

Document downloaded from:

<http://hdl.handle.net/10251/121418>

This paper must be cited as:

García Mora, MB.; Santamaria Navarro, C.; Rubio Navarro, G. (2018). Modeling dependence in the inter-failure times. An analysis in Reliability models by Markovian Arrival Processes. *Journal of Computational and Applied Mathematics*. 343:762-770.
<https://doi.org/10.1016/j.cam.2017.12.022>



The final publication is available at

<http://doi.org/10.1016/j.cam.2017.12.022>

Copyright Elsevier

Additional Information

Accepted Manuscript

Modeling dependence in the inter-failure times. An analysis in Reliability models by Markovian Arrival Processes

B. García-Mora, C. Santamaría, G. Rubio

PII: S0377-0427(17)30642-8
DOI: <https://doi.org/10.1016/j.cam.2017.12.022>
Reference: CAM 11439

To appear in: *Journal of Computational and Applied Mathematics*

Received date: 30 November 2016
Revised date: 17 October 2017

Please cite this article as: B. García-Mora, C. Santamaría, G. Rubio, Modeling dependence in the inter-failure times. An analysis in Reliability models by Markovian Arrival Processes, *Journal of Computational and Applied Mathematics* (2018), <https://doi.org/10.1016/j.cam.2017.12.022>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Modeling Dependence in the Inter-Failure Times. An Analysis in Reliability Models by Markovian Arrival Processes.

B. García-Mora^a, C. Santamaría^a, G. Rubio^a

^a*Instituto de Matemática Multidisciplinar. Universitat Politècnica de València*

Abstract

The most common assumptions in reliability studies are that *failures* occur independently and with the same distribution. However, these two assumptions are unrealistic in practice since *inter-failure* times are usually correlated and not identically distributed. In this sense the *Markov Arrival Process (MAP)* is an active research field for managing these features. We study two versions of the *MAP* approach. The first one is a model which we developed in previous works. The second one is based on a non-stationary MAP of second order. We compare the results of both models with simulated data.

Keywords: Non-stationary Markovian Arrival Process, Erlang distribution, Phase-Type distribution, Reliability, Inter-arrival times

1. Introduction

Multi-state stochastic processes are a convenient framework for modeling reliability problems in Engineering. In particular, Reliability and maintenance policies of systems considering failures of components have been studied by [1] and [2] among others. Within this framework Markov processes have been used but these models imply independence between the *inter-failure* times. However, in Engineering this assumption needs to be relaxed because frequently it is essential to take into account that the consecutive failures of a system show dependence between them [3]. Then another approach

Email addresses: magarmo5@imm.upv.es (B. García-Mora), crisanna@imm.upv.es (C. Santamaría), grubio@imm.upv.es (G. Rubio)

is required when the conditional independence assumption for waiting times does not hold. In this sense the *Markovian Arrival Process* model (MAP) has the relevant property of dependence between consecutive inter-arrival times in a process with multiple events, that constitutes an effective tool for modeling the dependence in a multi-state process. In the literature several analysis with *MAPs* are found in Reliability such as analysis of an n -system under shocks and repairs [4] and the study of failures of an electrical N -component framework [5] among others.

In this context, we study two methods of the *MAP* approach using simulated data. The first method was developed by the authors in previous works ([6], [7]). The second one is a recent model based on a non-stationary MAP of second order [5]. Our aim is to test our method comparing its performance versus some state of the art methodologies.

The paper is organized as follows: in section 2 we describe simulated data of devices undergoing three failures each one of them. In section 3 we introduce the *Markovian Arrival Process (MAP)* to handle the dependence between the inter-failure times of the system considered in this analysis. Section 4 deals with the authors' model. In section 5 we describe the *2-state non-stationary MAP* model. In Section 6 we fit the two models to our simulated data. Finally, in Section 7 the conclusions of the work are discussed.

2. The data: inter-failure times

In this study we simulate the operational random times corresponding to 100 devices in an engineering framework. In the interpretation of the data we assume that after each failure the device is repaired immediately and the repair time is negligible. The interest lies in the consecutive operational times of each device. We assume that each one of the 100 devices have three failures and so three inter-failure times represented by T_1 , T_2 and T_3 in the Figure 1. The Figure 2 shows the $N=100$ simulated sequences of such operational times, $t^{(1)}, t^{(2)}, \dots, t^{(100)}$, where

$$\begin{aligned} t^{(1)} &= (t_1^{(1)}, t_2^{(1)}, t_3^{(1)}) \\ t^{(2)} &= (t_1^{(2)}, t_2^{(2)}, t_3^{(2)}) \\ &\vdots \\ t^{(100)} &= (t_1^{(100)}, t_2^{(100)}, t_3^{(100)}) \end{aligned} \tag{1}$$

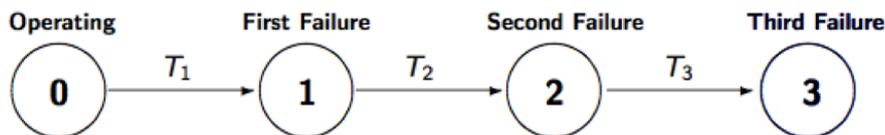


Figure 1: Device with three failures.

