Bayesian Analysis (2016) **11**, Number 1, pp. 151–190

Bayesian Inference for Partially Observed Multiplicative Intensity Processes

Sophie Donnet* and Judith Rousseau†

Abstract. Poisson processes are used in various applications. In their homogeneous version, the intensity process is a deterministic constant whereas it depends on time in their inhomogeneous version. To allow for an endogenous evolution of the intensity process, we consider multiplicative intensity processes. Inference methods for such processes have been developed when the trajectories are fully observed, that is to say, when both the sizes of the jumps and the jumps instants are observed. In this paper, we deal with the case of a partially observed process: we assume that the jumps sizes are non- or partially observed whereas the time events are fully observed. Moreover, we consider the case where the initial state of the process at time 0 is unknown. The inference being strongly influenced by this quantity, we propose a sensible prior distribution on the initial state, using the probabilistic properties of the process. We illustrate the performances of our methodology on a large simulation study.

MSC 2010 subject classifications: Primary 62F15, 62M09, 62P30; secondary 62N01.

Keywords: Bayesian analysis, counting process, latent variables, multiplicative intensity process.

1 Introduction

Counting processes (say $X(t)$) are commonly used in various fields of applications such as medicine (see Gusto and Schbath (2005), for instance), public health biology or reliability (see Chen (2011), for instance), or more generally in risk theory (see Ogata (1999), for instance). These processes are driven by their intensity process. The most simple counting processes are homogeneous Poisson processes, whose intensity process is a constant deterministic positive number. A classical generalization of the homogeneous Poisson process is the inhomogeneous Poisson process whose intensity process is a positive deterministic function. Although widely used in practice and flexible, these processes are limited by the fact they do not allow for endogenous evolution of the intensity function. Multiplicative intensity processes (see Aalen (1978)) allow for such an evolution. In this case the intensity process is expressed as $Y(t)\alpha(t)$, where α is a positive deterministic function – namely, the intensity function of the process – and $Y(t)$ is a positive predictable process – the exposure process – (see, for instance, Chen (2011)). These processes have been commonly used; they encompass in particular survival analysis and finite states Markov processes applications, see, for instance, Andersen

*UMR 518 AgroParisTech/INRA, AgroParisTech, 16, rue Claude Bernard, 75231 Paris CEDEX 05, France, sophie.donnet@agroparistech.fr

†Ceremade, Université Paris-Dauphine, Place du Maréchal DeLattre de Tassigny, 75016 Paris, France, judith.rousseau@ceremade.dauphine.fr

 DOI: 10.1214/15-BA940

et al. (1993). Parametric and nonparametric methods have been developed when the trajectories $(Y(t), X(t), t \in [0,1])$ are fully observed starting from the paper by Aalen (1978) (see, for instance, Chen (2011), Ishawaran and James (2004), Reynaud-Bouret and Schbath (2010) and references therein for nonparametric estimation, and Ogata (1999) for parametric estimation). It is, however, sometimes the case that the process, typically Y_t, is only partially observed. In this paper we propose a Bayesian analysis of a family of partially observed multiplicative intensity processes.

A particular multiplicative Poisson process with partial observations The processes we study are pure-birth processes with multi-size immigrations. More precisely, we consider a population of particles such that the particles give birth (randomly) to j_0 particles (or equivalently, divides into $j_0 + 1$ particles) with rate $\nu_0(t)$ and immigration groups of sizes $j_1, \ldots, j_K$ arrive with respective rates $\nu_1(t), \ldots, \nu_K(t)$. Let $X(t)$ be the number of particles at time t: $X(t)$ is a counting process with exposure process $Y(t)$ being

$$Y(t) = X(t^-)\nu_0(t) + \sum_{k=1}^{K} \nu_k(t) \quad \text{where} \quad X(t^-) = \lim_{s \to t, s<t} X(t),$$

$X(t^-)$ is predictable and $\nu_k(t)$ $(k = 0, \ldots, K)$ are positive functions.

Remark 1. *Note that contrariwise to compound Poisson processes, the above counting process is self-excited, in other words, the (conditional) intensity depends on the past trajectory of the process and not on some external stochastic process. It is thus strongly related to multivariate Hawkes processes as used in seismicity analysis, see Ogata (1999) or in DNA analysis Gusto and Schbath (2005). However, instead of considering only two point processes as in their cases we have K point processes. Also compared to Ogata (1999) we do not have the same parametric form for the intensity process, and compared to Gusto and Schbath (2005) we have infinite memory in our construction of the intensity point process. It can also be seen as a Markov jump process, as considered, for instance, in Rao and Teh (2013) and references therein, except that in our case the Markov process may be inhomogeneous.*

The aim is then to estimate the different rate functions from the observation of the counting process over a finite period $[\tau_0, \tau_0 + \tau]$. We consider a parametric context, i.e. $\exists\, \theta \in \Theta \subset \mathbb{R}^d$ such that

$$(\nu_0(t), \ldots, \nu_K(t)) = (\nu_0(t, \theta), \ldots, \nu_K(t, \theta)) \tag{1}$$

and θ is the parameter of interest. If the rates are constant in time, then $(\nu_0(t), \ldots, \nu_K(t)) = (\nu_0, \ldots, \nu_K) = \theta$. Time varying functions can be used to model particle aging, for instance, $\nu_0(t) = \alpha_0 t + \beta_0$ or $\nu_0(t) = \alpha_{01}\mathbb{1}_{t<t_1} + \alpha_{02}\mathbb{1}_{t\geq t_1}$, or an acceleration of the immigration process, $\nu_k(t) = \alpha_k t + \beta_k$.

Our aim is to estimate θ. We thus consider the following identifiability assumption:

Assumption $\mathcal{H}_0$: $\quad \nu_k(t;\theta) = \nu_k(t;\theta'), \quad \forall k = 0, \ldots, K \quad \forall t \in (\tau_0, \tau_0 + \tau) \Rightarrow \theta = \theta'.$

Estimating θ can be very different whether we observe the counting process completely or partially. In the following, observing X *completely* over $[\tau_0, \tau_0 + \tau]$ means not only observing the number of jumps $N^\star$ and the jump instants $T_1, \dots, T_{N^\star}$ of the process but also the types of the jumps (birth or immigration) and so their sizes ($\in \{j_0, \dots, j_K\}$).

In this paper, we provide a Bayesian estimation procedure based on a non-standard partial observation of the process. There are various meanings for the notion partial observations in counting processes. One can observe $X(t)$ only at discrete times $t_1, \dots, t_n$: see, for instance, the counting processes used in epidemiology where we observe the number of newly infected people weekly, or more generally, models based on Markov jump processes (see Rao and Teh, 2013, for instance). Another possibility is to observe a noisy version of X, as in state space models, where the counting process X represents the latent state of the dynamical system, see, for instance, Godsill (2007) or Whiteley et al. (2011). However, in this paper, we consider a different case of "partial observation" assuming that we have access – on a fixed period $[\tau_0, \tau_0 + \tau]$ – to all the jumps instants of the process $T_1, \dots, T_{N^\star}$ but we do not observe (or partially observe) the types (birth or immigration) / sizes of the jumps (possibly j_0, $j_1, \dots,$ or j_K). This can be framed into the setup of the state space model of Godsill (2007), where the distribution of the observation given the state is a point mass at the observed time points (which is part of the latent variable), but it is a rather artificial expression of the model, since our model is simpler. Note that in Godsill (2007), there are no parameters – apart from the latent states – to be estimated and the paper deals with the filtering problem, while here we are interested in recovering the parameters and possibly the process X.

Moreover, the feasibility and quality of the inference are also highly conditioned by the observation or not of the number of particles at the beginning of the observation period $X(\tau_0)$. Dealing with the rates estimation when $X(\tau_0)$ is unknown is a challenging issue that we propose to tackle here. In this special case, the Bayesian approaches are a natural way to handle this missing data framework (see Section 3.1).

Application This work is motivated by the analysis of an electrical network through time. To simplify the exposure, assume that the electrical network is composed of a cable (of constant length d) and accessories (such as joints, etc.). We observe the evolution of the network and, more precisely, the sequences of incidents (failures) taking place either on the cable itself or on the accessories. When an incident takes place on the cable, it is repaired by exchanging the damaged part (very small) of the cable by a new piece of cable, connected to the remaining network by two accessories. When an incident takes place on an accessory, a small part of the network containing the damaged accessory is removed and replaced by a new piece of cable connected to the network by two accessories. (see Figure 1 for a graphical illustration of the reparation process).

Let $X(t)$ be the number of accessories (i.e. particles) on the network. An incident on an accessory corresponds to the birth of one particle ($j_0 = 1$) whereas a breakdown on the cable corresponds to the immigration of two particles ($K = 1, j_1 = 2$). The cable incident rate is assumed to be proportional to the length of the cable ($\nu_1(t) = d\nu_c$)

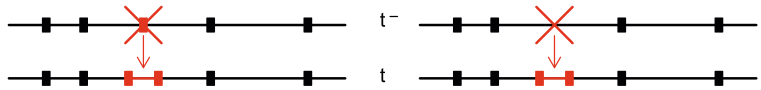

Figure 1: Electrical network. The horizontal line represents a cable, each node on the line represents an accessory. On the left (**failure on an accessory**) the accessory is replaced by two of them. On the right (**failure on the cable**) a new cable is connected to the remaining network by two additional accessories.

whereas the accessory incident rate is proportional to the number of accessories ($\nu_a X(t^-)$), leading to the following exposure process:

$$Y(t) = \nu_a X(t^-) + \nu_c d.$$

Remark 2. *The motivating example just described is related to repairable systems as described, for instance, in Gamiz et al. (2011), however, contrariwise to minimal repairable systems considered in Gamiz et al. (2011), which are counting processes with deterministic intensity functions, in our context the intensity function is modified through time by the state of the system.*

ν_a and ν_c are the parameters of interest since they will allow predicting the evolution of the network in the future. Moreover, ν_a and ν_c can be specific to the material or the geographical position of the network, and so their estimation can help to compare the types of material.

In this practical context, we have access to the instants of intervention (reparation) but only partially to the types of reparation (e.g. type of breakdown). More precisely, the instants of intervention are reported precisely whereas the types of intervention are reported with so much error that it is preferable to ignore that information (this point will be discussed at the end of the paper). As a consequence, we are in the situation described above: the observations are reduced to the jump instants denoted T_i; the heights of the jumps in the counting process – i.e. in this case, the cause of the incidents (cable or accessories) – are unobserved or partially observed.

Moreover, in general, the beginning of the observation period does not coincide with the installation of the network. More precisely the number of accessories is known at the installation of the electrical network, but the systematic collection of the incident times starts a long time after the installation instant.

$$\begin{array}{ccccc} 0 & \text{———} \cdots \longrightarrow & \tau_0 & \longrightarrow & \tau_0+\tau \\ X(0) & & X(\tau_0) & & X(\tau_0+\tau) \\ \text{Installation} & &] & \longleftarrow \text{ Observation } \longrightarrow &] \end{array} \tag{2}$$

Consequently, the number of accessories at the beginning of the observation period $X(\tau_0)$ is unknown and this quantity has to be inferred.

Hence, in this paper, we (i) propose a parametric Multiplicative Poisson process which is flexible and adapted to many problems including linear assets management and reliability evaluation for repairable systems and (ii) propose a general Bayesian procedure for inference from either complete or partial observation of the process. In the particular case where the number of particles at the beginning of the observation period $X(\tau_0)$ is unknown, we (iii) construct an ad-hoc prior distribution on $X(\tau_0)$, this semi-informative prior being based on the structure of the process X (and more precisely its asymptotic properties).

The paper is organized as follows. In Section 2, we introduce the notations and give the likelihood expression. Bayesian estimation is addressed in Section 3. In Section 3.1, we explain how we conduct the inference under a fully observed process. Section 3.2 describes how we extend the previous analysis to the setup of a partially observed process. We first consider the case where $X(\tau_0)$ is known and highlight the influence of $X(\tau_0)$ on the inference. Then, we treat the case with unknown $X(\tau_0)$ by constructing a prior distribution on $X(\tau_0)$ derived from the asymptotic behavior of the process. Section 4 presents an extensive numerical study. Finally, in Section 5, we discuss how the family of processes we have considered can be extended.

2 Notations and likelihood

We consider a pure birth process with multi-size immigrations. In the following, an "event" denotes either a birth or an arrival of immigrants. The particles (or individual) give birth to j_0 children (or equivalently, divides into $j_0 + 1$ particles), meaning that after one birth there are j_0 more particles. Groups of immigrants are of respective sizes $j_1 < j_2 < \cdots < j_K$.

Let $N(t)$ be the total number of events occurred in $[\tau_0, t]$. For the sake of simplicity, in the following, we use the following notation: $N^\star := N(\tau_0 + \tau)$, i.e. $N^\star$ is the total number of events occurring in the observation period.

For every $k = 1, \ldots, K$, we denote by $N_k(t)$ the number of immigration events of size j_k occurred in $[\tau_0, t]$, and $N_0(t)$ is the number of birth events occurred in $[\tau_0, t]$. Obviously, $N(t) = N_0(t) + N_1(t) + \cdots + N_K(t)$. $\{N_{0:K}(t), \tau_0 \leq t \leq \tau_0 + \tau\}$ is a multivariate counting process with multiplicative intensity $(\nu_0(t)X(t^-), \nu_1(t), \ldots, \nu_K(t))$ where $X(t^-) = \lim_{s \to t, s<t} X(t)$ and $X(t)$ is the number of particles at time t. We have

$$X(t) = X(\tau_0) + \sum_{k=0}^{K} j_k N_k(t). \tag{3}$$

Let $T_1, \ldots, T_{N^\star}$ be the occurrence times of the events during the observation period $[\tau_0, \tau_0 + \tau]$. Let Z_i be a discrete variable representing the type of the jth event: $Z_i \in \{0, \ldots, K\}$ is equal to k if the ith event is of type k, then we have

$$X(T_i) = X(T_{i-1}) + \sum_{k=0}^{K} j_k \mathbb{1}_{Z_i = k} = X(T_{i-1}) + j_{Z_i}.$$

Definition 1. *With these notations, the process is said to be* fully observed *if* $N_{0:K}(\cdot)$ *is continuously observed on* $[\tau_0, \tau_0 + \tau]$*, or equivalently, if the total number of events* $N^\star$*, the time events* $\{T_i\}_{i=1\dots N^\star}$ *and the nature of the events* $\{Z_i\}_{i=1\dots N^\star}$ *are observed.*

Remark 3. *Note that continuously observing* $N_{0:K}(\cdot)$ *on* $[\tau_0, \tau_0+\tau]$ *is not equivalent to continuously observing* $X(\cdot)$ *(the number of particles) unless* $j_0 \notin \{j_1, \dots, j_K\}$*. Indeed, if* j_0 *is equal to one of the* $j_1, \dots, j_K$*, then a birth and an immigration event will lead to the same increase of particles.*

In the fully observed setup, the likelihood is (see Andersen et al. (1993)):

$$\mathcal{L}(N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}; \theta, X(\tau_0)) = \prod_{k=0}^{K} \prod_{i=1}^{N^\star} \nu_k(T_i, \theta)^{\mathbb{1}_{Z_i=k}} \prod_{i=1}^{N^\star} X(T_{i-1})^{\mathbb{1}_{Z_i=0}} \times \exp\left[-\sum_{i=1}^{N^\star+1} X(T_{i-1}) \int_{T_{i-1}}^{T_i} \nu_0(t,\theta)dt - V(\tau,\theta)\right] \tag{4}$$

where $T_0 = \tau_0$, $T_{N^\star+1} = \tau + \tau_0$ and $V(\tau, \theta) = \sum_{k=1}^{K} \int_{\tau_0}^{\tau_0+\tau} \nu_k(t,\theta)dt$. This quantity is referred to as the *complete likelihood*. Note that in the case of time independent intensities the complete likelihood simplifies into

$$\mathcal{L}(N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}; \theta, X(\tau_0)) = \prod_{k=0}^{K} \nu_k^{N_k(\tau_0+\tau)} \prod_{i=1}^{N^\star} X(T_{i-1})^{\mathbb{1}_{Z_i=0}} \times \exp\left[-\nu_0 \sum_{i=1}^{N^\star} (T_i - T_{i-1}) X(T_{i-1}) - \nu_\bullet \tau\right] \tag{5}$$

where $\nu_\bullet = \sum_{k=1}^{K} \nu_k$

In Section 3.1, we consider the estimation of θ when the process is fully observed. Section 3.2 deals with the case where the time events $\{T_i\}_{i=1,\dots,N^\star}$ are observed but the nature of the events and the initial number of particles $X(\tau_0)$ are non or partially observed. The estimation procedures are detailed for the case where the intensities are constant but the cases of time dependent densities are discussed at each step.

3 Bayesian inference

3.1 Estimation from the complete observation

In this section, we assume that we completely observe the multivariate process $\{N_{0:K}(t), \tau_0 \le t \le \tau_0 + \tau\}$ and the initial number of particles $X(\tau_0)$ is known.

From (4), we can derive the identifiability of the parameter θ.

Proposition 1. *Assume that* $(\mathcal{H}_0)$ *holds. Let* θ *and* θ' *be two sets of parameters such that for any complete dataset* $(N^\star, (T_i, Z_i)_{i=1,\dots,N^\star})$*,*

$$\mathcal{L}(N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}; \theta) = \mathcal{L}(N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}; \theta').$$

Then $\theta = \theta'$.

