Introducing Covariates in Reliability Models by Markovian Arrival Processes

C. Santamaría[†] [*], B. García-Mora[†], and G. Rubio[†]

(†) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València

March 30, 2017

Our aim is to model *inter-failures* times that are correlated and not identically distributed, taking into account covariates. There is a constant need to get reliability models with the first two properties. And, to the best of our knowledge, the introduction of covariates is a pending task. The *Markovian Arrival Process (MAP)* [5] is an active research field for dealing with not identically distributed and correlated inter-failures times ([1], [2], [3]). Our approach is based on the work developed in [1] and [6] for the case without covariates.

We illustrate our approach with a set of simulated data of devices undergoing three failures each one (Figure 1). It is simulated a sample of operational random times for 100 devices with three failures for each one. The devices are independent. The sample is

$$t^{(1)} = (t_1^{(1)}, t_2^{(1)}, t_3^{(1)}), t^{(2)} = (t_1^{(2)}, t_2^{(2)}, t_3^{(2)}), \dots, t^{(100)} = (t_1^{(100)}, t_2^{(100)}, t_3^{(100)}) \quad (1)$$

where in each device we have *the times of the three failures*, and two covariates: age and device class, with two possibilities: class 1 and class 0. Now let T_k be the random variable representing the operational time between the (k–1)–th failure and the k–th failure. We have three variables T_1, T_2, T_3 of the *inter–failure times*, correlated and not identically distributed (Figure 1). If covariates are not taken into account, the methodology from [1] works, as

*e-mail: crisanna@imm.upv.es

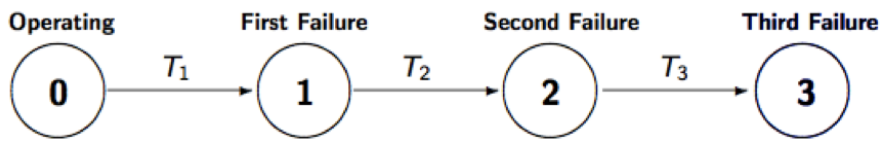

Figure 1: Device with three failures

expected. We apply a 2–state non–stationary Markovian Arrival Process to our data, denoted by MAP_2. It is a doubly stochastic process $\{J(t), N(t)\}$ where

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

The MAP_2 can be characterized by $\mathrm{M} = \{\alpha, D_0, D_1\}$ where D_0 and D_1 are rate matrices. $D = D_0 + D_1$ is the generator of $J(t)$, with stationary vector ϕ, calculated as $\phi P^* = \phi$. P^* is the transition probability matrix, given by $P^* = (-D_0)^{-1} D_1$. The cumulative density function (CDF) and the moments of the *variables* T_1, T_2 *and* T_3 are defined by the expressions

$$F_{T_k}(t) = 1 - \alpha_k e^{D_0 t} \mathbf{e}$$

where $\alpha_{\mathbf{k}} = \alpha (P^*)^{k-1}$ and $T_k \sim PH\{\alpha_{\mathbf{k}}, D_0\}$ represent different phase–type distributions for the correlated variables T_1, T_2 *and* T_3.

The goal is to estimate the model parameters in $\{\tilde{\alpha}, \tilde{D}_0, \tilde{D}_1\}$ in the MAP_2. For this it is used an optimization problem (P) [6]:

$$\begin{array}{rl} min & \gamma_\tau(\tilde{\alpha}, \tilde{D}_0, \tilde{D}_1) \\ s.t. & \tilde{x}, \tilde{u} \leq 0 \\ & \tilde{y}, \tilde{v} \leq 0 \\ & -\tilde{x} - \tilde{y} \geq 0 \end{array}$$