The aim is to estimate the performance of this system where the sequences of operational times, $t^{(1)}, t^{(2)}, \dots, t^{(100)}$, are independent among them.

Now let T_k be the random variable representing the operational time between the $(k-1)$ -th and the k -th failure (Figure 1). Then the three variables, T_1 , T_2 , and T_3 represent the three *inter-failure times* in columns. In this way according to the Figure 2, the three variables T_1 , T_2 and T_3 are

$$\begin{aligned} T_1 &= (t_1^{(1)}, t_1^{(2)}, \dots, t_1^{(100)}) \\ T_2 &= (t_2^{(1)}, t_2^{(2)}, \dots, t_2^{(100)}) \\ T_3 &= (t_3^{(1)}, t_3^{(2)}, \dots, t_3^{(100)}) \end{aligned} \quad (2)$$

T_1 , T_2 and T_3 are correlated and not identically distributed. We have considered three inter-failure times because there are enough arrivals to establish correlations between the inter-failure times. On the other hand there are studies with more operational inter-failure times as [5] among others.

In Section 6 we show a detailed explanation about the simulation procedures of this study.

3. The Markovian Arrival Process (*MAP*)

MAPs are a generalization of the *phase-type* distributions, so we start with a brief introduction to this kind of functions.

3.1. The Phase-type distributions (PHD)

PH-distributions generalize the exponential distribution and constitute a flexible class of probability models. They are a versatile class of distributions since a tractable point of view and have been used in many areas [4]. They were introduced by M. Neuts [8]. See [10] for more details about this topic.

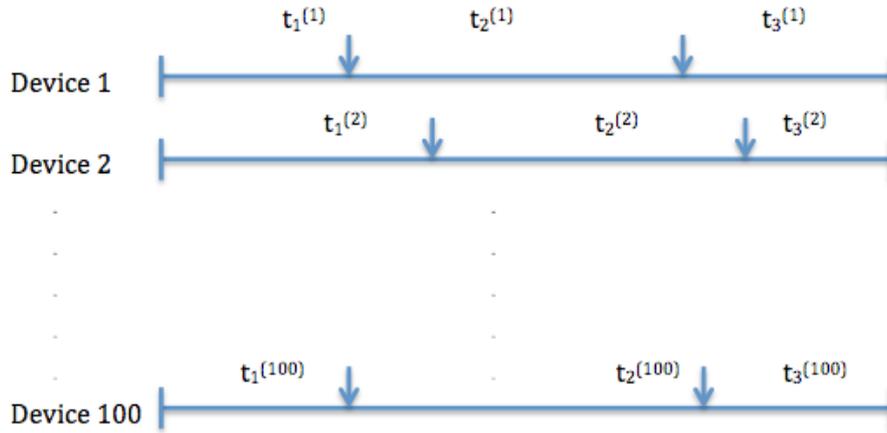


Figure 2: 100 Devices with three failures and their consecutive operational times.

Definition 1. The distribution $F(\cdot)$ on $[0, \infty[$ is a phase-type distribution (PH-distribution) with representation (α, T) if it is the distribution of the time until absorption in a Markov process on the states $\{1, \dots, m, m+1\}$ with generator

$$\begin{pmatrix} T & T^0 \\ 0 & 0 \end{pmatrix},$$

and initial probability vector (α, α_{m+1}) where α is a row m -vector. The matrix T is non-singular with negative diagonal entries and non-negative off-diagonal entries, and satisfies $-T\mathbf{e} = T^0 \geq 0$ where \mathbf{e} denotes a column vector with all components equal to one.

We need the following theorem [11] to build the first modeling approach of Section 4.

Theorem 1. If $F_1(\cdot)$ and $F_2(\cdot)$ are both continuous PH-distributions with representations (α, T) and (β, S) of orders m and n respectively, then their convolution $F(\cdot) = (F_1 * F_2)(\cdot)$ is a PH-distribution with representation (γ, L) , given by

$$\gamma = (\alpha, \alpha_{m+1}\beta) \quad \text{and} \quad L = \left[\begin{array}{c|c} T & T^0\beta \\ \hline 0 & S \end{array} \right] \quad (3)$$

3.2. Markovian Arrival Processes

Before setting up the formal definition, let us consider an approach that plays a fundamental role in the first model of this paper ([12], p. 186). The starting point is a PH renewal process (a renewal process with phase-type distributed renewal intervals). Let $PH(\alpha, T)$ be the phase-type distribution. Then the infinitesimal generator of the PH renewal process is given by

$$G = \begin{pmatrix} T & A & & \\ & T & A & \\ & & \ddots & \ddots \\ & & & \ddots & \ddots \end{pmatrix}$$

where $A := \eta\alpha$ with $\eta := -T\mathbf{e}$.

