Competing Risks Modeling by Extended Phase-type Semi-Markov Distributions

Brenda Garcia-Maya[1], *Nikolaos Limnios*[1] *Bo Henry Lindqvist*[2]

Abstract

We present competing risks models within a semi-Markov process framework via the semi-Markov phase-type distribution. We consider semi-Markov processes in continuous and discrete time with a finite number of transient states and a finite number of absorbing states. Each absorbing state represents a failure mode (in system reliability) or a cause of death of an individual (in survival analysis). This is an extension of the continuous-time Markov competing risks model presented in Lindqvist and Kjølen [2018]. We derive the joint distribution of the lifetime and the failure cause via the transition function of semi-Markov processes in continuous and discrete-time. Some examples are given for illustration.

Keywords: competing risks, semi-Markov process, extended semi-Markov Ph-distributions, survival analysis.

1 Introduction

In competing risks there are two random variables of interest, the time to failure T, and the cause of failure C, see, e.g., Crowder [2001], Aalen [1995], Lindqvist and Kjølen [2018]. For instance, we can consider that a person could die of different causes, lung cancer, heart attack, HIV, etc. If we are observing both the time to death and the cause of death, the model therefore has to include more than one absorbing state (failure state), see e.g., Crowder [2001, 2012]. Thus, if the interest is focused on a specific cause of failure in presence of different causes, we are in the case of a competing risks model. In engineering, competing risks refer to the lifetime of a machine and its cause of breakdown.

[1]Sorbonne Universités, Université de Technologie de Compiègne,

Compiègne, France.

(brgarcia, nikolaos.limnios)@utc.fr

[2]Norwegian University of Sciences and Technology, Trondheim, Norway

bo.lindvist@ntnu.no

For instance, if we consider a car, it can fails due to electrical problems, dead battery, malfunctioning sensors, etc. The idea of competing risks is to model a process where the system is exposed to several causes of failure and its eventual failure is attributed exactly to only one of them.

A natural extension of (Markov) phase-type distributions (Ph-distributions), see, e.g., Neuts [1981], Aalen [1995], Asmussen and O'Cinneide [2006], is the semi-Markov Ph-distribution in continuous or discrete-time. See, e.g., Limnios [2012], where the Ph-distribution is defined in semi-Markov processes for both continuous and discrete time. The aim here is to extend this to competing risks models (see, e.g., Crowder [2001], Beyersmann et al. [2011], Crowder [2012]). In particular we generate the Lindqvist and Kjømodel in the semi-Markov case.

Although Markov processes are able to model properly, and in a straightforward manner, different situations, they have some limitations. For instance, the Markov assumption imposes restrictions on the distribution of the sojourn time in a state, which is geometrically distributed in the case of a discrete time chain and exponentially distributed in a continuous time process. This is the main drawback when applying the Markov processes in real problems. By contrast, semi-Markov processes are generalizations of Markov processes. They relax the hypothesis of the sojourn time distribution in a state. In semi-Markov processes the sojourn time in a state can follow any distribution, see, e.g., Limnios and Oprişan [2001]. Our interest in semi-Markov processes comes from the fact that in many situations in modeling complex systems, the distribution of the holding time in some states of the system can be different from the exponential (or geometric in discrete-time) distributions. For example, in mechanical systems we have mostly Weibull distributed sojourn times and in some maintenance operations we have log-normally distributed times or even constant duration of some particular maintenance operations, etc.

The article is organized as follows: in Section 2, we introduce semi-Markov processes and extended ph-type distributions. In Section 3, we present semi-Markov processes and competing risks. In Section 4, we present the discrete-time competing risks by semi-

Markov chains. Finally, in Section 5, we give some concluding remarks.

2 Semi-Markov processes and extended ph-type distributions

Consider a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, on which we define all processes and random variables. Let $\mathbb{N} := \{0, 1, 2, ..\}$ be the set of natural numbers, $\mathbb{N}^* := \{1, 2, ...\}$ and $\mathbb{R}_+ = [0, \infty)$ the nonnegative real numbers. Let us consider a semi-Markov continuous process $Z = (Z_t, t \in \mathbb{R}_+)$ with state space $E = \{1, 2, ..., r+1\}$, where states $E_0 := \{1, 2, ..., r\}$ are the transient states and state $\{r+1\}$ is absorbing. Consider the jump times of Z, say $0 = S_0 < S_1 < \cdots < S_n < \cdots$. Let us consider the chain $(J_n)_{n\in\mathbb{N}}$ which records Z at the points (S_n), i.e., $J_n = Z_{S_n}, n \geq 0$. Notice that $(J_n, S_n), n \geq 0$, is the (embedded) Markov Renewal Process (MRP) of Z. Let i, j be two elements of E. Then the semi-Markov kernel $Q(t)$ is defined as follows,

$$Q_{ij}(t) := \mathbb{P}(J_{n+1} = j, S_{n+1} - S_n \leq t \mid J_n = i), \;\; n \geq 0, t \in \mathbb{R}_+. \tag{1}$$

