Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

APPROXIMATING THE LÉVY-FRAILTY MARSHALL-OLKIN MODEL FOR FAILURE TIMES

Javiera Barrera
Guido Lagos

Faculty of Engineering and Sciences
Universidad Adolfo Ibáñez
Av. Diagonal Las Torres 2640, Peñalolén
Santiago, CHILE

ABSTRACT

In this paper we approximate the *last*, *close-to-first*, and what we call *quantile* failure times of a system, when the system-components' failure times are modeled according to a Levy-frailty Marshall-Olkin (LFMO) distribution. The LFMO distribution is a fairly recent model that can be used to model components failing simultaneously in groups. One of its prominent features is that the failure times of the components are *conditionally iid*; indeed, the failure times are iid exponential when conditioned on the path of a given Lévy subordinator process. We are motivated by further studying the order statistics of the LFMO distribution, as recently Barrera and Lagos (2020) showed an atypical behavior for the upper-order statistics. We are also motivated by approximating the system when it has an astronomically large number of components. We perform computational experiments that show significative variations in the convergence speeds of our approximations.

1 INTRODUCTION

In this paper we propose and test computationally several asymptotic approximations for random failure times having a *Lévy-frailty Marshall-Olkin* (LFMO) distribution. The LFMO model is a multivariate distribution that is a particular subfamily of the Marshall-Olkin (MO) distribution, see, e.g., Mai and Scherer (2009), Mai and Scherer (2011). In turn, the MO distribution is a a cornerstone in reliability as a tool to model simultaneous failures in systems; see, e.g., Kvam and Peña (2005) for applications in software reliability and civil engineering, and Frostig and Pellerey (2015) for applications in population dynamics and insurance theory. The main attractive of the LFMO distribution is perhaps that the components of the vector are *conditionally iid*, a highly desirable property in statistical modeling and machine learning, see, e.g., Efron and Hastie (2016) and (Reiss and Thomas 2007, ch. 8). In simple terms, this property poses that conditional on the value of a certain latent random variable, the components of the random vector are distributed iid; in particular, in the LFMO case, the components are iid exponentials when conditioned on the path of a given *Lévy subordinator* stochastic process, as we will see in Section 2. An interesting interpretation is that it models the heterogeneous degradation of (random) homogeneous components, e.g., the components in an aircraft subject to the mechanical degradation due to vibrations, or a chain subject to the corrosion of a liquid medium where it is submerged. Additionally, as a particular case of the MO distribution, the LFMO model inherits the following hallmark properties: it generalizes to multiple dimensions the memoryless property of exponential random variables; it models the *simultaneous* occurrence of events; and these simultaneous occurrences can be modeled as triggered by the arrivals or happenings of independent "shocks" that hit multiple components at the same time. Overall, these properties of the MO distribution are particularly convenient in the field of reliability, since it allows (at least in principle, see Matus et al. (2018)) straightforward modeling and simulation of, e.g., earthquakes taking down simultaneously several transmission lines and buses of a power grid. See also L'Ecuyer and Tuffin (2011) and Botev et al. (2013) for works on rare-event simulation in the MO setting.

Specifically, in this paper we propose asymptotic approximations for the *close-to-first*, *last*, and what we call *quantile* failure times modeled by the LFMO distribution. In detail, we consider the setting where there is a finite collection of components that fail at random times —sometimes several components failing at the same time instant—, once a component fails it stays in that state onwards, and assume that the failure times of the components are jointly distributed as an LFMO random variable. In this setting, we propose approximations for the following: the time when the close-to-first failure occurs; the time when the last failure occurs; and the time when a proportion of the total of components have failed, for each possible proportion value; we call the latter *quantile failure times*. These random variables are both interesting from a theoretical perspective, as order statistics or a classical interest in statistics, see, e.g., Reiss (2012), and also from a practical perspective, as reliability engineers are usually interested in failure times such as "the time when the last working component fails", "the time when the first failure occurs", "the time when 80% of the components have failed", and so on.

Our interest in approximating the aforementioned failure times given by the LFMO distribution is essentially motivated by the finding in Barrera and Lagos (2020) of several *atypical* asymptotic regimes for the last failure times as the number of components grows. In this sense, deriving further asymptotic approximations for the LFMO distribution can lead to new structural insights on the model and can allow to fine-tune the distribution to model very large systems. We are also motivated by the case where there is an astronomically large number of components in the system, e.g., as considered in the computational experiments of Barrera and Lagos (2020). There, for the larger number of components considered, each simulation run of the distribution was burdensome in computational time despite the Lévy subordinator being very easy to simulate; see Mai and Scherer (2017) for further information on simulating LFMO random variables.

Main contributions

The main contributions of this paper are the following.

1. We propose approximations for the following failure times distributed according to an LFMO model: the close-to-first, last, and quantile failure times. Our approximations are based on asymptotic results when the number of components of the system grows to infinity. To the best of the authors' knowledge, up to now only Barrera and Lagos (2020) has explored this path, and they showed approximations for the last and close-to-last failure times; in contrast, in this paper we show further approximations for the last failure time, and moreover show approximations for the close-to-first, and for quantile failure times.
2. Our approximations help to significatively reduce the computation time needed to simulate the aforementioned failure times given by the LFMO distribution. Indeed, for each proposed approximation method, we replace the need to simulate n exponential random variables and then sorting them —where n is the number of components of the system— with the need to simulate only one random variable. Thus, our approximations allow to save roughly $O(n \log n)$ operations. This can lead to considerable improvements in efficiency in the case where n is very large.
3. We show computational experiments testing the accuracy of our proposed approximations, and present evidence pointing to them being fairly precise already for a small to very small number of components of the system. This is a positive and unexpected result since the approximations were derived for systems with a large number of components.
4. From a theoretical point of view, our approximations also contribute to the structural understanding of the order statistics of a sequence of (conditionally-iid) random variables distributed according to an LFMO model.