Let us write A explicitly:

$$A = \begin{pmatrix} \eta_1\alpha \\ \vdots \\ \eta_m\alpha \end{pmatrix}.$$

Now we relax the restriction that α be a constant vector, and consider a new matrix

$$A' = \begin{pmatrix} \eta_1\alpha_1 \\ \vdots \\ \eta_m\alpha_m \end{pmatrix} \quad (4)$$

with the requirement that $\alpha_i\mathbf{e} = 1$, for $i = 1, \dots, m$. If we denote $D_0 := T$ and $D_1 := A'$, the previous matrix G will be

$$Q = \begin{pmatrix} D_0 & D_1 & & \\ & D_0 & D_1 & \\ & & \ddots & \ddots \\ & & & \ddots & \ddots \end{pmatrix}.$$

A Markov process with such a generator is called a m -state Markovian Arrival Process (MAP).

A MAP_m is a matrix generalization of the non-homogeneous Poisson process with dependence between inter-failure times and non-exponential inter-failures distribution. They preserve an underlying Markovian structure with a finite state space $S = 1, 2, \dots, m$, and an initial vector π . The generator matrix Q which can be represented as $Q = D_0 + D_1$ where,

- $D_0 \geq 0$, $D_1 \neq 0$. D_1 is a non-negative $m \times m$ matrix.

- $D_0(i, j) \geq 0$ for $i \neq j$. D_0 has negative diagonal elements and non-negative off-diagonal elements.
- (π, D_0) is a PHD with π the initial vector probability.

The representation of a MAP_m is (π, D_0, D_1) and its joint density function generating k consecutive events with inter-event times x_i is given by the following expression

$$f(x_1, x_2, \dots, x_k) = \pi e^{D_0 x_1} D_1 e^{D_0 x_2} D_1 \dots e^{D_0 x_k} D_1 \mathbf{e} \quad (5)$$

for $x_1, \dots, x_k \geq 0$. $e^{D_0 x_j}$ is the exponential of $D_0 x_j$, where D_0 is a matrix and x_j is a number.

If (π, D_0) is a PHD, then

$$(\pi, D_0, d_1 \pi), \quad (6)$$

where $d_1 = -D_0 \mathbf{e}$, is a MAP with the same behavior [10]. This means that events are generated with independently and identically distributed inter-event times and distribution like the PHD (π, D_0) .

4. The first approach

Our aim is to perform a MAP model with representation (π, D_0, D_1) for modeling the process of this study $0 \rightarrow 1 \rightarrow 2 \rightarrow 3$ of three inter-failures times for the 100 independent devices (Figure 1).

Variables T_1 , T_2 and T_3 given in (2) represent the three dependent inter-arrival times between the states of each transition (see Figure 1). It is assumed that the time spent until the second failure (transition $1 \rightarrow 2$) depends on the spent time in the first transition ($0 \rightarrow 1$). That is to say the variable T_2 depends on T_1 , and in the same way for the variable T_3 respect to T_2 .

Expression (6) suggests the following approach. Let us begin building a PHD for the whole process as if waiting times of each transition were independent. Distributions of waiting times in each transition are modelled by means of a PHD made of a mixture of *Erlang* distributions [13]. The procedure is described in [14] and [15]. Then we make up the convolution of PHDs corresponding to the transitions $0 \rightarrow 1$ and $1 \rightarrow 2$, using Theorem 1. The obtained PHD is convoluted with the PHD of the third transition ($2 \rightarrow 3$) using again Theorem 1. The resulting PHD has an initial probability vector and a matrix D_0 .

Now the idea is to take into account dependence constructing matrix D_1 according to expression (4). The objective is to compute vectors α_i by maximum likelihood based on the proper joint density function for three consecutive events with three inter-failure times T_1, T_2 and T_3 that is

$$f(t_1^{(i)}, t_2^{(i)}, t_3^{(i)}) = \pi e^{D_0 t_1^{(i)}} D_1 e^{D_0 t_2^{(i)}} D_1 e^{D_0 t_3^{(i)}} D_1 \mathbf{e}, \quad i = 1, 2, \dots, 100. \quad (7)$$

Further details will be given in the numerical example of Section 6.

5. The 2-state non-stationary Markovian Arrival Process (MAP_2)

In order to compare the *MAP model* developed in the previous section with the state of the art we choose the most recently model using the *MAP* methodology [5]. We apply to our data the 2-state non-stationary Markovian Arrival Process, in its canonical representation, denoted by MAP_2 .

The MAP_2 is a doubly stochastic process $\{J(t), N(t)\}$ where

- $J(t)$ represents an irreducible and continuous Markov process with two states in the state space $S = \{1, 2\}$.
- $N(t)$ is the counting process and represents the number of failures in the interval $(0, t]$.
- The initial state $i_0 \in S$ is generated according to an initial probability $\boldsymbol{\alpha} = (\alpha, 1 - \alpha)$ where $\alpha \in (0, 1)$.

The operation of the MAP_2 is as follows: at the end of a sojourn time in state i with exponential distribution, with arrival rate parameter λ_i and mean $\frac{1}{\lambda_i}$, two types of transitions can occur (see Figure 3):

- No failure occurs with probability $0 \leq p_{ij0} \leq 1$, $j \in S$, and $J(t)$ jumps from state i to state j (different states, $i \rightarrow j$).
- A failure occurs with probability $0 \leq p_{ij1} \leq 1$, $j \in S$, and $J(t)$ jumps from state i to state j (the same or different states, $i \rightarrow i$ or $i \rightarrow j$).