It is worth noticing that the semi-Markov kernel considered here is independent of n, which means that the semi-Markov sequence is homogeneous in time, and that we use index $t \in \mathbb{R}_+$ for the calendar times, and index $n \in \mathbb{N}$ for the number of jumps of (Z_t). The stochastic process (J_n) is the embedded Markov chain of the MRP (J_n, S_n).
Let α be the initial distribution of the semi-Markov process Z, i.e., $\alpha(i) := \mathbb{P}(Z_0 = i) = \mathbb{P}(J_0 = i)$, $i \in E$. Define the transition function of the semi-Markov process by $P_t(i,j) := \mathbb{P}(Z_t = j \mid Z_0 = i)$, for $i \in E_0$, $j \in E$. Of course, we have $P_t(r+1, j) = 0$, $j \in E_0$ and $P_t(r+1, r+1) = 1$, for $t \geq 0$.

Consider now the partition of the semi-Markov kernel and of the initial law, following sets E_0 and $\{r+1\}$, as follows:

$$Q(t) = \begin{bmatrix} Q_0(t) & L(t) \\ 0_{1\times r} & 0 \end{bmatrix} \tag{2}$$

and $\alpha = (\alpha_0, 0)$, where α_0 is the sub-vector corresponding to transient states E_0. The matrix $Q_0(t)$ is the restriction of the semi-Markov kernel over the transient states $E_0 \times E_0$, an $r \times r$ matrix function, and $L(t)$ is an $r \times 1$ column vector function.

Consider also the matrix

$$\bar{H} := \begin{bmatrix} \bar{H}_0(t) & 0 \\ 0 & \bar{H}_1(t) \end{bmatrix} \tag{3}$$

where $\overline{H}_0(t) := diag(\overline{H}_i(t), i = 1, ..., r)$ is the restriction of the sojourn times survival functions on the transient states.

Let us consider a real-valued measurable function $\varphi : E \times \mathbb{R}_+ \longrightarrow \mathbb{R}$ and define its convolution by $Q(t)$ as follows

$$Q * \varphi(i,t) = \sum_{j \in E} \int_0^t Q_{ij}(ds) \varphi_j(t-s), \tag{4}$$

see Limnios [2012].

Let us define the absorption time T by

$$T = \inf\{t \geq 0 : Z_t = r+1\}$$

The closed form solution of a semi-Markov phase-type distribution, say F on $[0, \infty)$, where $F(t)$ is the distribution of T, is (see, e.g., Limnios and Oprişan [2001], Limnios [2012]),

$$\overline{F}(t) := 1 - F(t) = \alpha_0 (I - Q_0(t))^{(-1)} * \bar{H}_0(t) \tag{5}$$

where I is the identity matrix for $t \geq 0$, and the zero matrix for $t < 0$, and

$$\psi(t) := (I - Q_0(t))^{(-1)} = \sum_{n \geq 0} Q_0^{(n)}(t) \tag{6}$$

where $Q_0^{(n)}$ is the n-fold convolution of Q_0 (see, e.g., Limnios and Oprişan [2001]), i.e.,

$$Q_{0}{}_{ij}^{(n)}(t) = \begin{cases} \delta_{ij} \mathbf{1}_{\{t \geq 0\}} & n = 0 \\ Q_{0ij}(t) & n = 1 \\ \sum_{k \in E} \int_0^t Q_{0ik}(ds) Q_{0}{}_{kj}^{(n-1)}(t-s) & n \geq 2. \end{cases} \tag{7}$$

For the non singularity of this matrix see Section 3.

It is worth noticing here that the semi-Markov Ph-distributions on $[0, \infty)$, given by (5), is a dense class for the weak topology, in the set of all probability distributions on $[0, \infty)$, since this class includes as a particular case the dense class of Markov Ph-distributions (e.g., Neuts [1981]). It is also worth noticing that we are using here the convolution sense inversion of a matrix function A denoted by $A^{(-1)}$ which is different from the usual inversion of a matrix B, denoted by B^{-1}.