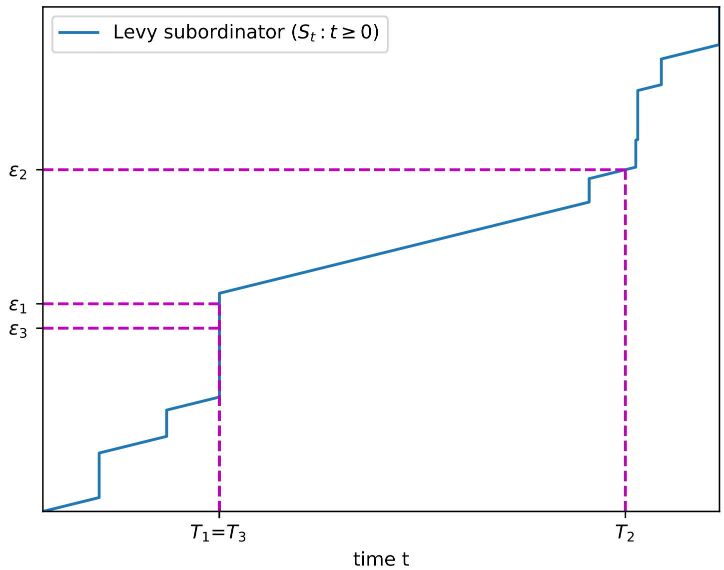

Figure 1: Simulation of a random vector (T_1, T_2, T_3) in $\mathbb{R}^3$ with an LFMO distribution: for each component i, T_i is the first time t the Lévy subordinator process S up-crosses the "trigger" ε_i, with $\varepsilon_1, \ldots, \varepsilon_n$ iid standard exponential random variables. In this simulation T_1 and T_3 are equal because an upward jump of the Lévy subordinator S up-crossed (or "killed") both triggers ε_1 and ε_3.

Organization of this paper This paper is organized as follows. In Section 2 we define the LFMO distribution and show our proposed approximations for the last, close-to-first, and quantile failure times. Finally, in Section 3 we show computational experiments testing the accuracy of our experiments.

2 ASYMPTOTIC APPROXIMATIONS OF THE LFMO DISTRIBUTION

In this section we show our proposed approximations. For that, first we give a brief description of the LFMO distribution in Definition 1, then in Section 2.1 we show our proposed approximations, and finally in Section 2.2 we give short proofs of our approximations.

Definition 1 A random vector T in $\mathbb{R}^n$ is said to have a *Lévy-frailty Marshall-Olkin (LFMO)* distribution if its components $(T_1, \ldots, T_n)$ can be jointly defined as

$$T_i := \min\{t \geq 0 \ : \ S_t \geq \varepsilon_i\}, \qquad i = 1, \ldots, n, \tag{1}$$

where $S = (S_t : t \geq 0)$ is a Lévy subordinator stochastic process with $S_0 = 0$, and $\varepsilon_1, \ldots, \varepsilon_n$ is a collection of n iid standard exponential random variables independent of S.

In system reliability, an LFMO distribution can be used to model the times at which the components of a system fail when several of the components can fail at the same time. Indeed, an intuitive interpretation of an LFMO distributed random vector $(T_1, \ldots, T_n)$ in a system with n components is, first, taking each T_i as the time at which component i fails, second, thinking that each component i has associated an exponential "trigger" ε_i, and third, noting that according to definition (1) component i fails the first time the Lévy subordinator S up-crosses (or "kills") its trigger ε_i. Recall now that a Lévy subordinator is a Lévy stochastic process with non-decreasing paths; in simple terms, it can be heuristically understood as a random trajectory of $S_t : t \geq 0$ that is non-decreasing in time t and that has upward jumps at random times — in total a countable number of such jumps. Importantly then, it is because of these upward jumps that several components can fail simultaneously; in other words, a jump of the subordinator S can take down multiple components by up-crossing several triggers at the same time. In Figure 1 we show an example of a system with $n = 3$ components where components 1 and 3 fail simultaneously because of an upward jump of the subordinator S.

Several interesting properties of the LFMO are in order. First, note that the components' lifetimes are *conditionally iid*, an important property e.g. in statistics, since the lifetimes T_i are iid when conditioned on the path of the Lévy subordinator S. In particular, the marginal distribution of each failure time T_i is exponential with rate $\psi(1)$, where $\psi(x) := -\log \mathbb{E}\exp(-xS_1)$ is the Laplace exponent of the Lévy subordinator S. Second, note that the LFMO distribution actually subsumes the model of iid exponentially distributed times: if the subordinator is deterministic with $S_t := \lambda t$ for some fixed $\lambda > 0$ then $T_1,\ldots,T_n$ are iid exponential random variables with rate λ. Third, the LFMO distribution is a particular case of the Marshall-Olkin distribution, a cornerstone in reliability modeling. In particular, it satisfies a multidimensional version of the memoryless property of exponential random variables, a key property for modeling and analysis. Lastly, an interesting interpretation of the construction (1) of the LFMO distribution is that the triggers $\varepsilon_1,\ldots,\varepsilon_n$ represent the *homogeneous* but random nature of the components, and the Lévy subordinator S represents a common "force" that degrades the components in an *heterogeneous* way. Think, e.g., about the failure of mechanical components in an airplane that may in principle be modeled as homogeneous (iid), but because of the vibrations and other stressors of the airplane's operation they degrade in a heterogenous fashion; here the Lévy subordinator represents the "overall stress" that degrades heterogeneously the, at first, homogenous components. See (Mai and Scherer 2017, ch. 3) and Barrera and Lagos (2020) for further properties of the LFMO distribution.