The non-stationary MAP_2 is characterized by $M = \{\alpha, \lambda, P_0, P_1\}$ where $\lambda = (\lambda_1, \lambda_2)$ and

$$P_0 = \begin{pmatrix} 0 & p_{120} \\ p_{210} & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} p_{111} & p_{121} \\ p_{211} & p_{221} \end{pmatrix} \quad (8)$$

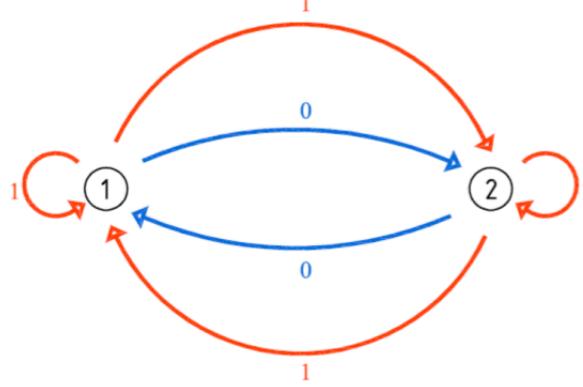


Figure 3: Transitions in the non-stationary MAP_2 .

On the other hand, the MAP_2 can be also characterized by $M = \{\boldsymbol{\alpha}, D_0, D_1\}$ where D_0 and D_1 are rate matrices corresponding to the transitions. They are squared of order 2. In this way $D = D_0 + D_1$ is the infinitesimal generator in the Markov process $J(t)$. As X_k represents the state of the $J(t)$ at the time of the k -th failure, $\{X_k\}_{k=1}^{\infty}$ is a Markov chain with the matrix P^* representing the transitions probabilities among states. This matrix is calculated as $P^* = (-D_0)^{-1}D_1$.

The representation of the MAP_2 in its canonical representation [16] is given by

- If $\gamma > 0$, with γ the eigenvalue of the stochastic matrix P^* different from 1, the representation is

$$\boldsymbol{\alpha} = (\alpha, 1 - \alpha), \quad D_0 = \begin{pmatrix} x & y \\ 0 & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} -x - y & 0 \\ v & -u - v \end{pmatrix} \quad (9)$$

- If, on the other hand, $\gamma \leq 0$, then the canonical representation is

$$\boldsymbol{\alpha} = (\alpha, 1 - \alpha), \quad D_0 = \begin{pmatrix} x & y \\ 0 & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & -x - y \\ -u - v & -v \end{pmatrix} \quad (10)$$

where $u \leq x < 0$, $y, v \geq 0$, $x + y \leq 0$ and $u + v \leq 0$. The relation among the parameters of the equations (8), (9) and (10) is

- $x = -\lambda_1, \quad y = \lambda_1 p_{120}, \quad w = \lambda_1 p_{111}$

- $z = \lambda_2 p_{210}, \quad u = -\lambda_2, \quad v = \lambda_2 p_{211}$

Now the main objective is to estimate the five model parameters, α , x , y , u and v in the MAP_2 .

5.1. Estimation of the parameters in the MAP_2

The time T_k , between the $(k-1)$ -th and k -th failures, has a *phase-type* distribution with representation $(\boldsymbol{\alpha}_k, D_0)$ with $\boldsymbol{\alpha}_k = \boldsymbol{\alpha}(P^*)^{k-1}$ the probabilities vector defining the state of the Markov chain $\{X_k\}_{k=1}^{\infty}$ at the time of the k -th failure. This implies that the inter-failure times have not the same distribution. In this sense the cumulative distribution function (*CDF*) for each T_k is defined by

$$F_{T_k}(t) = 1 - \boldsymbol{\alpha}_k e^{D_0 t} \mathbf{e} \quad (11)$$

Then in our study $T_k \sim PH\{\boldsymbol{\alpha}_k, D_0\}$, with $k = 1, 2, 3$, represent different *phase-type* distributions for the correlated variables T_1, T_2 and T_3 .

Now we focus on the estimation of the five parameters of the non-stationary MAP_2 , $\{\alpha, x, y, u, v\}$, for the rate matrices $\{D_0, D_1\}$ given in the expressions (9) and (10) from the set of operational times $t^{(1)}, t^{(2)}, \dots, t^{(100)}$ in (1).

As the non-stationary MAP_2 is fully characterized by the set of moments $\{\mu_{1,1}, \mu_{1,2}, \mu_{1,3}, \mu_{2,1}, \mu_{3,1}\}$ [16] where $\mu_{k,m} = \mu_{k,m}(\alpha, x, y, u, v)$ are the moments of the variable T_k , that can be computed as follows

$$\mu_{k,m} = E(T_k^m) = m! \boldsymbol{\alpha}_k (D_0)^{-m} \mathbf{e} \quad (12)$$

