Bayesian Inference for Partially Observed Branching Processes

Sophie Donnet, Judith Rousseau

Ceremade, Université Paris-Dauphine & CREST-ENSAE, France

January 2, 2013

Abstract

Poisson processes are used in various application fields applications (public health biology, reliability and so on). In their homogeneous version, the intensity process is a deterministic constant. In their inhomogeneous version, it depends on time. To allow for an endogenous evolution of the intensity process we consider multiplicative intensity processes. Inference methods have been developed when the trajectories are fully observed.

We deal with the case of a partially observed process. As a motivating example, consider the analysis of an electrical network through time. This network is composed of cables and accessories (joints). When a cable fails, the cable is replaced by a new cable connected to the network by two new accessories. When an accessory fails, the same kind of reparation is done leading to the addition of only one accessory. The failure rate depends on the stochastically evolving number of accessories. We only observe the times events; the initial number of accessories and the cause of the incident (cable or accessory) are only partially observed.

The aim is to estimate the different failure rates or to make predictions. The inference is strongly influenced by the initial number of accessories, which is typically an unknown quantity. We deduce a sensible prior on the initial number of accessories using the probabilistic properties of the process . We illustrate the performances of our methodology on a large simulation study.

Keywords: Counting Process, Bayesian analysis, Latent variables, Branching process

1 Introduction

Counting processes (say $X(t)$) are commonly used in various fields of applications such as medicine –see (5) for instance– public health biology or reliability –see (3) for instance– or more generally in risk theory, see (7) 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 called the intensity. A classical generalization of the homogeneous Poisson process is the inhomogeneous Poisson process whose intensity process is a positive deterministic function called the intensity 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 (1)) 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 also called the intensity function of the process and $Y(t)$ is a positive predictable process called the exposure process (see for instance (3)). Parametric and non parametric methods have been developed when the trajectories $(Y(t), X(t), t \in [0,1])$ are fully observed starting from the paper by (1) : see for instance (3), (6), (8) and references therein for nonparametric estimation and (7) for parametric estimation. It is however sometimes the case that the process is only partially observed. In this paper we propose a Bayesian analysis of a family of partially observed multiplicative intensity processes.

Motivating context This work is motivated by an analysis of an electrical network through time. To simplify the exposure, assume that this network is composed of cables and accessories (such as joints, etc). We observe the evolution of the network and more precisely the sequences of incidents (failures) taking place 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, so that two accessories are added to the network. When an incident takes place on an accessory the same kind of reparation is done but this leads only to the addition of one accessory (see Figure 1 for a graphical illustration of the reparation process).

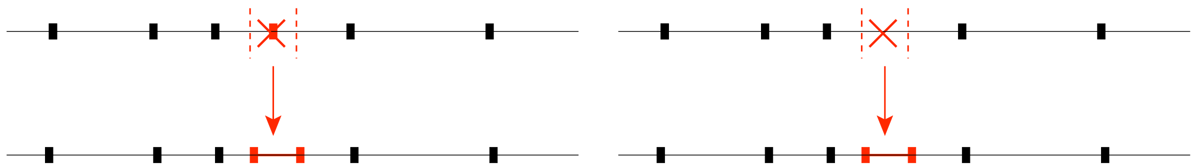


Figure 1: On the left, **failure on a cable** and reparation : the accessory is replaced by two of them. On the right, **failure on the cable** : a new cable is connected to the remaining network by two accessories.

The process can be interpreted in the population process context : an incident on an accessory corresponds to the birth of one particle (accessory) whereas a breakdown on the cable corresponds to the immigration of two particles (accessories). Assume that the inter-event periods are exponentially distributed with respective rates ν_{acc} and $\nu_{cab}d$ (where d is the length of the cable, which is assumed to be constant over time), then the counting process $X(t)$ is a Yule process with immigration whose exposure process is $Y(t) = \nu_{acc}X(t^-) + \nu_{cab}d$.

Remark 1. *The motivating example just described is related to repairable systems as described for instance in (4), however, contrariwise to minimal repairable systems considered in (4), 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_{acc} and ν_{cab} are the parameters of interest. In that practical context, we only have access to the instants of intervention (reparation) and not to the type of reparations (e.g. type of breakdown). As a consequence, from a point process point of view, the observations are reduced to the jump instants denoted T_j; the heights of the jumps in the counting process, i.e. in that case the cause of the incidents (cable or accessories) are unobserved.

Moreover, in general, the beginning of the observation period does not coincide with the installation of the network : in other words, the time events are systematically collected but only since a later date (say $t = \tau_0$ in the following).

$$
\begin{array}{llcccc}
\textit{Time} & 0 & \longrightarrow & \tau_0 & \longrightarrow & \tau_0 + \tau \\
 & X(0) & & [X(\tau_0) & - \textit{ Observation period } - & X(\tau + \tau_0)]
\end{array}
$$

As a consequence, $X(t)$ is a multiplicative intensity process whose parameters of interest have to be estimated from the observation of the jumps instants collected from a truncated period excluding the initial state of the process.