$$-\tilde{u} - \tilde{v} \geq 0$$
$$0 \leq \tilde{\alpha} \leq 1$$

where the *objective function* of the problem is given by

$$\begin{aligned}
\gamma_\tau(\tilde{\alpha}, \tilde{D_0}, \tilde{D_1}) &= \tau\Big\{\Big(\frac{r_1(\tilde{\alpha}, \tilde{D_0}, \tilde{D_1}) - \tilde{r_1}}{\tilde{r_1}}\Big)^2 + \Big(\frac{r_2(\tilde{\alpha}, \tilde{D_0}, \tilde{D_1}) - \tilde{r_2}}{\tilde{r_2}}\Big)^2 \\
&+\Big(\frac{r_3(\tilde{\alpha}, \tilde{D_0}, \tilde{D_1}) - \tilde{r_3}}{\tilde{r_3}}\Big)^2 + \Big(\frac{\mu_2(\tilde{\alpha}, \tilde{D_0}, \tilde{D_1}) - \tilde{\mu}_2}{\tilde{\mu}_2}\Big)^2 \\
&+\Big(\frac{\mu_3(\tilde{\alpha}, \tilde{D_0}, \tilde{D_1}) - \tilde{\mu}_3}{\tilde{\mu}_3}\Big)^2\Big\}
\end{aligned}$$

made of the model population moments and their empirical counterparts, and where τ is a penalty parameter that needs to be tuned, but setting $\tau = 1$ performs well in practice [6].

The problem is solved using the local search *MATLAB's routine* **fmincon** (Optimization toolbox). A *multistart approach* (100 different starting points randomly selected of the simulated data) is performed and we keep the solution with the minimum objective function $\gamma_\tau(\tilde{\alpha}, \tilde{D_0}, \tilde{D_1})$ in the optimization problem (P).

We solve the problem (P) for two canonical representations of the MAP_2

- The expression of the first canonical representation is

$$\tilde{\alpha} = (\tilde{\alpha}, 1 - \tilde{\alpha}), \;\; \tilde{D_0} = \begin{pmatrix} \tilde{x} & \tilde{y} \\ 0 & \tilde{u} \end{pmatrix}, \;\; \tilde{D_1} = \begin{pmatrix} -\tilde{x} - \tilde{y} & 0 \\ \tilde{v} & -\tilde{u} - \tilde{v} \end{pmatrix} \quad (2)$$

- The second canonical representation

$$\tilde{\alpha} = (\tilde{\alpha}, 1 - \tilde{\alpha}), \;\; \tilde{D_0} = \begin{pmatrix} \tilde{x} & \tilde{y} \\ 0 & \tilde{u} \end{pmatrix}, \;\; \tilde{D_1} = \begin{pmatrix} 0 & -\tilde{x} - \tilde{y} \\ -\tilde{u} - \tilde{v} & -\tilde{v} \end{pmatrix} \quad (3)$$

and we select the estimated *parameters* $\{\tilde{\alpha}, \tilde{x}, \tilde{y}, \tilde{u}, \tilde{v}\}$ under the *canonical representation* with the *highest log–likelihood* given in

$$\log f(\mathbf{t}^{(1)}, \mathbf{t}^{(2)}, \ldots \mathbf{t}^{(N)} | D_0, D_1) = \sum_{i=1}^{N} \log f(\mathbf{t}^{(i)} | D_0, D_1). \quad (4)$$

To introduce covariates, we have been inspired by our previous work [4] that worked successfully. The way to the present work will be justified in detail during the conference and in an eventual extended paper. The essential idea is to use the covariate information in order to modify the matrices $\tilde{D}_0$ and $\tilde{D}_1$ in a suitable manner.

References

[1] Rodríguez J., Lillo R.E., Ramírez–Cobo P. Failure modeling of an electrical N–component framework by the non–stationary Markovian arrival process. *Reliability Engineering and System Safety*, (134): 126–133, 2015.

[2] Santamaría C., García–Mora B., Rubio G and Pérez–Ocón R. Managing dependence in Flowgraphs models. An application to Reliability Engineering. Modelling for Engineering & Human Behaviour 2015, Valencia, Instituto Universitario de Matemática Multidisciplinar, 2015.

[3] Rubio G, García–Mora B., Santamaría C., Pontones J.L. Modeling dependence in multistate processes. International Work–Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2016. 20–22 April, 2016. Granada (Spain).

[4] 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.

[5] Buchholz P., Kriege J. and Felko I. Input Modeling with Phase–Type Distributions and Markov Models. Theory and Applications, Springer Cham Heidelberg New York Dordrecht London, 2014.

[6] Carrizosa E., Ramírez–Cobo P. Maximum likelihood estimation for the two–state Markovian arrival. arXiv: 1401.3105v1.