Then, the parameters are estimated by means of the *moment matching estimation approach* [17]. This consists of matching the population moments, $\mu_{k,m}$, with the empirical counterparts

$$\bar{\mu}_{k,m} = \frac{1}{N} \sum_{i=1}^N (t_k^{(i)})^m \quad (13)$$

This leads to the following nonlinear system of equations

$$\begin{aligned} \mu_{1,m}(\alpha, x, y, u, v) &= \bar{\mu}_{1,m}, & m &= 1, 2, 3 \\ \mu_{k,1}(\alpha, x, y, u, v) &= \bar{\mu}_{k,1}, & k &= 2, 3 \end{aligned} \quad (14)$$

On the other hand the system of equations (14) have not a feasible solution and, in order to obtain the estimation, we use the proposal given in [17]

which consists of solving the following the *optimization problem* (P) instead solving the system (14).

$$\begin{aligned}
\min \quad & \gamma_\tau(\alpha, x, y, u, v) \\
\text{s.t.} \quad & x, u \leq 0 \\
& y, v \leq 0 \\
& -x - y \geq 0 \\
& -u - v \geq 0 \\
& 0 \leq \alpha \leq 1
\end{aligned} \tag{15}$$

where γ_τ is the following *objective function*

$$\begin{aligned}
\gamma_\tau(\alpha, x, y, u, v) = \tau \left\{ \left(\frac{\mu_{1,1} - \bar{\mu}_{1,1}}{\bar{\mu}_{1,1}} \right)^2 + \left(\frac{\mu_{1,2} - \bar{\mu}_{1,2}}{\bar{\mu}_{1,2}} \right)^2 + \left(\frac{\mu_{1,3} - \bar{\mu}_{1,3}}{\bar{\mu}_{1,3}} \right)^2 \right. \\
\left. + \left(\frac{\mu_{2,1} - \bar{\mu}_{2,1}}{\bar{\mu}_{2,1}} \right)^2 + \left(\frac{\mu_{3,1} - \bar{\mu}_{3,1}}{\bar{\mu}_{3,1}} \right)^2 \right\},
\end{aligned} \tag{16}$$

with $\tau > 0$ a penalty parameter that needs to be tuned, but setting $\tau = 1$ performs well in practice [17]. Then $\{\hat{\alpha}, \hat{x}, \hat{y}, \hat{u}, \hat{v}\}$ will be the solution of (14) if and only if it is an optimal solution in (P).

Notice the *optimization problem* (P) must be solved twice, one per each canonical representation, (9) and (10), from the sample (1). In a second step we select the estimated parameters with the highest value of the log-likelihood function, whose expression for the sample is

$$\log f(t^{(1)}, t^{(2)}, \dots, t^{(N)} | D_0, D_1) = \sum_{i=1}^N \log f(t^{(i)} | D_0, D_1) \tag{17}$$

where

$$f(t_1^{(i)}, t_2^{(i)}, t_3^{(i)}) = \pi e^{D_0 t_1^{(i)}} D_1 e^{D_0 t_2^{(i)}} D_1 e^{D_0 t_3^{(i)}} D_1 \mathbf{e}, \quad i = 1, 2, \dots, 100 \tag{18}$$

Finally with the estimated parameters in the matrices D_0 and D_1 we can obtain a quantity of interest concerning the counting process $N(t)$, the probability of n failures in the interval $(0, t]$. For it, $P(n, t) = \{P_{ij}(n, t)\}_{n \in N, t \geq 0}$ represent the set of 2×2 matrices whose (i, j) -th element is the following expression

$$P_{ij}(n, t) = P(N(t) = n, J(t) = j | N(0) = 0, J(0) = i) \tag{19}$$

for $1 \leq i, j \leq 2$. Then, as the matrix $P(n, t)$ represents the probability of n failures in the interval $(0, t]$, so

$$P(N(t) = n | N(0) = 0) = \alpha P(n, t) \mathbf{e} \quad (20)$$

The computation of the matrices $P(n, t)$ is based on the uniformization method developed by [18]. In this algorithm the matrices $P(n, t)$ are written as

$$P(n, t) = \sum_{r=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^r}{r!} V(n, r), \quad (21)$$

where λ is a constant, such that $\lambda \geq \max_j \{-(D_0)_{jj}\}$ and $V(n, r)$ are 2×2 matrices computed as follows

1. Define $C0$ and $C1$ matrices as $C0 = I + \frac{D_0}{\lambda}$ and $C1 = I + \frac{D_1}{\lambda}$
2. Find the smallest index for which $\sum_{n=N+1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^r}{r!} \leq \epsilon$ with ϵ is a fixed tolerance parameter.
3. For $n = 0$, $V(0, 0) = I$, $P(0, t) \leftarrow V(0, 0)e^{-\lambda t}$
4. For $n \geq 1$, $V(n, 0) \leftarrow 0$, $P(n, t) \leftarrow 0$
5. For $1 \leq k \leq N$ and $0 \leq i \leq k$,

$$V(i, k) = \sum_{j=\max(0, i-(k-1))}^{\min(i, 1)} V(i-j, k-1)C_j \quad (22)$$

$$P(i, t) \leftarrow P(i, t) + V(i, k)e^{-\lambda t} \frac{(\lambda t)^r}{r!} \quad (23)$$

6. Application with a simulated data set

In this section we show the estimation of the parameters with both versions of MAP approach. Data are the operational times of the 100 devices described in section 2. In order to obtain the correlated transition times, we have used *Copulas* functions that provide a correlation structure between variables [9]. We have simulated dependent multivariate uniform data using a "Gaussian" and "t" copula functions (elliptical copulas) with a specified rank correlation among variables (Spearman's). In a second step the correlated interarrival times are obtained with a Weibull distribution by means of the Inversion method. The Weibull distributions are the marginal distributions obtained for each variable (transition time) and it is very used in Reliability.

