Simulation assessment of Expectation-Maximization algorithm in pseudo-convex mixtures generated by the exponential distribution

RUI SANTOS[1,3], MIGUEL FELGUEIRAS [1,3,4], JOÃO MARTINS [2,3,5]
[1] ESTG, Polytechnic Institute of Leiria, PORTUGAL
[2] ESS, Polytechnic Institute of Porto, PORTUGAL
[3] CEAUL, Faculdade de Ciências, Universidade de Lisboa, PORTUGAL
[4] CIDMA, University of Aveiro, PORTUGAL
[5] CEISUC/CIBB, Coimbra, PORTUGAL

Abstract: The use of pseudo-convex mixtures generated from stable distributions for extremes offers a valuable approach for handling reliability-related data challenges. This framework encompasses pseudo-convex mixtures stemming from exponential distribution. However, precise parameter estimation, particularly in cases where the weight parameter ω is negative, remains a challenge. This work assesses the performance of the Expectation-Maximization algorithm in estimating parameters for pseudo-convex mixtures generated by the exponential distribution through simulation.

Key-Words: Expectation-Maximization algorithm, exponential distribution, generalized mixtures, parameter estimation, simulation.

Received: November 27, 2023. Revised: March 19, 2024. Accepted: April 13, 2024. Published: May 10, 2024.

1 Introduction

The exponential (Exp) distribution plays a pivotal role in reliability analysis owing to its constant hazard rate, signifying a consistent probability of an event occurring within a specific time interval, regardless of elapsed time. This characteristic harmonizes seamlessly with scenarios where failure rates are time-independent, rendering it a fundamental model across diverse fields such as engineering, medicine, finance, among others (see, e.g., [1], [2], [3], [4]). The hazard function within the exponential distribution framework plays a crucial role in predicting and addressing risks tied to system reliability. It enables proactive planning to boost performance and mitigate potential failures.

On another note, generalized mixtures distributions emerge as valuable tools in statistics for achieving more flexible distributions to better model random phenomena. These mixtures, characterized by a distribution function that is a weighted average of other distribution functions, allow for the incorporation of negative weights, expanding the scope of modeling possibilities. Preliminary work on this subject has explored non-convex mixtures of exponentials (e.g., [5], [6], [7]) and Gaussian mixtures, [8], with recent applications in various domains such as cluster analysis, bioinformatics, biology, epidemiology, social sciences, and finance (e.g., [9], [10], [11], [12]).

Further advancements include pseudo-convex mixtures generated by the exponential distribution (see, [13], [14]), which offer increased flexibility in hazard functions while converging to the exponential distribution's hazard function. However, estimation techniques such as the method of moments or maximum likelihood may exhibit limitations, prompting an evaluation of estimation performance using the Expectation-Maximization (EM) algorithm. This work aims to delve into such assessments, presenting parameter estimators and conducting a simulation study to compare their performance.

Hence, Section 2 provides some preliminary concepts and notations concerned with stable distributions for extremes, generalized mixtures and pseudo-convex mixtures (PCM) generated by shape-extended stable distributions for extremes. Afterwards, Section 3 delineates the pseudo-convex mixtures generated by the exponential distribution and furnishes estimators for the parameters derived through the method of moments (MM), maximum likelihood (ML), and Expectation-Maximization (EM) algorithm. In Section 4, a simulation study is conducted to assess and compare the performance of the provided estimators. Lastly, Section 5 encapsulates the key findings and provides final remarks.

2 Pseudo-convex Mixtures Generated by Shape-extended Stable Distributions for Extremes

To establish pseudo-convex mixtures generated by shape-extended stable distributions for extremes, this section first outlines the definitions of min-stable and max-stable distributions. Subsequently, it introduces the concept of shape-extended stable distributions to broaden the spectrum of available distributions, [14].

2.1 Distributions Stable for Extremes

Consider a sequence of independent and identically distributed (i.i.d.) absolutely continuous random variables (r.v.) denoted as $X_1, \ldots, X_n$, with distribution function (d.f.) F and survival function (s.f.) $\overline{F}$, i.e., $\overline{F}(x) := 1 - F(x)$. Furthermore, let $X_{i:n}$ represent the i-th ascending order statistic associated with these random variables. Consequently, $X_{1:n}$ denotes the minimum of $X_1, ..., X_n$, while $X_{n:n}$ denotes the maximum of $X_1, ..., X_n$.

A r.v. X with d.f. F is stable for minima or min-stable (minS) if there exist normalizing sequences $\{\alpha_n \in \mathbb{R}^+\}$ and $\{\beta_n \in \mathbb{R}\}$ such that the equality in distribution $X_{1:n} \overset{d}{=} \alpha_n X + \beta_n$ holds $\forall n \in \mathbb{N}$, with $X \sim F$. This is equivalent to stating that the s.f. $\overline{F}$ satisfies

$$\overline{F}_{X_{1:n}}(x) = \overline{F}^n(x) = \overline{F}\left(\frac{x - \beta_n}{\alpha_n}\right),$$

for all $x \in \mathbb{R}$ where $\overline{F}_{X_{1:n}}$ denotes the s.f. of $X_{1:n}$. Therefore, if F is minS, the minima of n independent copies of $X \sim F$ also follow the F distribution (potentially with a scale and location adjustment). The Extreme Value Distribution for minima (EVm$_\gamma$), with s.f. given by

$$\overline{F}_{\text{GEVm}_\gamma}(x) = \begin{cases} \exp\left\{-\left[1 - \gamma x\right]^{-1/\gamma}\right\},\ 1 + \gamma x > 0 & \gamma \neq 0 \\ \exp\left\{-\exp(x)\right\},\ x \in \mathbb{R} & \gamma = 0, \end{cases}$$

represents the sole potential min-stable distribution. This distribution applies $\alpha_n = n^\gamma$ and $\beta_n = \gamma^{-1}(1 - n^\gamma)$ if $\gamma \neq 0$ or $\alpha_n = 1$ and $\beta_n = -\ln(n)$ if $\gamma = 0$.