The proof is given in Appendix A.

Bayesian estimation Inference under such models obviously depends on the parametric form of the $\nu_j(t;\theta)$. In order to understand the role of $X(\tau_0)$ in the estimation, we detail the simplest case where $\nu_k(t) = \nu_k$. In this case $\theta = (\nu_0, \dots, \nu_K)$.

Constant intensity rates In the case of constant intensity rates, we set Gamma prior distributions on these parameters:

$$\nu_k \sim \Gamma(\alpha_k, \beta_k), \quad \forall k = 0, \dots, K, \quad \text{independently}$$

with $(\alpha_k, \beta_k) \in (\mathbb{R}^{*+})^2$ such that $E[\nu_k] = \frac{\alpha_k}{\beta_k}$.

In the fully observed case, we deduce from (5) that the Gamma distributions are conjugate and the posterior distributions on the ν_k's are given by:

$$\begin{array}{l} \nu_0 | N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}, X(\tau_0) \sim \Gamma\left(\alpha_0 + N_0(\tau_0+\tau), \beta_0 + \sum_{i=1}^{N^\star+1}(T_i - T_{i-1})X(T_{i-1})\right), \\ \nu_k | N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}, X(\tau_0) \sim \Gamma\left(\alpha_k + N_k(\tau_0+\tau), \beta_k + \tau\right) \quad \forall k = 1, \dots, K. \end{array} \tag{6}$$

As a consequence we obtain the following posterior expectation estimators:

$$\begin{array}{rclcl} \widehat{\nu_0} &=& E[\nu_0 | N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}, X(\tau_0)] &=& \frac{\alpha_0 + N_0(\tau_0+\tau)}{\beta_0 + \sum_{i=1}^{N^\star+1}(T_i - T_{i-1})X(T_{i-1})}, \\ \widehat{\nu_k} &=& E[\nu_k | N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}, X(\tau_0)] &=& \frac{\alpha_k + N_k(\tau_0+\tau)}{\beta_k+\tau}, \quad \forall k = 1, \dots, K. \end{array} \tag{7}$$

Role of $X(\tau_0)$ in the estimators From (7), we note that the estimators $\{\widehat{\nu_k}, k = 1, \dots, K\}$ only depend on the number of events of type k:

$$\widehat{\nu_k} = \frac{\alpha_k + N_k(\tau_0+\tau)}{\beta_k + \tau}, \quad \forall k = 1, \dots, K.$$

As a consequence, even if $X(\tau_0)$ is unobserved, we are able to estimate the immigration rates $(\nu_k)_{k=1\dots K}$, provided we observe the number of events of each type $(N_k(\tau_0 + \tau))_{k=1\dots K}$. It is not the case for ν_0. Indeed, using the following reformulation:

$$\sum_{i=1}^{N^\star+1}(T_i - T_{i-1})X(T_{i-1}) = \sum_{i=1}^{N^\star} T_i j_{Z_i} + \tau X(\tau_0) + (\tau+\tau_0)\sum_{k=0}^{K} N_k(\tau_0+\tau) j_k,$$

we obtain

$$\widehat{\nu_0} = \frac{\alpha_0 + N_0(\tau_0+\tau)}{\beta_0 + \sum_{i=1}^{N^\star} T_i j_{Z_i} + (\tau+\tau_0)\sum_{k=0}^{K} N_k(\tau_0+\tau) j_k + \tau X(\tau_0)}. \tag{8}$$

Expression (8) enlightens the influence of $X(\tau_0)$. We will see in Sections 4.2 and 4.3 how partial observation of either N_k or $X(\tau_0)$ impacts the quality of the inference.

Remark 4. *If the intensities are time dependent, the conditional distribution is likely to be non-conjugate, and we would have to resort to an ad-hoc Metropolis–Hastings algorithm to sample the posterior distribution.*

3.2 Estimation from the partial observation of the process

We now consider the case where we partially observe the process: more precisely, we observe all the instants of occurrences $T_{1:N^\star}$ and partially the types of the events $(Z_j)_{j=1\dots N^\star}$.

Let $\mathbf{Z}$ denote $(Z_1,\dots,Z_{N^\star})$. We introduce n_{nobs} and n_{obs} as the numbers of non-observed and observed event types, respectively. Obviously, we have $n_{nobs}+n_{obs}=N^\star$ and $0\le n_{obs}\le N^\star$. Let $\{i_1,\dots,i_{n_{nobs}}\}$ be the non-observed indices, $\mathbf{Z}_{nobs}=(Z_{i_1},\dots,Z_{i_{n_{nobs}}})$ is the vector composed of the non-observed Z_i's and $\mathbf{Z}_{obs}=\mathbf{Z}\setminus\mathbf{Z}_{nobs}$ is the vector composed of the observed Z_i's. We first consider the case where we estimate the parameter from the partial observation $(N^\star, T_{1:N^\star},\mathbf{Z}_{obs})$, $X(\tau_0)$ being known.

Case 1. $X(\tau_0)$ is known

The likelihood of the observations $(N^\star, T_{1:N^\star},\mathbf{Z}_{obs})$ is

$$\mathcal{L}(N^\star, T_{1:N^\star},\mathbf{Z}_{obs};\theta) = \sum_{\mathbf{z}\in\{0,\dots,K\}^{n_{nobs}}} \mathcal{L}(N^\star, T_{1:N^\star},\mathbf{Z}_{obs},\mathbf{z};\theta). \tag{9}$$

Interestingly, even if $n_{obs}=0$, i.e. if none of the types of events are observed, the parameter θ can still be identified (proof given in Appendix A). We consider the following assumption:

$\mathcal{H}_1$: $\nu_0(\cdot,\theta)$ is such that

$$\frac{e^{-j_0\int_t^{\tau_0+\tau}\nu_0(u,\theta)du}-1}{\nu_0(\tau+\tau_0;\theta)} = \frac{e^{-j_0\int_t^{\tau_0+\tau}\nu_0(u,\theta')dt}-1}{\nu_0(\tau+\tau_0;\theta')}, \quad \forall t\in[\tau_0,\tau+\tau_0] \Rightarrow \nu_0(\cdot,\theta)=\nu_0(\cdot,\theta').$$

Then

Proposition 2. *Assume that $\mathcal{H}_0$ and $\mathcal{H}_1$ hold. Let θ and θ' be two sets of parameters such that for any partial dataset $(N^\star, T_{1:N^\star},\mathbf{Z}_{obs})$,*

$$\mathcal{L}(N^\star, T_{1:N^\star},\mathbf{Z}_{obs};\theta)=\mathcal{L}(N^\star, T_{1:N^\star},\mathbf{Z}_{obs};\theta').$$

Then $\theta=\theta'$.

Remark 5. *$\mathcal{H}_1$ holds, for instance, when ν_0 is polynomial (see Appendix A Subsection A.4).*

Bayesian estimation For the ease of simplicity, we detail the Bayesian estimation for constant rates: $\nu_k(t)=\nu_k;\forall k=0,\dots,K$. As soon as n_{nobs} becomes moderately large, the sum in the likelihood (9) is intractable. Moreover, the conjugacy of the prior distributions is no more ensured. However, we can use a Gibbs algorithm to sample the posterior distribution which consists in sampling the latent types $\mathbf{Z}_{nobs}$, in the usual data augmentation scheme as proposed by Tanner and Wong (1987). This makes the use of the conjugate Gamma priors considered in Section 3.1 particularly useful, in the case where the intensities ν_j are assumed constant.

The non-observed types are simulated and updated iteratively. To generate a non-observed type Z_{i_l} conditionally on all the other quantities, we proceed in the following way. For each possible value of $Z_{i_l} \in \{0, \dots, K\}$, we compute its conditional probability $p_{i_l,k} = P(Z_{i_l} = k | N^\star, T_{1:N^\star}, \mathbf{Z}^{(\ell-1)} \setminus \{Z_{i_l}\}; \theta^{(\ell)})$, this probability being proportional to the complete likelihood evaluated for the $\mathbf{Z}$ where the missing type Z_{i_l} has been replaced by k. Once the probabilities have been computed, we sample $Z_{i_l}^{(\ell)} \sim \sum_{k=0}^{K} p_{i_l,k}\delta_{\{k\}}$ (where $\delta_\bullet$ is the Dirac distribution).

We thus have the following pseudo-code:

Posterior distribution sampling for partial observation with $X(\tau_0)$ known

- **At iteration** (0), initialize the algorithm on $\mathbf{Z}_{nobs}^{(0)}$ chosen arbitrarily.
- **At iteration** $(\ell \geq 1)$

Set $\mathbf{Z}^{(\ell-1)} = (\mathbf{Z}_{obs}, \mathbf{Z}_{nobs}^{(\ell-1)})$ and compute the following statistics:

$$\begin{array}{rcll} X^{(\ell-1)}(T_i) & = & X^{(\ell-1)}(T_{i-1}) + j_{Z_i^{(\ell-1)}} & \forall i = 1, \dots, N^\star, \\ N_k^{(\ell-1)} & = & \sum_{i=1}^{N^\star} \mathbb{1}_{Z_i^{(\ell-1)}=k} & \forall k = 0, \dots, K. \end{array}$$

[1.] Generate the parameters conditionally on $N^\star, X(\tau_0), T_{1:N^\star}, \mathbf{Z}^{(\ell-1)}$:

$$\nu_0^{(\ell)} | N^\star, T_{1:N^\star}, \mathbf{Z}^{(\ell-1)} \sim \Gamma\left(\alpha_0 + N_0^{(\ell-1)}(\tau_0 + \tau), \beta_k + \sum_{j=0}^{N^\star+1} (T_j - T_{j-1}) X^{(\ell-1)}(T_{j-1})\right),$$

$$\nu_k^{(\ell)} | N^\star, T_{1:N^\star}, \mathbf{Z}^{(\ell-1)} \sim \Gamma\left(\alpha_k + N_k^{(\ell-1)}(\tau_0 + \tau), \beta_k + \tau\right) \quad \forall k = 1, \dots, K.$$

[2.] Generate the non-observed event types $\mathbf{Z}_{nobs}^{(\ell)}$ conditionally on $(N^\star, T_{1:N^\star}, \mathbf{Z}_{obs}, \nu_0^{(\ell)}, \dots, \nu_K^{(\ell)})$:

Set $\widetilde{\mathbf{Z}} = \mathbf{Z}^{(\ell-1)}$, $\forall l = 1, \dots, n_{nobs}$.

[2.1] For any of the possible type $k = 0, \dots, K$, compute the conditional probability

$$p_{i_l,k} = P(Z_{i_l} = k | N^\star, T_{1:N^\star}, \mathbf{Z}^{(\ell-1)} \setminus \{Z_{i_l}\}; \theta^{(\ell)}) \propto \mathcal{L}(N^\star, T_{1:N^\star}, \widetilde{\mathbf{Z}}^{k,l}; \nu_0^{(\ell)}, \dots, \nu_K^{(\ell)})$$

with $\widetilde{\mathbf{Z}}_i^{k,l} = \widetilde{\mathbf{Z}}_i$, for $i \neq i_l$ and $\widetilde{Z}_{i_l}^{k,l} = k$.

[2.2] Generate $Z_{i_l}^{(\ell)} \sim \sum_{k=0}^{K} p_{i_l,k}\delta_{\{k\}}$.

[2.3] In $\widetilde{\mathbf{Z}}$ replace its i_lth component by $Z_{i_l}^{(\ell)}$ and return to [2.1] with $l := l + 1$ until $l = n_{nobs}$ and set $\mathbf{Z}^{(\ell)} = \widetilde{\mathbf{Z}}$.

Remark 6. *Note that in practice, the order of simulation of the latent types Z_{i_l} is chosen randomly at each iteration. For the sake of simplicity in the previous pseudo-code, this point has not been detailed.*

Remark 7. *In case of time dependent rates, the Gibbs algorithm has to be adapted. Indeed, for a general form of the rates, the various conditional distributions may not be explicit anymore. Consequently, we may have to resort to the use of Metropolis–Hastings kernels which have to be designed for each particular case.*

In Section 4.2, we illustrate the influence of the partial or non-observation of $\mathbf{Z}$ on the quality of estimation of θ. $X(\tau_0)$ characterizes the state of the system at the beginning of the study. However, in situations where the Z_j's are only partially observed, it is often the case that $X(\tau_0)$ is not observed either. Inference in this case can be dramatically impacted by a miss-specification of $X(\tau_0)$ (see Figures 4 and 5). We present in the following section our inference procedure when $X(\tau_0)$ is not observed.

Case 2. $X(\tau_0)$ is unknown

In this section, we assume that $X(\tau_0)$ is not observed and has to be inferred. We first prove the identifiability of the model:

Proposition 3. *Let $(\theta, X(\tau_0))$ and $(\theta', X'(\tau_0))$ be two sets of parameters such that for any partial dataset $(N^\star, T_{1:N^\star}, \mathbf{Z}_{obs})$,*

$$\mathcal{L}(N^\star, T_{1:N^\star}, \mathbf{Z}_{obs}; \theta, X(\tau_0)) = \mathcal{L}(N^\star, T_{1:N^\star}, \mathbf{Z}_{obs}; \theta', X'(\tau_0)).$$

Then $\theta = \theta'$ as soon as $\mathcal{H}_1$ and $\mathcal{H}_0$ hold.

The proof is given in the Appendix A for the least favourable case when $n_{obs} = 0$.

Prior derivation on $X(\tau_0)$ Since $X(\tau_0)$ has a strong influence on the inference, the choice of its prior π is a key issue. A first solution is to propose a uniform distribution on $\{x(\tau_0)^-, \ldots, x(\tau_0)^+\} \subset \mathbb{N}$: $X(\tau_0) \sim \mathcal{U}_{\{x(\tau_0)^-, \ldots, x(\tau_0)^+\}}$. In practice, $x(\tau_0)^-$ and $x(\tau_0)^+$ would typically be elicited using expert knowledge. However, when $x(\tau_0)^-$ is much smaller than $x(\tau_0)^+$, posterior inference on the other parameters can become too diffuse to be of any practical use, see Section 4.3, Figure 4 for a numerical illustration of this phenomenon.

An alternative is to use the probabilistic structure of the counting process $N_{0:K}$ to construct a coherent prior distribution on $X(\tau_0)$. It is often the case (see, for instance, linear assets, as in our motivating example based on the electrical network) that although $X(\tau_0)$ is not known, the state of the network at its installation – several decades prior to the beginning of the study at time τ_0 – is known. When the observation period starts, the system has evolved until a certain number $X(\tau_0)$ of particles. As a consequence we propose to derive the prior distribution on $X(\tau_0)$ from the asymptotic distribution of the number of particles.

Proposition 4 gives the exact distribution of $X(t)$ for all $t \geq 0$ through its moment generating function, in terms of $X(0) = x_0$, the $\{j_k, k = 0, \dots, K\}$ and the intensity rates $\{\nu_k(\cdot), k = 0, \dots, K\}$. Theorem 1 provides a more explicit expression of its asymptotic distribution as t goes to infinity under some conditions on the j_k's and ν_k's.

Proposition 4. *Let $X(t)$ be the number of particles issued from the pure birth multi-immigration process described in Section 2. We assume that $X(0) = x_0$. We set $V(t) = \sum_{k=1}^K \int_0^t \nu_k(u)du$ and $V_0(t) = \int_0^t \nu_0(u)du$. Then we have*

$$\Phi(s,t) = E[s^{X(t)}] = \left[1 - e^{j_0 V_0(t)}(1 - s^{-j_0})\right]^{-x_0/j_0} e^{-V(t)} \exp\{\mathcal{J}(s,t)\}$$

where

$$\mathcal{J}(s,t) = \sum_{k=1}^K \int_0^t \nu_k(u) \left[1 - (1 - s^{-j_0}) \exp(j_0 V_0(t-u))\right]^{-j_k/j_0} du.$$

A developed expression of $\Phi(s,t)$ for the electrical network is given in Section 4, see (10).

A power series development supplies $X(t)$'s probability distribution. As a consequence, a first way to propose a prior distribution on $X(\tau_0)$ would be to use that exact distribution. However, the calculations can be burdensome. In case where τ_0 is large enough, we propose to use the asymptotic distribution instead of the exact distribution. In some cases, this asymptotic distribution is quite easy to handle and can be used as prior distribution on $X(\tau_0)$. This asymptotic distribution is given in Theorem 1.

Theorem 1. *Let $X(t)$ be the number of particles issued from the pure birth multi-immigration process described in Section 2. We assume that $X(0) = x_0$ and the following two conditions:*

(i) $\forall k = 1, \dots, K$, $j_k/j_0 = r_k \in \mathbb{N}^*$.

(ii) For all $k \geq 1$ $\nu_k(t) = \nu_k$ and there exists $t_1 > 0$ such that $\nu_0(t) = \nu_{0,1}\mathbb{1}_{t \leq t_1} + \nu_{0,2}\mathbb{1}_{t > t_1}$ with $0 < \nu_{0,1} \leq \nu_{0,2}$.