Variation Coefficient

Copula, k, λ	Transition 0 \rightarrow 1			Transition 1 \rightarrow 2			Transition 2 \rightarrow 3		
	α	μ_1	μ_2	α	μ_1	μ_2	α	μ_1	μ_2
"Gaussian", 1, 1	0.191	0.424	2.378	0.299	0.779	1.231	0.441	1.603	0.687
"Gaussian", 2, 1	0.240	0.635	0.150	0.346	0.484	0.136	0.332	0.344	0.715
"t", 1, 1	0.274	0.898	2.138	0.271	1.340	1.138	0.309	0.900	0.148
"t", 2, 1	0.341	0.819	0.129	0.264	0.451	0.132	0.331	0.559	0.625

Table 1: Variation coefficient of the estimated parameters of the mixture of the two Erlangs in each transition of the inter-arrival times. The interarrival-times variables are random Weibull variables generated by the inverse method. The parameters scale and shape are $k = 1, 2$ and $\lambda = 1$

We have generated series of simulated data (size $n = 100$ devices in each group of simulated data) for the three inter-arrival times as follows: we calculate the rank correlation coefficient of Spearman's from a initial specified correlation between variables. In a second step we use the copula function (Gaussian or t) to construct a new multivariate distribution for these three dependent variables. These three variables have a uniform distribution and finally we transform each of the variables in a Weibull distribution by means of the Inversion method. In these simulations we have used different scale and shape parameters for the Weibull distribution and the range of the specified initial correlation was $[0.1, 0.3]$.

Previously we have generated four different groups of simulated data from a Gaussian or t copula function and with different parameters of a Weibull distribution (Table 1). In each group (size $n = 10$) we have calculated the variation coefficient of the estimated parameters of the mixture of the two Erlangs in each transition of the inter-arrival times. The parameters scale and shape of the Weibull distribution are $k = 1, 2$ and $\lambda = 1$. Since the Gaussian copula function ($k = 2, \lambda = 1$) show small variation coefficients we generate three correlated inter-arrival time from this copula function.

Then In the modelling we use two simulations of correlated inter-arrival times: in the first we simulate these data from the *2-state non-stationary MAP* [10] and the second simulation is obtained from three Weibull distributed variables by the Gaussian copula function described above.

Then with these simulated data the aim of both modeling approaches is to compute the probability vector π and the matrices D_0 and D_1 . Then

any quantity of interest can be obtained. In order to compare the analyzed approaches we computed the estimated and empirical distributions in each model.

As stated above, we consider a system of 100 devices, each one of them with three failures where the state 0 represents the starting point of the process (the device is operating). State 1 is the first failure in the system (the device fails), the state 2 represents the second failure and the state 3 represents the third failure (the device is off).

6.1. Estimation of the first approach

PHDs of each transition are mixtures of two Erlangs (Theorem 9.11, [12]), one of dimension 1×1 and another of 2×2 . So the dimension of D_0 is 9×9 . The probability vector π and D_1 are computed by maximum likelihood using the local search MATLAB's routine **fmincon** (Optimization toolbox). The representation of the estimated Cumulative Distribution Functions of the *MAP* model versus the empirical interarrival times T_1, T_2, T_3 is shown in the Figure 4.

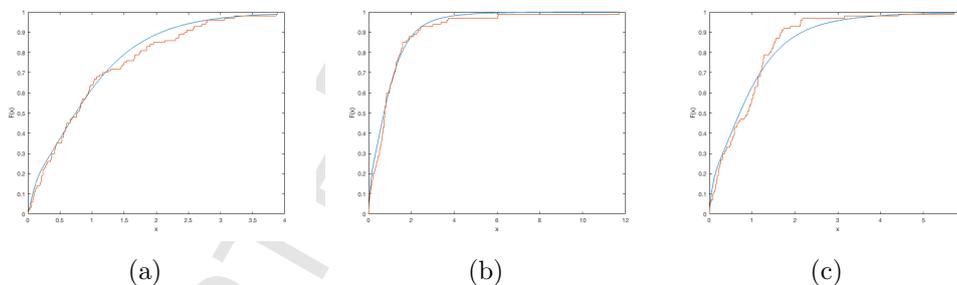


Figure 4: Estimated CDF (*smooth line*) under the *MAP* and the empirical distributions (*step function*) for the correlated inter-failure Weibull distributed times generated from the Gaussian copula function: (a) T_1 (time until the first failure), (b) T_2 (inter-failure time between the first and second failure) and (c) T_3 (inter-failure time between the second and the third failure).

The fits of the estimated distributions and the consecutive inter-failure times have been validated with the one-sample Kolmogorov-Smirnov test using the MATLAB function **kstest**.