3 Semi-Markov process and competing risks

In this section we are going to extend the semi-Markov Ph-distributions to the competing risks setting, as it has been done for the Markov case by Lindqvist and Kjølen [2018]. Let us consider a continuous-time semi-Markov process $(Z_t, t \in \mathbb{R}_+)$, with state space E and initial distribution α. We shall decompose the state space E into the transient subset E_0 (good performance states) and absorbing subset E_1 (failures states), i.e., $E = E_0 \cup E_1$. We shall consider $r \geq 1$ transient states and $m \geq 2$ absorbing states. Under these conditions, we shall give the main results for the extended semi-Markov Ph-distribution in continuous time. The time that the process has to wait until reaching the set E_1 (failure states), the absorption time T is this time defined as follows,

$$T := \inf\{k \geq 0 : Z_k \in E_1\}. \tag{8}$$

The lifetime T and the cause of failure, C, with values in the set $\{1, ..., m\}$, depend on Z_t. More precisely, we have $\{T \leq t, C = j\} = \{Z_t = r + j\}$. This is the key relation of the connection between competing risks and the extended Semi-Markov Ph-distributions.

Consider now the partition of the semi-Markov kernel Q, and the initial law α, in this new situation following the partition E_0, E_1 of E, as follows:

$$Q(t) = \begin{bmatrix} Q_0(t) & L(t) \\ 0_{m\times r} & 0_{m\times m} \end{bmatrix} \tag{9}$$

and $\alpha = (\alpha_0, \alpha_1)$, notice that, in this particular case α_0 is the r-dimensional vector. The function $L(t)$ is now an $r \times m$ matrix.

Consider also the diagonal matrices $\overline{H}_0(t) := diag(\overline{H}_i(t), i = 1, ..., r)$ and $\overline{H}_1(t) := diag(\overline{H}_i(t), i = r + 1, ..., r + m)$ is the restriction of the sojourn times survival functions on the absorbing states, where $H_1(t) = 0$, so $\overline{H}_1(t) = I$ the identity matrix, for $t \geq 0$, and $0_{m\times m}$ otherwise.

Let us denote the distribution function of (T, C) by $F_{ij}(t) := \mathbb{P}_i(T \leq t, C = j)$. This is the cumulative incidence function in the competing risks terminology. Let further the failure rate $\lambda_{ij}(t)$, for initial state $i \in E_0$, and cause $j \in E_1$, which is the cause-specific hazard in competing risks terminology, be defined by

$$\lambda_{ij}(t) := \lim_{h\downarrow 0} \frac{\mathbb{P}_i(t < T \leq t + h, C = j \mid T > t)}{h}.$$

It is worth noticing here that for fixed $i \in E_0$, and $j \in E_1$, $F_{ij}(t)$ is a sub-distribution function.

Let us define the matrix functions $F(t) := (F_{ij}(t); i \in E_0, j \in E_1)$ and $\lambda(t) := (\lambda_{ij}(t); i \in E_0, j \in E_1)$.

Proposition 3.1. *Suppose that the entries of the matrix function L, in the semi-Markov kernel (2), have Radon-Nikodym derivatives. Then the distribution function matrix $F(t)$ and the cause specific, failure rate function $\lambda(t)$, are given by*

$$F(t) = (I - Q_0)^{(-1)} * L(t)$$

and

$$\lambda_{ij}(t) = \frac{e_i(I - Q_0)^{(-1)} * \ell(t)e_j}{e_i(I - Q_0)^{(-1)} * \overline{H}_0(t)\mathbf{1}_r},$$

where $\ell(t) := L'(t)$, the element derivatives of L with respect to t and $e_i := (0, ..., 0, 1, 0, ..., 0)$, with 1 in the i-th entry.