2.1 Proposed approximations

We now consider the *last*, *close-to-first*, and what we call *quantile* failure times: the time when a given fraction q in $(0,1)$ of the total components have failed. For that, we use the notation $T_{1:n}$, $T_{2:n}$, ..., $T_{n:n}$ for the time when, respectively, the first, second, ..., last failure occurs. That is, $T_{1:n},\ldots,T_{n:n}$ is the sorting of the times $T_1,\ldots,T_n$ in increasing order: $\{T_{1:n},T_{2:n},\ldots,T_{n:n}\} = \{T_1,T_2,\ldots,T_n\}$ and $T_{1:n} \le T_{2:n} \le \ldots \le T_{n:n}$. Analogously, $\varepsilon_{1:n},\ldots,\varepsilon_{n:n}$ denotes the sorting, in increasing order, of the "triggers" $\varepsilon_1,\ldots,\varepsilon_n$ of the components of the system. Note that because of the definition (1) of the failure times it holds that

$$T_{k:n} = \min\{t \ge 0 \ : \ S_t \ge \varepsilon_{k:n}\}, \qquad k = 1,\ldots,n. \tag{2}$$

With this, our proposed approximations are the following.

Last failure times: Consider the last failure time $T_{n:n}$. We propose approximating it by $\widetilde{T}_{n:n}$ defined as

$$\widetilde{T}_{n:n} := \min\{t \ge 0 \ : \ S_t \ge G_n\}, \tag{3}$$

where G_n is distributed as a *Gumbel*$(\log n, 1)$ random variable. That is, we approximate $T_{n:n}$ by replacing the random variable $\varepsilon_{n:n}$ in the definition (2) of $T_{n:n}$ by G_n. The approximation is based on the following lemma; see Section 2.2 below for its proof.

Lemma 1 For $T_{n:n}$ and $\widetilde{T}_{n:n}$ defined according to (2) and (3), respectively, it holds that

$$\sup_{t>0} \left| \mathbb{P}(T_{n:n} > t) - \mathbb{P}(\widetilde{T}_{n:n} > t) \right| \to 0 \qquad \text{as } n \to +\infty. \tag{4}$$

Close-to-first failure times: Consider the close-to-first failure times $T_{k:n}$ for $k = k_n$ such that $k_n/n \to 0$ as $n \to +\infty$. We propose to approximate $T_{k:n}$ by $\widetilde{T}_{k:n}$ defined as

$$\widetilde{T}_{k:n} = \widetilde{T}_{1:n} := \min\{t \ge 0 \ : \ S_t \ge e_n\}, \tag{5}$$

where e_n is distributed as an *exponential*(n) random variable. That is, we approximate $T_{k:n}$ by replacing the random variable $\varepsilon_{k:n}$ in the definition (2) by e_n. The approximation is based on the following lemma; see Section 2.2 below for its proof.

Lemma 2 For $T_{k:n}$ and $\widetilde{T}_{k:n}$ defined according to (2) and (5), respectively, it holds that

$$\left|\mathbb{P}(T_{k:n} > t) - \mathbb{P}(\widetilde{T}_{k:n} > t)\right| \to 0 \qquad \text{as } n \to +\infty \tag{6}$$

for all $t > 0$ and any fixed integer $k \geq 1$.

Quantile failure times: Let q in $(0,1)$ and consider the $\lceil qn \rceil$-th failure time out of the total of n, i.e., $T_{\lceil qn \rceil:n}$. We call this the *q-quantile failure time*. We propose approximating it by $\widetilde{T}_{\lceil qn \rceil:n}$ defined as

$$\widetilde{T}_{\lceil qn \rceil:n} := \min\left\{t \geq 0 \ : \ S_t \geq N_{q,n}\right\}, \tag{7}$$

where $N_{q,n}$ is distributed as a $Normal\left(-\log(1-q), \left(\frac{q}{1-q}\right)/n\right)$ random variable. That is, we approximate $T_{\lceil qn \rceil:n}$ by replacing the random variable $\varepsilon_{\lceil qn \rceil:n}$ in the definition (2) by $N_{q,n}$. The approximation is based on the following lemma; see Section 2.2 below for its proof.

Lemma 3 Let q in $(0,1)$. For $T_{\lceil qn \rceil:n}$ and $\widetilde{T}_{\lceil qn \rceil:n}$ defined according to (2) and (7), respectively, it holds that

$$\sup_{t>0}\left|\mathbb{P}(T_{\lceil qn \rceil:n} > t) - \mathbb{P}(\widetilde{T}_{\lceil qn \rceil:n} > t)\right| \to 0 \qquad \text{as } n \to +\infty. \tag{8}$$

2.2 Asymptotic analysis: proofs of Lemmas 1, 2 and 3

We now prove Lemmas 1, 2 and 3, which are the basis of the approximations we propose. These results are essentially based on the limits (9), (10) and (11) below, respectively, which are asymptotic approximations for the order-statistics of a collection of iid exponential random variables.