6.2. Estimation of the non-stationary MAP_2

We fit the non-stationary MAP_2 to the same sample of operational times (1) of the 100 devices. As we have to solve the *optimization problem* (P) we

use also the MATLAB's routine **fmincon**. A multistart approach (200 different starting points randomly selected of the simulated data) is performed and we keep the solution with the minimum objective function $\gamma_\tau(\hat{\alpha}, \hat{x}, \hat{y}, \hat{u}, \hat{v})$ in (P).

We solve (P) for two canonical representations of the MAP_2 . The parameter estimations under the two canonical representations of the model are

- First canonical estimated representation:

$$\hat{\alpha}^1 = (0.8425, 0.1575), \hat{D}_0^1 = \begin{pmatrix} -1.5605 & 1.1308 \\ 0 & -1.5606 \end{pmatrix}, \hat{D}_1^1 = \begin{pmatrix} 0.4297 & 0 \\ 1.1308 & 0.4297 \end{pmatrix}$$

- Second canonical estimated representation:

$$\hat{\alpha}^2 = (0.5989, 0.4011), \hat{D}_0^2 = \begin{pmatrix} -1.5543 & 1.5543 \\ 0 & -1.5543 \end{pmatrix}, \hat{D}_1^2 = \begin{pmatrix} 0 & 0.0000 \\ 0.8863 & 0.6680 \end{pmatrix}$$

Finally we compute the **log-likelihoods** for the two canonical estimated representation:

- $\log f(t^{(1)}, t^{(2)}, \dots, t^{(100)} | \tilde{D}_0^1, \tilde{D}_1^1) = 0.0201$
- $\log f(t^{(1)}, t^{(2)}, \dots, t^{(100)} | \tilde{D}_0^2, \tilde{D}_1^2) = 0.0204$

which provides evidence in favor of the estimation given by the second canonical estimated representation.

The representation of the estimated Cumulative Distribution Functions of the non-stationary MAP_2 versus the empirical interarrival times T_1, T_2, T_3 is shown in the Figure 5.

The fits of the estimated distributions and empirical inter-failure times have been validated with the one-sample Kolmogorov-Smirnov test using the MATLAB function **kstest**.

Now we can obtain quantities of interest concerning the counting process $N(t)$. In this sense, Figure 6 shows the probabilities of 1, 2 and 3 failures over time. Probability of having 0 failures is high at the beginning but it decreases slowly over time. On the other hand, the probability of having 1, 2 and 3 failures increases very slowly over time. In this analysis we have considered 20 units of time.

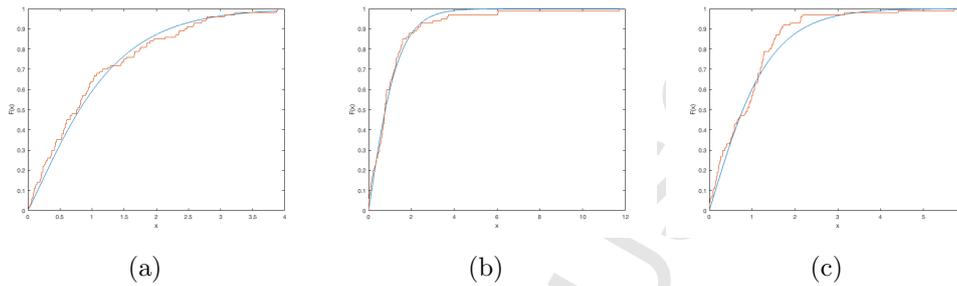


Figure 5: Estimated CDF (*smooth line*) under the MAP_2 and the empirical distributions (*step function*) for the correlated inter-failure Weibull distributed times generated from the Gaussian copula function: (a) T_1 (time until the first failure), (b) T_2 (inter-failure time between the first and second failure) and (c) T_3 (inter-failure time between the second and the third failure).

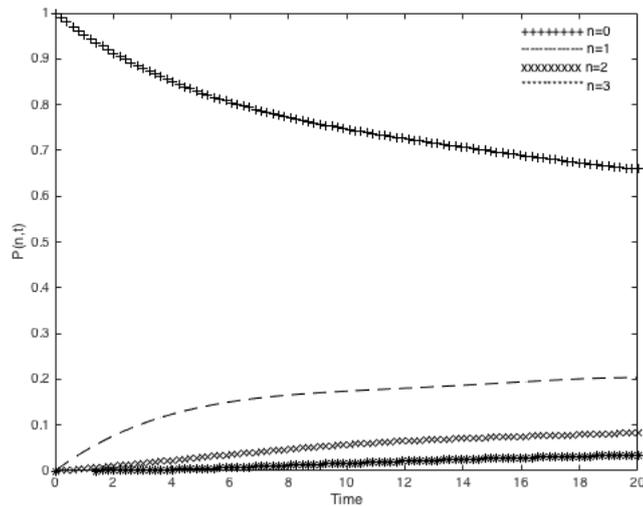


Figure 6: Probabilities $P(n, t)$ for $n = 1, 2, 3$ and $t > 0$

7. Discussion

The hypotheses of considering i.i.d. inter-failure times are not realistic in many situations. It is essential to take into consideration that failures of a system generate dependence among them. For this purpose we have introduced the *Markovian Arrival Processes* to incorporate this underlying dependency. Two models have been compared, in which maximum likelihood methodology is used differently.