Inference for multiplicative Poisson process with observation reduced to the jumps instants In section 2 we present a generalization of the model considered in the motivating example. Let $X(t)$ be a so-called j_0-Yule process with multi-size immigrations : each particle give birth to j_0 particles with a rate ν_0 and immigration groups of sizes $j_1, \ldots, j_K$ arrive with respective rates $\nu_1, \ldots, \nu_K$. The exposure process $Y(t)$ of $X(t)$ is $Y(t) = X(t^-)\nu_0 + \sum_{j=1}^{K} \nu_j$ where $X(t^-)$ is predictable and ν_j $(j = 0, \cdots, K)$ are positive constants. The aim is then to either estimate the different failure rates or to make predictions on the future number of incidents. When X is completely observed we have a fully observed multiplicative Poisson process and the estimation is easy to handle (see Section 3.1). In that paper, we deal with the estimation issue when the observations are reduced to the jumps instants. In the following, we call this *a partially observed multiplicative Poisson process.*

Remark 2. *Note that contrariwise to other contexts such as count panel data models, the time of events are here fully observed. This model can be extended in various directions, which we discuss in Section 5.*

When the time $t = \tau_0$, corresponding to the beginning of the observations, does not correspond to the time of the installation of the system (or time at the origin $(t = 0)$) $X(\tau_0)$ is then also unknown. This is in practice a challenging issue since inference on the parameters ν_j, $j = 0, \dots, K$, is strongly influenced by the value of $X(\tau_0)$. This issue is tackle in Section 3.2.1. In that case, it then becomes necessary to construct a sensible prior on $X(\tau_0)$. We use this Yule process representation to construct automatic semi-informative prior distributions on $X(\tau_0)$ when it is not observed.

Hence, in this paper we : (1) propose a parametric Multiplicative Poisson process which is flexible and adapted to many problems including linear assets management and reliability evaluation for repairable systems; (2) Propose a general bayesian procedure for inference in such models and (3) construct a semi-informative prior on the unobserved quantity $X(\tau_0)$ based on the asymptotic behaviour of X.

The paper is organized as follows. In Section 2 we describe precisely the model, in Section 3.1 we explain how we conduct the inference under a fully observed process, Section 3.2.1 describes how we extend the previous analysis to the setup of a partially observed process when $X(\tau_0)$ is known and we show the influence of $X(\tau_0)$ on the inference. In Section 3.2.2 we treat the case with unknown $X(\tau_0)$ by constructing a prior distribution on $X(\tau_0)$ using the structure of the process. Section 4 presents a large numerical study. Finally in Section 5 we explain how the family of processes we have considered can be extended .

2 Model and notations

We consider a j_0-Yule process with multi-size immigrations. More precisely, let $j_0, j_1, j_2, \dots, j_K$ be positive integers such that $(\mathcal{C}) : j_1 < j_2 < \dots < j_K$ and $j_0 \notin \{j_1, \dots, j_K\}$. Each particle (or individual) gives birth to a fixed number j_0 of children within an exponentially distributed (with parameter ν_0) time interval, this means that after one split there are j_0 more particles. Moreover, groups of immigrants of respective sizes $j_1, j_2, \dots, j_K$ arrive with exponentially distributed time intervals of respective parameters $\nu_1, \nu_2, \dots, \nu_K$.

In the following, an "event" denotes either a split or an arrival of immigrants. $N(\tau)$ is the number of events occurred in $[\tau_0, \tau_0 + \tau]$. Let $T_1, \ldots, T_{N(\tau)}$ be the occurrence times of the events during the observation period $[\tau_0, \tau_0 + \tau]$. Assuming that the particles are independent, we have $T_j - T_{j-1} | T_{j-1} = t, X(t) \sim \mathcal{E}(\nu_0 X(t) + \nu_1 + \cdots + \nu_K)$ where $X(t)$ is the number of particles at time t and $\mathcal{E}(\gamma)$ denotes an exponential distribution with mean $1/\gamma$. $\{X(t), \tau_0 \leq t \leq \tau_0 + \tau\}$ is a counting process and so is a right continuous step function. Its jumps have (positive) sizes j_0 (split) or j_1, $j_2 \ldots$ or j_K (immigrations). Its infinitesimal probabilities verify : for all $t \geq 0, i \in \mathbb{N}^*$,

$$P(X(t{+}h) = j | X(t) = i) = \begin{cases} \nu_0 i h + o(h) & \text{if } j = i + j_0 \\ \nu_k h + o(h) & \text{if } j = i + j_k, \text{ for } k \in \{1, \ldots, K\} \\ 1 - (\nu_0 i + \sum_{k=1}^K \nu_k) h + o(h) & \text{if } j = i \\ o(h) & \text{otherwise} \end{cases} \tag{1}$$

For every $k = 1, \ldots K$, let $\{N_k(t), \tau_0 \leq t \leq \tau_0 + \tau\}$ be the counting process of immigration events of size j_k. Let $\{N_0(t), \tau_0 \leq t \leq \tau_0 + \tau\}$ be the counting process of split events. We have $N(t) = N_0(t) + N_1(t) + \cdots + N_K(t)$. $\{(N_0(t), N_1(t), N_2(t), \ldots, N_K(t)), \tau_0 \leq t \leq \tau_0 + \tau\}$ is a multivariate counting process with multiplicative intensity $(\nu_0 X(t^-), \nu_1, \ldots, \nu_K)$ where $X(t^-) = \lim_{s \to t, s < t} X(t)$. We adopt the following notation $N_{0:K}(t) = (N_0(t), N_1, (t), N_2(t) \ldots N_K(t))$. Note that our process is strongly related to Hawkes processes as used in seismicity analysis, see (7) or in DNA analysis (5) except that instead of considering only two point processes as in their cases we have K point processes, also compared to (7) we do not have the same parametric form for the intensity process and compared to (5) we have infinite memory in our construction of the intensity of our point process.