The EVm$_\gamma$ encompasses the Gumbel ($\gamma = 0$), Fréchet ($\gamma > 0$), and Weibull ($\gamma < 0$) minimum distributions. The parameter γ serves as the extreme value index, gauging the heaviness of the left tail function F. Introducing location (μ) and scale (σ) parameters allows for the generalization of EVm$_\gamma$ through $F_{\text{EVm}_\gamma}(x; \mu, \sigma) = F_{\text{EVm}_\gamma}((x - \mu)/\sigma)$. Moreover, this distribution holds paramount importance in Extreme Value Theory (EVT), as per the Extreme Value Theorem (Fisher-Tippett-Gnedenko): if the minima of n random variables converge to a non-degenerate distribution as n increases to infinity, it must converge to the EVm$_\gamma$ distribution.

All results pertaining to the minima of a sequence of i.i.d. continuous r.v. can be similarly applied to the maxima due to the relationship $Y_{1:n} = -X_{n:n}$, and also $Y_{n:n} = -X_{1:n}$, if $Y = -X$. Therefore, a r.v. X with a d.f. F is stable for maxima, or max-stable (maxS), if there exist normalizing sequences $\{\alpha_n \in \mathbb{R}^+\}$ and $\{\beta_n \in \mathbb{R}\}$ such that the equality in distribution $X_{n:n} \overset{d}{=} \alpha_n X + \beta_n$ holds for all $n \in \mathbb{N}$, meaning the d.f. F satisfies

$$F_{X_{n:n}}(x) = F^n(x) = F\left(\frac{x - \beta_n}{\alpha_n}\right),$$

for all $x \in \mathbb{R}$ where $F_{X_{n:n}}$ denotes the d.f. of $X_{n:n}$. The only possible max-stable distribution is the Extreme Value Distribution for maxima (EVM$_\gamma$), with its d.f. given by $F_{\text{GEVM}_\gamma}(x) = \overline{F}_{\text{GEVm}_\gamma}(-x)$. EVM$_\gamma$ includes the Gumbel ($\gamma = 0$), Fréchet ($\gamma > 0$), and Weibull ($\gamma < 0$) maximum distributions, and can also incorporate location and scale parameters through $F_{\text{GEVM}_\gamma}(x; \mu, \sigma) = F_{\text{GEVM}_\gamma}((x - \mu)/\sigma)$.

Indeed, in many statistical applications, the focus lies not on studying typical occurrences (events with higher probability) but on modelling extreme events, which tend to have lower probabilities. Therefore, the primary objective of Extreme Value Theory is to characterize the minimum and/or maximum of a set of random variables. Fundamental concepts in this domain include order statistics, distributions stable for extremes, and the Extreme Value Theorem. Key results and advancements in this theory are documented in various sources (see, e.g., [15], [16], [17], [18], [19], [20]). Presently, this theory finds numerous applications in fields like biostatistics, climatology, finance, hydrology, industry and insurance (see, e.g., [20], [21], [22], [23], [24]), and continues to be an active area of research, as evidenced by works like, [25], [26], [27], and their associated references.

2.2 Shape-extended Stable Distributions

The class of stable distributions can be expanded to accommodate variations in the shape parameter, cf., [13], [14]. Consequently, F qualifies as a shape-extended min-stable (SEminS) distribution if there exist normalizing sequences $\{\alpha_n \in \mathbb{R}^+\}$, $\{\beta_n \in \mathbb{R}\}$, and $\{\gamma_n \in \mathbb{R}\}$ such that the equality in distribution $X_{1:n} \overset{d}{=} \alpha_n X + \beta_n$ holds for all $n \in \mathbb{N}$, where $X \sim F_{\gamma_n}$, and F_{γ_n} signifies the same distribution as F but with a modified shape parameter value (γ_n denotes the new shape parameter value). Therefore, this equivalence in distribution can be expressed as:

$$F_{X_{1:n}}(x) = 1 - \overline{F}^n(x) = F_{\gamma_n}\left(\frac{x - \beta_n}{\alpha_n}\right),$$

for all $x \in \mathbb{R}$. Apart from the EVm$_\gamma$ distribution, further examples of SEminS distributions encompass the generalized logistic type II (GL2) distribution and the Generalized Pareto (GP) distribution. For example, considering the sequence $X_1, \ldots, X_n$ of i.i.d. random varibles with Generalized Pareto distribution, GP(μ, σ, γ), where $\mu \in \mathbb{R}$, $\sigma, \gamma \in \mathbb{R}^+$, and

$$\overline{F}(x) = \left[1 + \frac{x-\mu}{\gamma\sigma}\right]^{-\gamma}, \; x > \mu,$$

then the s.f. of the minimum $X_{1:n}$ is given by

$$\overline{F}_{X_{1:n}}(x) = \left[1 + \frac{nx + (1-n)\mu - \mu}{n\gamma\sigma}\right]^{-n\gamma}.$$

Thus, GP is a SEminS distribution with $\alpha_n = n^{-1}$, $\beta_n = n^{-1}(n-1)\mu$ and $\gamma_n = n\gamma$, or analogously $nX_{1:n} + (1-n)\mu \sim$ GP $(\mu, \sigma, n\gamma)$. The GP distribution indeed holds significance in EVT, particularly in modelling excesses, [28].

Similarly, F is regarded as a shape-extended max-stable (SEmaxS) distribution if there exist normalizing sequences $\{\alpha_n \in \mathbb{R}^+\}$, $\{\beta_n \in \mathbb{R}\}$, and $\{\gamma_n \in \mathbb{R}\}$ such that the equality in distribution $X_{n:n} \stackrel{d}{=} \alpha_n X + \beta_n$, with $X \sim F_{\gamma_n}$, holds for all $n \in \mathbb{N}$, i.e.,

$$F_{X_{n:n}}(x) = F^n(x) = F_{\gamma_n}\left(\frac{x-\beta_n}{\alpha_n}\right),$$

for all $x \in \mathbb{R}$. In addition to the EVM$_\gamma$ distribution, other examples of SEmaxS distributions include the Generalized Logistic (type I) distribution and the Power function distribution.

The shape-extended stable class of distributions allows the generalization of stable distributions. However, this shape-extended definition does not retain the same properties. Another drawback is the absence of a precise definition of a shape parameter (unlike the location and scale parameters that have precise meanings). Nevertheless, this generalization provides a richer family of distributions able to generate the pseudo-convex mixtures (PCM).