Then using the same notations as in Proposition 4, with $\nu_\bullet = \sum_{k=1}^K \nu_k$,

$$e^{-j_0 V_0(t)} X(t) \xrightarrow[t \to \infty]{\mathcal{L}} \Gamma\left(\frac{x_0}{j_0}, \frac{1}{j_0}\right) + \sum_{l=0}^{r_K - 1} Z_l$$

where the Z_l's are independent random variables with $Z_0 \sim \Gamma(\frac{\nu_\bullet}{\nu_{0,2} j_0}, \frac{1}{j_0})$ and for $l = 1, \dots, r_K - 1$,

$$Z_l \sim \sum_{j=1}^\infty \omega_{j,l} \Gamma\left(jl, \frac{1}{j_0}\right)$$

with $\omega_{j,l} = e^{\lambda_l} \frac{\lambda_l^j}{j!}$, $\lambda_l = \frac{\alpha_l}{l \nu_{0,2} j_0}$, $\alpha_l = \nu_\bullet$, $\forall l \in \{1, \dots, r_1 - 1\}$ and $\alpha_l = \nu_l + \cdots + \nu_K$, $\forall l \in \{r_{k-1}, \dots, r_k - 1\}$, $\forall k = 2, \dots, K$.

The proof is given in Appendix B. Theorem 1 shows that as τ_0 becomes large, conditional on the ν_j's and x_0, $X(\tau_0)$'s distribution can be approximated by $e^{j_0 V_0(\tau_0)}$ times the sum of infinite mixtures of Gamma random variables. In other words, it increases exponentially quickly with τ_0. A numerical illustration of the precision of this approximation is illustrated in Section 4.3, in the case where $\nu_{0,1} = \nu_{0,2}$. It is interesting to note that in the simple case where only ν_0 is allowed to vary by taking two possible values, the limiting rate $j_0 V_0(\cdot)$ depends on the whole function $\nu_0(\cdot)$, but the remaining part of the distribution only depends on $\nu_{0,2}$.

Neglecting the modification of the system through time can lead to strongly biased estimation, as soon as $V_0(\tau_0) j_0$ is not negligible. For an intermediate value of τ_0 it is possible to improve the approximation by re-centring the distribution using the true mean of $X(\tau_0)$ which can be deduced from the Laplace transform given in Proposition 4. We denote by π_∞^R the re-centred asymptotic distribution.

Remark 8. *Note that the result applies for the electrical context described previously.*

Posterior sampling for Bayesian inference We now detail the posterior sampling algorithm for $X(\tau_0)$ and θ in the case of constant rates. In this case, the parameter of interest is $(\theta, X(\tau_0)) = (\nu_0, \ldots, \nu_K, X(\tau_0))$, and we set the hierarchical prior distribution:

$$\begin{aligned} X(\tau_0)|(\nu_0, \ldots, \nu_K) &\sim \pi_\infty^R(X(\tau_0); \nu_0, \ldots, \nu_K), \\ \nu_k &\sim \Gamma(\alpha_k, \beta_k), k = 0, \ldots, K. \end{aligned}$$

With this new prior distribution on $X(\tau_0)$ the model is not fully conjugate (see (5) with $\pi_\infty^R(X(\tau_0); \theta)$ equal to the infinite Poisson mixture of Gamma distributions). As a consequence, we have to resort to a Metropolis–Hastings algorithm to sample the posterior distribution. The proposal distributions on $X(\tau_0)$ and $(\nu_0, \ldots, \nu_K)$ are detailed hereafter and have proved their efficiency in the simulation study.

Posterior distribution sampling for partial observation with $X(\tau_0)$ unknown

- **At iteration** (0), initialize the algorithm on $(\mathbf{Z}_{nobs}^{(0)}, \theta^{(0)}, X^{(0)}(\tau_0))$.
- **At iteration** (ℓ)

[1.] For $k = 0, \ldots, K$

- Propose $\log \widetilde{\nu_k} = \log \nu_k^{(\ell-1)} + \sigma_k \mathcal{N}(0,1)$ and set $\widetilde{\theta} = (\nu_0^{(\ell)}, \ldots, \nu_{k-1}^{(\ell)}, \widetilde{\nu_k}, \nu_{k+1}^{(\ell-1)}, \ldots, \nu_K^{(\ell-1)})$.
- Compute

$$\begin{aligned} \alpha_k \quad &= \min \Bigg\{ 1, \frac{\mathcal{L}\left(N^\star, T_1, \ldots, T_{N^\star}, \mathbf{Z}_{nobs}^{(\ell-1)}, \mathbf{Z}_{obs}; \widetilde{\theta}, X^{(\ell-1)}(\tau_0)\right)}{\mathcal{L}\left(N^\star, T_1, \ldots, T_{N^\star}, \mathbf{Z}_{nobs}^{(\ell-1)}, \mathbf{Z}_{obs}; \theta^{(\ell-1)}, X^{(\ell-1)}(\tau_0)\right)} \frac{\pi_\infty^R\left(X^{(\ell-1)}(\tau_0)|\widetilde{\theta}\right)}{\pi_\infty^R\left(X^{(\ell-1)}(\tau_0)|\theta^{(\ell-1)}\right)} \\ &\qquad \frac{\pi(\widetilde{\nu_k})}{\pi\left(\nu_k^{(\ell-1)}\right)} \frac{q\left(\nu_k^{(\ell-1)}|\widetilde{\nu_k}\right)}{q\left(\widetilde{\nu_k}|\nu_k^{(\ell-1)}\right)} \Bigg\}. \end{aligned}$$

- Set

$$\theta^{(\ell)} = \begin{cases} \widetilde{\theta} & \text{with probability} \quad \alpha_k, \\ \theta^{(\ell-1)} & \text{with probability} \quad 1-\alpha_k. \end{cases}$$

[2.] Generate non-observed event types $\mathbf{Z}_{nobs}$ using step [2.] in the algorithm presented in Section 3.2.
[3.] Generate $X(\tau_0)$

- Propose $\widetilde{X}(\tau_0) \sim \pi_\infty^R(\cdot|\theta^{(\ell)})$.
- Compute

$$\alpha(X^{(\ell-1)}(\tau_0), \widetilde{X}(\tau_0)) \quad = \min\left\{1, \frac{\mathcal{L}(N^\star, T_1, \dots, T_{N^\star}, \mathbf{Z}_{nobs}^{(\ell)}, \mathbf{Z}_{obs}; \widetilde{X}(\tau_0), \theta^{(\ell)})}{\mathcal{L}(N^\star, T_1, \dots, T_{N^\star}, \mathbf{Z}_{nobs}^{(\ell)}, \mathbf{Z}_{obs}; X^{(\ell-1)}(\tau_0), \theta^{(\ell)})}\right\}.$$

- Set

$$X(\tau_0)^{(\ell)} = \begin{cases} \widetilde{X}(\tau_0) & \text{with probability} \quad \alpha(X^{(\ell-1)}(\tau_0), \widetilde{X}(\tau_0)), \\ X^{(\ell-1)}(\tau_0) & \text{with probability} \quad 1-\alpha(X^{(\ell-1)}(\tau_0), \widetilde{X}(\tau_0)). \end{cases}$$

The σ_k's are tuned to achieve an reasonable acceptance rate. As described above, $\widetilde{X}(\tau_0)$ is proposed using the prior distribution π_∞^R. This choice has proved its efficiency (for instance, with respect to a random walk).

4 Numerical studies

We now illustrate and study our model and procedure using an extensive simulation study. All the simulations take place in the electrical network context evoked in the introduction. We first suppose that the initial state of the process $X(\tau_0)$ is known and we study the influence of the non-observation of $\mathbf{Z}$ on the quality of estimation of the parameters. In a second part, we assume that $X(\tau_0)$ is unknown and has to be estimated too: we first describe various prior distributions derived from the asymptotic behaviour of the process and then we compare the results obtained when a uniform prior on $X(\tau_0)$ is chosen with those obtained with its asymptotic distribution. Finally, we conduct a study on a pseudo-real dataset.

4.1 Notations and useful quantities

Recall that in our framework the electrical network is composed of electrical cables and accessories. A failure of a joint implies the birth of a new accessory and a failure of the cable implies the immigration of two accessories. As a consequence, we have

$$j_0 = 1, \quad K = 1, \quad j_1 = 2.$$

From now on, we use the following notations: $\nu_0 := \nu_a$ is the failure rate on the accessories; $\nu_1 := \nu_c d$ is the failure rate of the cable and so the immigration rate where d is the length of the cable; ν_a and ν_c are the parameters of interest. We set Gamma prior distributions, $\nu_a \sim \Gamma(\alpha_a, \beta_a)$ and $\nu_c \sim \Gamma(\alpha_c, \beta_c)$. The number of accessories at the time of installation $X(0)$ is known and equal to x_0. Applying Proposition 4 to this case, we have the following expression for the probability generating function:

$$\begin{aligned}\Psi(s,t) &= E[s^{X(t)}] \\ &= e^{-\nu_c dt}\left[1 - s(1-e^{-\nu_a t})\right]^{-\rho} e^{-\rho s} e^{\rho(1-e^{\nu_a t}(1-1/s))^{-1}} (1-e^{\nu_a t}(1-1/s))^{-x_0}\end{aligned} \tag{10}$$

where $\rho = \nu_c d/\nu_a$. From (10) we can deduce

$$\begin{array}{rcl} E[X(\tau_0)] &=& e^{\tau_0\nu_a}\left[2\rho(1-e^{-\tau_0\nu_a}) + x_0\right], \\ V[X(\tau_0)] &=& (e^{\tau_0\nu_a}-1)\left[x_0 e^{\tau_0\nu_a} + \rho(3e^{\tau_0\nu_a}-1)\right]. \end{array} \tag{11}$$

From Theorem 1, we derive the asymptotic distribution of the number of accessories

$$e^{-\tau_0\nu_a} X(\tau_0) \xrightarrow[t\to\infty]{\mathcal{L}} Z \quad \text{where} \quad Z \overset{\mathcal{L}}{=} \textstyle\sum_{k=0}^{\infty} e^{-\rho}\frac{\rho^k}{k!}\Gamma\left(k+\rho+x_0, 1\right). \tag{12}$$

We denote by π_∞ the asymptotic distribution of $X(\tau_0)$ given by (12). The corresponding asymptotic expectation and variance are

$$E_{\pi_\infty}[X(\tau_0)] = e^{\tau_0\nu_a}\left[2\rho + x_0\right] \quad V_{\pi_\infty}[X(\tau_0)] = e^{2\tau_0\nu_a}\left[3\rho + x_0\right]. \tag{13}$$

These quantities will be used in the following numerical experiments.

4.2 $X(\tau_0)$ is known, influence of partial observation of Z

In this first numerical experiment, we suppose that $X(\tau_0)$ is known and study the influence of the amount of missing data on the estimation of ν_a and ν_c. The data are simulated with

$$\nu_a = 10^{-4}, \quad \nu_c = 2\times 10^{-6}, \quad X(\tau_0) = 400, \quad \tau = 10 \text{ years}, \quad d = 8000,$$

which are realistic values in an electrical network. With these parameter values we simulate 100 datasets whose summary statistics are given in Table 1. For each dataset,

	min	mean	max
$N^\star$	199	254.08	308
$N_a(\tau_0+\tau)$	150	197.18	240
$N_c(\tau_0+\tau)$	34	56.90	75

Table 1: Simulation study 1 ($X(\tau_0)$ known and fixed). Statistics of the datasets: number of events $N^\star$, number of failures on the accessories $N_a(\tau_0+\tau)$, number of failures on the cable $N_c(\tau_0+\tau)$.

we sample from the posterior distribution of ν_a and ν_c in the following 4 scenarios.

		Scenario 1	Scenario 2	Scenario 3	Scenario 4
ν_a	Relative Bias (%)	–0.85	–0.99	–1.46	–3.36
	RMSE (%)	6.58	7.14	8.31	8.66
ν_c	Relative Bias (%)	–2.12	–3.09	–1.47	4.76
	RMSE (%)	12.34	14.06	18.48	11.48

Table 2: Simulation study 1 ($X(\tau_0)$ known and fixed): relative bias and RMSE (percentages) for $\widehat{\nu}_a$ and $\widehat{\nu}_c$ in the 4 scenarios.

- *Scenario 1.* We suppose that the whole sequence $Z_1, \dots, Z_{N^\star}$ is observed. In this context, the posterior distribution of (ν_a, ν_c) has an explicit expression given by:

$$\begin{aligned} \nu_a | T_1, \dots, T_{N^\star}, Z_1, \dots, Z_{N^\star}, X(\tau_0) &\sim \Gamma\Big(\alpha_a + N_a(\tau_0+\tau), \\ &\qquad \beta_a + \textstyle\sum_{i=0}^{N^\star+1}(T_i - T_{i-1})X(T_{i-1})\Big), \\ \nu_c | T_1, \dots, T_{N^\star}, Z_1, \dots, Z_{N^\star}, X(\tau_0) &\sim \Gamma\left(\alpha_c + N_c(\tau_0+\tau), \beta_c + d\tau\right). \end{aligned} \tag{14}$$

- *Scenario 2.* One third of the $Z_1, \dots, Z_{N^\star}$ are unobserved (the unobserved Z_j are randomly chosen). In that case, the posterior distribution of (ν_a, ν_c) is sampled by a Gibbs algorithm described in Section 3.2, performed with 10.000 iterations and a burn-in period of 5000 iterations.

- *Scenario 3.* Two thirds of $Z_1, \dots, Z_{N^\star}$ are unobserved. The observed Z_j are randomly chosen among those of Scenario 2.

- *Scenario 4.* $Z_1, \dots, Z_{N^\star}$ are completely unobserved.

Remark 9. *In order to avoid too many figures in the paper, we do not provide trajectories of the Gibbs outputs for Scenarios* 2*,* 3 *and* 4*. Convergence assessment tools of the MCMC / Gibbs algorithms will be provided in Subsection 4.4, in the least favourable case when not only* $\boldsymbol{Z}$ *but also* $X(\tau_0)$ *have to be sampled.*

In Figure 2, we plot the prior and the four marginal posterior distributions of ν_a (top) and ν_c (bottom), one per scenario, for one typical dataset. As expected, the smaller n_{obs}, the more spread the posterior is. This phenomenon is enhanced when the sequence $Z_1, \dots, Z_{N^\star}$ is completely non-observed.

Denoting by $\widehat{\nu}_a^{(m)}$ and $\widehat{\nu}_c^{(m)}$ the posterior means of ν_a and ν_c respectively associated to dataset m, we compute the relative bias and relative root-mean-square-error respectively given by

$$Bias = \sum_{m=1}^{100} \frac{\widehat{\nu}^{(m)} - \nu}{\nu}, \quad RMSE = 10\sqrt{\sum_{m=1}^{100} \frac{(\widehat{\nu}^{(m)} - \nu)^2}{\nu^2}}$$

and displayed in Table 2 in percentages. As expected, the quality of estimation decreases when the number of observations decreases.

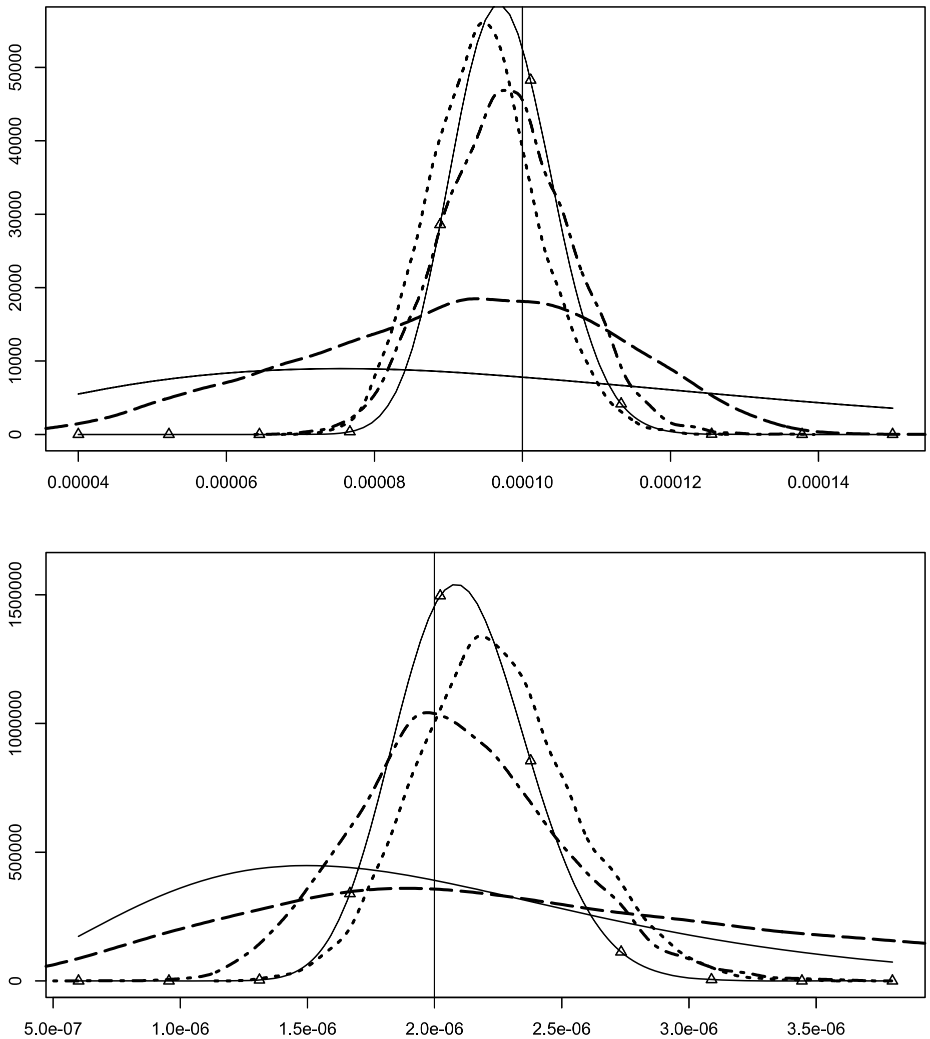

Figure 2: Influence of the non-observation of $Z_1, \dots, Z_{N^\star}$ on the posterior distributions of ν_a (upper figure) and ν_c (bottom figure). Prior distribution (plain line), posterior distribution in Scenario 1 (plain line with triangles), posterior distribution in Scenario 2 $(\cdots)$, posterior distribution in Scenario 3 $(\cdot - \cdot)$, posterior distribution in Scenario 4 (dashed line).