Both models fit well to our simulated data. The question we raise is whether both approaches will be adequate to deal with covariates and censored data, two usual features in Reliability. The first model is a generalization of [6] and [19], and their *covariates* and *censored data* handling is also applicable to the MAP case. However, the dimension of the problem increases with the number of events [6]. In fact, the work with a lot of events can be complicated because of the dimension of the matrix Q increase, although we can introduce lumpable MAP to reduce the dimension [10]. On the other hand MAP_2 model works perfectly [5] with many events, but until the date there is any work with *covariates* or *censored data*. In fact, to the best of our knowledge the inclusion of censored data and covariates in the context of Reliability models with MAPS is still a pending task.

We have illustrated our approach in the context of Engineering. We may apply this methodology in other situations with recurrent events, such as chronic diseases in medicine where multiple recurrences are registered. In this context successive recurrences of the same patient are not actually independent events and data are usually censored.

References

- [1] D. Montoro-Cazorla, R. Pérez-Ocón, A shock and wear system with memory of the phase of failure, *Mathematical and Computer Modelling* 54(9-10) (2011) 2155–2164.
- [2] M. F. Neuts, M. C. Bhattacharjee, Shock models with phase type survival and shock resistance, *Naval Res Logist* 28 (1981) 213–219.
- [3] L. Fiondella, S. Rajasekaran, SS. Gokhale (2011). Efficient System Reliability with Correlated Component Failures. *IEEE 13th International Symposium on High-Assurance Systems Engineering (HASE)*, 269-276.

- [4] D. Montoro-Cazorla, R. Pérez-Ocón, A redundant n -system under shocks and repairs following markovian arrival processes, *Mathematical and Computer Modelling* 134 (2014) 69–75.
- [5] J. Rodríguez, R. E. Lillo, P. Ramírez-Cobo, Failure modeling of an electrical N -component framework by the non-stationary markovian arrival process, *Reliability Engineering and System Safety* 134 (2015) 126–133.
- [6] C. Santamaría, B. García-Mora, G. Rubio, R. Pérez-Ocón, Managing dependence in flowgraphs models. an application to reliability engineering, in: *Modelling for Engineering & Human Behaviour 2015*, Instituto Universitario de Matemática Multidisciplinar, 2015.
- [7] G. Rubio, M. B. García-Mora, C. Santamaría, J. L. Pontones, Incorporating multiple recurrences in a flowgraph model for bladder carcinoma, in: *International Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2015*.
- [8] M. F. Neuts, Probability distributions of phase type. In: R. Holvoet and H. Florin (eds.) *Liber Amicorum Prof. Emeritus H. Florin*, University of Louvain. Belgium, pp 173-206, 1975.
- [9] R.B. Nelsen (1999), *An Introduction to Copulas*, New York: Springer, ISBN 0-387-98623-5.
- [10] P. Buchholz, J. Kriege, I. Felko, *Input Modeling with Phase-Type Distributions and Markov Models. Theory and Applications*, Springer Cham Heidelberg New York Dordrecht London, 2014.
- [11] M. F. Neuts, *Matrix Geometric Solutions in Stochastic Models. An Algorithmic Approach*, The Johns Hopkins University Press, Baltimore, 1981.
- [12] L. Breuer, D. Baum, *An Introduction to Queueing Theory and Matrix-Analytic Methods*, Springer, Dordrecht, 2005.
- [13] R. Pérez-Ocón, M. C. Segovia, Modeling lifetimes using phase-type distributions, *Risk, Reliability and Societal Safety*, *Proceedings of the European Safety and Reliability Conference 2007 (ESREL 2007)*, pp. 1–5.

- [14] G. Rubio, M. B. García-Mora, C. Santamaría, J. L. Pontones, A flow-graph model for bladder carcinoma, *Theoretical Biology and Medical Modelling* (2014), 11(Suppl 1):S3.
- [15] M. B. García-Mora, C. Santamaría, G. Rubio, J. L. Pontones, Bayesian prediction for flowgraph models with covariates. An application to bladder carcinoma, *Journal of Computational and Applied Mathematics* 291 (2016) 85–93.
- [16] J. Rodríguez, R. E. Lillo, P. Ramírez-Cobo, Analytical issues regarding the lack of identifiability of the non-stationary MAP2, *Performance Evaluation*, 102 (2016) 1-20.
- [17] Carrizosa E., Ramírez-Cobo P. Maximum likelihood estimation for the two-state Markovian arrival. arXiv: 1401.3105v1.
- [18] M. F. Neuts, J. M. Li, *Matrix-Analytic Methods in Stochastic Models*, Chakravarthy SR, Alfa AS, editors, 1996, Ch. An Algorithm for the $P(n,t)$ Matrices of a Continuous BMAP, pp. 7–19, nY: Marcel Dekker.
- [19] G. Rubio, M. B. García-Mora, C. Santamaría, J. L. Pontones, Modeling dependence in multistate processes, *International Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2016*.