Remark 3. *This model is equivalent to a Yule process with immigration events of random size, each size j_k, $(k = 1 \ldots K)$ arriving with probability $\frac{\nu_k}{\sum_{k=1}^K \nu_k}$.*

We are interested in the estimation of the parameter $\theta = (\nu_0, \nu_1, \ldots, \nu_K) \in (\mathbb{R}^{*+})^{K+1}$. We consider π a prior on θ, possible priors will be given in Section 3.1. We say that the process is fully observed if $(N(t), X(t), \tau_0 \leq t \leq \tau_0 + \tau)$ are observed, or equivalently if both the time events $\{T_i\}_{i=1\ldots N(\tau)}$ and the nature of the events observed through $\{X(T_j)\}_{j=1,\cdots,N(\tau)}$ are observed. 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 $\{T_i\}_{i=1,\cdots,N(\tau)}$ are observed but the nature of the events and the initial number of particles $X(\tau_0)$ are non or partially observed.

3 Bayesian inference

3.1 Estimation from the complete observation

In this section we assume that we completely observe the multivariate process $\{N_{0:K}(t), \tau_0 \leq t \leq \tau_0+\tau\}$ and the initial number of particle $X(\tau_0)$. In the fully observed setup we use the fact that $\{N_{0:K}(t), \tau_0 \leq t \leq \tau_0+\tau\}$ is a multivariate counting process with multiplicative intensity $(\nu_0 X(t-), \nu_1, \ldots, \nu_K)$ to give an expression of the likelihood (see (2), page 98). Let Z_j be a discrete variable representing the type of the j-th event : Z_j takes its values in $\{0, \ldots, K\}$ such that Z_j is equal to k if the j-th event is of type k. The likelihood is :

$$\begin{aligned}\mathcal{L}(X(\tau_0), \{N_{0:K}(t), \tau_0 \leq t \leq \tau_0+\tau\}; \theta) = \mathcal{L}(X(\tau_0); \theta) \prod_{k=0}^{K} \nu_k^{N_k(\tau)} \prod_{j=1}^{N(\tau)} X(T_{j-1})^{\mathbb{1}_{Z_j=0}} \\ \times \exp\left[-\nu_0 \sum_{j=1}^{N(\tau)+1} (T_j - T_{j-1}) X(T_{j-1}) - \mu\tau\right]\end{aligned} \tag{2}$$

where $T_0 = \tau_0$, $T_{N(\tau)+1} = \tau + \tau_0$ and $\mu = \sum_{k=1}^{K} \nu_k$. This quantity is referred to as the *complete likelihood*. From equation (2) we can deduce easily discuss the identifiability of the parameters θ, the proof being given in the supplementary material.

Proposition 1. *Let θ and θ' be two sets of parameters such that for any complete data set $X(\tau_0), \{N_{0:K}(t), \tau_0 \leq t \leq \tau_0+\tau\}$, $\mathcal{L}\left(X(\tau_0), \{N_{0:K}(t), \tau_0 \leq t \leq \tau_0+\tau\}; \theta\right) =$ $\mathcal{L}\left(X(\tau_0), \{N_{0:K}(t), \tau_0 \leq t \leq \tau_0+\tau\}; \theta'\right)$, then $\theta = \theta'$.*

Remark 4. *Note that the identifiability of θ from the counting trajectory $\{X(t), \tau_0 \leq t \leq \tau_0 + \tau\}$ is not ensured if $\mathcal{C}$ is not verified.*

In a Bayesian context, we specify a prior distribution on the parameter $\theta = (\nu_0, \ldots, \nu_K)$. A standard choice in that case is to use Gamma distributions on these parameters : $\nu_k \sim \Gamma(\alpha_k, \beta_k), \quad \forall k = 0, \ldots K$ 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 work conditionally on $X(\tau_0)$, then from equation (2) we deduce that the Gamma distributions are conjugate and the posterior distributions on the ν_k's are given by :

$$\begin{aligned} \nu_0 | X(\tau_0), \{N_{0:K}(t), \tau_0 \leq t \leq \tau_0 + \tau\} &\sim \Gamma\left(\alpha_0 + N_0(\tau), \beta_0 + \textstyle\sum_{j=1}^{N(\tau)+1} (T_j - T_{j-1}) X(T_{j-1}))\right) \\ \nu_k | X(\tau_0), \{N_{0:K}(t), \tau_0 \leq t \leq \tau_0 + \tau\} &\sim \Gamma\left(\alpha_k + N_k(\tau), \beta_k + \tau\right) \quad \forall k = 1 \ldots K \end{aligned} \tag{3}$$

with $T_0 = \tau_0$ and $T_{N(\tau)+1} = \tau_0 + \tau$. As a consequence we obtain the following posterior expectation estimators :

$$\begin{aligned} \widehat{\nu_0} &= E[\nu_0 | X(\tau_0), \{N_{0:K}(t), \tau_0 \leq t \leq \tau_0 + \tau\}] = \frac{\alpha_0 + N_0(\tau)}{\beta_0 + \sum_{j=1}^{N(\tau)+1} (T_j - T_{j-1}) X(T_{j-1})} \\ \widehat{\nu_k} &= E[\nu_k | X(\tau_0), \{N_{0:K}(t), \tau_0 \leq t \leq \tau_0 + \tau\}] = \frac{\alpha_k + N_k(\tau)}{\beta_k + \tau}, \forall k = 1 \ldots K \end{aligned} \tag{4}$$

Role of $X(\tau_0)$ in the estimators Note from equation (4) that the estimators $\{\widehat{\nu_k}, k = 1 \ldots K\}$ only depend on the number of events of type k, $N_k(\tau)$. However, the quantity $\sum_{j=0}^{N(\tau)+1} (T_j - T_{j-1}) X(T_{j-1})$ in $\widehat{\nu_0}$ can be reformulated as $\sum_{j=1}^{N(\tau)+1} (T_j - T_{j-1}) X(T_{j-1}) = \tau X(\tau_0) + \tau \sum_{k=0}^{K} j_k N_k(\tau) - \sum_{j=1}^{N(\tau)} T_j (X(T_j) - X(T_{j-1}))$ leading to

$$\widehat{\nu_0} = \frac{\alpha_0 + N_0(\tau)}{\beta_0 + \tau X(\tau_0) + \tau \sum_{k=0}^{K} j_k N_k(\tau) - \sum_{j=1}^{N(\tau)} T_j (X(T_j) - X(T_{j-1}))} \tag{5}$$

which enlightens the influence of $X(\tau_0)$. We 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.

