulation results show that parameter estimation works well. Nevertheless, since the variance is large when the sample size is small, the initial value problem in the parameter estimation is considered. Furthermore, it is necessary to discuss the identifiability of the parameters in the future.

3.3 Real data analysis

We compare the Akaike information criterion (AIC) with the extended exponential distribution of Gómez et al. (2014) (i.e., the case of $n = 1$, EE1) and f_γ with $n = 2, 3$ (EE2, EE3), Gamma distribution $G(\alpha_1, \alpha_2)$, and Weibull distribution $W(\alpha_1, \alpha_2)$ based on the maximum likelihood approach.

We consider two data sets of the life of the fatigue fracture of Kevlar 49/epoxy, which is a widely used synthetic fiber, given in Glaser (1983) as follows:

Dataset 1:

0.0251, 0.0886, 0.0891, 0.2501, 0.3113, 0.3451, 0.4763, 0.5650, 0.5671, 0.6566, 0.6748, 0.6751, 0.6753, 0.7696, 0.8375, 0.8391, 0.8425, 0.8645, 0.8851, 0.9113, 0.9120, 0.9836, 1.0483, 1.0596, 1.0773, 1.1733, 1.2570, 1.2766, 1.2985, 1.3211, 1.3503, 1.3551, 1.4595, 1.4880, 1.5728, 1.5733, 1.7083, 1.7263, 1.7460, 1.7630, 1.7746, 1.8275, 1.8375, 1.8503, 1.8808, 1.8878, 1.8881, 1.9316, 1.9558, 2.0048, 2.0408, 2.0903, 2.1093, 2.1330, 2.2100, 2.2460, 2.2878, 2.3203, 2.3470, 2.3513, 2.4951, 2.5260, 2.9911, 3.0256, 3.2678, 3.4045, 3.4846, 3.7433, 3.7455, 3.9143, 4.8073, 5.4005, 5.4435, 5.5295, 6.5541, and 9.0960.

Dataset 2:

0.7367, 1.1627, 1.8945, 1.9340, 2.3180, 2.6483, 2.8573, 2.9918, 3.0797, 3.1152, 3.1335, 3.2647, 3.4873, 3.5390, 3.6335, 3.6541, 3.7645, 3.8196, 3.8520, 3.9653, 4.2488, 4.3017, 4.3942, 4.6416, 4.7070, 4.8885, 5.1746, 5.4962, 5.5310, 5.5588, 5.6333, 5.7006, 5.8730, 5.8737, 5.9378, 6.1960, 6.2217, 6.2630, 6.3163, 6.4513,

Table 5 Parameter estimates for various models

Distribution	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\alpha}_3$	$\hat{\beta}_3$	AIC
Dataset 1							
EE1	0.954	6.365	–	–	–	–	247.3
EE2	0.709	0.000	2.296	0.817	–	–	253.2
EE3	156.3	1.320	99.54	118.6	1.007	15.91	256.8
Gamma	1.641	0.838	–	–	–	–	248.5
Weibull	1.326	2.133	–	–	–	–	249.0
Dataset 2							
EE1	0.281	835.7	–	–	–	–	405.1
EE2	0.474	10.77	0.476	0.348	–	–	402.5
EE3	1.898	0.498	0.462	1.064	0.465	0.201	406.5
Gamma	3.071	0.432	–	–	–	–	398.5
Weibull	1.877	8.039	–	–	–	–	400.0
Dataset 3							
EE1	0.050	7.038	–	–	–	–	597.6
EE2	0.335	2.5×10^6	0.032	0.003	–	–	594.8
EE3	0.715	4.125	0.030	0.001	0.519	4.823	596.9
Gamma	1.621	0.041	–	–	–	–	595.8
Weibull	1.274	43.21	–	–	–	–	597.8