Proof of Lemma 1. For the case of last failure times, to check the limit (4), i.e.,

$$\sup_{t>0}\left|\mathbb{P}(T_{n:n} > t) - \mathbb{P}(\widetilde{T}_{n:n} > t)\right| \to 0 \qquad \text{as } n \to +\infty,$$

first note that $\mathbb{P}(\varepsilon_{n:n} - \log n > s) \to \mathbb{P}(G_1 > s)$ as $n \to +\infty$ for all s, where G_1 is a *Gumbel*$(0,1)$ distributed random variable. In fact, since G_1 has a continuous distribution then it holds that

$$\sup_x |\mathbb{P}(\varepsilon_{n:n} - \log n > x) - \mathbb{P}(G_1 > x)| \to 0 \qquad \text{as } n \to +\infty; \tag{9}$$

see (Chung and Zhong 2001, S. 4.3, Example 4). Hence, since $\{T_{k:n} > t\} = \{\varepsilon_{k:n} > S_t\}$ and $\{\widetilde{T}_{k:n} > t\} = \{G_n > S_t\}$, we get that for G_n a *Gumbel*$(\log n, 1)$ distributed random variable and all $t > 0$ it holds that

$$\begin{aligned}\left|\mathbb{P}(T_{n:n} > t) - \mathbb{P}(\widetilde{T}_{n:n} > t)\right| &= |\mathbb{P}(\varepsilon_{n:n} > S_t) - \mathbb{P}(G_n > S_t)| \\ &= |\mathbb{P}(\varepsilon_{n:n} - \log n > S_t - \log n) - \mathbb{P}(G_1 > S_t - \log n)| \leq \sup_x |\mathbb{P}(\varepsilon_{n:n} - \log n > x) - \mathbb{P}(G_1 > x)|,\end{aligned}$$

the latter bound holding uniformly on $t > 0$, from which the limit (4) follows. □

Proof of Lemma 2. For the close-to-first failure times, the approximation (6) is essentially based on the limit

$$|\mathbb{P}(\varepsilon_{k:n} > s) - \mathbb{P}(\varepsilon_{1:n} > s)| \to 0 \qquad \text{as } n \to +\infty \tag{10}$$

for all $s > 0$ and every fixed $k \geq 1$. Indeed, for such s and k we have

$$\begin{aligned}|\mathbb{P}(\varepsilon_{k:n} > s) - \mathbb{P}(\varepsilon_{1:n} > s)| &\leq |\mathbb{P}(\varepsilon_{k:n} > s) - \mathbb{P}(E_{k,1}/n > s)| + |\mathbb{P}(E_{k,1}/n > s) - \mathbb{P}(\varepsilon_{1:n} > s)| \\ &\leq \sup_{x>0} |\mathbb{P}(\varepsilon_{k:n} > x) - \mathbb{P}(E_{k,1}/n > x)| + |\mathbb{P}(E_{k,1}/n > s) - \mathbb{P}(\varepsilon_1/n > s)| \\ &= \sup_{x>0} |\mathbb{P}(n\varepsilon_{k:n} > x) - \mathbb{P}(E_{k,1} > x)| + |\mathbb{P}(E_{k,1} > ns) - \mathbb{P}(\varepsilon_1 > ns)|,\end{aligned}$$

where $E_{k,\gamma}$ is as an Erlang(k,γ) distributed random variable and where we used basic properties of the Erlang distribution. It follows that $\sup_{x>0} |\mathbb{P}(n\varepsilon_{k:n} > x) - \mathbb{P}(E_{k,1} > x)| \to 0$ as $n \to +\infty$ from (Reiss 2012, p. 162) and the fact the cumulative distribution function of $E_{k,1}$ has no jumps; and also $\mathbb{P}(E_{k,1} > ns) - \mathbb{P}(\varepsilon_1 > ns) = \sum_{i=0}^{k-1} e^{-ns}(ns)^i/i! - e^{-ns} = \sum_{i=1}^{k-1} e^{-ns}(ns)^i/i! \to 0$ as $n \to +\infty$. Lastly, since $\{T_{k:n} > t\} = \{\varepsilon_{k:n} > S_t\}$ and $\{\widetilde{T}_{1:n} > t\} = \{\varepsilon_{1:n} > S_t\}$ then

$$\left|\mathbb{P}(T_{k:n} > t) - \mathbb{P}(\widetilde{T}_{1:n} > t)\right| = |\mathbb{P}(\varepsilon_{k:n} > S_t) - \mathbb{P}(\varepsilon_{1:n} > S_t)|,$$

from which (6) follows. □

Proof of Lemma 3. For the quantile failure times, the limit (8) is based on (Reiss 2012, Theorem 4.1.3), that establishes

$$\sup_x \left|\mathbb{P}\left(\varepsilon_{\lceil qn \rceil:n} > x\right) - \mathbb{P}(N_{q,n} > x)\right| \to 0 \qquad \text{as } n \to +\infty, \tag{11}$$

for $N_{q,n}$ having a *Normal*$\left(-\log(1-q), (\tfrac{q}{1-q})/n\right)$ distribution. Indeed, since $\{T_{\lceil qn \rceil:n} > t\} = \{\varepsilon_{\lceil qn \rceil:n} > S_t\}$ and $\{\widetilde{T}_{\lceil qn \rceil:n} > t\} = \{N_{q,n} > S_t\}$ then

$$\begin{aligned}\sup_{t>0} \left|\mathbb{P}(T_{\lceil qn \rceil:n} > t) - \mathbb{P}(\widetilde{T}_{\lceil qn \rceil:n} > t)\right| &= \sup_{t>0} \left|\mathbb{P}\left(\varepsilon_{\lceil qn \rceil:n} > S_t\right) - \mathbb{P}(N_{q,n} > S_t)\right| \\ &\leq \sup_x \left|\mathbb{P}\left(\varepsilon_{\lceil qn \rceil:n} > x\right) - \mathbb{P}(N_{q,n} > x)\right|\end{aligned}$$

from which (8) follows. □

3 COMPUTATIONAL EXPERIMENTS

In this section we show the computational experiments we perform to test our proposed approximations.

Setup. To evaluate the accuracy of our approximations we use a large number of Kolmogorov-Smirnov tests to check if the distribution of our proposed approximation is "close" to the distribution to be approximated.