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 events occurrences $T_1, \dots, T_{N(\tau)}$ and partially the types of the events. In the following we introduce Z_j a discrete variable representing the type of the j-th event: Z_j takes its values in $\{0, \dots, K\}$ such that Z_j is equal to k if the j-th event is of type k.

Remark 5. *Fully observing the process $\{N_{0:K}(\ell), \tau_0 \leq t \leq \tau_0+\tau\}$ is the same as observing the number of events $N(\tau)$, the time events $T_1 \dots T_{N(\tau)}$ and $Z_1, \dots, Z_{N(\tau)}$.*

Let $\mathbf{Z}$ denote $(Z_1, \dots, Z_{N(\tau)})$. We introduce n_{nobs} and n_{obs} the numbers of non-observed and observed event types respectively. Obviously we have $n_{nobs} + n_{obs} = N(\tau)$ and n_{obs} can vary from 0 (if $\mathbf{Z}$ is completely unobserved) to $N(\tau)$ (if $\mathbf{Z}$ is completely observed). Let $\{i_1, \dots, i_{n_{nobs}}\}$ be the non-observed indices, $\mathbf{Z}_{nobs} = (Z_{i_1}, \dots, Z_{i_{n_{nobs}}})$ the vector composed of the non-observed Z_i's and $\mathbf{Z}_{obs} = \mathbf{Z} \setminus \mathbf{Z}_{nobs}$ the vector composed of the observed Z_i's. We first consider the case where we estimate the parameter from the partial observation $(N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}_{obs})$.

3.2.1 Case 1: $X(\tau_0)$ is known

The likelihood of the observations $(N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}_{obs})$ is

$$\mathcal{L}(N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}_{obs}; \theta) = \sum_{\mathbf{z} \in \{0,\dots,K\}^{n_{nobs}}} \mathcal{L}(N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}_{obs}, \mathbf{z}; \theta)$$

Interestingly, even in $n_{obs} = 0$, i.e. if the types of events are now observed at all, the parameter θ can still be identified (proof given in the supplementary material):

Proposition 2. *Let θ and θ' be two sets of parameters such that for any partial data set $(N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}_{obs})$, $\mathcal{L}\left(N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}_{obs}; \theta\right) = \mathcal{L}\left(N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}_{obs}; \theta'\right)$. Then $\theta = \theta'$*

As soon as n_{nobs} becomes reasonably large, this sum in the likelihood 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 $\mathbf{Z}_{nobs}$. This makes the use of the conjugate priors Γ considered in Section 3.1 particularly useful.

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

---

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

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

$$\begin{array}{rcll} X^{(\ell-1)}(T_j) & = & X^{(\ell-1)}(T_{j-1}) + \sum_{k=0}^{K} j_k \mathbb{1}_{Z_j^{(\ell-1)}=k} & \forall j = 1 \dots N(\tau) \\ N_k^{(\ell-1)} & = & \sum_{j=1}^{N(\tau)} \mathbb{1}_{Z_j^{(\ell-1)}=k} & \forall k = 0 \dots K \end{array}$$

[**2.**] Generate the parameters conditionally to $N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}^{(\ell-1)}$:

$$\begin{array}{rcl} \nu_0^{(\ell)} | N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}^{(\ell-1)} & \sim & \Gamma\left(\alpha_0 + N_0^{(\ell-1)}(\tau), \beta_k + \sum_{j=0}^{N(\tau)+1} (T_j - T_{j-1}) X^{(\ell-1)}(T_{j-1})\right) \\ \nu_k^{(\ell)} | N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}^{(\ell-1)} & \sim & \Gamma\left(\alpha_k + N_k^{(\ell-1)}(\tau), \beta_k + \tau\right) \quad \forall k = 1 \dots K \end{array}$$

[**3.**] Generate the non-observed event types $\mathbf{Z}^{(\ell)}_{nobs}$ conditionally to $(N(\tau), X(\tau_0), T_1, \dots, T_{N(\tau)}, \theta^{(\ell)}, \mathbf{Z}_{obs})$:

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

[3.1] for $k = 0 \dots K$, set $\widetilde{\mathbf{Z}}_i^{k,l} = \widetilde{\mathbf{Z}}_i$, for $i \neq i_l$ and $\widetilde{Z}_{i_l}^{k,l} = k$. Compute

$$p_{i_l,k} = P(Z_{i_l} = k | X(\tau_0), T_1, \dots, T_{N(\tau)}, \mathbf{Z}^{(\ell-1)} \setminus \{Z_{i_l}\}) \propto \mathcal{L}(X(\tau_0), T_1, \dots, T_{N(\tau)}, \widetilde{\mathbf{Z}}^{k,l})$$

[3.2] Generate $Z_{i_l}^{(\ell)} | X(\tau_0), T_1, \dots, T_{N(\tau)}, \widetilde{\mathbf{Z}}^{(\ell-1)} \setminus \{Z_{i_l}\} \sim (p_{i_l,0}, \dots p_{i_l,K})$

[3.3] In $\widetilde{\mathbf{Z}}$ replace its i_l-th component by $Z_{i_l}^{(\ell)}$ and return to [3.1] with $l := l+1$ until $l = n_{nobs}$ and set $\mathbf{Z}^{(\ell)} = \widetilde{\mathbf{Z}}$.