Remark. In the case when we consider a general initial distribution α_0 on E_0, then the above formulas can be written as $F_j(t) := \mathbb{P}(T \leq t, C = j) = \alpha_0(I - Q_0)^{(-1)} * L(t)$ and

$$\lambda_j(t) = \lim_{h \downarrow 0} \frac{\mathbb{P}(t < T \leq t + h, C = j \mid T > t)}{h} = \frac{\alpha_0(I - Q_0)^{(-1)} * \ell(t)e_j}{\alpha_0(I - Q_0)^{(-1)} * \overline{H}_0(t)\mathbf{1}_r}.$$

Proof. We have:

$$\begin{aligned}
F_{ij}(t) &:= \mathbb{P}_i(T \leq t, C = j) \\
&= \sum_{k \in E_0} \int_0^t \mathbb{P}_i(T \leq t, C = j \mid S_1 = s, J_1 = k)\mathbb{P}_i(S_1 \in ds, J_1 = k) \\
&\quad + \sum_{k \in E_1} \int_0^t \mathbb{P}_i(T \leq t, C = j \mid S_1 = s, J_1 = k)\mathbb{P}_i(S_1 \in ds, J_1 = k), \\
&= \sum_{k \in E_0} \int_0^t Q_{ik}(ds)F_{kj}(t - s) + \sum_{k \in E_1} \int_0^t Q_{ik}(ds)\delta_{kj}
\end{aligned}$$

Hence

$$F_{ij}(t) = Q_{ij}(t) + \sum_{k \in E_0} \int_0^t Q_{ik}(ds)F_{kj}(t - s)$$

$i \in E_0, j \in E_1$

$$F(t) = L(t) + Q_0 * F(T)$$

$$F(t) = (I - Q_0)^{(-1)} * L(t)$$

and δ_{kj} as the Kroner's symbol, i.e.,

$$\delta_{kj} = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j. \end{cases}$$

Let us now consider the probabilities $R_{ik} = \mathbb{P}_i(Z_T = r + k) = \mathbb{P}_i(C = k)$, for starting in state $i \in E_0$ and being absorbed in state $r + k$ that means it died by cause $k = 1, ..., m$, and define the matrix $R := (R_{ik}; i = 1, ..., r; k = 1, ..., m)$. Consider also the transition probability matrix P of the embedded Markov chain (J_n) of the semi-Markov process (Z_t) and its partition following sets E_0, E_1, i.e.,

$$P = \begin{bmatrix} P_0 & P_1 \\ 0_{m \times r} & I \end{bmatrix}.$$

Proposition 3.2. *We have*

$$R = (I - P_0)^{-1} P_1.$$

Proof. Since this probability depends only on the transition probabilities of the embedded Markov chain, the proof of the result is straightforward by Markov chain theory (see, e.g., Girardin and Limnios [2018]).

4 The discrete-time competing risk

Let $(Z_k), k \in \mathbb{N}$, be a semi-Markov discrete-time process, i.e., a semi-Markov chain (SMC) with state space E, and $(J_n, S_n), n \in \mathbb{N}$, its embedded Markov renewal chain, see e.g., Barbu and Limnios [2008]. The semi-Markov kernel of (Z_k), is defined as follows,

$$q_{ij}(k) := \mathbb{P}(J_{n+1} = j, S_{n+1} - S_n = k \mid J_n = i), \;\; n \geq 0, k \in \mathbb{N}, \text{ for } i, j \in E. \tag{10}$$

It is worth noticing that the semi-Markov kernel considered here is independent of n, which means that the SMC is homogeneous in time, and that we use index $k \in \mathbb{N}$ for the calendar times, and index $n \in \mathbb{N}$ for the number of jumps of (Z_k). The stochastic process (J_n) is the embedded Markov chain of the MRP (J_n, S_n).

Let us denote by $(X_n)_{n\in\mathbb{N}}$ the process which determines the successive sojourn times in the visited states, where by convention $X_0 := S_0 := 0$, and

$$X_{n+1} = S_{n+1} - S_n, \ \ n \in \mathbb{N}.$$

We shall denote the conditional distribution of X_{n+1}, $n \in \mathbb{N}$ by

$$f_{ij}(k) := \mathbb{P}(X_{n+1} = k \mid J_n = i, J_{n+1} = j), \ \ k \in \mathbb{N}. \tag{11}$$

The cumulative distribution function of the sojourn time in state $i \in E$ is defined by the following relation

$$H_i(k) := \sum_{l=0}^{k} \sum_{j\in E} q_{ij}(l).$$

Let $\varphi(i,k), \ i,j \in E, k \in \mathbb{N}$, be a measurable function and define the convolution of φ by q by

$$(q * \varphi)_{ij}(k) := \sum_{r\in E} \sum_{l=0}^{k} q_{ir}(l)\phi_{rj}(k-l).$$