Indeed, first, for each approximation method of failure time, and for several possible values of the number of components n, we sample 10^3 values of the actual LFMO failure times, and also 10^3 simulations of our proposed approximations. For example, in the case of the last failure times, we sample 10^3 times the random variables $T_{n:n}$ and $\widetilde{T}_{n:n}$. It follows that a two-sample Kolmogorov-Smirnov test should reflect on both samples coming from the same distribution: if the p-value is high then we cannot reject the hypothesis that the distributions of the two samples are the same. Now, the result obtained in one run of the test is inherently random, so to hedge against this noise we repeat the process 1,000 times and report the average p-value. This experiment is used, e.g., in Lachaud and Ycart (2006) to test convergence of Markov chains.

For the Lévy subordinator process $S = (S_t : t \geq 0)$ underlying the LFMO distribution we choose to use a compound Poisson process (CPP), possibly with drift, i.e., a process of the type

$$\mu t + \sum_{k=1}^{N_t} J_k, \qquad \text{for } t \geq 0, \tag{12}$$

where $\mu \geq 0$ is the *drift* term, $N = (N_t : t \geq 0)$ is a Poisson process, say with rate $\lambda > 0$, and $J_1, J_2, \dots$ is a sequence of iid strictly positive random variables —the jumps— that are independent of N. We base our choice on the fact that, in a way, any Lévy subordinator can be approximated arbitrarily close by a compound Poisson process, see, e.g., (Feller 1971, Ch. XVII S. 2). More generally, though, any Lévy subordinator can be decomposed into the independent sum of a compound Poisson process with jumps larger than any fixed positive value, say 1, and a Lévy subordinator with Lévy measure having jumps less than 1; see, e.g., (Asmussen and Glynn 2007, Ch. XII) and (Mai and Scherer 2017, Appendix A.2).

More specifically, for the compound Poisson process we use as Lévy subordinator S, we test two type of distributions for the jumps J_k in (12): a Pareto$(\alpha = 1.5)$ distribution (i.e., with mean 3 and infinite variance), representing heavy-tailed jumps, and a uniform$(0,1)$ distribution (i.e., with mean $1/2$ and variance $1/12$), representing light-tailed jumps. We also examine the cases with drift $\mu = 1$, and the drift-less case $\mu = 0$.

Lastly, we remark that we run these computational experiments on `Python v3.7.3` and perform the two-sample Kolmogorov-Smirnov tests using the command `scipy.stats.ks_2samp` of the package `scipy v1.2.3`.

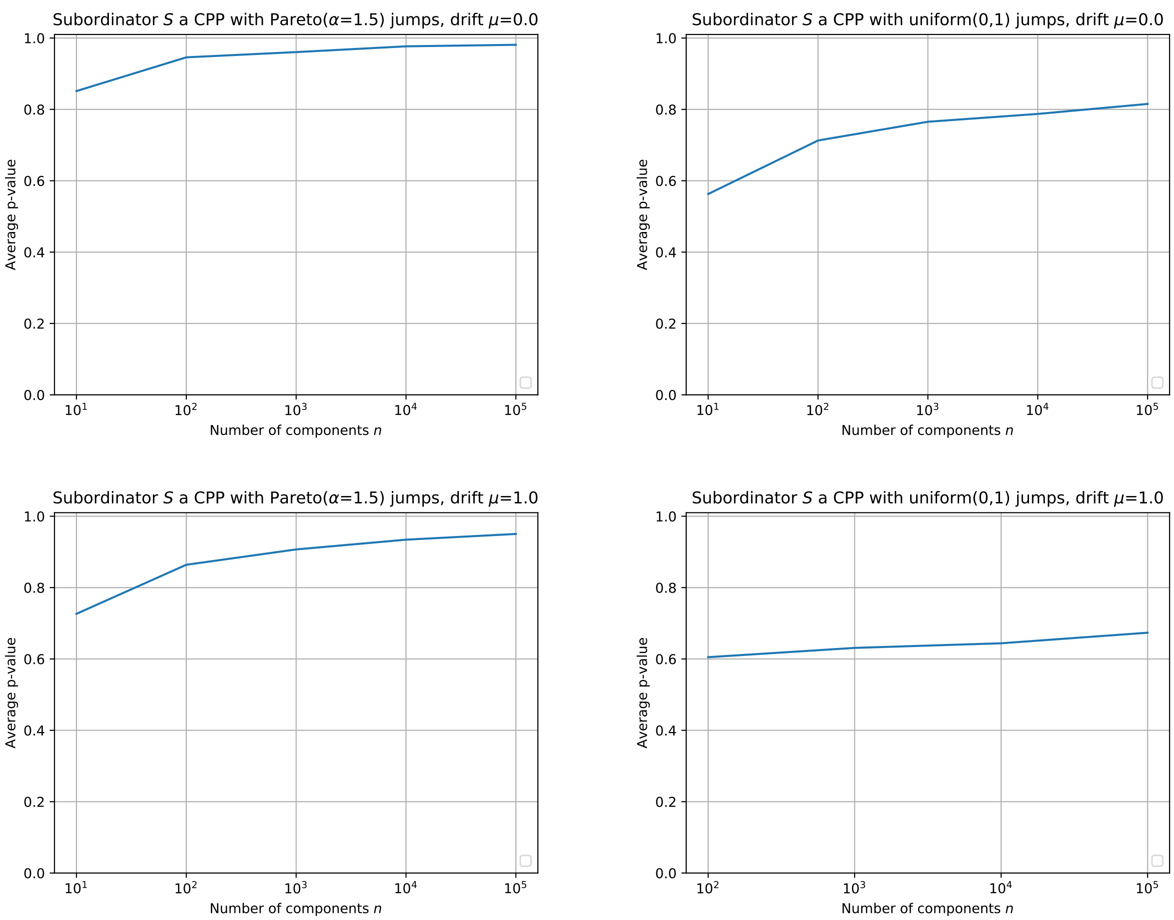

Figure 2: Average p-value over 1,000 K-S tests comparing our proposed approximation (3) for the last failure times, i.e., approximating $\varepsilon_{n:n}$ by a *Gumbel*$(\log(n), 1)$ random variable in the definition (2) of $T_{n:n}$.