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). Inference in this case is described in the following section.

3.2.2 Case 2 : $X(\tau_0)$ is unknown

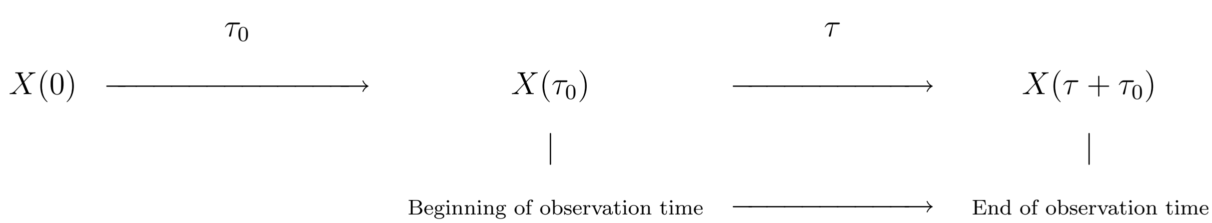
In this section, we assume that $X(\tau_0)$ is not observed, we thus consider a prior distribution π on $X(\tau_0)$. In the following proposition we prove that θ is still identifiable, the proof is given in the supplementary material.

Proposition 3. *Let θ and θ' be two sets of parameters and let π and π' be two distributions on $X(\tau_0)$ possibility depending on θ and θ' respectively, such that for any partial data set $(N(\tau)T_1, \dots, T_{N(\tau)}, \mathbf{Z}_{obs})$, $\mathcal{L}\left(N(\tau), T_1, \dots, T_{N(\tau)}, \mathbf{Z}_{obs}; \theta, \pi\right) = \mathcal{L}\left(N(\tau), T_1, \dots, T_{N(\tau)}, \mathbf{Z}_{obs}; \theta', \pi'\right)$. Then $\theta = \theta'$*

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 $\{x(\tau_0)^-, \dots, x(\tau_0)^+\} \subset \mathbb{N}$: $X(\tau_0) \sim \mathcal{U}_{\{x(\tau_0)^-, \dots, x(\tau_0)^+\}}$. In practice, $x_{\tau_0 -}$ and $x_{\tau_0^+}$ would typically be elicited using experts 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 4.4, figure **??** for a numerical illustration of that remark.

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 time $t = \tau_0$ which is the beginning of the study, is known. When the observation period starts (time τ_0), the system has evolved leading to a certain number of $X(\tau_0)$. As a consequence we propose to derive the prior distribution on $X(\tau_0)$ from the asymptotic distribution of the number of individuals.

Assume that the observation starts at a given time (called $t = \tau_0$ in this paper) but the system has existed prior to τ_0, for some time with a known initial state $X(0)$, so that $t = 0$ corresponds to the installation of the system.



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

Proposition 4. *Let $X(t)$ be the number of particles issued from the multi immigration j_0-Yule process described in Section 2. We assume that $X(0) = x_0$. We set $\mu = \sum_{k=1}^{K} \nu_k$. We have:*

$$\Phi(s,t) = E[s^{X(t)}] = \left[1 - e^{\nu_0 j_0 t}(1 - s^{-j_0})\right]^{-x_0/j_0} e^{-\mu t} \exp\{\mu t J(s,t)\}$$

where $J(s,t) = \sum_{k=1}^{K} \frac{\nu_k}{\mu} \frac{1}{t\nu_0 j_0} \int_{1-\frac{1}{s^{j_0}}}^{(1-\frac{1}{s^{j_0}})\exp(\nu_0 j_0 t)} \frac{1}{(1-v)^{j_k/j_0} v} dv$

A power serie development supplies $X(t)$ 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 multi immigration Yule process described in Section 2. We assume that $X(0) = x_0$.*

Assumption $\mathcal{A}$: $\forall k = 1 \dots K$, $j_k/j_0 = r_k \in \mathbb{N}^*$.

Then

$$e^{-\nu_0 j_0 t} X(t) \xrightarrow[t \to \infty]{\mathcal{L}} \Gamma\left(\frac{x_0}{j_0}, \frac{1}{j_0}\right) + \sum_{l=0}^{r_K - 1} Z_l$$

where $Z_0 \sim \Gamma\left(\frac{\mu}{\nu_0 j_0}, \frac{1}{j_0}\right)$ and $(Z_l)_{l=0 \dots r_K - 1}$ are independent variables distributed as infinite mixture of Gamma distributions. More precisely, for $l = 1 \dots r_K - 1$,

$$Z_l \sim \sum_{k=0}^{\infty} \omega_{k,l} \Gamma\left(kl, \frac{1}{j_0}\right)$$

with $\omega_{k,l} = e^{\lambda_l} \frac{\lambda_l^k}{k!}$, $\lambda_l = \frac{\mu \alpha_l}{l \nu_0 j_0}$ $\alpha_l = 1$, $\forall l = 0 \dots r_1 - 1$ and $\alpha_l = (\nu_k + \dots + \nu_K)/\mu$, $\forall l = r_{k-1} \dots r_k - 1$, $\forall k = 2 \dots K$

The proof is given in the supplementary material. Theorem 1 shows that as τ_{past} becomes large, conditional on the ν_j's and $X(0)$, $X(\tau_0)$ can be approximated in distribution as $e^{\nu_0 j_0 \tau_0}$ times the sum of iniinite mixtures of Gamma random variables. In other words it increases exponentially quickly with τ_0. A numerical illustration of the precision of this approximation if illustrated in Section 4.3. Hence, neglecting the modification of the system through time can lead to strongly biased estimation, as soon as $\nu_0 \tau_0 j_0$ is not negligible. For intermediate value of $\nu_0 \tau_0 j_0$ it is possible to improve the approximation by re-centering the distribution using the true mean of $X(\tau_0)$ which can be deduced from the Laplace transform given in proposition 4.