The n-fold convolution of q by itself is defined recursively by

$$q_{ij}^{(n)}(k) := \sum_{r\in E} \sum_{l=0}^{k} q_{ir}(l) q_{rj}^{(n-1)}(k-l).$$

We shall make the same considerations for the semi-Markov chain $(Z_k)_{k\in\mathbb{N}}$ as in continuous time, i.e., we shall decompose the state space in transient (good performance states) and absorbing states (failures states), i.e., $E = E_0 \cup E_1$. We shall consider $r \geq 1$ transient states and $m \geq 2$ absorbing states. We shall also partition the semi-Markov kernel following the states E_0 and E_1, i.e.,

$$q(k) = \begin{pmatrix} q_0(k) & q_1(k) \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{12}$$

Observe that the first zero in the second line is the $m \times r$ zero matrix and the second one is the $m \times m$ matrix; $q_0(k)$ and $q_1(k)$ are the restriction of $q(k)$ on $E_0 \times E_0$ and $E_0 \times E_1$ respectively. The next proposition gives the main result for the extended Semi-Markov Ph-distribution in discrete time.

Proposition 4.1. *For a semi-Markov chain (Z_k), $k \in \mathbb{N}$, with state space E and initial distribution α as described above, we have:*

$$g_j(k) := \mathbb{P}(T = k, C = j) = \begin{cases} 0, & k = 0; \\ \alpha_0(I - q_0)^{(-1)} * Q_1(k)e_j, & k \in \mathbb{N}^*; \end{cases}$$

where

$$Q(k) = \sum_{l=0}^{k} q_1(l).$$

Therefore

$$G_j(k) = \mathbb{P}(T \leq k, C = j) = \sum_{l=0}^{k} \alpha_0(I - q_0)^{(-1)} * q_1(l)e_j,$$

where e_j is a column vector of size $|E_1|$ where all its coordinates are zero except the coordinate which correspond to state j.

Proof: Set

$$g_{ij}(k) = \mathbb{P}_i(T = k, C = j), \;\; i \in E_0, j + r \in E_1.$$

Obviously, we have:

$$g_j(k) = \sum_{i \in E_0} \alpha_i g_{ij}(k). \tag{13}$$

Now, we can write, for $i \in E_0$ and $j \in E_1$:

$$\begin{aligned} g_{ij}(k) &= \mathbb{P}_i(T = k, C = j) \\ &= \mathbb{P}_i(T = k, C = j, S_1 \leq k) \\ &= \sum_{r \in E_1} \sum_{l=1}^{k} \mathbb{P}_i(T = k, C = j \mid S_1 = l, J_1 = r)\mathbb{P}_i(J_1 = r_1, S_1 = l) \\ &\quad + \sum_{r \in E_0} \sum_{l=1}^{r} \mathbb{P}_i(T = k, C = j \mid S_1 = l, J_1 = r)\mathbb{P}_i(J_1 = r_1, S_1 = l) \\ &= \sum_{r \in E_1} \sum_{l=1}^{k} q_{ir}(l)\delta_{rj} + \sum_{r \in E_0} \sum_{l=1}^{k} q_{ir}(l) g_{rj}(k - l) \end{aligned}$$

Hence

$$g_{ij}(k) = Q_1(k) + \sum_{l=1}^{k} \sum_{r \in E_0} q_{ir}(l) g_{ij}(k - l) \tag{14}$$

The Markov renewal Equation (14) can be written in its matrix form as follows

$$g(k) = Q_1(k) + q_0 * g(k)$$

and its solution is

$$g(k) = (I - q_0)^{(-1)} * Q_1(k) \blacksquare$$

5 Numerical Applications

5.1 Continuous-time case

Let us consider a four states semi-Markov process, i.e., let $E = \{1, 2, 3, 4\}$, where states $1, 2$ are transient and states $3, 4$ are absorbing states, i.e., $E_0 = \{1, 2\}$ and $E_1 = \{3, 4\}$. Assume that the semi-Markov kernel of this process is $Q(t)$,

$$Q(t) = \begin{bmatrix} 0 & Q_{12}(t) & Q_{13}(t) & 0 \\ Q_{21}(t) & 0 & 0 & Q_{24}(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

with the following blocks of its partition:

$$Q_0(t) = \begin{bmatrix} 0 & Q_{12}(t) \\ Q_{21}(t) & 0 \end{bmatrix}, \qquad L(t) = \begin{bmatrix} Q_{13}(t) & 0 \\ 0 & Q_{24}(t) \end{bmatrix}.$$