2.3 PCM Generated by Shape-extended Stable Distributions for Extremes

Let F be SEminS distribution, then the r.v. X_m with d.f. F_{X_m} defined by

$$F_{X_m}(x) = (1+\omega)\, F(x) - \omega F_{X_{1:2}}(x),$$

with $\omega \in [-1, 1]$, is a pseudo-convex mixture (PCM) generated by the SEminS distribution F. F_{X_m} is a mixture between F and $F_{X_{1:2}}$, which is convex for $\omega < 0$ and non-convex for $\omega > 0$. The same reasoning can be applied to the maximum. Let F be a SEmaxS distribution, then the r.v. X_M with d.f. F_{X_M} defined by

$$F_{X_M}(x) = (1-\omega)\, F(x) + \omega F_{X_{2:2}}(x),$$

with $\omega \in [-1, 1]$, is a PCM generated by the SEmaxS distribution F. Hence, F_{X_M} is a mixture between F and $F_{X_{2:2}}$, convex for $\omega > 0$ and non-convex for $\omega < 0$. The formulas of F_{X_m} and F_{X_M} can be simplified to

$$F_{X_m}(x) = F_{X_M}(x) = F(x)\left[1 - \omega\overline{F}(x)\right], \quad (1)$$

with $\omega \in [-1, 1]$, which only depends on $F(x)$ and ω.

Note that in generalized mixtures, when there is one negative weight, as in equation (1), F_{X_m} is not guaranteed to be a d.f., [29]. Nevertheless, [13], proves that if F is a shape-extended stable distributions for extremes then F_{X_m} defined by equation (1) is a d.f.. Thus, PCM have the same parameters as F plus the ω parameter. Consequently, it is more flexible than the convex mixtures without raising the estimation cost. Figure 1 and Figure 2 illustrate the remarkable flexibility inherent in this distribution family, showing the density function of PCM generated by the standard Gumbel and the standard Logistic II distributions for different omega values. The main properties of PCM generated by shape-extended stable distributions are provided in [14].

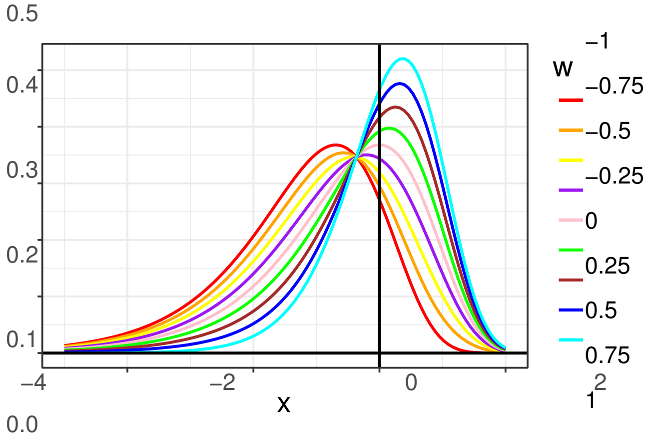

Fig. 1: Density function of PCM generated by the standard Gumbel distribution with $\omega = -1 + 0.25k$, $k = 0, 1, \ldots, 8$.

In this study, we confine our focus to a specific scenario: PCM generated by the exponential distribution. The exponential distribution, as an ESminS distribution, serves as the foundation for our investigation. It's worth noting that the exponential distribution represents a particular case of the Weibull distribution and holds significance across various domains of reliability analysis due to its flexibility and simplicity, [30].

3 PCM Generated by the Exponential Distribution

Let X be a r.v. with exponential (Exp) distribution with parameter $\lambda \in \mathbb{R}^+$ and d.f. $F(x) = 1 - e^{-\lambda x}$, $x \in \mathbb{R}^+$, which is a SEminS distribution as

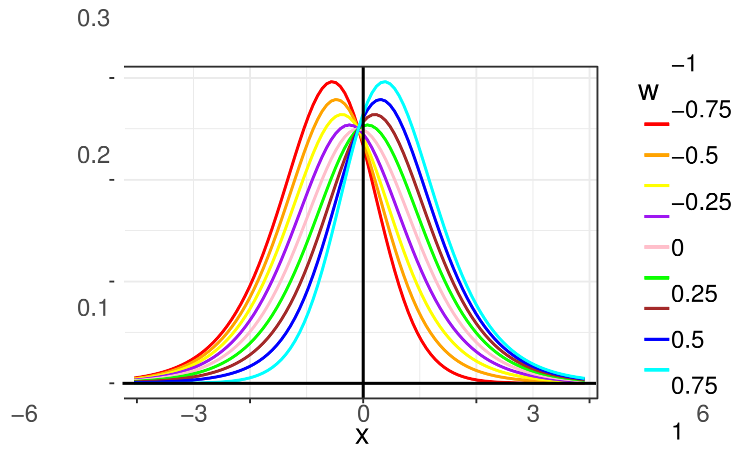

Fig. 2: Density function of PCM generated by the standard Logistic II distribution with $\omega = -1 + 0.25k$, $k = 0, 1, \ldots, 8$.

$X_{1:n} \sim \text{Exp}\,(n\lambda)$. The density function and the d.f. of the PCM generated by the exponential distribution (PCM_{Exp}) X_m are given by

$$F_{X_m}(x) = 1 - \left[1 + \omega\left(1 - e^{-\lambda x}\right)\right] e^{-\lambda x}$$

and

$$f_{X_m}(x) = (1 + \omega)\,\lambda e^{-\lambda x} - \omega 2\lambda e^{-2\lambda x}.$$

Figure 3 shows the shape of density functions of the PCM generated by the standard exponential distribution for different values of ω, with $\omega = -1 + 0.25k$, $k = 0, 1, \ldots, 8$.

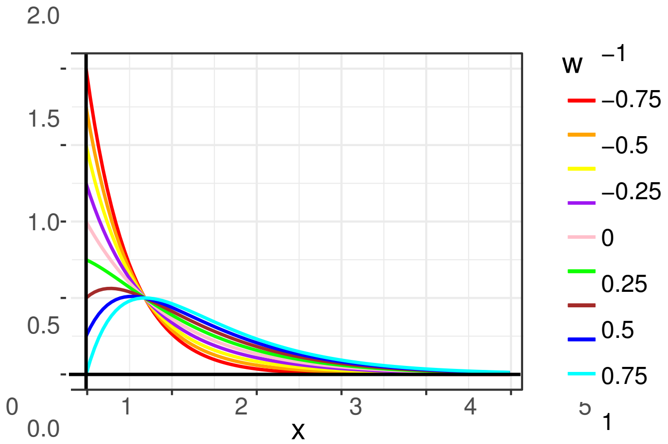

Fig. 3: Density function of PCM_{Exp} with $\omega = -1 + 0.25k$, $k = 0, 1, \ldots, 8$.