Remark 6. *Note that this result applies for instance for a standard Yule process that is to say $j_0 = 1$ i.e. when each particle divides into two particles.*

Estimation With that new prior distribution on $X(0)$ the model is not fully conjugate (see equation (2) with $\mathcal{L}(X(0);\theta)$ equal to the infinite Poisson mixture of Gamma distributions) and the full conditional distributions of the rates ν_k are not Gamma distributions any longer. As a consequence the posterior distribution is sampled using a Metropolis-Hastings algorithm with a random walk on ν_k.

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

- **At iteration** (0), initialize the algorithm on $\mathbf{Z}_{nobs}^{(0)}$ arbitrarly chosen
- **At iteration** (ℓ),

[1.] Set $\mathbf{Z}^{(\ell-1)} = (\mathbf{Z}_{obs}, \mathbf{Z}_{nobs}^{(\ell-1)})$ and compute

$$X^{(\ell)}(T_j) = X^{(\ell)}(T_{j-1}) + \sum_{k=1}^{K} j_k \mathbb{1}_{Z_j^{(\ell-1)}=k} \quad \text{and} \quad N_k^{(\ell)} = \sum_{j=1}^{N(\tau)} \mathbb{1}_{Z_j^{(\ell-1)}=k} \quad \forall k = 0 \dots K$$

[2.] For $k = 0 \dots K$

- Propose $\widetilde{\nu_k} \sim q(\widetilde{\nu_k} | \nu_k^{(\ell-1)})$ and set $\widetilde{\theta} = (\nu_0^{(\ell)}, \dots \nu_{k-1}^{(\ell)}, \widetilde{\nu_k}, \nu_{k+1}^{(\ell-1)} \dots \nu_K^{(\ell-1)})$
- Compute

$$\alpha_k = \min \left\{ 1, \frac{\mathcal{L}(X^{(\ell-1)}(\tau_0), \{N_{0:K}^{(\ell-1)}(t), \tau_0 \le t \le \tau_0 + \tau\}; \widetilde{\theta}) \pi(\widetilde{\theta})}{\mathcal{L}(X^{(\ell-1)}(\tau_0), \{N_{0:K}^{(\ell-1)}(t), \tau_0 \le t \le \tau_0 + \tau\}; \theta^{(\ell-1)}) \pi(\theta^{(\ell-1)})} \frac{q(\nu_k^{(\ell-1)} | \widetilde{\nu_k})}{q(\widetilde{\nu_k} | \nu_k^{(\ell-1)})} \right\}$$

- Set

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

[3.] Generate non-observed event types $\mathbf{Z}_{nobs}$ using step [3.] in the algorithm presented in Section 3.2.1

[4.] Generate $X(\tau_0)$:

- Propose $\widetilde{X}(\tau_0) \sim q(\cdot|X(\tau_0)^{(\ell-1)})$ using a proposal distribution
- Compute

$$\alpha(X(\tau_0)^{(\ell-1)}, \widetilde{X}(\tau_0)) = \min\left\{1, \frac{\mathcal{L}(\{N_{0:K}^{(\ell)}, \tau_0 \leq t \leq \tau_0+\tau\}|\widetilde{X}(\tau_0);\theta^{(\ell)})}{\mathcal{L}(\{N_{0:K}^{(\ell)}, \tau_0 \leq t \leq \tau_0+\tau\}|X(\tau_0)^{(\ell-1)};\theta^{(\ell)})} \frac{\pi(\widetilde{X}(\tau_0))}{\pi(X(\tau_0)^{(\ell-1)})} \frac{q(X(\tau_0)^{(\ell-1)}|\widetilde{X}(\tau_0))}{q(\widetilde{X}(\tau_0)|X(\tau_0)^{(\ell-1)})}\right\}$$

- Set

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

---

4 Numerical studies

We now conduct a large 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 study the precision the asymptotic approximation of $X(\tau_0)$. 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 data set.

4.1 Study of an electrical network through time

We consider the analysis of an electrical network through time. An electrical network is composed of electrical cables and accessories (such as joints). We study the occurrences of breakdowns on the network. The incidents can derive either from a cable or an accessory failure. In case of incident, the network is repaired as follows: if an accessory breaks down, the network is cut and replaced by two other accessories; if the failure comes from the cable itself, the damaged part is removed and a new cable is reconnected to the network thanks to two accessories (see Figure 1 for an illustration). We make the realistic assumption that the various reparations do not significantly modify the length of the network.

The evolution of this network can be seen as a branching process. Using notations of Section 2, an accessory is a particle , a failure on an accessory corresponds to a birth of one child ($j_0 = 1$) whereas a failure on the cable corresponds to an immigration of two particles ($K = 1$ and $j_K = 2$). As before we introduce the discrete variable Z_j such that Z_j is equal to 0 if the j-th failure takes place on an accessory and is equal to 1 if the j-th failure takes place on the cable.