4.3 Estimation of (ν_a, ν_c) when $X(\tau_0)$ is unknown

In case where $X(\tau_0)$ is unknown, we have to set its prior distribution. As exposed before, we propose to deduce this prior distribution from the asymptotic properties of the counting process. To assess this choice, in a first step, we compare the true distribution of the number of accessories $X(t)$ with three other distributions deduced from the asymptotic behaviour. We then study the influence of this choice on the estimation of ν_a and ν_c.

Properties of the asymptotic approximation of the distribution of $(X(t))_{t\geq 0}$

To assess the validity of Theorem 1 and propose an adequate prior distribution, we conduct the following numerical study. We fix the parameter values to the following values:

$$\nu_a = 4.10^{-4}, \quad \nu_c = 4.10^{-6}, \quad d = 4000, \quad x_0 = 10,$$

and consider 3 values for τ_0: 5, 25 and 130 years. For each τ_0, we simulate 10 000 trajectories of our branching process starting at x_0 and store the number of accessories at the end of the period. The true distribution of the number of accessories is estimated from these samples through a kernel estimation method.

1. We first compare it to the asymptotic distribution π_∞ given by (12) (plotted in Figure 3 with squares $\square$).

2. By comparing the true and asymptotic expectations – given in (11) and (13) – we deduce that the asymptotic distribution has an inflated expectation with respect to the true one. As a consequence, we propose to correct the asymptotic distribution by re-centring it around the true expectation. We denote by π^R_∞ the re-centred asymptotic distribution (plotted with triangles $\triangle$ in Figure 3).

3. Finally, one could be attracted by the use of a simple Poisson distribution with mean $E[X(\tau_0)]$ given by (11) (plotted with circles $\circ$ in Figure 3). This choice could be considered as an easier alternative prior to avoid the use of the asymptotic distribution (12).

The density functions are plotted in Figure 3. We observe that for a long elapsed time ($\tau_0 = 130$ years, bottom figure) the asymptotic, the true and the re-centred distributions overlap and cannot be distinguished. The Poisson distribution is far much narrower and has not been plotted in the bottom panel. For an intermediate time period ($\tau_0 = 25$ years), the asymptotic distribution overestimates the number of accessories and cannot be used as a good approximation. However, the re-centred asymptotic distribution is a much better approximation, still retaining heavier tails than the true one. In the perspective of its use as a prior distribution on $X(\tau_0)$, this makes it a reasonable option. On the contrary, the Poisson distribution is far too narrow to be used as a prior distribution. This phenomenon could have been deduced from the comparison of the variances (given in (11) and (13)) but is well illustrated on the plot. When the time period is really short ($\tau_0 = 5$ years) the re-centred asymptotic distribution is much larger than the true one. As a consequence, this choice of prior distribution can appear to be less interesting. However, taking into account the fact that the asymptotic distribution has an explicit density expression and does not require the tuning of any hyper-parameter, it finally stays competitive with respect to the exact distribution or a uniform prior distribution, from an implementation point of view. We will see in the following that this choice also leads to better estimation results when compared to the uniform distribution.

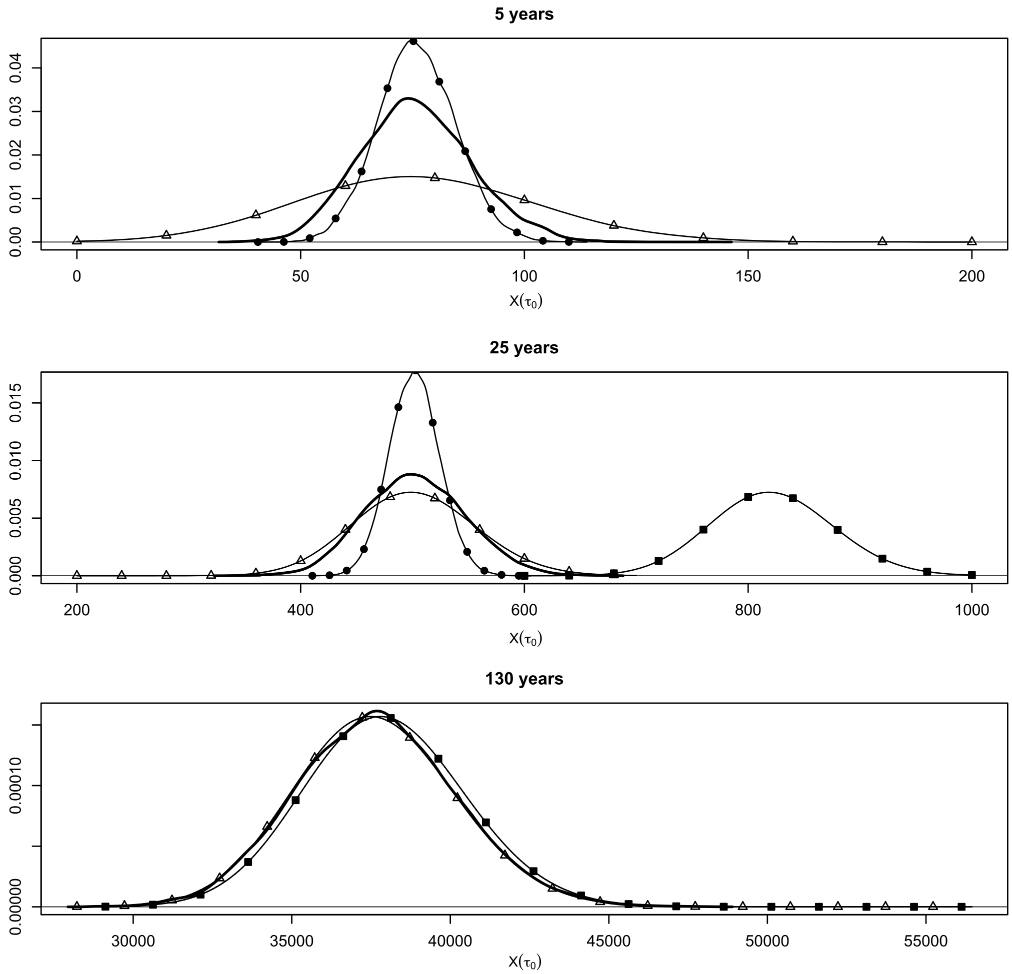

Figure 3: Distribution of $X(\tau_0)$: the estimated true distribution (plain line), asymptotic distribution (□), re-centred asymptotic distribution (△), Poisson distribution (○); (top) $\tau_0 = 5$ years, (middle) $\tau_0 = 25$ years, (bottom) $\tau_0 = 130$ years.

Influence of $X(\tau_0)$ and estimation

We now pay attention to the way to infer $X(\tau_0)$ (either fixing or estimating it using the previously described priors) and will study how the strategy influences the estimation of ν_a and ν_c.

When $X(\tau_0)$ is unknown we consider several solutions: fixing it at some arbitrary value or estimating it using either a uniform prior distribution or using the re-centred asymptotic distribution as a prior. Using the following parameter values:

$$\nu_a = 1.10^{-5}, \quad \nu_c = 4.10^{-6}, \quad d = 4000,$$
$$x_0 = 0, \quad \tau_0 = 40 \text{ years}, \quad \tau = 15 \text{ years},$$

we simulate 100 trajectories starting at $x_0 = 0$ during a time interval of length $\tau + \tau_0$. For these datasets, the statistics are summarized below, in Table 3.

	min	mean	max
$X(\tau_0)$	432	504.95	650
$N^\star$	87	121.26	151
$N_a(\tau_0+\tau)$	19	32.77	48
$N_c(\tau_0+\tau)$	66	88.49	112

Table 3: Simulation study 2 (influence of $X(\tau_0)$), summary statistics of the datasets.

Now, for each dataset we estimate the parameters ν_a and ν_c using 4 scenarios. Note that in order to separate the sources of imprecision, in this study, we suppose that $Z_1, \dots, Z_{N^\star}$ are observed.

• *Scenario 0.* Scenario 0 will refer to the case where $X(\tau_0)$ is known (and so set to its true value) and will be our reference. In that case, the model is conjugate and the posterior distributions on ν_a and ν_c have been given by (14).

• *Scenario 1.* The naive solution is to set $X(\tau_0)$ to its value at the instant of installation of the network (x_0), neglecting the evolution of the process between the installation and the beginning of the study. However, in our simulation study x_0 is equal to 0 and $X(\tau_0)$ is around 500, leading to dramatically bad estimations. Instead, we set $X(\tau_0)$ to $X(\tau_0)/2$.

• *Scenario 2.* We consider a uniform prior distribution: $X(\tau_0) \sim \mathcal{U}_{\{x(\tau_0)^- \cdots x(\tau_0)^+\}}$ with $x(\tau_0)^- = 100$ and $x(\tau_0)^+ = 1000$.

• *Scenario 3.* We consider the re-centred asymptotic distribution π_∞^R on $X(\tau_0)$ given by (12).

Results In Figure 4, we plot the posterior densities of ν_a (upper) and ν_c (bottom) for one arbitrarily chosen dataset. As expected, $X(\tau_0)$ does not influence the posterior distribution of ν_c and the posterior densities corresponding to the 4 scenarios nearly overlap. On the contrary, the posterior density for ν_a clearly depends on $X(\tau_0)$. If $X(\tau_0)$ is under-evaluated (Scenario 1), the posterior density of ν_a (dashed line) is shifted to the right: this phenomenon was clearly expected from (8). When a prior on $X(\tau_0)$ is considered, the re-centred asymptotic prior distribution clearly outperforms the uniform prior distribution: first of all, the asymptotic prior distribution does not require the elicitation of the support $\{x(\tau_0)^- \cdots x(\tau_0)^+\}$ and above all the posterior distribution for ν_a is clearly narrower and closer to the reference posterior distribution (Scenario 0) when π_∞^R is used.

Remark 10. *Note that the implementation of the Metropolis–Hastings algorithm for $X(\tau_0)$ requires the evaluation of $\pi_\infty^R(X(\tau_0)|\nu_a, \nu_c)$:*

$$\pi_\infty^R(X(\tau_0)|\nu_a,\nu_c) = e^{-\tau_0\nu_a}\sum_{k=0}^{\infty} e^{-\rho}\frac{\rho^k}{k!} f_{\Gamma(k+\rho+x_0,1)}(e^{-\tau_0\nu_a}(X(\tau_0)+2\rho))$$

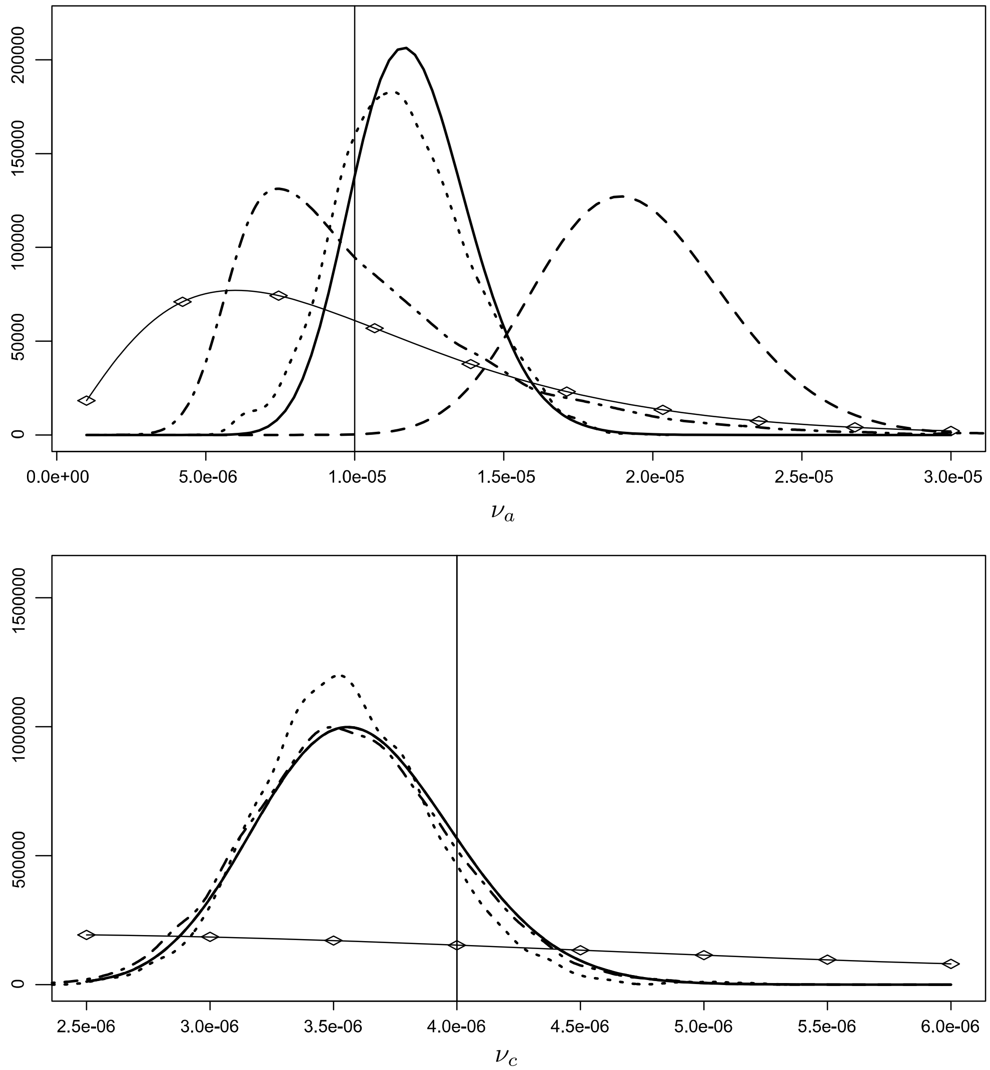

Figure 4: Influence of the non-observation of $X(\tau_0)$ on the posterior distributions of ν_a (upper figure) and ν_c (bottom figure) for one dataset: prior distribution (plain line with diamonds), posterior distribution with the true $X(\tau_0)$ (Scenario 0) (plain line), posterior distribution with under-evaluated $X(\tau_0)$ (Scenario 1) (dashed line), posterior distribution with a uniform prior distribution on $X(\tau_0)$ (Scenario 2) ($\cdot - \cdot$) and posterior distribution with asymptotic prior distribution on $X(\tau_0)$ (dotted line).

where $f_{\Gamma(a,b)}$ is the density function of the Gamma distribution of parameters (a, b). In practice, this infinite summation has to be truncated, depending on the current value of θ: in our algorithm, we have used a truncated version of the prior π^R_∞ defined by

$$\pi_\infty^R(X(\tau_0)|\nu_a,\nu_c) = e^{-\tau_0\nu_a}\sum_{k=0}^{K(\theta)} e^{-\rho}\frac{\rho^k}{k!S_\theta} f_{\Gamma(k+\rho+x_0,1)}(e^{-\tau_0\nu_a}(X(\tau_0)+2\rho)),$$

with $S_\theta = \sum_{k=0}^{K(\theta)} e^{-\rho}\rho^k/k!$ *and*

$$K(\theta) = \inf\left\{k \geq \rho | e^{-\rho}\frac{\rho^k}{k!} < 10^{-324}\right\}.$$

We compute for each dataset and each scenario the posterior means and variances of ν_a and ν_c which we denote by denote by $E[\nu_a^{\ell,s}|\mathbf{Y}^\ell]$, $E[\nu_c^{\ell,s}|\mathbf{Y}^\ell]$, $Var[\nu_a^{\ell,s}|\mathbf{Y}^\ell]$ and $Var[\nu_c^{\ell,s}|\mathbf{Y}^\ell]$, respectively, with $s \in \{0,\dots,3\}$ and $\ell \in \{1,\dots,100\}$ indexing the scenario and the dataset, respectively. In Figure 5, a summary of these quantities is presented in the form of their posterior densities.

As already remarked, the posterior variance and expectation of ν_c are not influenced by $X(\tau_0)$ (Figure 5, bottom left and right). On the contrary, an under-estimated $X(\tau_0)$ leads to a large positive bias on ν_a (Figure 5, top left, density in dashed line). When a uniform prior distribution is used on $X(\tau_0)$ we observe a large posterior variance on ν_a (Figure 5, top right, density in $\cdot - \cdot$ line) whereas the use of the re-centred asymptotic prior distribution on $X(\tau_0)$ leads to a much more sensible posterior variance (Figure 5, top right, density in dotted line).