The hazard rate $r_X(x) := f_X(x)\,\overline{F}_X^{-1}(x)$, of the PCM_{Exp} is given by

$$\begin{aligned} r_{X_m}(x) &= \lambda\left(1 - \frac{\omega e^{-\lambda x}}{1 + \omega - \omega e^{-\lambda x}}\right) \\ &= r(x)\left(1 - \omega \frac{\overline{F}(x)}{1 + \omega F(x)}\right). \end{aligned}$$

Additionally, when $\omega = -1$, the PCM hazard rate becomes equal to $2r(x)$, where $r(x) = \lambda$ represents the hazard rate of a exponential distribution. It's important to note that when $\omega = -1$, this implies that X_m equals $X_{1:2}$, and consequently, $r_{X_{1:2}}(x) = 2r(x)$. Conversely, if ω is not equal to -1, then the PCM hazard rate will tend to converge to $r(x) = \lambda$ as x approaches infinity. Figure 4 illustrates the variations in the shape of the hazard rate functions of the PCM, which are generated by the standard exponential distribution, across different values of ω, with $\omega = 1 + 0.25k$, $k = 0, 1, \ldots, 8$.

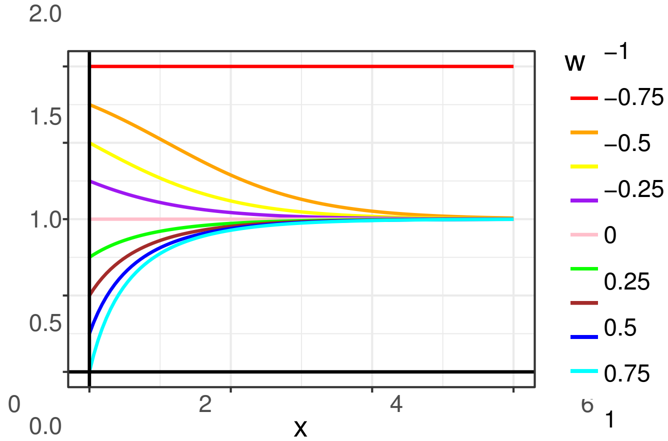

Fig. 4: Hazard rate of PCM_{Exp} with $\omega = -1 + 0.25k$, $k = 0, 1, \ldots, 8$.

3.1 Method of Moments Estimation

The k-th order raw moment of X_m, with $k \in \mathbb{N}$, is given by

$$\mathbb{E}\left(X_m^k\right) = \tfrac{k!}{\lambda^k}\left[1 + \omega\left(1 - \tfrac{1}{2^k}\right)\right].$$

Thus, the method of moments (MM) estimators can be given by

$$\widetilde{w} = 2\left(\lambda\overline{X} - 1\right)$$

and

$$\widetilde{\lambda} = \frac{3\overline{X} + \sqrt{9\overline{X}^2 - 4m_2}}{2m_2},$$

with

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

3.2 Maximum Likelihood Estimation

The log-likelihood function of λ and ω given the random sample $X = (X_1, \cdots, X_n)$ is

$$\begin{aligned} & \ell(\lambda, \omega|X) = \ln \mathcal{L}(\lambda, w|X) = \\ = \; & n \ln(\lambda) - n\lambda\overline{X} + \\ & \sum_{i=1}^{n} \ln\left(1 + w - 2w \exp\left(-\lambda X_i\right)\right), \end{aligned}$$

and its first partial derivatives are

$$\frac{\partial \ell(\lambda, \omega|X)}{\partial \lambda} = \\ \frac{n}{\lambda} - n\overline{X} + \sum_{i=1}^{n} \frac{2\omega X_i \exp\left(-\lambda X_i\right)}{1 + \omega - 2\omega \exp\left(-\lambda X_i\right)}$$

and

$$\frac{\partial \ell(\lambda, \omega|X)}{\partial \omega} = \sum_{i=1}^{n} \frac{1 - 2\exp\left(-\lambda X_i\right)}{1 + \omega - 2\omega \exp\left(-\lambda X_i\right)}.$$

Hence, it is not straightforward to find the vector $(\lambda_{\text{EMV}}, \omega_{\text{EMV}})$ that maximizes the likelihood function. Nevertheless, iterative methods for numerical approximation can be applied in order to achieve (an approximate value of) the maximum likelihood estimates (ML).

3.3 Expectation-maximization Algorithm

The expectation-maximization (EM) algorithm, [31], can be applied to estimate the unknown parameter $\theta = (\omega, \lambda) \in [-1, 1] \times]0, +\infty[$ in the PCM_{Exp}. In this case, for $\omega \leq 0$,

$$f_{X_m}(x) = (1 + \omega)\lambda e^{-\lambda x} - \omega 2\lambda e^{-2\lambda x}$$

is a convex mixture between $\lambda e^{-\lambda x}$ (Exp(λ) distribution) and $2\lambda e^{-2\lambda x}$ (Exp(2λ) distribution). Thus, the expectation step (E-step) in the k-th iteration can be obtained by

$$\gamma_0\left(x_i, \theta^{(k)}\right) = \\ \frac{(1 + \widehat{\omega}^{(k)}) \exp(-\widehat{\lambda}^{(k)} x_i)}{(1 + \widehat{\omega}^{(k)}) \exp(-\widehat{\lambda}^{(k)} x_i) + 2\widehat{\omega}^{(k)} \exp(-2\widehat{\lambda}^{(k)} x_i)},$$

where $\widehat{\theta}^{(k)} = (\widehat{\omega}^{(k)}, \widehat{\lambda}^{(k)})$. For the maximization step (M-step) in the k-th iteration we get

$$Q\left(\theta, \theta^{(k)}\right) = \\ \sum_{i=1}^{n} \gamma_0\left(x_i, \theta^{(k)}\right)\left[\ln(1+\omega) + \ln(\lambda) - \lambda x_i\right] + \\ \sum_{i=1}^{n}\left[1 - \gamma_0\left(x_i, \theta^{(k)}\right)\right]\left[\ln(-\omega) + \ln(2\lambda) - 2\lambda x_i\right],$$

which is maximized by

$$\widehat{\omega}_i^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} \gamma_0\left(x_i, \theta^{(k)}\right) - 1$$

and

$$\widehat{\lambda}_i^{(k+1)} = \frac{-n}{\sum_{i=1}^{n} x_i\left[\gamma_0\left(x_i, \theta^{(k)}\right) - 2\right]}.$$