From now on, we use the following notations: $\nu_0 := \nu_{acc}$ is the failure rate on the accessories. The cable failure rate depends on its length. Let d denote the length of the cable, $\nu_1 := \nu_{cab} d$ is the failure rate of the cable and so the immigration rate. ν_{acc} and ν_{cab} are the parameters of interest with prior distributions: $\nu_{acc} \sim \Gamma(\alpha_{acc}, \beta_{acc})$ and $\nu_{cab} \sim \Gamma(\alpha_{cab}, \beta_{cab})$. The number of accessories at the installation instant is known and denoted $X(0)$. At the beginning of the observation period, the networks are realistically 25 years old and the study lasts around 4 years.

In this particular model, (theorem 1) leads to the asymptotic distribution of the number of accessories:

$$\begin{array}{rcl} X(\tau_0) & \sim & e^{\tau_0 \nu_{acc}} \sum_{k=0}^{\infty} e^{-\rho} \frac{\rho^k}{k!} \Gamma\left(k+\rho+X(0),1\right) \\ \rho & = & \nu_{cab} d / \nu_{acc}, \end{array} \tag{6}$$

which we denote by π_∞. We derive from this distribution the asymptotic expectation of $X(\tau_0)$:

$$E_\infty[X(\tau_0)] = e^{\tau_0 \nu_{acc}} \left[2\rho + X(0)\right]. \tag{7}$$

Moreover, from the Laplace transform of the process, see proposition 4, we can derive the exact expectation of $X(\tau_0)$:

$$E[X(\tau_0)] = e^{\tau_0 \nu_{acc}} \left(2\rho(1 - e^{-\tau_0 \nu_{acc}}) + X(0)\right) \tag{8}$$

In the following, we denote by π_∞^R the asymptotic distribution re-centered around $E[X(\tau_0)]$ using equations (7) and (8).

In practice, the observations consist in the times of the breakdowns on the network, their types (cable or accessories) are non and partially reported. We now study the influence of this partial observation on the estimation of the two failure rates ν_{acc} and ν_{cab}.

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

In this first section, we suppose that $X(\tau_0)$ is known. We study the influence of the amount of missing data on the estimation of ν_{acc} and ν_{cab}. Here, we fix the parameters to be equal to:

$$\nu_{acc} = 10^{-5}, \quad \nu_{cab} = 2 \times 10^{-6}, \quad X(\tau_0) = 400 \quad \tau = 10 \text{ years} \ , \quad d = 8000$$

which are realistic values in an electrical network. With these parameter values we simulate 100 datasets. For these datasets, the number of observations $N(\tau)$ varies between 199 and 308 with a mean value equal to 254.08.

For each dataset, we sample from the posterior distribution of ν_{acc} and ν_{cab} in the following 4 scenarios.

- *Scenario 1*: we suppose that the whole sequence $Z_1, \dots, Z_{N(\tau)}$ is observed. In that context, the posterior distribution of (ν_{acc}, ν_{cab}) has an explicit expression given by:

$$\begin{aligned} \nu_{acc}|X(\tau_0), Z_1, \dots, Z_{N(\tau)} &\sim \Gamma\left(\alpha_{acc} + N_{acc}(\tau), \beta_{acc} + \sum_{j=0}^{N(\tau)+1}(T_j - T_{j-1})X(T_{j-1}))\right) \\ \nu_{cab}|X(\tau_0), Z_1, \dots, Z_{N(\tau)} &\sim \Gamma\left(\alpha_{cab} + N_{cab}(\tau), \beta_{cab} + d\tau\right) \end{aligned} \tag{9}$$

where $N_{acc}(\tau)$ and $N_{cab}(\tau)$ are the number of failures on the accessories and cable respectively.

- *Scenario 2* : one third of the $Z_1, \dots, Z_{N(\tau)}$ are unobserved (the unobserved Z_j are randomly chosen). In that case, the posterior distribution of (ν_{acc}, ν_{cab}) is sampled by a Gibbs algorithm described in Section 3.2.1.

- *Scenario 3* : two thirds of $Z_1, \dots, Z_{N(\tau)}$ are unobserved. The observed Z_j are randomly chosen among those of scenario 2.

- *Scenario 4* : $Z_1, \dots, Z_{N(\tau)}$ are completely unobserved.

On Figure 2, we plot the prior and the 4 marginal posterior distributions of ν_{acc} (top) and ν_{cab} (bottom), one per scenario, for one typical dataset. As expected, the smaller n_{obs} the more spread the posterior. This phenomenon is enhanced when the sequence $Z_1, \ldots, Z_{N(\tau)}$ is completely non-observed.

Denoting $\widehat{\nu}_{acc}^{(d)}$ and $\widehat{\nu}_{cab}^{(d)}$ the posterior mean estimators for dataset d, we compute the relative bias and relative root mean square error respectively given by

$$Biais = \sum_{d=1}^{100} \frac{\widehat{\nu}_{acc}^{(d)} - \nu_{acc}}{\nu_{acc}} \quad RMSE = 10\sqrt{\sum_{d=1}^{100} \frac{(\widehat{\nu}_{acc}^{(d)} - \nu_{acc})^2}{\nu_{acc}^2}}$$

The relative bias and RMSE are given in Table 1 in percentage. As expected, the quality of estimation decreases when n_{nobs} increases.

		Scenario 1	Scenario 2	Scenario 3	Scenario 4
ν_{acc}	Relative Biais	-0.85	-0.99	-1.46	-3.36
	RMSE	6.58	7.14	8.31	8.66
ν_{acc}	Relative Biais	-2.12	-3.09	-1.47	4.76
	RMSE	12.34	14.06	18.48	11.48

Table 1: Simulation study 1 ($X(\tau_0)$ known and fixed): relative bias and RMSE for $\widehat{\nu}_{acc}$ and $\widehat{\nu}_{cab}$ in the 4 scenarii.