4.4 Estimation on pseudo-real dataset

To illustrate the performance of our prior distribution and our estimation method, we propose to consider a pseudo-real dataset, that is to say, a realistic context with $X(\tau_0)$ unknown and $\mathbf{Z}$ partially unobserved.

We consider a $d = 40000$ metre network, $\tau_0 = 25$ year old at the beginning of the study. The study lasts $\tau = 4$ years, and half of the breakdowns types are reported whereas the other ones are unknown. x_0 is known to be equal to 10. Among the $N^\star = 276$ occurred breakdowns, half of them are observed ($n_{obs} = 138$). Among the observed breakdowns, 114 are due to the cable and 24 take place on the accessories.

We use the centred asymptotic prior distribution π_∞^R for $X(\tau_0)$.

Initialization and tuning of the MCMC algorithm We sample the posterior distribution using the MCMC algorithm described in Section 3.2. Since ν_c does not depend on $X(\tau_0)$ but only on the number of events taking place on the cable, we initialize

$$\nu_c^{(0)} = \frac{N^\star}{n_{obs}}\frac{1}{\tau d}\sum_{i=1}^{n_{obs}} \mathbb{1}_{Z_{obs,i}=1}$$

using the fraction of event types we observe. No easy way to initialize ν_a could be found: in the following, we generate 10 chains starting from starting from 10 different

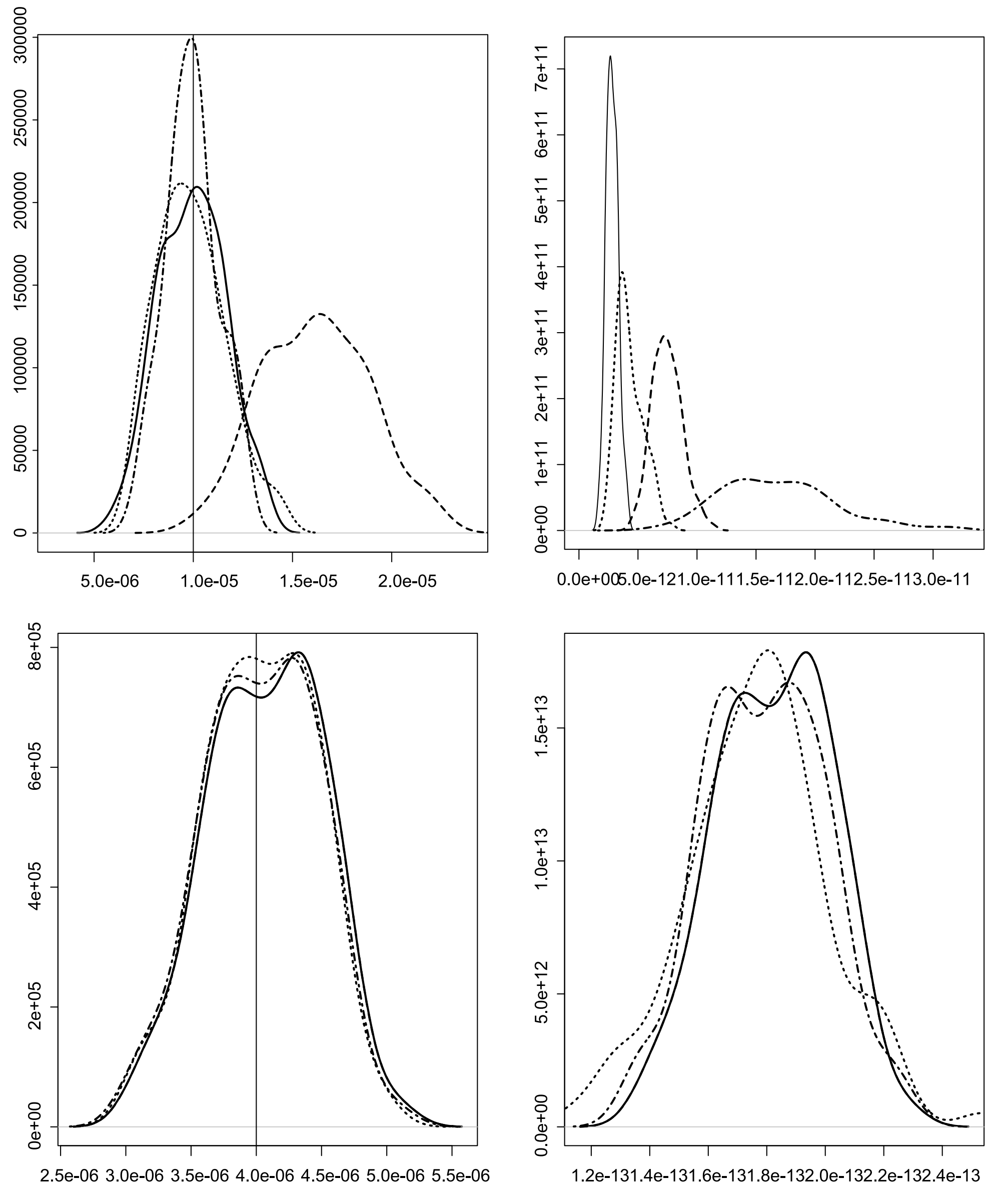

Figure 5: Influence of the non-observation of $X(\tau_0)$ on the posterior expectation (left) and variance (right) of ν_a (top) and ν_c (bottom) for the 100 datasets: estimated density with the true $X(\tau_0)$ (Scenario 0) (plain line), with under-evaluated $X(\tau_0)$ (Scenario 1) (dashed line, $(----)$), with a uniform prior distribution on $X(\tau_0)$ (Scenario 2) $(\cdot - \cdot)$ and with asymptotic prior distribution on $X(\tau_0)$ (dotted line $(\cdots)$).

$\nu_a^{(0)}$ regularly distributed between 10^{-4} and 2.10^{-2}. Conditionally on $\nu_a^{(0)}$, we set

$$X^{(0)}(\tau_0) = e^{\tau_0 \nu_a^{(0)}} \left(2 \frac{\nu_c^{(0)} d}{\nu_a^{(0)}} (1 - e^{-\tau_0 \nu_a^{(0)}}) + x_0 \right).$$

The $Z_{nobs,i}$'s are all initialized at 0. Some trials with other initializations on $Z_{nobs,i}$ ($Z_{nobs,i}$ randomly chosen to be 0 or 1) were run without showing any interesting effect. The random walks on $\log \nu_a$ and $\log \nu_c$ are implemented with $\sigma_a = 0.2$ and $\sigma_c = 0.05$ leading to acceptation rates around 60%.

Convergence assessment of the MCMC algorithm

The algorithm is implemented with 50000 iterations. A period of burn-in of 20000 iterations is removed. As described before, we generate 10 MCMC chains, starting from 10 different $\nu_a^{(0)}$ regularly distributed between 10^{-4} and 2.10^{-2}. We use these 10 chains to assess the convergence of the MCMC algorithm. More precisely, for each of them, we compute the variance intra-chain and compare it to the inter-chains variance through the potential scale reduction factor (Gelman and Rubin, 1992). We obtain

$$R_{\nu_a} = 1.014351 \quad R_{\nu_c} = 1.003598 \quad R_{X(\tau_0)} = 1.0021661,$$

all of them being smaller than 1.2. We also provide two graphical tools. For one arbitrarily chosen chain, we plot the auto-correlation function for the three parameters of interest (see Figure 7). The autocorrelations quickly decrease for $X(\tau_0)$ and ν_c, the decrease is slower for ν_a but remains acceptable. Finally, in Figure 6, we plot – for the same chain – the trajectories produced by the MCMC, assessing that we have reached the stationary distribution.

In Figure 8, we plot the posterior distributions of $X(\tau_0)$, ν_a and ν_c estimated by kernel density estimation from the concatenation of one over 5 values from the last 25000 iterations of the 10 chains. The prior distribution is plotted in dashed line. It is interesting to note that even when $X(\tau_0)$ is unknown, the posterior distributions of ν_a and ν_c lead to accurate estimation of these parameters.

5 Discussion and possible extensions of the model

In this section, we discuss some directions in which the model can be extended.

First, in this model, we assume that the $Z_1, \dots, Z_{N(\tau)}$ are partially observed. An other interesting scenario would be to consider a mis-reporting of the event types Z_j's. More precisely we observe types of events $Z_1^r, \dots, Z_{N(\tau)}^r$ which are reported with error, defined by a probabilistic model $P[Z^r|Z]$. Writing the new full likelihood

$$\tilde{\mathcal{L}}(N^*, (T_i, Z_i, Z_i^r)_{i=1,\dots,N^*}, X(\tau_0)) = \mathcal{L}(N^*, (T_i, Z_i)_{i=1,\dots,N^*}, X(\tau_0)) \times \prod_{i=1}^{N^*} P[Z_i^r|Z_i]$$

we obtain a tractable posterior distribution on $(\boldsymbol{Z}, X(\tau_0), \theta)$ which we can simulate using a Gibbs sampler.

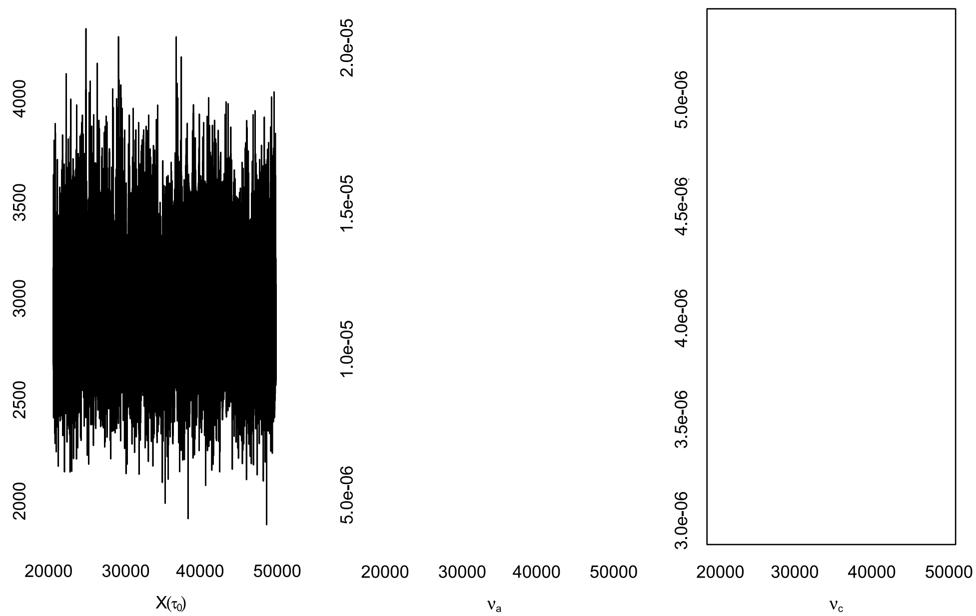

Figure 6: *Pseudo-real dataset.* Trajectories of the MCMC algorithm: (left) $X(\tau_0)$, (middle) ν_a, (right) ν_c.

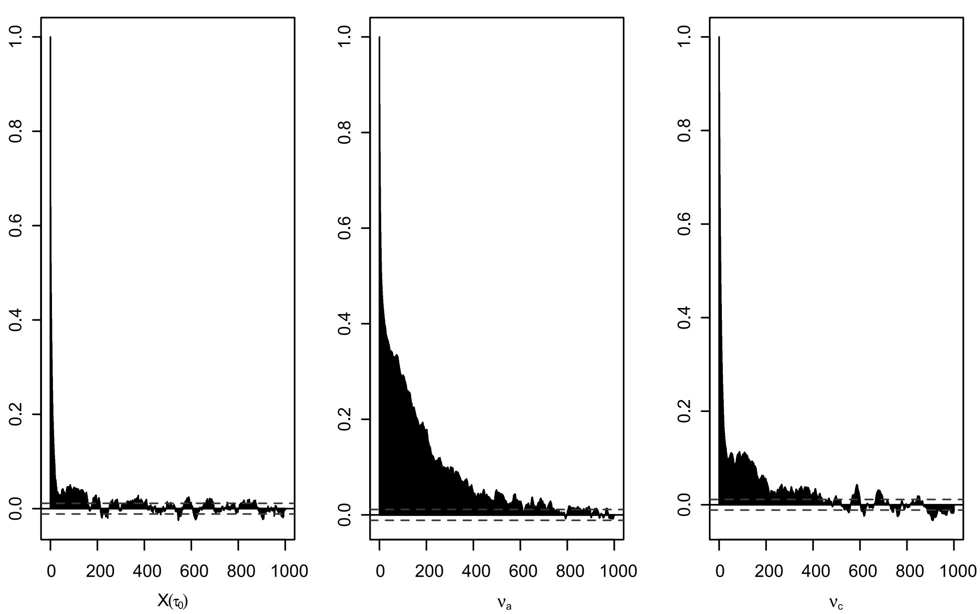

Figure 7: *Pseudo-real dataset.* Autocorrelation functions of the output of one particular chain for the 3 parameters: (left) $X(\tau_0)$, (middle) ν_a, (right) ν_c.

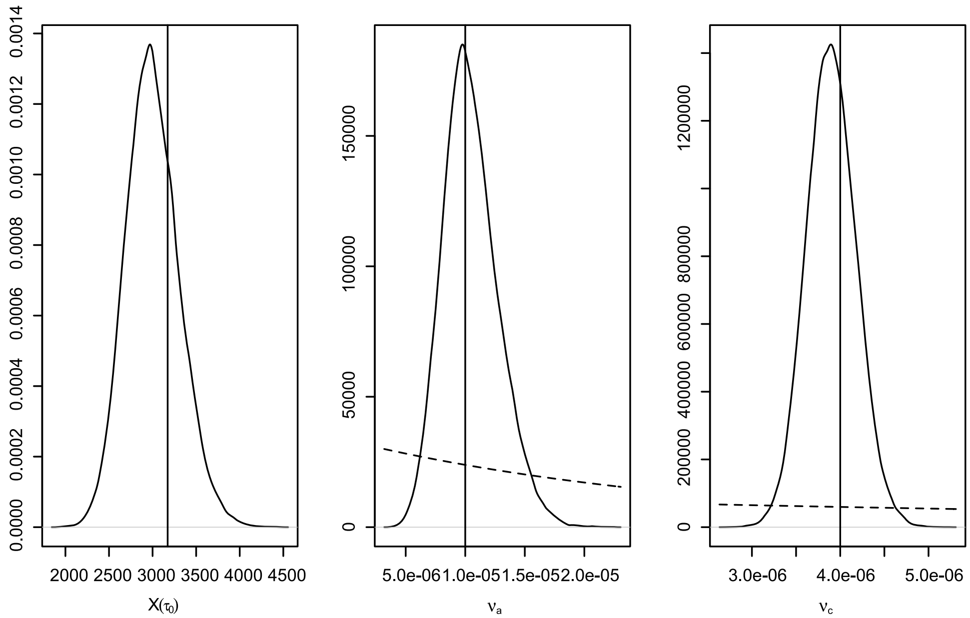

Figure 8: *Pseudo-real dataset*. Posterior distribution (plain) line and prior distribution in dashed line: (left) $X(\tau_0)$, (middle) ν_a, (right) ν_c.

In our description of the model and methodology, emphasis has been put on the case where the rates ν_j are constant. We have already explained how the methodology can be extended to the case where they depend on time in a parametric way. The structure of the algorithms would remain the same, apart from the possible loss in conjugacy so that Metropolis–Hasting steps within Gibbs might have to be considered in such situations, depending on the parametric form of the function $\nu_j(t;\theta)$. Depending of the form of $\nu_j(t;\theta)$, and using Proposition 4, extensions of Theorem 1 to cases where the ν_k's are allowed to vary could be obtained.

Another direct extension from our model is to consider covariates which do not vary with time. In that case, a hierarchical formulation of our Bayesian model can be stated as follows. Let C denote the covariate taking values in a set $\mathcal{C}$ (typically $\mathcal{C}$ would be finite). Then given C, define a process $(N_C(t), X_C(t), t \in [0,T])$ as in Section 2 with parameters $\nu_C = (\nu_{C,0}, \dots, \nu_{C,K})$, assume that the parameters ν_C are independent and identically distributed from the prior distribution proposed in Section 3.1.

An easier way to consider aging in the system is to say that after a given time $\tau^\star$, the accessories are replaced by a new type of material with their proper failure rate $\nu^\star$. In that context, we would have a multi-type counting process. Let $X^\star(t)$ denote the number of new type-accessories and $X(t)$ the number of old type accessories. After τ^*, at each event (immigration or birth) $X(t)$ decreases and $X^\star(t)$ increases conjointly. The study of that process and the estimation of the parameters would remain essentially the same as the one presented in the paper.