However, EM algorithm does not converge with negative weights, [32], as when $\omega > 0$ in the PCM_{Exp}. Therefore, whenever $\widehat{\omega}^{(k)} > 0$, the density mixture was rewritten in the following convex mixture

$$f_{X_m}(x) = \omega 2\lambda e^{-\lambda x}\left(1 - e^{-\lambda x}\right) + (1-\omega)\lambda e^{-\lambda x}.$$

Thus, for positive values of ω, f_{X_m} can also be seen as a convex mixture between $\lambda e^{-\lambda x}$ (Exp(λ) distribution) and $2\lambda e^{-\lambda x}\left(1 - e^{-\lambda x}\right)$ (density of the maximum of two independent Exp(λ) distributions).

Therefore, in these cases ($\widehat{\omega}^{(k)} > 0$), the E-step in the k-th iteration is given by

$$\gamma_0'\left(x_i, \theta^{(k)}\right) = \frac{2\omega\left(1 - \exp(-\lambda x_i)\right)}{2\omega\left(1 - \exp(-\lambda x_i)\right) + 1 - \omega},$$

and for the M-step in the k-th iteration

$$Q'\left(\theta, \theta^{(k)}\right) = \\ \sum_{i=1}^{n} \gamma_0'\left(x_i, \theta^{(k)}\right)\left[\ln(2\omega\lambda) - \lambda x_i + \right. \\ \left. \ln\left(1 - \exp(-\lambda x_i)\right)\right] + \\ \sum_{i=1}^{n}\left(1 - \gamma_0'\left(x_i, \theta^{(k)}\right)\right)\left[\ln(1-\omega) + \ln(\lambda) - \lambda x_i\right]$$

which is maximized by

$$\widehat{\omega}_i^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} \gamma_0'\left(x_i, \theta^{(k)}\right)$$

and

$$\widehat{\lambda}_i^{(k+1)} = \left[\overline{x} - \frac{1}{n} \sum_{i=1}^{n} \gamma_0' \left(x_i, \theta^{(k)} \right) x_i \left(\exp(\lambda x_i) - 1 \right)^{-1} \right]^{-1} .$$

The EM algorithms repeat the E-step and the M-step until a fixed point is reached, i.e.,

$$\left\| \widehat{\theta}_i^{(k+1)} - \widehat{\theta}_i^{(k)} \right\| < \varepsilon,$$

for some fixed small enough $\varepsilon > 0$.

The EM algorithm's sensitivity to initial values is a well-known phenomenon, [31]. In this scenario, where the PCM is divided into two different convex mixtures, the problem is even worse as the sign of the initial omega value will almost surely define the sign of the final omega estimate. Hence, to address this issue, two estimates were computed, each initiated with different omega values: one with $\omega_0 = -0.5$ and the other with $\omega_0 = 0.5$. Regarding the initial λ value, as $\widetilde{w} = 2\left(\lambda\overline{X} - 1\right)$ by the MM, it follows that

$$\lambda = \frac{1 + 0.5\,\widetilde{w}}{\overline{X}}.$$

Thus, the chosen initial values (λ_0, ω_0) are $\left(0.75\,\overline{x}^{-1}, -0.5\right)$ and $(1.25\,\overline{x}^{-1}, 0.5)$. Ultimately, the two resulting estimates are compared using the Akaike Information Criterion (AIC), [33]. The estimate yielding the best fit (lowest AIC value) will be designated as the final EM estimate.

4 Simulations

In this section, the performance of parametric estimators for PCM_{Exp} through Monte Carlo simulation (10^4 replicas) is analysed. This evaluation was carry out in software R version 4.3.1, a language and environment for statistical computing, [34]. To this end, PCM_{Exp} were simulated with $\lambda \in \{1, 10\}$, $\omega \in \{-.75, -.50, -.25, 0, .25, .50, .75\}$ and $n \in \{100, 1000\}$. The parameters have been estimated using the MM, the ML based on numerical iterative methods using package `maxLik`, [35], on R (Newton-Raphson algorithm) with starting points $(\lambda_0, \omega_0) = \left(\overline{x}^{-1}, 0\right)$, and on the EM algorithm using as starting points $(\lambda_0, \omega_0) = \left(0.75\,\overline{x}^{-1}, -0.5\right)$ and $(\lambda_0, \omega_0) = (1.25\,\overline{x}^{-1}, 0.5)$, cf. Section 3.3. The EM algorithm stops when $\left\| \widehat{\theta}_i^{(k+1)} - \widehat{\theta}_i^{(k)} \right\| < 10^{-6}$. To assess the performance of the estimators, the bias (Bias), the absolute relative bias (ARB) and the mean square error (MSE) were used. The results obtained are presented in Table 1 and Table 2, and Figure 5.