Now we have the following block matrix of the transition function

$$P_{12}(t) = (I - Q_0)^{(-1)} * L(t) = M * \begin{bmatrix} Q_{13}(t) & Q_{12} * Q_{24}(t) \\ Q_{21} * Q_{13}(t) & Q_{24}(t) \end{bmatrix}$$

where $M(t) := (1 - Q_{21} * Q_{13})^{(-1)}(t) = 1 + \sum_{k=1}^{\infty} (Q_{21} * Q_{13})^{(k)}(t)$. This is a usual renewal type function.

So, we have

$$F_1(t) = \alpha_0 P_{12}(t) e_1 = \alpha(1) M * Q_{13}(t) + \alpha(2) M * Q_{21} * Q_{13}(t)$$

and

$$F_2(t) = \alpha_0 P_{12}(t) e_1 = \alpha(1) M * Q_{12} * Q_{24}(t) + \alpha(2) M * Q_{24}(t).$$

The primes here mean derivatives with respect to t. In figure 1 we can observe an example of the function $F_j(t)$ where $j = 1, 2$ and the sojourn time in a state is modeled by a Weibull distribution. The matrix value function ψ, see Equation (6) is computed using the method valued proposed by Wu et al. [2020].

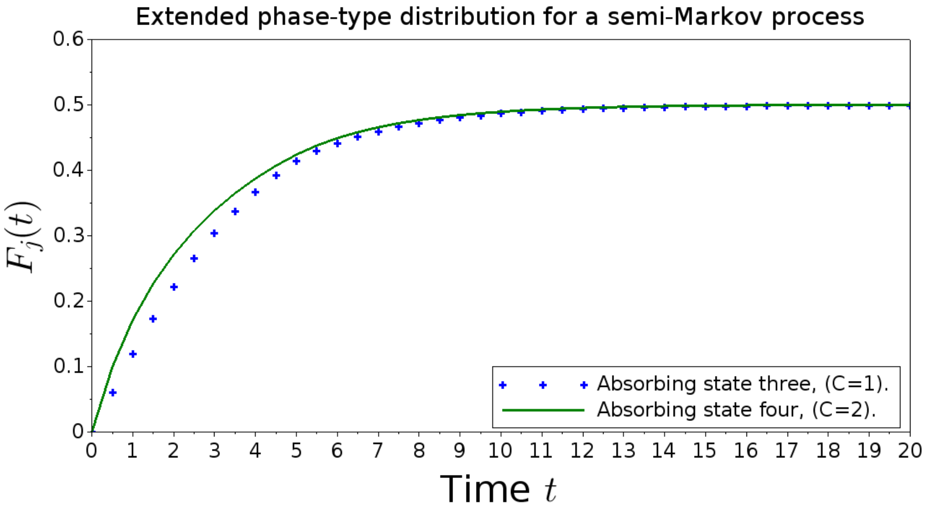

Figure 1: Probability of absorption before time t by the first and the second cause of failure in a semi-Markov process

We also calculate the cause specific failure rates $\lambda_j(t)$, for $j = 1, 2$, as follows:

$$\lambda_1(t) = \frac{\alpha(1)M * Q'_{13}(t) + \alpha(2)M * Q_{21} * Q'_{13}(t)}{M * [\alpha(1)Q_{13} * \overline{H}_1(t) + \alpha(1)Q_{12} * Q_{24}\overline{H}_2(t) + \alpha(2)Q_{12} * Q_{13} * \overline{H}_1(t) + \alpha(2)Q_{24} * \overline{H}_2(t)]}$$

$$\lambda_2(t) = \frac{\alpha(1)M * Q_{12} * Q_{24}(t)' + \alpha(2)M * Q_{24}(t)'}{M * [\alpha(1)Q_{13} * \overline{H}_1(t) + \alpha(1)Q_{12} * Q_{24}\overline{H}_2(t) + \alpha(2)Q_{12} * Q_{13} * \overline{H}_1(t) + \alpha(2)Q_{24} * \overline{H}_2(t)]}$$

where $\overline{H}_1(t) = 1 - (Q_{12}(t) + Q_{13}(t))$ and $\overline{H}_2(t) = 1 - (Q_{21}(t) + Q_{24}(t))$, for $t \geq 0$, and $\alpha(i) := \mathbb{P}(Z_0 = i)$, for $i = 1, 2$. In figure 2 we can observe the failure rate by the first and the second cause of failure.