Appendix A: Proof of identifiability (Propositions 1, 2 and 3)

Let us recall the expression of the complete likelihood:

$$\mathcal{L}(N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}; \theta, X(\tau_0)) = \left[\prod_{k=0}^{K}\prod_{i=1}^{N^\star} \nu_k(T_i,\theta)^{\mathbb{1}_{Z_i=k}} \prod_{i=1}^{N^\star} X(T_{i-1})^{\mathbb{1}_{Z_i=0}}\right] \times \exp\left[-\sum_{j=1}^{N^\star+1} X(T_{j-1})\int_{T_{j-1}}^{T_j}\nu_0(t,\theta)dt - \sum_{k=1}^{K}\int_{\tau_0}^{\tau_0+\tau}\nu_k(t,\theta)dt\right] \tag{15}$$

where $T_0 = \tau_0$, $T_{N^\star+1} = \tau + \tau_0$.

A.1 Proof of Proposition 1

We prove the identifiability of the model from the complete observation of the process (Proposition 1) and $X(\tau_0)$ known. Let θ and θ' be such that for any complete observation of the process, we have

$$\mathcal{L}(N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}; \theta, X(\tau_0)) = \mathcal{L}(N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}; \theta', X(\tau_0)). \tag{16}$$

Note that (16) has to be verified for any possible dataset $N^\star, (T_i, Z_i)_{i=1,\dots,N^\star}$. As a consequence, we propose to deduce the identifiability from two particular cases, namely $N^\star = 0$ and $N^\star = 1$. These special values present the advantage to make the calculus easy.

- For the particular case where $N^\star = 0$, applying (15) gives

$$\exp\left[-X(\tau_0)\int_{\tau_0}^{\tau+\tau_0}\nu_0(t,\theta)dt - \sum_{k=1}^{K}\int_{\tau_0}^{\tau+\tau_0}\nu_k(t,\theta)dt\right] = \exp\left[-X(\tau_0)\int_{\tau_0}^{\tau+\tau_0}\nu_0(t,\theta')dt - \sum_{k=1}^{K}\int_{\tau_0}^{\tau+\tau_0}\nu_k(t,\theta')dt\right],$$

or equivalently,

$$\begin{aligned} R(\theta, X(\tau_0)) &= \exp\left[-\textstyle\sum_{k=0}^{K}\int_{\tau}^{\tau+\tau_0}\nu_k(t,\theta)X(\tau_0)^{\mathbb{1}_{k=0}}dt\right] \\ &= \exp\left[-\textstyle\sum_{k=0}^{K}\int_{\tau}^{\tau+\tau_0}\nu_k(t,\theta')X(\tau_0)^{\mathbb{1}_{k=0}}dt\right] = R(\theta', X(\tau_0)). \end{aligned} \tag{17}$$

- Now, if we observe $N^\star = 1$ event, which is a birth ($Z_1 = 0$) occurring at time T_1, we can write the likelihood as (using the fact that $X(T_1) = X(\tau_0) + j_0$)

$$\begin{aligned}
&\mathcal{L}(1, T_1, Z_1; \theta, X(\tau_0)) \\
&\quad = \nu_0(T_1, \theta) X(\tau_0) \exp\left[-X(\tau_0) \int_{\tau_0}^{T_1} \nu_0(t,\theta)dt - X(T_1) \int_{T_1}^{\tau_0+\tau} \nu_0(t,\theta)dt \right. \\
&\qquad \left. - \sum_{k=1}^{K} \int_{\tau_0}^{\tau+\tau_0} \nu_k(t,\theta)dt \right] \\
&\quad = \nu_0(T_1, \theta) X(\tau_0) \exp\left[-j_0 \int_{T_1}^{\tau_0+\tau} \nu_0(t,\theta)dt - \sum_{k=0}^{K} \int_{\tau_0}^{\tau+\tau_0} \nu_k(t,\theta) X(\tau_0)^{\mathbb{1}_{k=0}} dt \right].
\end{aligned}$$

Applying (17), the equality of likelihoods implies

$$X(\tau_0)\nu_0(T_1,\theta) \exp\left[-j_0 \int_{T_1}^{\tau_0+\tau} \nu_0(t,\theta)dt \right] = X(\tau_0)\nu_0(T_1,\theta') \exp\left[-j_0 \int_{T_1}^{\tau_0+\tau} \nu_0(t,\theta')dt \right] \tag{18}$$

$\forall T_1 \in [\tau_0, \tau_0 + \tau]$, which is equivalent to having $\forall x \in [\tau_0, \tau_0 + \tau]$

$$\frac{\partial \exp\left[-j_0 \overline{V}_0(x;\theta) \right]}{\partial x} = \frac{\partial \exp\left[-j_0 \overline{V}_0(x;\theta') \right]}{\partial x}$$

where $\overline{V}_0(x;\theta) = \int_x^{\tau_0+\tau} \nu_0(t,\theta)dt$. Hence,

$$\exp[-j_0 \overline{V}_0(x;\theta)] = \exp[-j_0 \overline{V}_0(x;\theta')] + C, \quad \forall x \in [\tau_0, \tau + \tau_0].$$

Using the fact that $\overline{V}_0(\tau + \tau_0, \theta) = \overline{V}_0(\tau + \tau_0, \theta') = 0$, we obtain $C = 0$ and, by derivation,

$$\nu_0(t,\theta) = \nu_0(t,\theta') \quad \forall t \in [\tau_0, \tau + \tau_0].$$

• Assuming now that $N^\star = 1$ and $Z_1 = k$ for any $k = 1, \ldots, K$, we once again write the corresponding likelihood. Using the fact that $X(T_1) = X(\tau_0) + j_k$, we obtain

$$\nu_k(t,\theta) \exp\left[-j_k \overline{V}_0(t;\theta) \right] = \nu_k(t,\theta') \exp\left[-j_k \overline{V}_0(t;\theta') \right].$$

Using $\nu_0(t,\theta) = \nu_0(t,\theta')$, we obtain

$$\nu_k(t,\theta) = \nu_k(t,\theta') \quad \forall k = 1 \ldots K, \forall t \in [\tau_0, \tau + \tau_0].$$

Finally, we have $\nu_k(t,\theta) = \nu_k(t,\theta')$, $\forall k = 0, \ldots, K, \forall t \in [\tau_0, \tau + \tau_0]$ and so by condition $\mathcal{H}_0$, $\theta = \theta'$.

A.2 Proof of Proposition 2

We now assume that we observe $N^\star$, $(T_j)_{j=1,\ldots,N^\star}$, $X(\tau_0)$ is known, but we do not observe $Z_1, \ldots, Z_{N^\star}$.

- As before, we set $N^\star = 0$, which leads to $R(\theta, X(\tau_0)) = R(\theta', X(\tau_0))$.

- Now, we set $N^\star = 1$. In this case, the likelihood is written as a marginal version of (15) (integrated over Z_1),

$$\begin{aligned}
\mathcal{L}(1, T_1; \theta, X(\tau_0)) &= \textstyle\sum_{k=0}^K \nu_k(T_1, \theta)(X(\tau_0))^{\mathbb{1}_{k=0}} \\
&\quad \exp\Big[-X(\tau_0)\int_{\tau_0}^{T_1} \nu_0(t,\theta)dt - (X(\tau_0) + j_k)\int_{T_1}^{\tau_0+\tau} \nu_0(t,\theta)dt \\
&\quad -\textstyle\sum_{k=1}^K \int_{\tau_0}^{\tau+\tau_0} \nu_k(t,\theta)dt\Big] \\
&= \exp\Big[-\textstyle\sum_{k=0}^K \int_{\tau_0}^{\tau+\tau_0} X(\tau_0)^{\mathbb{1}_{k=0}}\nu_k(t,\theta)dt\Big] \\
&\quad \textstyle\sum_{k=0}^K \nu_k(T_1,\theta)(X(\tau_0))^{\mathbb{1}_{k=0}} \exp\Big[-j_k\int_{T_1}^{\tau_0+\tau} \nu_0(t,\theta)dt\Big] \\
&= R(\theta, X(\tau_0))\textstyle\sum_{k=0}^K \nu_k(T_1,\theta)(X(\tau_0))^{\mathbb{1}_{k=0}} e^{-j_k\int_{T_1}^{\tau_0+\tau}\nu_0(t,\theta)dt}.
\end{aligned}$$

Let us introduce for $n \geq 0$ and $t \in [\tau_0, \tau_0 + \tau]$

$$M^n(t, \theta, X(\tau_0)) = \sum_{k=0}^K j_k^n \nu_k(T_1,\theta)(X(\tau_0))^{\mathbb{1}_{k=0}} e^{-j_k\int_{T_1}^{\tau_0+\tau}\nu_0(t,\theta)dt}.$$

$\mathcal{L}(1, T_1; \theta, X(\tau_0)) = \mathcal{L}(1, T_1; \theta', X(\tau_0))$ implies, using (17),

$$M^0(t, \theta, X(\tau_0)) = M^0(t, \theta', X(\tau_0)) \quad \forall t \in [\tau_0, \tau + \tau_0]. \tag{19}$$

- Now, set $N^\star = 2$. Then the likelihood is

$$\begin{aligned}
&\mathcal{L}(2, T_1, T_2; \theta, X(\tau_0)) \\
= \;& e^{-\sum_{k=1}^K \int_{\tau_0}^{\tau+\tau_0}\nu_k(t,\theta)dt} \sum_{k_1,k_2=0}^K \nu_{k_1}(T_1,\theta)(X(\tau_0))^{\mathbb{1}_{k_1=0}}\nu_{k_2}(T_2,\theta)(X(T_1))^{\mathbb{1}_{k_2=0}} \\
&\exp\Bigg[-X(\tau_0)\int_{\tau_0}^{T_1}\nu_0(t,\theta)dt - (X(\tau_0)+j_{k_1})\int_{T_1}^{T_2}\nu_0(t,\theta)dt \\
&-(X(\tau_0)+j_{k_1}+j_{k_2})\int_{T_2}^{\tau_0+\tau}\nu_0(t,\theta)dt\Bigg] \\
= \;& R(\theta, X(\tau_0))\sum_{k_1=0}^K \nu_{k_1}(T_1,\theta)(X(\tau_0))^{\mathbb{1}_{k_1=0}} e^{-j_{k_1}\int_{T_1}^{\tau_0+\tau}\nu_0(t,\theta)dt} \\
&\left[\sum_{k_2=1}^K \nu_{k_2}(T_2,\theta)e^{-j_{k_2}\int_{T_2}^{\tau_0+\tau}\nu_0(t,\theta)dt} + \nu_0(T_2,\theta)(X(\tau_0)+j_{k_1})e^{-j_0\int_{T_2}^{\tau_0+\tau}\nu_0(t,\theta)dt}\right].
\end{aligned}$$

Denoting

$$\phi_0(t,\theta) \;=\; \nu_0(t,\theta)e^{-j_0\int_t^{\tau_0+\tau}\nu_0(u,\theta)u},$$

we have

$$\mathcal{L}(2, T_1, T_2; \theta, X(\tau_0)) = R(\theta, X(\tau_0))\sum_{k_1=0}^K \nu_{k_1}(T_1,\theta)(X(\tau_0))^{\mathbb{1}_{k_1=0}} e^{-j_{k_1}\int_{T_1}^{\tau_0+\tau}\nu_0(t,\theta)dt}$$

$$\begin{aligned}
& \times \left[\sum_{k_2=0}^{K} \nu_{k_2}(T_2,\theta) X(\tau_0)^{\mathbb{1}_{k_2=0}} e^{-j_{k_2} \int_{T_2}^{\tau_0+\tau} \nu_0(t,\theta)dt} + \nu_0(T_2,\theta) j_{k_1} e^{-j_0 \int_{T_2}^{\tau_0+\tau} \nu_0(t,\theta)dt} \right] \\
= \; & R(\theta, X(\tau_0)) \left[M^0(T_1,\theta,X(\tau_0)) M^0(T_2,\theta,X(\tau_0)) \right] \\
& + R(\theta, X(\tau_0)) \left[\phi_0(T_2,\theta) \sum_{k_1=0}^{K} j_{k_1} \nu_{k_1}(T_1,\theta) X(\tau_0)^{\mathbb{1}_{k_1=0}} e^{-j_{k_1} \int_{T_1}^{\tau_0+\tau} \nu_0(t,\theta)dt} \right] \\
= \; & R(\theta, X(\tau_0)) \left[M^0(T_1,\theta,X(\tau_0)) M^0(T_2,\theta,X(\tau_0)) + \phi_0(T_2,\theta) M^1(T_1,\theta,X(\tau_0)) \right].
\end{aligned} \tag{20}$$

Let θ and θ' such that for any $\tau_0 < T_1 < T_2 < \tau_0 + \tau$ we have

$$\mathcal{L}(2, T_1, T_2; \theta, X(\tau_0)) = \mathcal{L}(2, T_1, T_2; \theta', X(\tau_0)).$$

From (17) and (19) we have $R(\theta, X(\tau_0)) = R(\theta', X(\tau_0))$ and $M^0(T_1, \theta, X(\tau_0)) = M^0(T_1, \theta', X(\tau_0))$. As a consequence, for any T_1, T_2

$$\phi_0(T_2,\theta) M^1(T_1,\theta,X(\tau_0)) = \phi_0(T_2,\theta') M^1(T_1,\theta',X(\tau_0)), \tag{21}$$

or equivalently,

$$\frac{\phi_0(T_2,\theta)}{\phi_0(T_2,\theta')} = \frac{M^1(T_1,\theta',X(\tau_0))}{M^1(T_1,\theta,X(\tau_0))}. \tag{22}$$

So each ratio is a constant with respect to its variable, i.e. $\exists C_{\theta,\theta'}$ independent of T_2 so that

$$\begin{aligned}
\phi_0(T_2,\theta) &= \nu_0(T_2,\theta) e^{-j_0 \int_{T_2}^{\tau_0+\tau} \nu_0(t,\theta)dt} \\
&= C_{\theta,\theta'} \nu_0(T_2,\theta') e^{-j_0 \int_{T_2}^{\tau_0+\tau} \nu_0(t,\theta')dt} = C_{\theta,\theta'} \phi_0(T_2,\theta').
\end{aligned}$$

Setting $T_2 = \tau + \tau_0$, we can compute $C_{\theta,\theta'}$

$$\nu_0(\tau+\tau_0;\theta) = C_{\theta,\theta'} \nu_0(\tau+\tau_0;\theta').$$

So

$$\frac{\nu_0(T_2,\theta)}{\nu_0(\tau+\tau_0;\theta)} e^{-j_0 \int_{T_2}^{\tau_0+\tau} \nu_0(t,\theta)dt} = \frac{\nu_0(T_2,\theta')}{\nu_0(\tau+\tau_0;\theta')} e^{-j_0 \int_{T_2}^{\tau_0+\tau} \nu_0(t,\theta')dt}.$$

Integrating out this equality, there exists $D_{\theta,\theta'}$ such that

$$\frac{1}{\nu_0(\tau+\tau_0;\theta)} e^{-j_0 \int_{T_2}^{\tau_0+\tau} \nu_0(t,\theta)dt} = \frac{1}{\nu_0(\tau+\tau_0;\theta')} e^{-j_0 \int_{T_2}^{\tau_0+\tau} \nu_0(t,\theta')dt} + D_{\theta,\theta'}.$$

Once again, setting $T_2 = \tau + \tau_0$, we obtain

$$\frac{1}{\nu_0(\tau+\tau_0;\theta)} = \frac{1}{\nu_0(\tau+\tau_0;\theta')} + D_{\theta,\theta'}.$$

So $\nu_0(t,\theta)$ and $\nu_0(t,\theta')$ verify

$$\frac{e^{-j_0 \int_{T_2}^{\tau_0+\tau} \nu_0(t,\theta)dt} - 1}{\nu_0(\tau+\tau_0;\theta)} = \frac{e^{-j_0 \int_{T_2}^{\tau_0+\tau} \nu_0(t,\theta')dt} - 1}{\nu_0(\tau+\tau_0;\theta')}, \quad \forall t \in [\tau_0, \tau+\tau_0].$$

From assumption $\mathcal{H}_1$, we obtain

$$\nu_0(t;\theta) = \nu_0(t;\theta'), \quad \forall t \in [\tau_0, \tau + \tau_0]. \tag{23}$$

Combining (21) and (23), we get $M^1(T_1, \theta, X(\tau_0)) = M^1(T_1, \theta', X(\tau_0))$, i.e. for all $t \in [\tau_0, \tau + \tau_0]$,

$$\sum_{k_1=1}^{K} j_{k_1} \nu_{k_1}(t,\theta) e^{-j_{k_1} \int_t^{\tau_0+\tau} \nu_0(u,\theta)du} = \sum_{k_1=1}^{K} j_{k_1} \nu_{k_1}(t,\theta') e^{-j_{k_1} \int_t^{\tau_0+\tau} \nu_0(u,\theta')du}.$$

- The same type of calculus can be done for $N^\star = 3, 4, \dots$ and we obtain (for all $n \geq 2$)

$$\sum_{k=1}^{K} j_k^n \nu_k(t,\theta) e^{-j_k \int_t^{\tau_0+\tau} \nu_0(u,\theta)du} = \sum_{k=1}^{K} j_k^n \nu_k(t,\theta') e^{-j_k \int_t^{\tau_0+\tau} \nu_0(u,\theta)du}.$$

Moreover, since $0 < j_1 < \cdots < j_K$, letting n go to infinity in the above equation leads to

$$\nu_K(t,\theta) = \nu_K(t,\theta') \quad \forall t \in [\tau_0, \tau_0 + \tau],$$

and subsequently, $\forall k$

$$\nu_k(t,\theta) = \nu_k(t,\theta') \quad \forall t \in [\tau_0, \tau_0 + \tau]$$

which combined with $\mathcal{H}_0$ leads to $\theta = \theta'$.