Table 1. λ estimation in PCM_{Exp} with 10^4 replicas

ω	$-.75$	$-.50$	$-.25$	$.00$	$.25$	$.50$	$.75$
			MM, with $\lambda = 1$, $n = 100$				
Bias	.3683	.1291	.0557	.0332	.0306	.0222	.0235
ARB	.4249	.2741	.2235	.1768	.1438	.1232	.1088
MSE	.3005	.1170	.0739	.0493	.0338	.0244	.0191
			ML, with $\lambda = 1$, $n = 100$				
Bias	.4017	.1523	.0509	.0096	.0047	.0047	.0059
ARB	.4219	.2380	.1936	.1602	.1272	.1008	.0819
MSE	.2896	.0973	.0578	.0404	.0273	.0169	.0107
			EM, with $\lambda = 1$, $n = 100$				
Bias	.3649	.1366	.0331	.0054	−.0027	.0033	.0056
ARB	.3950	.2352	.1936	.1584	.1310	.1037	.0812
MSE	.2662	.0970	.0574	.0405	.0293	.0189	.0108
			MM, with $\lambda = 1$, $n = 1000$				
Bias	.1953	.0150	−.0035	.0030	.0024	.0023	.0026
ARB	.2031	.1199	.0833	.0567	.0453	.0387	.0342
MSE	.0693	.0196	.0131	.0052	.0033	.0024	.0018
			ML, with $\lambda = 1$, $n = 1000$				
Bias	.1708	.0198	−.0127	−.0012	.0002	.0009	.0007
ARB	.1973	.1125	.0862	.0516	.0381	.0306	.0253
MSE	.0705	.0190	.0135	.0046	.0023	.0015	.0010
			EM, with $\lambda = 1$, $n = 1000$				
Bias	.1507	.0145	−.0119	−.0026	.0002	.0008	.0007
ARB	.1846	.1122	.0852	.0515	.0379	.0299	.0250
MSE	.0635	.0188	.0131	.0047	.0023	.0014	.0010
			MM, with $\lambda = 10$, $n = 1000$				
Bias	1.957	.1473	−.0160	.0288	.0343	.0287	.0194
ARB	.2038	.1208	.0830	.0571	.0454	.0388	.0342
MSE	6.946	1.971	1.130	.5196	.3253	.2371	.1854
			ML, with $\lambda = 10$, $n = 1000$				
Bias	1.776	.2268	−.1001	−.0098	.0108	.0056	.0035
ARB	.2014	.1114	.0841	.0518	.0383	.0305	.0250
MSE	7.247	1.852	1.263	.4506	.2310	.1463	.0984
			EM, with $\lambda = 10$, $n = 1000$				
Bias	1.516	.1652	−.1157	−.0028	−.006	.0108	.0004
ARB	.1857	.1112	.0858	.0516	.0386	.0304	.0248
MSE	6.408	1.842	1.324	.4662	.2380	.1448	.0973

The accuracy of estimating the parameter λ is intricately tied to the precision of estimating ω; when one achieves precision, so does the other. In smaller samples ($n = 100$), MM notably demonstrates the poorest performance, evidenced by higher MSE. Moreover, EM outperforms ML when $\omega < 0$, although ML and EM display similar performances whenever $\omega > 0$.

As anticipated, increasing the sample size to $n = 1000$ enhances estimation quality across all estimators, resulting in more comparable performances. Nonetheless, MM continues to exhibit inferior performance compared to ML and EM, albeit showing similarities when ω is negative (mainly with ML). The performance of ML and EM continue to shows no significant differences for $n = 1000$ when $\omega > 0$, but maintains some differences in the performance of these estimators for $\omega < 0$. Additionally, altering the parameter value (for $\lambda = 10$) appears to have min-

Table 2. ω estimation in $\mathrm{PCM}_{\mathrm{Exp}}$ with 10^4 replicas

ω	$-.75$	$-.50$	$-.25$	$.00$	$.25$	$.50$	$.75$
	MM, with $\lambda = 1$, $n = 100$						
Bias	.4312	.1705	.0733	.0434	.0496	.0300	.0103
ARB	.6291	.7406	1.386	–	1.098	.4898	.2618
MSE	.3675	.1921	.1675	.1451	.1211	.0936	.0591
	ML, with $\lambda = 1$, $n = 100$						
Bias	.4621	.1948	.0580	–.0055	–.0071	–.0066	–.0008
ARB	.6329	.6148	1.179	–	.9350	.3698	.1827
MSE	.3527	.1532	.1246	.1183	.0973	.0601	.0305
	EM, with $\lambda = 1$, $n = 100$						
Bias	.4191	.1698	.0274	–.0091	–.0228	.0011	.0037
ARB	.6261	.6175	1.155	–	.9172	.3821	.1856
MSE	.3261	.1521	.1205	.1171	.0936	.0582	.0310
	MM, with $\lambda = 1$, $n = 1000$						
Bias	.2410	.0202	–.0082	.0050	.0042	.0041	.0051
ARB	.3265	.3422	.5408	–	.3496	.1608	.0991
MSE	.0997	.0385	.0304	.0167	.0121	.0102	.0086
	ML, with $\lambda = 1$, $n = 1000$						
Bias	.2076	.0242	–.0254	–.0035	–.0007	.0007	.0003
ARB	.3180	.3230	.5624	–	.2751	.1127	.0573
MSE	.1018	.0382	.0375	.0146	.0075	.0050	.0029
	EM, with $\lambda = 1$, $n = 1000$						
Bias	.1832	.0155	–.0233	–.0070	–.0004	.0005	.0002
ARB	.2992	.3206	.5541	–	.2781	.1111	.0577
MSE	.0924	.0377	.0359	.0149	.0079	.0049	.0030
	MM, with $\lambda = 10$, $n = 1000$						
Bias	.2405	.0193	–.0060	.0043	.0065	.0049	.0034
ARB	.3260	.3437	.5374	–	.3533	.1599	.1008
MSE	.0995	.0385	.0301	.0165	.0124	.0101	.0090
	ML, with $\lambda = 10$, $n = 1000$						
Bias	.2151	.0281	–.0215	–.0034	.0013	–.0007	–.0007
ARB	.3242	.3178	.5456	–	.2782	.1117	.0575
MSE	.1044	.0368	.0346	.0138	.0076	.0049	.0029
	EM, with $\lambda = 10$, $n = 1000$						
Bias	.1833	.0187	–.0237	–.0029	–.0020	.0004	–.0004
ARB	.3002	.3201	.5574	–	.2802	.1113	.0568
MSE	.0922	.0372	.0364	.0143	.0079	.0048	.0027

imal relative impact on estimation quality across all estimators.

Moreover, results tend to enhance with higher values of ω, particularly when dealing with non-convex mixtures, indicating superior outcomes. Specifically, for low values of ω, such as $\omega = -0.75$, all methods tend to overestimate ω, though this overestimation tends to diminish with larger sample sizes (albeit remaining significant even with $n = 1000$). Consequently, for these ω values, estimates may still lack precision.