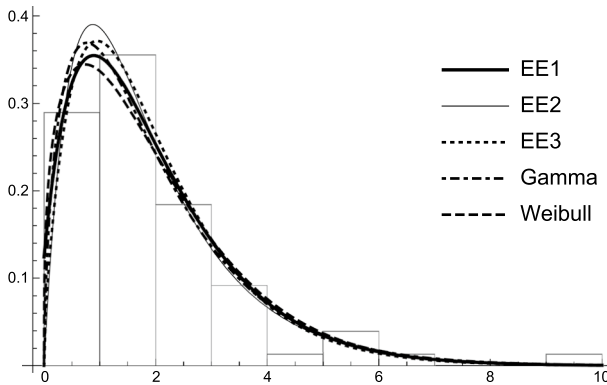


Fig. 1 Fitted pdfs of the distribution for Dataset 1

6.8320, 6.9447, 7.2595, 7.3183, 7.3313, 7.7587, 8.0393, 8.0693, 8.1928, 8.4166, 8.7558, 8.8398, 9.2497, 9.2563, 9.5418, 9.6472, 9.6902, 9.9316, 10.018, 10.4028, 10.4188, 10.7250, 10.9411, 11.7962, 12.075, 12.6933, 13.5307, 13.8105, 14.5067, 15.3013, 16.2742, 18.2682, and 19.2033.

In addition, we consider another type of data set consisting of the waiting times between 65 consecutive eruptions of the Kiama Blowhole. These data can be obtained at <http://www.statsci.org/data/oz/kiama.html>.

Dataset 3:

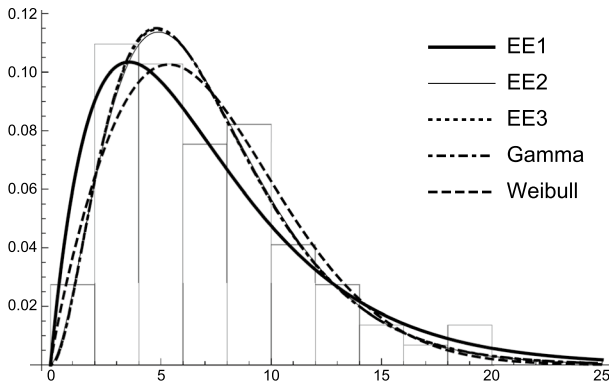


Fig. 2 Fitted pdfs of the distribution for Dataset 2

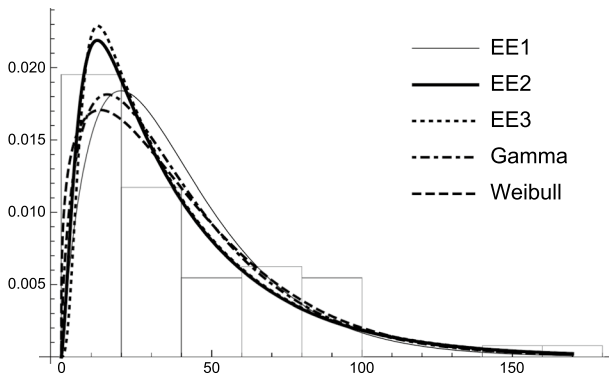


Fig. 3 Fitted pdfs of the distribution for Dataset 3

83, 51, 87, 60, 28, 95, 8, 27, 15, 10, 18, 16, 29, 54, 91, 8, 17, 55, 10, 35, 47, 77, 36, 17, 21, 36, 18, 40, 10, 7, 34, 27, 28, 56, 8, 25, 68, 146, 89, 18, 73, 69, 9, 37, 10, 82, 29, 8, 60, 61, 61, 18, 169, 25, 8, 26, 11, 83, 11, 42, 17, 14, 9 and 12.

Results of the parameter estimation of various models are shown in Table 5.

Moreover, Figs. 1, 2 and 3 reveal data fitting for various models.

In Dataset 1, as same as Gómez et al. (2014), AIC indicated that EE1 is the most suitable models. Another example, EE2 is more suitable than EE1 while the gamma model is a better fit in Dataset 2. However, EE2 is the best model to Dataset 3. As a whole, the sum of the extended exponential models has versatility for these data sets.

4 Conclusion and discussion

In this study, we obtained the exact distribution of the sum of the i.n.i.d. extended exponential n random variables using the simple gamma series and recursive formula. Numerical simulation showed that the saddlepoint approximation is the most appropriate for the cumulative distribution function, as well as in terms of the calculation time. Distribution of the sum of the extended exponential random variables was a suitable model for real data based on the AIC.