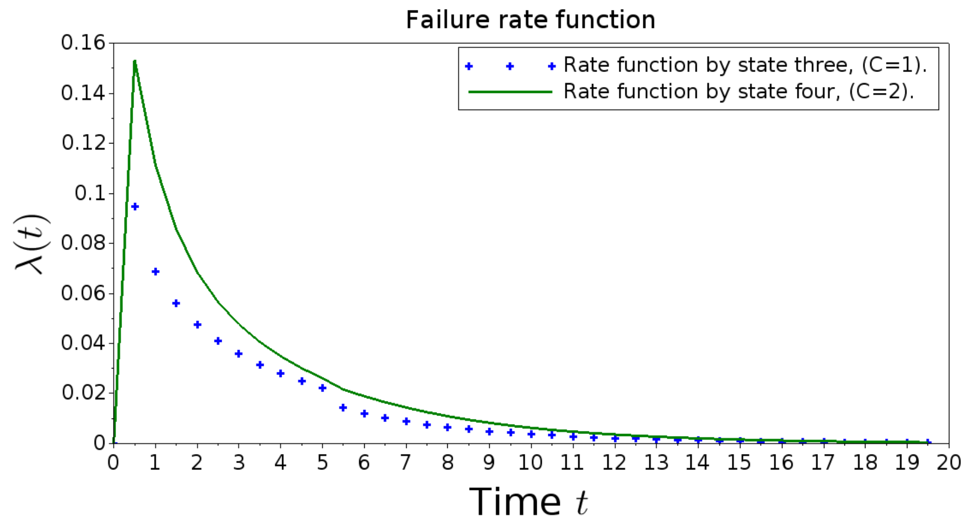

Figure 2: Failure rate by the first and the second cause of failure

Finally, the matrix R is

$$R = (1 - p_{12}p_{21})^{-1} \begin{bmatrix} p_{13} & p_{12}p_{24} \\ p_{21}p_{13} & p_{24} \end{bmatrix}$$

where $p_{ij} := Q_{ij}(\infty)$, for $i, j \in E_0$ and $p_{ij} := \delta_{ij}$ for $i, j \in E_1$ (Kronecker's δ).

It is worth noticing that from $p_{12} + p_{13} = 1$ and $p_{21} + p_{24} = 1$, we can see that R is a stochastic matrix.

5.2 Discrete-time case

We present an example of a semi-Markov chain with two absorbing states, i.e., we consider two causes of failure. For this example the state space is $E = \{1, 2, 3, 4\}$, with up states: 1 and 2; and down states: 3 and 4, i.e., the first cause of failure is the state 3, and the second cause of failure is the state 4.

In a semi-Markov chain, $f_{ij}(k)$ could be any distribution on $\mathbb{N}$. In this example every $f_{ij}(\cdot)$ is a discrete-time Weibull distribution, i.e.,

$$f_{ij}(k) := W_{a,b}(k)$$

where

$$W_{a,b}(0) := 0 \quad \text{for all } i, j$$

and

$$W_{a,b}(k) := a^{(k-1)^b} - a^{k^b}, \qquad k \geq 1.$$

see Nakagawa and Osaki [1975]. For this particular example

$$q(k) = \big(q_{ij}(k)\big)_{1 \leq i,j \leq 4} = \begin{pmatrix} 0 & p_{12}f_{12}(k) & p_{13}f_{13}(k) & p_{14}f_{14}(k) \\ p_{21}f_{21}(k) & 0 & 0 & p_{24}f_{24}(k) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad k \in \mathbb{N},$$

where $p_{ij} := \mathbb{P}(J_{n+1} = j \mid J_n = i)$, $i, j \in E$, $n \in \mathbb{N}$ and the value for $a = 0.5$ and $b = 0.5$. In the next figure we show the extended Phase-type distribution of the random pair (T, C) for a semi-Markov chain. In Figure 3, the evolution of the system is modeled by a semi-Markov chain. In this figure we can observe the two distribution $G_3(k)$ and $G_4(k)$, that the process enters the absorbing state three (first cause of failure) and, the absorbing state four (second cause of failure).