A.3 Proof of Proposition 3

We now consider that $X(\tau_0)$ is unknown. The calculus are the same, except that $X(\tau_0)$ is now included in the parameters. We now obtain the following equations:

$$\begin{aligned} R(\theta, X(\tau_0)) &= \sum_{k=0}^{K} \int_{\tau_0}^{\tau+\tau_0} X(\tau_0)^{\mathbb{1}_{k=0}} \nu_k(t,\theta) dt \\ &= \sum_{k=0}^{K} \int_{\tau_0}^{\tau+\tau_0} X'(\tau_0)^{\mathbb{1}_{k=0}} \nu_k(t,\theta') dt = R(\theta', X'(\tau_0)), \end{aligned} \tag{24}$$

$$M^0(t, \theta, X(\tau_0)) \quad = \quad M^0(t, \theta', X'(\tau_0)) \tag{25}$$

and

$$\frac{\phi_0(T_2, \theta)}{\phi_0(T_2, \theta')} = \frac{M^1(T_1, \theta', X'(\tau_0))}{M^1(T_1, \theta, X(\tau_0))} \tag{26}$$

where $\phi_0(\cdot, \theta)$ and $M^1(\cdot, \theta, X(\tau_0))$ have been defined before. The same deductions can be done as before, and we obtain $\nu_0(t,\theta) = \nu_0(t,\theta')$. Exactly as in the proof of Proposition 2, we derive that for all $t \in [\tau_0, \tau + \tau_0]$,

$$M^n(T_1, \theta, X(\tau_0)) = M^n(T_1, \theta', X'(\tau_0)), \quad \forall n \geq 0,$$

or equivalently, for all $n \geq 0$, for all $t \in [\tau_0, \tau + \tau_0]$,

$$\sum_{k=1}^{K} j_k^n (X'(\tau_0))^{\mathbb{1}_{k=0}} \nu_k(t,\theta) e^{-j_k \int_t^{\tau_0+\tau} \nu_0(u,\theta)du} = \sum_{k=0}^{K} j_k^n X(\tau_0)^{\mathbb{1}_{k=0}} \nu_k(t,\theta') e^{-j_k \int_t^{\tau_0+\tau} \nu_0(u,\theta)du}$$

which leads, using the same arguments, to $X'(\tau_0) = X(\tau_0)$ and $\theta = \theta'$.

A.4 Comments on the assumption $\mathcal{H}_1$

We now verify assumption $\mathcal{H}_1$ when $\nu_0(\cdot,\theta)$ is polynomial in t. Recall that, under $\mathcal{H}_1$, $\nu_0(\cdot,\theta)$ is such that

$$\frac{e^{-j_0 \int_t^{\tau_0+\tau} \nu_0(u,\theta)du} - 1}{\nu_0(\tau+\tau_0;\theta)} = \frac{e^{-j_0 \int_t^{\tau_0+\tau} \nu_0(u,\theta')du} - 1}{\nu_0(\tau+\tau_0;\theta')}, \quad \forall t \in [\tau_0, \tau+\tau_0] \Rightarrow \nu_0(\cdot,\theta) = \nu_0(\cdot,\theta').$$

The assumption can be transformed into:

$$\begin{aligned} e^{-j_0 \int_t^{\tau_0+\tau} \nu_0(u,\theta)du} &= \frac{\nu_0(\tau+\tau_0;\theta)}{\nu_0(\tau+\tau_0;\theta')}(e^{-j_0 \int_t^{\tau_0+\tau} \nu_0(u,\theta')du} - 1) + 1 \quad \Rightarrow \\ -j_0 \int_t^{\tau_0+\tau} \nu_0(u,\theta)du &= \log\left[\frac{\nu_0(\tau+\tau_0;\theta)}{\nu_0(\tau+\tau_0;\theta')}(e^{-j_0 \int_t^{\tau_0+\tau} \nu_0(u,\theta')du} - 1) + 1\right]. \end{aligned} \quad (27)$$

If $\nu_0(\cdot,\theta)$ is a polynomial function, i.e. $\nu_0(t,\theta) = \sum_{k=0}^{P} \theta_k(\tau_0+\tau-t)^k$, then $\nu_0(\tau+\tau_0,\theta) = \theta_0$ and (27) becomes: $\forall u \in [0,\tau]$,

$$-j_0 \sum_{k=1}^{P+1} \frac{\theta_{k-1}}{k} u^k = \log\left[\frac{\theta_0}{\theta'_0}(e^{-j_0 \sum_{k=1}^{P+1} \frac{\theta'_{k-1}}{k} u^k} - 1) + 1\right].$$

The left-hand side is a polynomial function of degree $P+1$. Having a look, for instance, at the power series expression of the right-hand term of the equality, we clearly see that it cannot be a polynomial function of degree $P+1$, unless $\theta_0 = \theta'_0$. If $\nu_0(\tau+\tau_0,\theta) = \theta_0 = \theta'_0 = \nu_0(\tau+\tau_0,\theta')$, then, going back to the first equation, we obtain $\nu_0(t,\theta) = \nu_0(t,\theta')$ for all $t \in [\tau_0, \tau+\tau_0]$.

The result is directly extended for ν_0 being exponential. For any other function, the previous demonstration has to be adapted.

Appendix B: Asymptotic study of $(X(t))_{t\geq 0}$. Proof of Proposition 4 and Theorem 1

In Proposition 4, we give an integral expression of the generating function of the j_0-Yule process with multi-size immigration described in Section 2. More precisely, let $X(t)$ be a branching process such that each particle gives birth to j_0 particles at random time distributed with $\mathcal{E}(\nu_0)$ and such that groups of j_k immigrants arrive at random times distributed with $\mathcal{E}(\nu_k)$, for $k = 1, \ldots, K$. We assume that $X(0) = x_0$.

B.1 Proof of Proposition 4

The proof of Proposition 4 can be divided into two parts. First, the x_0 particles existing at time 0 will give birth to x_0 independent pure birth processes (each particle giving birth to j_0 particles). Once the processes derived from those particles have been taken into account, we can reduce the study to a pure birth process with multi-size immigration starting from 0 particles.

In Lemma 1, we recall the expression of the generating function of a pure birth process starting with one particle and study its asymptotic distribution. The generating function of the process of interest starting with $X(0) = 0$ particle is then detailed. The proof is similar to the one given in Shonkwiler (1980) but adapted to our particular case. In the particular case where the sizes of the immigration groups are proportional to j_0, we derive an explicit expression of the limiting distribution in Section B.2.

Lemma 1. *Let $Y(t)$ be a branching process starting with $Y(0) = y_0$ particle such that each particle gives birth to j_0 particles within a non-homogeneous Poisson process distribution of intensity $\nu_0(t)$. Then we have*

$$\Psi(s,t) = E[s^{Y(t)}] = \left[1 - (1 - s^{-j_0}) \exp(V_0(t) j_0)\right]^{-y_0/j_0}.$$

Moreover,

$$\lim_{t\to\infty} e^{-j_0 V_0(t)} Y(t) = \Gamma\left(\frac{y_0}{j_0}, \frac{1}{j_0}\right) \quad (\mathcal{L})$$

where $V_0(t) = \int_0^t \nu_0(u)du$.

Proof of Lemma 1. We first assume that $Y(0) = 1$. By construction of the process $Y(t)$, $Q_{i,j}(t,h) = P(Y(t+h) = j | Y(t) = i)$ only depends on t, h, i, j. When h is small, $Q_{i,j}(t,h)$ verifies

$$Q_{i,j}(t,h) = \begin{cases} \nu_0(t) i h + o(h) & \text{if} \quad j = i + j_0, \\ 1 - \nu_0(t) i h + o(h) & \text{if} \quad j = i, \\ o(h) & \text{if} \quad j \notin \{i, i + j_0\}. \end{cases} \tag{28}$$

We now derive a partial differential equation fulfilled by the probability-generating function. Using the fact that $Y(t)$ take its values in $\{j_0 k + 1, k \in \mathbb{N}\}$ we have

$$\Psi(s,t) = E[s^{Y(t)}] = \sum_{k\in\mathbb{N}} P(Y(t) = j_0 k + 1 | Y(0) = 1) s^{j_0 k + 1} = \sum_{k\in\mathbb{N}} Q_{1, j_0 k+1}(0,t) s^{j_0 k+1}. \tag{29}$$

Using a backward-equation, we derive an expression for $Q_{1,j_0 k+1}(0,t) = P(Y(t) = j_0 k + 1 | Y(0) = 1)$. Indeed,

$$Q_{1,j_0 k+1}(0, t+h) = P(Y(t+h) = j_0 k + 1 | Y(0) = 1)$$

$$= \sum_{l=0}^{\infty} P(Y(t+h) = j_0k+1 | Y(t) = j_0 l + 1) P(Y(t) = j_0 l + 1 | Y(0) = 1)$$
$$= \sum_{l=0}^{\infty} Q_{j_0 l+1, j_0 k+1}(t,h) Q_{1, j_0 l+1}(0,t)$$
$$= (1 - \nu_0(t)(j_0 k+1)h) Q_{1,j_0 k+1}(0,t) + Q_{1,j_0(k-1)+1}(0,t)\nu_0(t)(j_0(k-1)+1)h + o(h),$$

using (28). We directly obtain the following ODE:

$$\begin{aligned} Q'_{1,j_0k+1}(0,t) &= \lim_{h\to 0} \tfrac{Q_{1,j_0k+1}(0,t+h) - Q_{1,j_0k+1}(0,t)}{h} \\ &= -\nu_0(t)(j_0k+1)Q_{1,j_0k+1}(0,t) + \nu_0(t)(j_0(k-1)+1)Q_{1,j_0(k-1)+1}(0,t). \end{aligned} \quad (30)$$

We now derive from (30) and (29) a partial differential equation for $\Psi(s,t)$,

$$\frac{\partial}{\partial t}\Psi(s,t) = \sum_{k\in\mathbb{N}} Q'_{1,j_0k+1}(0,t)s^{j_0k+1} = \sum_{k=1}^{\infty} Q'_{1,j_0k+1}(0,t)s^{j_0k+1}$$
$$= \sum_{k=0}^{\infty} s^{j_0k+1}\left[-\nu_0(t)(j_0k+1)Q_{1,j_0k+1}(0,t) + \nu_0(t)(j_0(k-1)+1)Q_{1,j_0(k-1)+1}(0,t)\right]$$
$$= -\nu_0(t)\sum_{k=0}^{\infty} s^{j_0k+1}(j_0k+1)Q_{1,j_0k+1}(0,t)$$
$$+ \nu_0(t)\sum_{k=1}^{\infty}(j_0(k-1)+1)Q_{1,j_0(k-1)+1}(0,t)s^{j_0k+1}$$
$$= -\nu_0(t)\sum_{k=0}^{\infty} s^{j_0k+1}(j_0k+1)Q_{1,j_0k+1}(0,t) + \nu_0(t)\sum_{k=0}^{\infty}(j_0k+1)Q_{1,j_0k+1}(0,t)s^{j_0(k+1)+1}$$
$$= -\nu_0(t)s\frac{\partial}{\partial s}\Psi(s,t) + \nu_0(t)s^{j_0+1}\frac{\partial}{\partial s}\Psi(s,t)$$
$$= \nu_0(t)s(s^{j_0}-1)\frac{\partial}{\partial s}\Psi(s,t).$$

As a consequence, $\Psi(s,t)$ satisfies the following equation:

$$\frac{\partial}{\partial t}\Psi(s,t) = \nu_0(t)s(s^{j_0}-1)\frac{\partial}{\partial s}\Psi(s,t), \quad \Psi(s,0) = s. \quad (31)$$

The solution of this equation is

$$\Psi(s,t) = \left[1 - e^{V_0(t)j_0}(1 - s^{-j_0})\right]^{-1/j_0} \qquad \text{where } V_0(t) = \int_0^t \nu_0(u)du.$$

We now study the asymptotic distribution of $\tilde{Z}(t) = e^{-V_0(t)j_0}Y(t)$ through its moment-generating function

$$\Phi_{\tilde{Z}(t)}(\theta) = E[e^{\theta\tilde{Z}(t)}] = E[e^{\theta e^{-V_0(t)j_0}Y(t)}] = E[(e^{\theta e^{-V_0(t)j_0}})^{Y(t)}] = E[s_t^{Y(t)}] = \Psi(s_t,t)$$

where $s_t = e^{\theta e^{-V_0(t) j_0}} \simeq_{t\to\infty} 1 + \theta e^{-V_0(t) j_0}$ if $\lim_{t\to\infty} V_0(t) = \infty$.

We easily obtain the following limit, for all θ, with $|\theta| < 1/j_0$,

$$\lim_{t\to\infty} \Phi_{\tilde{Z}(t)}(\theta) = \frac{1}{(1 - j_0\theta)^{1/j_0}}$$

and recognise the moment generating function of the $\Gamma(\frac{1}{j_0}, \frac{1}{j_0})$.

Now, if the process starts with $Y(0) = y_0$ particles, each of them initiates a j_0-Yule process which is independent of the other ones, leading to

$$\Psi(s,t) = \left[1 - (1 - s^{-j_0}) \exp(V_0(t) j_0)\right]^{-y_0/j_0}$$

and

$$\lim_{t\to\infty} e^{-j_0 V_0(t)} Y(t) = \Gamma\left(\frac{y_0}{j_0}, \frac{1}{j_0}\right). \qquad (\mathcal{L}) \qquad \square$$

We now use Lemma 1 to study the distribution of $X(t)$, the number of particles issued from the multi-immigration j_0-Yule process described in Section 2. We first assume that $X(0) = 0$. Let $\phi(s,t)$ denote the probability-generating function of $(X(t))_{t\geq 0}$,

$$\phi(s,t) = E[s^{X(t)}] = \sum_{n=0}^{\infty} P_n(t) s^n$$

where $P_n(t) = P(X(t) = n)$ is the probability to have n particles at time t. This probability can be decomposed into

$$P_n(t) = P_{n|0}(t) m_0(t) + \sum_{k=1}^{\infty} P_{n|k}(t) m_k(t) \tag{32}$$

where $m_k(t)$ is the probability that k immigration groups arrived in the time interval $[0,t)$ and $P_{n|k}(t)$ denotes the probability there are n particles at time t given that k immigration groups arrived during $[0,t)$. Moreover, $P_{n|0}(t) = \delta_{n0}$ because $X(0) = 0$.

Using the independence of the immigration events, $P_{n|k}(t)$ can also be decomposed as

$$P_{n|k}(t) = \sum_{i_1+\cdots+i_k=n} U_{i_1}(t) \cdots U_{i_k}(t) \tag{33}$$

where $U_m(t)$ denotes the probability that an immigration group leads (by the branching mechanism) to m particles at time t given that the group immigrates during the interval $[0,t)$. Combining (33) and (32), we can rewrite the probability-generating function $\phi(s,t)$ as

$$\phi(s,t) \quad = \quad \sum_{n=0}^{\infty} P_n(t) s^n = m_0(t) + \sum_{n=0}^{\infty} s^n \sum_{k=1}^{\infty} m_k(t) \sum_{i_1+\cdots+i_k=n} U_{i_1}(t) \cdots U_{i_k}(t)$$

$$= \sum_{k=0}^{\infty} m_k(t) \left(\sum_{n=0}^{\infty} s^n U_n(t) \right)^k .$$

Denoting $J(s,t) = \sum_{n=0}^{\infty} s^n U_n(t)$, we obtain

$$\phi(s,t) = \sum_{k=0}^{\infty} m_k(t) J^k(s,t) \tag{34}$$

where $m_k(t)$ is the probability that k immigration groups arrived in the time interval $[0,t)$. Using the Poisson properties of our immigration process we have

$$m_k(t) = e^{-\mu(t)} \frac{(\mu(t))^k}{k!}, \quad \text{with} \quad \mu(t) = \sum_{j=1}^{K} \int_0^t \nu_j(u) du$$

and so by (34),

$$\phi(s,t) \quad = \quad \sum_{k=0}^{\infty} e^{-\mu(t)} \frac{\mu(t)^k}{k!} J^k(s,t) = e^{-\mu(t)} \exp\{\mu(t) J(s,t)\}. \tag{35}$$

We now compute $J(s,t)$ and to that purpose we first study $U_n(t)$. Recall that $U_n(t)$ is the probability that an immigration group leads (by the branching mechanism) to n particles at time t given that the group immigrates during the interval $[0,t)$. As a consequence, using the infinitesimal probabilities, $U_n(t)$ can be decomposed into

$$U_n(t) = \int_0^t \sum_{k=1}^{K} r_k(u) Q(n,t|j_k,u) d_u N(u|t) \tag{36}$$

where

- $d_u N(u|t)$ is the conditional infinitesimal immigration rate, i.e. the probability that there is exactly one immigration group during the infinitesimal interval $[u, u+du) \subset [0,t)$ given there is exactly one immigration group in the interval $[0,t)$. In the case of an inhomogeneous Poisson process, $d_u N(u|t) = \frac{d\mu(u)}{\mu(t)} = \frac{\nu_\bullet(u)du}{\mu(t)}$ where $\nu_\bullet(t) = \sum_{k=1}^{K} \nu_k(t)$.