Results. In Figure 2 we show the results for approximating the last failure times $T_{n:n}$ with $\widetilde{T}_{n:n}$ of (3): we show how the average p-value varies when n grows exponentially fast, and we do this for several choices of compound Poisson process. Recall that p-values close to 1 suggest that the samples compared come from the same distribution. The plots show that the p-value grows with n, however it does so slowly. Nonetheless, in most cases the p-value is far from 1, and including a drift term makes the convergence slower. Note though that the convergence is guaranteed by Lemma 1, so overall these experiments suggest a very slow rate of convergence.

In Figure 3 we show the results for approximating the close-to-first failure times $T_{k:n}$ with $\widetilde{T}_{k:n}$ of (5) when $k = k_n = \lfloor \log n \rfloor$; specifically, we show how the average p-value varies when n grows, and we do this for several choices of compound Poisson process. The plots show two behaviors: convergence is fast when there is no drift ($\mu = 0$), and moreover in the case of Pareto jumps the two distributions are indistinguishable already for $n = 10$; however there is no glimpse of convergence —even when n grows exponentially fast— when there is a drift term. Importantly, though, note that Lemma 2 guarantees convergence for k fixed, but for the results in Figure 3 we have taken $k = k_n$ growing with n such that $k_n/n \to 0$.

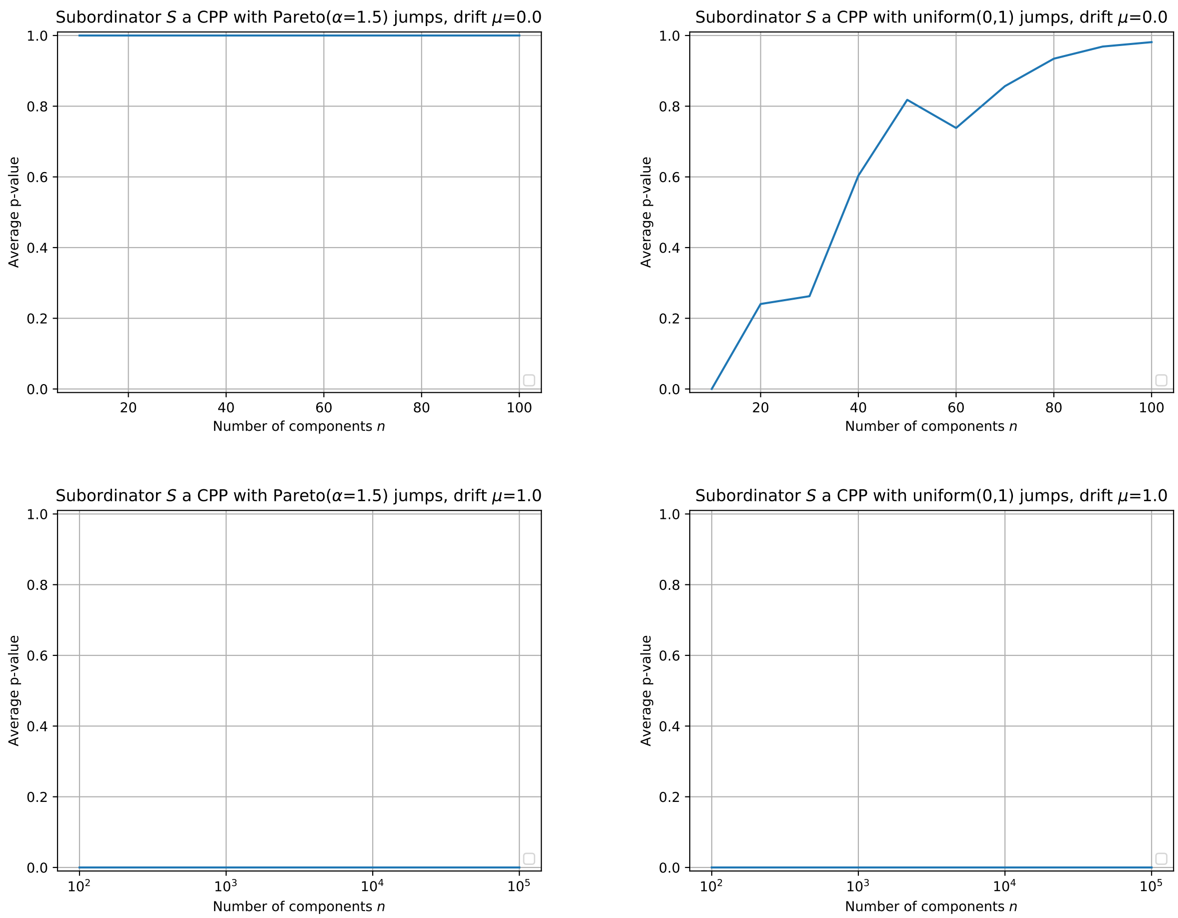

Figure 3: Average p-value over 1,000 K-S tests comparing our proposed approximation (5) for the close-to-first failure times, i.e., approximating $\varepsilon_{k:n}$ by an $exponential(n)$ random variable in the definition (2) of $T_{k:n}$, for $k = k_n = \lfloor \log n \rfloor$. Importantly, note that Lemma 2 guarantees convergence for k fixed, but here we have taken $k = k_n$ growing with n such that $k_n/n \to 0$. Note also the change in scale of the horizontal axes.