We need to consider the initial value problem of the parameter estimation. In addition, the upper bound in (5) is just one example. Then, it is necessary to determine the minimum upper bound of $f_Y(y)$. The future challenge is to obtain the distribution of the sum of the non-independent and non-identically extended exponential random variables.

References

- Almarashi, A. M., Elgarhy, M., Elsehetry, M. M., Golam Kibria, B. M., & Algarni, A. (2019). A new extension of exponential distribution with statistical properties and applications. *Journal of Nonlinear Sciences and Applications*, *12*, 135–145.
- Alouini, M.-S., Abdi, A., & Kaveh, M. (2001). Sum of gamma variates and performance of wireless communication systems over Nakagami-fading channels. *IEEE Transactions on Vehicular Technology*, *50*, 1471–1480.
- Amari, S. V., & Misra, R. B. (1997). Closed-form expressions for distribution of sum of exponential random variables. *IEEE Transactions on Reliability*, *46*, 519–522.
- Bradley, D. M., & Gupta, C. R. (2002). On the distribution of the sum of n non-identically distributed uniform random variables. *Annals of the Institute of Statistical Mathematics*, *54*, 689–700.
- Butler, R. W. (2007). *Saddlepoint Approximations with Applications*. Cambridge: Cambridge University Press.
- Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, *25*, 631–650.
- Daniels, H. E. (1987). Tail probability approximations. *International Statistical Review*, *55*, 37–48.
- Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). Saddlepoint approximation for the sum of independent non-identically distributed binomial random variables. *Statistica Neerlandica*, *67*, 190–201.
- Glaser, R. E. (1983). *Statistical analysis of Kevlar 49/epoxy composite stress-rupture data*. Livermore: Lawrence Livermore National Laboratory. (Report UCID-19849).
- Gómez, Y. M., Bolfarine, H., & Gómez, H. W. (2014). A new extension of the exponential distribution. *Revista Colombiana de Estadística*, *37*, 25–34.
- Gupta, R. D., & Kundu, D. (2001). Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biometrical Journal*, *43*, 117–130.
- Huzurbazar, S. (1999). Practical saddlepoint approximations. *American Statistician*, *53*, 225–232.
- Jensen, J. L. (1995). *Saddlepoint approximations*. New York: Oxford University Press.
- Khuong, H. V., & Kong, H.-Y. (2006). General expression for pdf of a sum of independent exponential random variables. *IEEE Communications Letters*, *10*, 159–161.
- Kolassa, J. E. (2006). *Series approximation methods in statistics*. New York: Springer.
- Lemonte, A. J., Cordeiro, G. M., & Moreno Arenas, G. (2016). A new useful three-parameter extension of the exponential distribution. *Statistics*, *50*, 312–337.
- Lugannani, R., & Rice, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, *12*, 475–490.
- Mathai, A. M. (1982). Storage capacity of a dam with gamma type inputs. *Annals of the Institute of Statistical Mathematics*, *34*, 591–597.
- Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics*, *37*, 541–544.

- Murakami, H. (2014). A saddlepoint approximation to the distribution of the sum of independent non-identically uniform random variables. *Statistica Neerlandica*, *68*, 267–275.
- Murakami, H. (2015). Approximations to the distribution of sum of independent non-identically gamma random variables. *Mathematical Sciences*, *9*, 205–213.
- Nadarajah, S., & Haghighi, F. (2011). An extension of the exponential distribution. *Statistics*, *45*, 543–558.
- Nadarajah, S., Jiang, X., & Chu, J. (2015). A saddlepoint approximation to the distribution of the sum of independent non-identically beta random variables. *Statistica Neerlandica*, *69*, 102–114.
- Olds, E. G. (1952). A note on the convolution of uniform distributions. *The Annals of Mathematical Statistics*, *23*, 282–285.
- Sadooghi-Alvandi, S. M., Nematollahi, A. R., & Habibi, R. (2009). On the distribution of the sum of independent uniform random variables. *Statistical Papers*, *50*, 171–175.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.