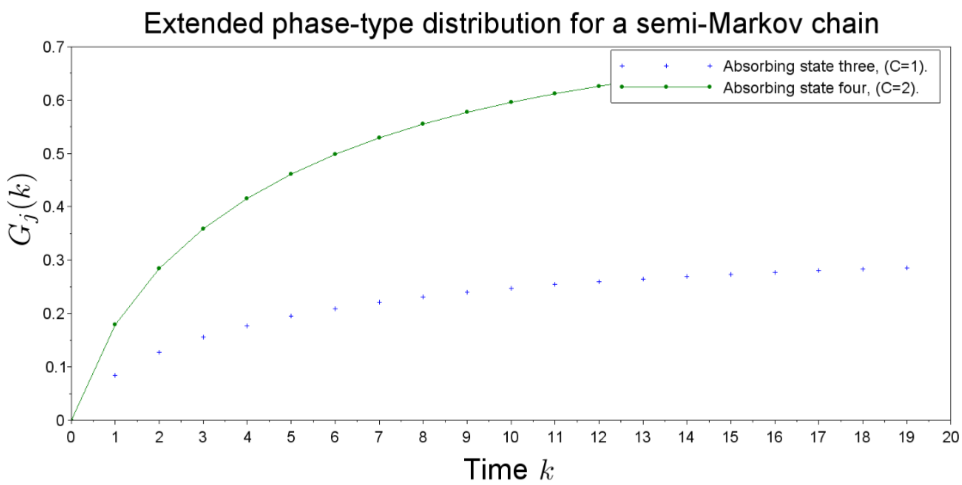

Figure 3: Probability of absorption before time k for states $j = 3$ and $j = 4$ for a semi-Markov chain

6 Concluding remarks

In this paper, we have presented an extension of the competing risks models based on absorbing Markov chains to semi-Markov type competing risks models. This generalization is important since we can now consider any distribution for the sojourn times in a state, instead of merely the exponential (or geometric) distribution. We have calculated the joint distribution of the time to failure T and cause of failure in presence of different causes using a semi-Markov approach. This is the probability of absorption by time t in a specified absorbing state $j \in E_1$, and corresponds to the cumulative incidence function in standard competing risks terminology. We further provide expressions for the cause specific hazard rate $\lambda_{ij}(t)$ for initial state $i \in E_0$ and $j \in E_1$, as well as the general cause-specific hazard $\lambda_j(t)$. Both continuous and discrete time semi-Markov processes are considered. The closed form formulas provided in the paper can be used for statistical inference for competing risks data.

Acknowledgments

We are indebted to an anonymous referee for his useful comments that improved the presentation of this paper.

This work was supported by a PhD scholarship funding (to the first author), granted by the Mexican *Consejo Nacional de Ciencia y Tecnologia* (CONACYT).

References

O. O. Aalen. Phase type distributions in survival analysis. *Scandinavian Journal of Statistics*, pages 447–463, 1995.

S. Asmussen and C. O'Cinneide. Matrix-exponential distributions. *Encyclopedia of Statistical Sciences*, 3, 2006.

V. S. Barbu and N. Limnios. *Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis*, volume 191. Springer Science & Business Media, 2008.

J. Beyersmann, A. Allignol, and M. Schumacher. *Competing risks and multistate models with R*. Springer Science & Business Media, 2011.

M. Crowder. *Classical competing risks*. Chapman & Hall/CRC, 2001.

M. J. Crowder. *Multivariate survival analysis and competing risks*. Chapman and Hall/CRC, 2012.

V. Girardin and N. Limnios. *Applied Probability: From Random Sequences to Stochastic Processes*. Springer, 2018.

N. Limnios. Reliability measures of semi-markov systems with general state space. *Methodology and Computing in Applied Probability*, 14(4):895–917, 2012.

N. Limnios and G. Oprişan. *Semi-Markov processes and reliability*. Springer Science & Business Media, 2001.

B. H. Lindqvist and S. H. Kjølen. Phase-type models and their extension to competing risks. In *Recent Advances in Multi-state Systems Reliability*, pages 107–120. Springer, 2018.

T. Nakagawa and S. Osaki. The discrete weibull distribution. *IEEE Transactions on Reliability*, 24(5):300–301, 1975.

M. Neuts. *Matrix-geometric solutions an algorithmic approach*. The Johns Hopkins University Press, Baltimore, MD, 1981.

B. Wu, B. I. G. Maya, and N. Limnios. Using semi-markov chains to solve semi-markov processes. *Methodology and Computing in Applied Probability*, pages 1–13, 2020.