- $r_k(u)$ is the probability that the immigration group is of size j_k given that it arrived at time u, $k = 1, \dots, K$. Using the Poisson properties of our immigration process, we have
$$r_k(u) = \frac{\nu_k(u)}{\nu_1(u) + \cdots + \nu_K(u)} = \frac{\nu_k(u)}{\nu_\bullet(u)}.$$

- $Q(n,t|j_k,u)$ denotes the probability that an immigration occurring at time u and consisting of j_k particles leads to n particles at time t. $Q(n,t|j_k,u)$ only relies on the branching part of the process and can be decomposed as previously

$$Q(n,t|j_k,u) = \sum_{i_1+\cdots+i_{j_k}=n} Q_{1,i_1}(t-u) \cdots Q_{1,i_{j_k}}(t-u) \tag{37}$$

where $Q_{1,i}(t)$ is the probability that one particle leads to i particles by division process in a period of length t. Indeed, once arrived, each particle is the initial point of a branching process which evolves independently during the remaining time $t-u$.

As a consequence, we can express $J(s,t)$ as

$$\begin{aligned} J(s,t) &= \sum_{n=0}^{\infty} s^n U_n(t) = \sum_{n=0}^{\infty}\sum_{k=1}^{K} \frac{1}{\mu(t)} s^n \int_0^t \nu_k(u) Q(n,t|j_k,u) du \\ &= \sum_{k=1}^{K} \frac{1}{\mu(t)} \int_0^t \nu_k(u) \sum_{n=0}^{\infty} s^n \sum_{i_1+\cdots+i_{j_k}=n} Q_{1,i_1}(t-u)\cdots Q_{1,i_{j_k}}(t-u) du \\ &= \sum_{k=1}^{K} \frac{1}{\mu(t)} \int_0^t \nu_k(u) \left[\sum_{n=0}^{\infty} s^n Q_n(t-u)\right]^{j_k} du. \end{aligned}$$

Let $\Psi(s,t) = \sum_{n=0}^{\infty} s^n Q_n(t)$ be the probability-generating function of a j_0-Yule process without immigration, starting with one particle, then

$$J(s,t) = \sum_{k=1}^{K} \frac{1}{\mu(t)} \int_0^t \nu_k(u)(\Psi(s,t-u))^{j_k} du \tag{38}$$

which, combined with Lemma 1, leads to

$$J(s,t) = \sum_{k=1}^{K} \frac{1}{\mu(t)} \int_0^t \nu_k(u) \left[1-(1-s^{-j_0})\exp(j_0 V_0(t-u))\right]^{-j_k/j_0} du.$$

Defining $\mathcal{J}(s,t) = \mu(t)J(s,t)$, we obtain the result of Proposition 4.

We now use Proposition 4 to prove Theorem 1.

B.2 Proof of Theorem 1

Let us set $\tilde{X}(t) = e^{-j_0 V_0(t)} X(t)$. We study the limit of its moment-generating function $\Phi_{\tilde{X}(t)}(\theta) = E[e^{-\theta \tilde{X}(t)}]$ as t tends to ∞ under the assumptions given in Theorem 1, namely:

(i) $\forall k = 1,\ldots,K$, $j_k/j_0 = r_k \in \mathbb{N}^*$.
(ii) For all $k \geq 1$ $\nu_k(t) = \nu_k$ and there exists $t_1 > 0$ such that $\nu_0(t) = \nu_{0,1}\mathbb{1}_{t\leq t_1} + \nu_{0,2}\mathbb{1}_{t>t_1}$ with $0 < \nu_{0,1} \leq \nu_{0,2}$.

Defining

$$s_t = \exp(\theta e^{-j_0 V_0(t)})$$

which converges to 1 as t tends to ∞, we have

$$\Phi_{\tilde{X}(t)}(\theta) = E[(e^{\theta e^{-j_0 V_0(t)}})^{X(t)}] = \Phi(s_t,t). \tag{39}$$

Using Proposition 4, we have

$$\begin{aligned}\Phi(s_t,t) &= \left[1-e^{j_0V_0(t)}(1-s_t^{-j_0})\right]^{-x_0/j_0} e^{-\mu(t)}\exp\{\mathcal{J}(s_t,t)\}\\ &= \left[1-e^{j_0V_0(t)}(1-s_t^{-j_0})\right]^{-x_0/j_0}\exp\{\mathcal{J}(s_t,t)-\mu(t)\} \qquad (40)\end{aligned}$$

with

$$\mathcal{J}(s_t,t)-\mu(t)=\sum_{k=1}^{K}\int_0^t \nu_k(u)\left[1-(1-s_t^{-j_0})\exp(j_0V_0(t-u))\right]^{-j_k/j_0}du-\nu_\bullet t.$$

• By definition of $s_t=\exp(\theta e^{-j_0V_0(t)})$, we have

$$1-(1-s_t^{-j_0})\exp(j_0V_0(t-u))=1-\theta j_0\exp(j_0[V_0(t-u)-V_0(t)])(1+o(1)).$$

• The study of $\mathcal{J}(s,t)$ depends on the forms of $\nu_j(t)$. Under conditions (i) and (ii), for all $t\geq t_1$, $V_0(t)=\nu_{0,1}t_1+\nu_{0,2}(t-t_1)$ and if $u\leq t-t_1$ $V_0(t-u)=\nu_{0,1}t_1+\nu_{0,2}(t-u-t_1)$ and if $u>t-t_1$, $V_0(t-u)=\nu_{0,1}(t-u)$. This leads to

$$\begin{aligned}\mathcal{J}(s_t,t)-\mu(t) &= -\nu_\bullet t+\sum_{k=1}^{K}\nu_k\left(\int_0^{t-t_1}[1-\theta j_0\exp(-j_0\nu_{0,2}u)(1+o(1))]^{-j_k/j_0}du\right.\\ &\quad+\int_{t-t_1}^{t}[1-\theta j_0\exp(-j_0\nu_{0,1}u\\ &\quad\left.-j_0(\nu_{0,2}-\nu_{0,1})(t-t_1))(1+o(1))]^{-j_k/j_0}du\right)\\ &= \sum_{k=1}^{K}\nu_k\int_0^{t-t_1}[1-\theta j_0\exp(-j_0\nu_{0,2}u)(1+o(1))]^{-j_k/j_0}du\\ &\quad+t_1\sum_{k=1}^{K}\nu_k+o(1)-\nu_\bullet t\\ &= \sum_{k=1}^{K}\nu_k\int_0^{t-t_1}[1-\theta j_0\exp(-j_0\nu_{0,2}u)(1+o(1))]^{-j_k/j_0}du\\ &\quad+(t_1-t)\nu_\bullet+o(1).\end{aligned}$$

In that case, we can make the following variable change. Substituting $v=\theta j_0\exp(-j_0\nu_{0,2}u)$ into the integral gives

$$\int_0^{t-t_1}[1-\theta j_0\exp(-j_0\nu_{0,2}u)]^{-j_k/j_0}du = \frac{1}{\nu_{0,2}j_0}\int_{\theta j_0\exp(-j_0\nu_{0,2}(t-t_1))}^{\theta j_0}\frac{1}{(1-v)^{j_k/j_0}v}dv.$$

The usual decomposition of $\frac{1}{(1-v)^{r_k}v}$ into fractions leads to

$$\int_0^{t-t_1}[1-\theta j_0\exp(-j_0\nu_{0,2}u)]^{-j_k/j_0}du$$

$$= \frac{1}{\nu_{0,2}j_0}\left[\log(v) - \log(1-v) + \sum_{l=1}^{r_k-1}\frac{1}{l(1-v)^l}\right]_{\theta j_0 \exp(-j_0\nu_{0,2}(t-t_1))}^{\theta j_0}$$
$$= \frac{1}{\nu_{0,2}j_0}\left[j_0\nu_{0,2}(t-t_1) - \log(1-\theta j_0) + \sum_{l=1}^{r_k-1}\frac{1}{l}\left(\frac{1}{(1-\theta j_0)^l} - 1\right) + o(1)\right].$$

We obtain the following for $\mathcal{J}(s_t,t) - \mu(t)$:

$$\begin{aligned}
&\mathcal{J}(s_t,t) - \mu(t)\\
&= \sum_{k=1}^{K}\nu_k\frac{1}{\nu_{0,2}j_0}\left[j_0\nu_{0,2}(t-t_1) - \log(1-\theta j_0) + \sum_{l=1}^{r_k-1}\frac{1}{l}\left(\frac{1}{(1-\theta j_0)^l} - 1\right)\right]\\
&\quad + (t_1 - t)\nu_\bullet + o(1)\\
&= \sum_{k=1}^{K}\nu_k\frac{1}{\nu_{0,2}j_0}\left[-\log(1-\theta j_0) + \sum_{l=1}^{r_k-1}\frac{1}{l}\left(\frac{1}{(1-\theta j_0)^l} - 1\right)\right] + o(1)\\
&= -\frac{\nu_\bullet}{\nu_{0,2}j_0}\log(1-\theta j_0) + \sum_{k=1}^{K}\frac{\nu_k}{\nu_{0,2}j_0}\sum_{l=1}^{r_k-1}\frac{1}{l}\left(\frac{1}{(1-\theta j_0)^l} - 1\right) + o(1).
\end{aligned}$$

We define

$$\mathcal{J}_0 = -\frac{\nu_\bullet}{\nu_{0,2}j_0}\log(1-\theta j_0)$$

and for $1 \le l \le R_k - 1$

$$\mathcal{J}_l(s_t,t) = \frac{1}{\nu_{0,2}j_0}\frac{1}{l}\left(\frac{1}{(1-\theta j_0)^l} - 1\right).$$

Then we can write $\mathcal{J}(s_t,t) = \mathcal{J}_0(s_t,t) + \sum_{k=1}^{K}\nu_k\sum_{l=1}^{r_k-1}\mathcal{J}_l(s_t,t)$. A rearrangement of the sums (using the fact that $r_1 < r_2 < \cdots < r_K$) leads to

$$\begin{aligned}
\mathcal{J}(s_t,t) - \mu(t) &= \mathcal{J}_0(s_t,t) + (\nu_1 + \cdots + \nu_K)(\mathcal{J}_1(s_t,t) + \cdots + \mathcal{J}_{r_1-1}(s_t,t))\\
&+ (\nu_2 + \cdots + \nu_K)(\mathcal{J}_{r_1}(s_t,t) + \cdots + \mathcal{J}_{r_2-1}(s_t,t))\\
&+ \cdots\\
&+ \nu_K(\mathcal{J}_{r_{K-1}}(s_t,t) + \cdots + \mathcal{J}_{r_K-1}(s_t,t)).
\end{aligned}$$

Setting

$$\alpha_l = \begin{cases} \nu_\bullet & \text{for } 0 \le l \le r_1 - 1,\\ \nu_k + \cdots + \nu_K & \text{for any } r_{k-1} \le l \le r_k - 1 \text{ and for any } k = 2, \dots, K,\end{cases}$$

we obtain

$$\mathcal{J}(s_t,t) - \mu(t) = \mathcal{J}_0(s_t,t) + \sum_{l=1}^{r_K-1}\alpha_l\mathcal{J}_l(s_t,t).$$

Going back to $\Phi(s_t, t)$ given in (40), we obtain

$$\begin{aligned}\phi(s_t, t) &= (1-\theta j_0)^{-x_0/j_0}(1-\theta j_0)^{-\nu_\bullet/(j_0\nu_{0,2})}\prod_{l=1}^{r_K-1}\exp\left[\frac{\alpha_l}{l\nu_{0,2}j_0}\left((1-j_0\theta)^{-l}-1\right)\right] \\ &= E[e^{\theta Y_0^*}]E[e^{\theta Y_0}]\prod_{l=1}^{r_K-1}E[e^{\theta Y_l}]\end{aligned}$$

where $Y_0^* \sim \Gamma(\frac{x_0}{j_0}, \frac{1}{j_0})$, $Y_0 \sim \Gamma(\frac{\nu_\bullet}{\nu_{0,2}j_0}, \frac{1}{j_0})$, and Y_l is such that its moment-generating function is $\exp[\frac{\alpha_l}{l\nu_{0,2}j_0}((1-j_0\theta)^{-l}-1)]$. This generating function can be reformulated as

$$\begin{aligned}\exp\left[\frac{\alpha_l}{l\nu_{0,2}j_0}\left((1-j_0\theta)^{-l}-1\right)\right] &= \sum_{k=0}^{\infty}\exp\left[-\frac{\alpha_l}{l\nu_{0,2}j_0}\right]\left(\frac{\alpha_l}{l\nu_{0,2}j_0}\right)^k\frac{1}{k!}\frac{1}{(1-j_0\theta)^{kl}} \\ &= \sum_{k=0}^{\infty}\rho_{kl}\frac{1}{(1-j_0\theta)^{kl}}\end{aligned}$$

where $\rho_{kl} = \exp[-\frac{\alpha_l}{l\nu_{0,2}j_0}](\frac{\alpha_l}{l\nu_{0,2}j_0})^k\frac{1}{k!}$ is the probability that a Poisson random variable of parameter $\frac{\alpha_l}{l\nu_{0,2}j_0}$ is equal to k. So Y_l is distributed as an infinite mixture of $\{\Gamma(kl, \frac{1}{j_0})\}_{k\geq 0}$ with Poisson weights.

Finally, $e^{-j_0V_0(t)}X(t)$ converges in distribution to $\Gamma(\frac{x_0}{j_0}, \frac{1}{j_0})$ plus a sum of r_K independent variables $\sum_{l=0}^{r_K-1}Y_l$ where the $Y_0 \sim \Gamma(\frac{\nu_\bullet}{\nu_{0,2}j_0}, \frac{1}{j_0})$ and $Y_l \sim \sum_{k=0}^{\infty}\rho_{kl}\Gamma(kl, \frac{1}{j_0})$ with $\rho_{kl} = \exp[-\frac{\alpha_l}{l\nu_{0,2}j_0}](\frac{\alpha_l}{l\nu_{0,2}j_0})^k\frac{1}{k!}$. The theorem is proved.

References

Aalen, O. (1978). "Nonparametric inference for a family of counting processes." *The Annals of Statistics*, 6: 701–726. MR0491547. doi: http://dx.doi.org/10.1214/aos/1176344247. 151, 152

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics. New York: Springer-Verlag. MR1198884. doi: http://dx.doi.org/10.1007/978-1-4612-4348-9. 151, 156

Chen, F. (2011). "Maximum local partial likelihood estimators for the counting process intensity function and its derivatives." *Statistica Sinica*, 21: 107–128. MR2796855. 151, 152

Gamiz, L., Kulasekera, K. B., Limnios, N., and Lindqvist, B. (2011). *Applied Nonparametric Statistics in Reliability*. Springer Series in Reliability Engineering. New York: Springer-Verlag. MR3186312. doi: http://dx.doi.org/10.1007/978-0-85729-118-9. 154

Gelman, A. and Rubin, D. B. (1992). "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science*, 7(4): 457–472. doi: http://dx.doi.org/10.1214/ss/1177011136. 173

Godsill, S. (2007). "Particle filters for continuous-time jump models in tracking applications." In: *ESAIM: Proceedings of Oxford Workshop on Particle Filtering*. MR2405648. doi: http://dx.doi.org/10.1051/proc:071907. 153

Gusto, G. and Schbath, S. (2005). "FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes model." *Statistical Applications in Genetics and Molecular Biology*, 4: 1–26. MR2170440. doi: http://dx.doi.org/10.2202/1544-6115.1119. 151, 152

Ishawaran, H. and James, L. (2004). "Computational methods for multiplicative intensity models using weighted Gamma processes: proportional hazards, marked point processes, and panel count data." *Journal of the American Statistical Association*, 99: 175–190. MR2054297. doi: http://dx.doi.org/10.1198/016214504000000179. 152

Ogata, Y. (1999). "Seismicity Analysis through Point-process Modelling: A review." *Pure and Applied Geophysics*, 155: 471–507. doi: http://dx.doi.org/10.1007/s000240050275. 151, 152

Rao, V. A. and Teh, Y. W. (2013). "Fast MCMC sampling for Markov jump processes and extensions." *Journal of Machine Learning Research*, 14: 3207–3232. MR3144463. 152, 153

Reynaud-Bouret, P. and Schbath, S. (2010). "Adaptive estimation for Hawkes processes; application to genome analysis." *The Annals of Statistics*, 38(5): 2781–2822. MR2722456. doi: http://dx.doi.org/10.1214/10-AOS806. 152

Shonkwiler, R. (1980). "On age-dependent branching processes with immigration." *Computers & Mathematics with Applications*, 6(3): 289–296. MR0604093. doi: http://dx.doi.org/10.1016/0898-1221(80)90036-X. 182

Tanner, M. R. and Wong, W. H. (1987). "The calculation of posterior distributions by data augmentation." *Journal of the American Statistical Association*, 82(398): 528–550. MR0898357. 158

Whiteley, N., Johansen, A., and Godsill, S. (2011). "Monte Carlo filtering of piecewise deterministic processes." *Journal of Computational and Graphical Statistics*, 20(1): 119–139. MR2816541. doi: http://dx.doi.org/10.1198/jcgs.2009.08052. 153

Acknowledgments

The authors are grateful to Prof. Jean-Michel Marin for his helpful contribution at the beginning of the project.