Lastly, in Figure 4 we show the results for approximating the quantile failure times $T_{\lceil qn \rceil:n}$ with $\widetilde{T}_{\lceil qn \rceil:n}$ of (7), for several values of q. Indeed, we show how the average p-value varies when n grows and for several values of q in $(0,1)$, and we do this for four types of compound Poisson process. We see again two type of behaviors depending on the drift term μ: there is fast convergence when there is no drift, however convergence becomes unclear —even when n grows exponentially fast— when there is a drift term. Note also that, intriguingly, the inclusion of a drift term inverts how "easy" the convergence is with q: when there is no drift, convergence becomes (roughly) easier as q decreases to 0; and when there is a drift term convergence becomes easier when q increases to 1. In any case, recall that Lemma 3 guarantees convergence of the approximation, so the experiments suggest that the speed of convergence can depend on q and the drift term μ.

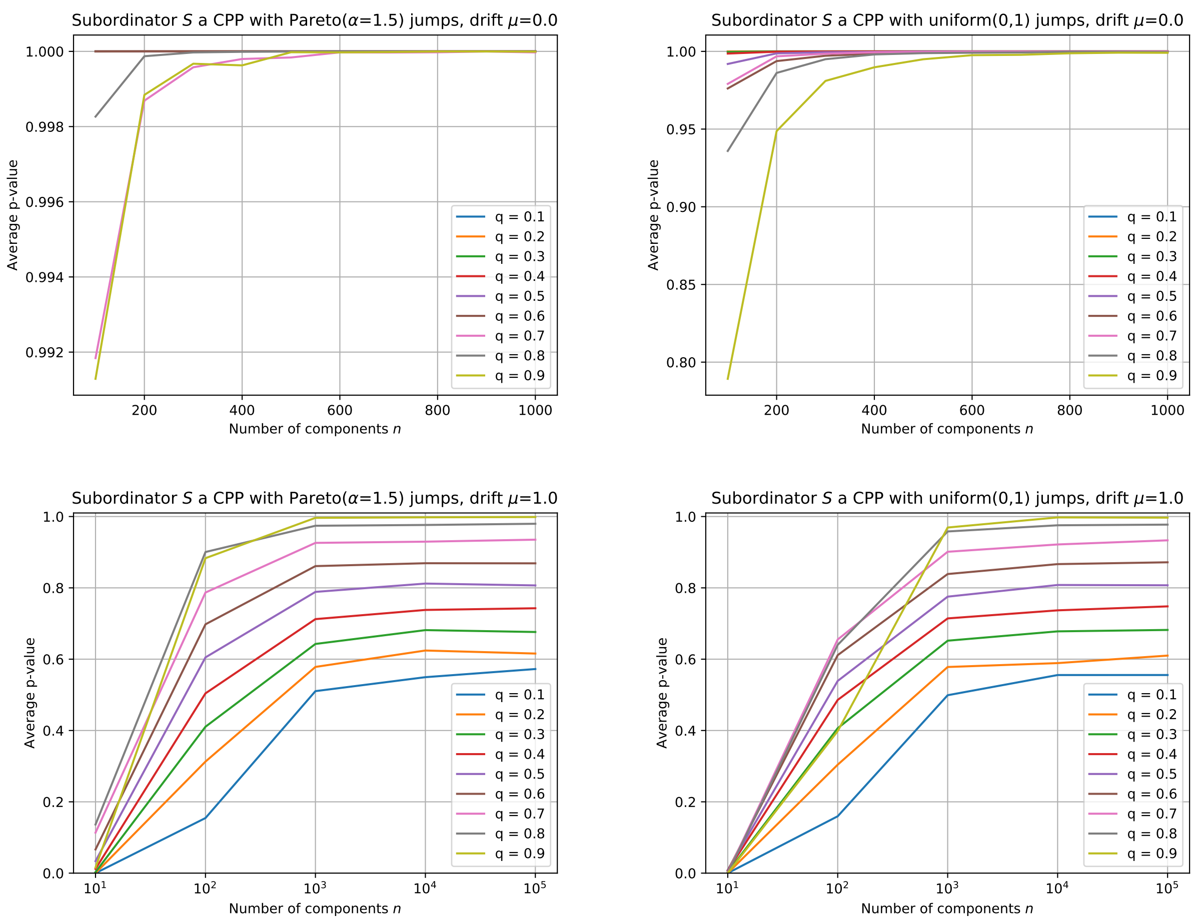

Figure 4: Average p-value over 1,000 K-S tests comparing our proposed approximation (7) for the quantile failure times, i.e., approximating $\varepsilon_{\lceil qn \rceil:n}$ by a *Normal* $\left(-\log(1-q), (\frac{q}{1-q})^2/(n+1)\right)$ random variable in the definition (2) of $T_{\lceil qn \rceil:n}$, for several values of q in $(0,1)$. Note the change in scale of both the vertical and horizontal axes.

4 DISCUSSION

In this work we have presented approximations for the last, close-to-first, and quantile failure times of the components of a system whose lifetimes are modeled according to a Lévy-frailty Marshall-Olkin (LFMO) distribution. The approximations we propose are based on asymptotic results we derive, which are in turn essentially based on classical results on the approximation of the order-statistics of a collection of iid exponential random variables.