The boxplots depicted in Figure 5 clearly illustrate that estimation precision notably increases when ω is positive. Additionally, bias tends towards zero or its proximity, a trend notably absent when $\omega = -0.75$. Noteworthy is the presence of outliers, indicating significantly lower estimation precision. Even employing EM, instances arise, particularly ev-

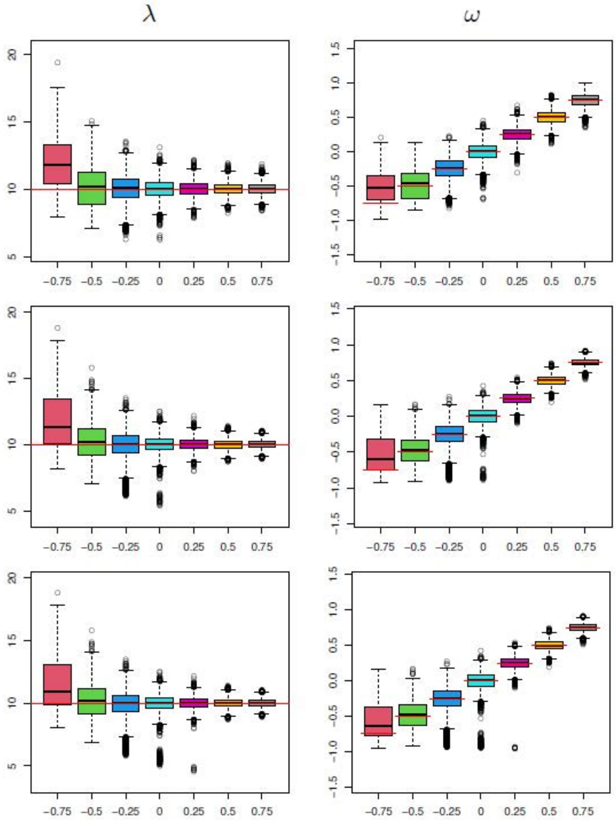

Fig. 5: λ and ω estimation in $\mathrm{PCM}_{\mathrm{Exp}}$ with 10^4 replicas and $\lambda = 1$ for MM (top), ML (middle) and EM (below).

ident when $\omega = 0.25$, where the estimate of ω nears -1 (the furthest value within the support of ω), resulting in similarly inaccurate estimates for λ (approximately 4.5 when $\lambda = 10$). It's worth noting that when $(\omega, \lambda) = (0.25, 10)$, $\mathbb{E}(X) = 0.1125$, and conversely, when $(\omega, \lambda) = (-1, 4.(4))$, $\mathbb{E}(X)$ remains 0.1125. Equivalently, the same expected value for X is obtained when $(\omega, \lambda) = (0, 10)$ and $(\omega, \lambda) = (-1, 5)$, or when $(\omega, \lambda) = (-0.25, 10)$ and $(\omega, \lambda) \approx (-1, 5.7143)$; representing some of the less precise scenarios observed in the simulations. Despite clear differences in the distribution functions in these cases, it appears that AIC occasionally struggles to select the optimal solution. Hence, it becomes pertinent to employ alternative measures of model selection or employ a combination of different metrics. However, it's crucial to acknowledge that such instances of very low precision in estimation, while impacting the overall metrics presented in Tables 1 and 2, are infrequent (less than 0.5%) and predominantly occur when the estimate of ω approaches -1. Consequently, in practical applications, exercising caution and employing a broader range of initial values is advisable when encountering such cases ($\widehat{\omega} \approx -1$) to ascertain the presence of significantly disparate estimates.

It's worth noting that different sample sizes (n) and λ parameter values were evaluated, and the results remained consistent with those reported, although there is a slight decrease in the number of cases where the estimate becomes less precise as the sample size increases.

Additionally, while variations in initial values were examined in ML, there were no noticeable differences observed, although these results were not detailed in the provided tables. Furthermore, the EM estimator were also assessed using different starting points, such as the MM estimates, i.e., considering $(\lambda_0, \omega_0) = (\lambda_{\text{MM}}, \omega_{\text{MM}})$ as it is straightforward computed. In this cases, only one estimate were evaluated and, therefore, the results were slightly worse. Nevertheless, probably the reason for this proximity is the fact that the sign of the MM estimate of ω (ω_{MM}) is the same as the true sign of ω with hight probability, namely whenever $|\omega| \geq 0.25$. Although this probability is low, in this cases the λ estimate can be quite different. For ω values in the neighbourhood of zero, the percentage of opposite signs is higher, but in these scenarios the density functions are quite similar, so the difference in λ estimates is not so significant. Furthermore, this percentages clearly decreases when the sample size increases, being quite lower when the sample size is $n = 1000$ than when $n = 100$.

5 Conclusion

Any PCM_{Exp} can be conceptualized as two separate convex mixtures, delineated for positive and negative values of ω. Hence, the final EM estimate for PCM_{Exp} will be the best of these EM estimates obtained under these two scenarios. Thus, this structure allows the application of the EM algorithm to be carried out only under convex mixtures, wherein the algorithm typically yields favourable outcomes. However, although yield superior estimates compared to other methods previously used, such as the maximum likelihood estimator, employing this algorithm doesn't appear to yield precise estimates across the entire support of (ω, λ). Hence, we plan to incorporate additional fit measures alongside AIC to evaluate potential disparities in the obtained results and explore alternative parameter estimation methods for cases requiring enhanced precision. In addition to other information criteria, it can be used goodness-of-fit statistics to determine the best estimate among the EM estimates, such as Kolmogorov-Smirnov, Anderson-Darling or Cramér-von Mises statistics, cf., [36], [37]. Furthermore, we aim to adopt a similar methodology to analyse other PCM generated by shape-extended stable distributions for extremes, with the goal of assessing the suitability of this approach.

References:

[1] O'Connor, A.N., Modarres, M. Mosleh, A., *Probability Distributions Used in Reliability Engineering*, Center for Risk and Reliability, University of Maryland, 2016.

[2] Elsayed, A.E., *Reliability Engineering*, John Wiley & Sons, Inc, 2021.

[3] Hosmer, D.W., Lemeshow, S, and May, S., *Applied Survival Analysis: Regression Modeling of Time to Event Data*, John Wiley & Sons, Inc., 2008.