Our motivations to approximate failure times modeled according to an LFMO distribution are mainly the following. First, we aim to extend the work of Barrera and Lagos (2020), that found several *atypical* asymptotic regimes for the failure times as the number of components grows. Second, we are lead by the need to derive (very) efficient methods to simulate failure times when the system has an astronomically large number of components. For example, in Barrera and Lagos (2020), Monte Carlo estimation was performed for systems with a number n of components of 10^{10}, 10^{40}, 10^{90} and even 10^{160}. In these types of settings, crude Monte Carlo simulation of the LFMO distribution can be impractical, so more efficient methods are needed. In that sense, the approximations we propose in this paper replace the need to simulate and sort n iid exponential random variables, requiring roughly $O(n\log n)$ operations, by simulating only one random variable. This improvement can make a significant difference when simulating systems with a very large number of components n. Lastly, our work is ultimately motivated by the need to approximate the failure times *of the whole system* by using the Samaniego Signature result, see, e.g., Samaniego (2007). This is a promising avenue of research that can allow us to tackle more general definitions for the failure time of the system, as long as the system satisfies certain monotonicity conditions; see the aforementioned reference.

We also present computational experiments testing the accuracy of our approximations. Our experiments are largely based on the two-sample Kolmogorov-Smirnov test. The results, however, are inherently random, so to hedge against this noise our tests take the average p-value of 1,000 of these tests. The results of our experiments suggest that the convergence of our approximations are affected by the inclusion of a drift term in the Lévy subordinator, with the convergence being slower when there is drift. Moreover, the experiments suggest that in the close-to-first approximations, there may not be convergence when k grows with n and there is a drift term.

Overall, our results suggest that the speed of convergence of our approximations can vary considerably with the parameters of the subordinators and the parameters of the failure times to be approximated — say, k in the close-to-first failure times, and q in the quantile failure times. Hence, an important future step is to analyze the rate of convergence of the limits in Lemmas 1, 2 and 3.

ACKNOWLEDGMENTS

The authors thank fruitful discussions and suggestions from Héctor Cancela; particularly, the interpretation of the LFMO distribution as model for failures of homogeneous components in an heterogeneous medium. Javiera Barrera acknowledges the financial support of FONDECYT grant 1161064, Programa Iniciativa Científica Milenio NC120062, and by ANILLO ACT192024. Guido Lagos acknowledges the financial support of FONDECYT grant 3180767.

REFERENCES

Asmussen, S., and P. W. Glynn. 2007. *Stochastic simulation: algorithms and analysis*, Volume 57. Springer Science & Business Media.

Barrera, J., and G. Lagos. 2020. "Limit distributions of the upper order statistics for the Lévy-frailty Marshall-Olkin distribution". *Extremes*. To appear.

Botev, Z. I., P. L'Ecuyer, G. Rubino, R. Simard, and B. Tuffin. 2013. "Static network reliability estimation via generalized splitting". *INFORMS Journal on Computing* 25(1):56–71.

Chung, K., and K. Zhong. 2001. *A Course in Probability Theory*. Elsevier Science.

Efron, B., and T. Hastie. 2016. *Computer age statistical inference*, Volume 5. Cambridge University Press.

Feller, W. 1971. *An Introduction to Probability theory and its application Vol II*. John Wiley and Sons.

Frostig, E., and F. Pellerey. 2015. "General Marshall–Olkin Models, Dependence Orders, and Comparisons of Environmental Processes". In *Marshall-Olkin Distributions - Advances in Theory and Applications*, 51–64. Springer.

Kvam, P. H., and E. A. Peña. 2005. "Estimating Load-Sharing Properties in a Dynamic Reliability System". *Journal of the American Statistical Association*:262–272.

Lachaud, B., and B. Ycart. 2006. "Convergence times for parallel Markov chains". *Lecture notes in control and information sciences* 341:169.

L'Ecuyer, P., and B. Tuffin. 2011. "Approximating zero-variance importance sampling in a reliability setting". *Annals of Operations Research* 189(1):277–297.

Mai, J.-F., and M. Scherer. 2009. "Lévy-frailty copulas". *Journal of Multivariate Analysis* 100(7):1567–1585.

Mai, J.-F., and M. Scherer. 2011. "Reparameterizing Marshall-Olkin copulas with applications to sampling". *Journal of Statistical Computation and Simulation* 81(1):59–78.

Mai, J.-F., and M. Scherer. 2017. *Simulating copulas: stochastic models, sampling algorithms, and applications*, Volume 6. World Scientific.

Matus, O., J. Barrera, E. Moreno, and G. Rubino. 2018. "On the Marshall-Olkin Copula Model for Network Reliability Under Dependent Failures". *IEEE Transactions on Reliability* 68(2):451–461.

Reiss, R.-D. 2012. *Approximate distributions of order statistics: with applications to nonparametric statistics*. Springer Science & Business Media.

Reiss, R.-D., and M. Thomas. 2007. *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields*. Springer Science & Business Media.

Samaniego, F. J. 2007. *System signatures and their applications in engineering reliability*, Volume 110. Springer Science & Business Media.

AUTHOR BIOGRAPHIES

JAVIERA BARRERA is an associate professor of the Faculty of Engineering and Sciences at Universidad Adolfo Ibáñez. Her current research interests include: modeling of stochastic systems, reliability modeling and analysis, and network design. She obtained her PhD in Mathematics from the Université Paris Descartes, France, and at the School of Engineering of Universidad de Chile in 2006, specializing in the area of Stochastic Models. Her research interests lie in the asymptotic behavior of stochastic models and, more recently, in stochastic optimization. Applications in network design, computer science, physics and finance motivate the models addressed in her research. Her email address is javiera.barrera@uai.cl.

GUIDO LAGOS is a FONDECYT postdoctoral researcher at the Faculty of Engineering and Sciences at Universidad Adolfo Ibáñez. He received his Ph.D. in Operations Research from Georgia Institute of Technology, a M.Sc. in Operations Management from Universidad de Chile, and a Mathematical Engineering title (a USA equivalent to a M.Sc. and B.S. in Applied Math) from Universidad de Chile. He has done research in reliability, applied probability, methodology of simulation, adaptive optimization and stochastic programming. His email address is guido.lagos.barrios@gmail.com.