[4] Zeng, K., Xu, X., Zhou, P. *et al.*, Financing the newsvendor with vendor credit line, *Oper Manag Res*, 2024. Published online. https://doi.org/10.1007/s12063-024-00475-3

[5] Bartholomew, D., Sufficient conditions for a mixture of exponentials to be a probability density function, *Ann. Math. Stat.* 40, 1969, pp. 2189–2194. https://doi.org/10.1214/AOMS%2F1177697296

[6] Steutel, F., Note on the infinite divisibility of exponential mixtures, *Ann. Math. Stat.*, 38, 1967, pp. 1303–1305. https://doi.org/10.1214/AOMS%2F1177698806

[7] Steutel, F., Preservation of infinite divisibility under mixing and related topics, *Mathematical Center Tracts* 33, Mathematisch Centrum, Amsterdam, 1970. https://doi.org/10.2307/2556202

[8] Zhang, B., Zhang, C., Finite mixture models with negative components, *MLDM 2005*, 2005, pp. 31–41. https://link.springer.com/chapter/10.1007/11510888_4

[9] McLachlan, G, Peel, D., *Finite Mixture Models*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 2000.

[10] Murphy, K.P., *Machine Learning A Probabilistic Perspective*, Massachusetts Institute of Technology, 2012.

[11] Klüppelberg, C., Seifert, M.I., *Explicit results on conditional distributions of generalized exponential mixtures*, Journal of Applied Probability, 2020, pp. 760–774. https://doi.org/10.1017/jpr.2020.26

[12] Yang Y., Tian W., Tong T., Generalized mixtures of exponential distribution and associated inference, *Mathematics*, Vol.9, No.12, 2021. https://doi.org/10.3390/math9121371

[13] Felgueiras, M., Martins, J.P., and Santos, R., Pseudo-convex Mixtures, *AIP Conf. Proc.*, 1479, 2012, pp. 1125-1128. https://doi.org/10.1063/1.4756346

[14] Santos, R., Felgueiras, M., and Martins, J.P., Pseudo-convex Mixtures Generated by Shape-extended Stable Distributions for Extremes, *Journal of Statistical Theory and Practice*, Vol.10, No.2, 2016, pp. 357–374. https://doi.org/10.1080/15598608.2016.1146929

[15] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., *Statistics of Extremes: Theory and Applications*, Wiley, England, 2004.

[16] David, H.A., Nagaraja, H.N., *Order Statistics*, John Wiley & Sons, New Jersey, 2003.

[17] De Haan, L., Ferreira, A., *Extreme Value Theory: An Introduction*, Springer, New York, 2006.

[18] Kotz, S., Nadarajah, S., *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London, 2000.

[19] Pickands, J., Statistical inference using extreme order statistics, *Ann. Stat.* 3, 1975, pp. 119–131. https://doi.org/10.1214/AOS%2F1176343003

[20] Reiss, R.D., Thomas, M., *Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields* 3rd edition, Birkhäuser Verlag, 2007.

[21] Beirlant, J., Caeiro, F., Gomes, M.I., An overview and open research topics in statistics of univariate extremes, *Revstat* 10, 2012, pp. 1–31. https://doi.org/10.57805/revstat.v10i1.109

[22] Castillo, E., Hadi, A.S., Balakrishnan, N., Sarabia, J.M., *Extreme Value and Related Models with Applications in Engineering and Science*, John Wiley & Sons, Hoboken, New Jersey, 2005.

[23] Embrechts, P., Klüppelberg, C., Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 2001.

[24] Ferreira, M., Clustering of extreme values: estimation and application, *AStA Advances in Statistical Analysis*, Vol.108, No.1, 2024, pp. 101–125. https://doi.org/10.1007/s10182-023-00474-y

[25] Albrecher, H., Beirlant, J., Statistics of Extremes for the Insurance Industry, In book:*Handbook of Statistics of Extremes*, Chapman & Hall, 2024.

[26] Dey, D.K., Yan, J., *Extreme Value Modeling and Risk Analysis*, Chapman and Hall/CRC, 2016.

[27] Longin, F., *Extreme Events in Finance: A Handbook of Extreme Value Theory and its Applications*, John Wiley & Sons, Inc, 2016.

[28] Zhao X., Zhang Z., Cheng W., Zhang P., A New Parameter Estimator for the Generalized Pareto Distribution under the Peaks over Threshold Framework. *Mathematics*, Vol.7, No.5, 2019. https://doi.org/10.3390/math7050406

[29] Wu, J.W., Characterizations of generalized mixtures of geometric and exponential distributions based on upper record values, *Stat. Pap.*, Vol42, 2001, pp. 123–133. https://doi.org/10.1007/s003620000045

[30] Das, S., Kundu, D., On Weighted Exponential Distribution and Its Length Biased Version, *J Indian Soc Probab Stat*, Vol17, 2016, pp. 57–77. https://doi.org/10.1007/s41096-016-0001-9

[31] Dempster, A.P., Laird, N.M., Rubin, D.B., Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol.39, No.1, 1977, pp. 1–38. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

[32] Frühwirth-Schnatter, S., *Finite Mixture and Markov Switching Models*, Springer, New York, 2012.

[33] Akaike, H., Factor analysis and AIC, *Psychometrika* 52, 1987, pp. 317–332. https://doi.org/10.1007/BF02294359

[34] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, 2023. https://www.R-project.org/

[35] Toomet, O., Henningsen, A., Graves, S., Croissant, Y., Hugh-Jones, D., Scrucca, L., *Package 'maxLik': Maximum Likelihood Estimation and Related Tools*, Publishing House, 2022. https://cran.r-project.org/web//packages/maxLik/

[36] Dickhaus, T., Goodness-of-Fit Tests. In: *Theory of Nonparametric Tests*, Springer, 2018. https://doi.org/10.1007/978-3-319-76315-6_3

[37] Darling, D.A., The Kolmogorov-Smirnov, Cramer-von Mises Tests, *The Annals of Mathematical Statistics*, Vol. 28, No. 4, 1957, pp. 823–38. https://doi.org/10.1214/AOMS%2F1177706788

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This work is partially financed by national funds through FCT – Fundação para a Ciência e a Tecnologia under the project UIDB/00006/2020. DOI: 10.54499/UIDB/00006/2020 (https://doi.org/10.54499/UIDB/00006/2020)

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.