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

In this part we illustrate the practical interest of Theorem 1 by comparing the true distribution of the number of accessories $X(t)$ with its asymptotic approximation. In the numerical study we fix the parameter values to the following values:

$$\nu_{acc} = 4.10^{-4} \quad , \quad \nu_{cab} = 4.10^{-6}, \quad d = 4000, \quad X(0) = 10$$

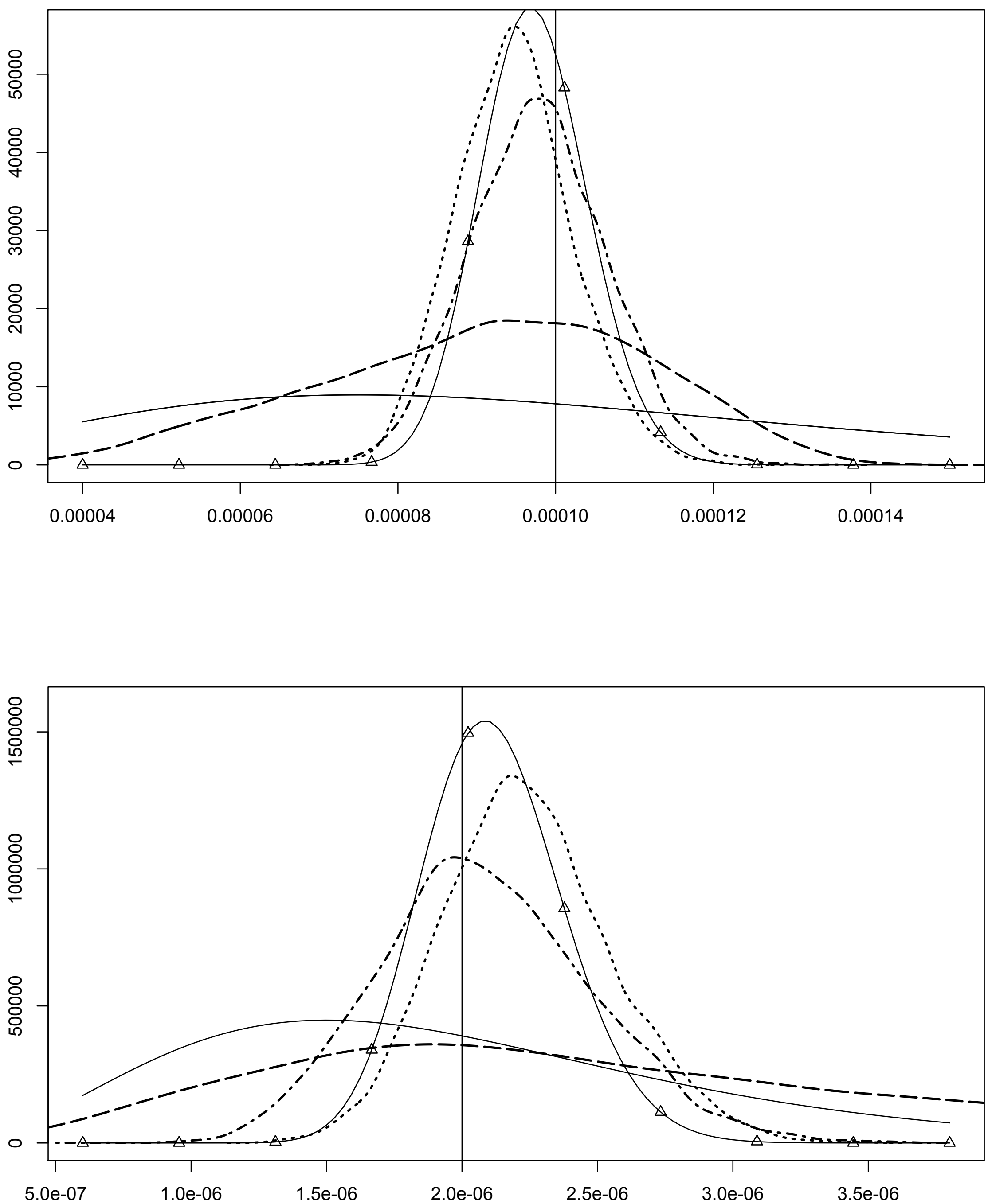


Figure 2: Influence of the non-observation of $Z_1, \ldots, Z_{N(\tau)}$ on the posterior distributions of ν_{acc} (upper figure) and ν_{cab} (bottom figure). Prior distribution (plain line), posterior distribution in scenario 1 (plain line with triangles), posterior distribution in scenario 2 ($\cdots$), posterior distribution in scenario 3 ($\cdot - \cdot$), posterior distribution in scenario 4 (dashed line)

and we consider 3 possible values for τ_0: $\tau_0 = 2$ years, $\tau_0 = 25$ years and $\tau_0 = 130$ 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. We compare it to

1. the asymptotic distribution given by expression (6) (plotted in dotted line)

2. the re-centered asymptotic distribution π^R_∞ (plotted in $\cdot - -$) . More precisely, from equations (8) and (7), we propose a correction and recenter the asymptotic distribution around the true expectation.

3. the Poisson distribution of mean $E[X(\tau_0)]$ given by equation (8) (dashed line).

The density functions are plotted on Figure 3. We observe that for a long elapsed time ($\tau_0 = 130$ years, bottom figure) the asymptotic, the true and the re-centered distributions overlap. 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-centered 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. When the time period is really short ($\tau_0 = 5$ years) the re-centered asymptotic distribution is much larger than the true one. As a consequence, this choice of prior distribution is less interesting but stays competitive with respect to a uniform distribution, in particular because of its automatic definition (no parameter has to be tuned or elicited).

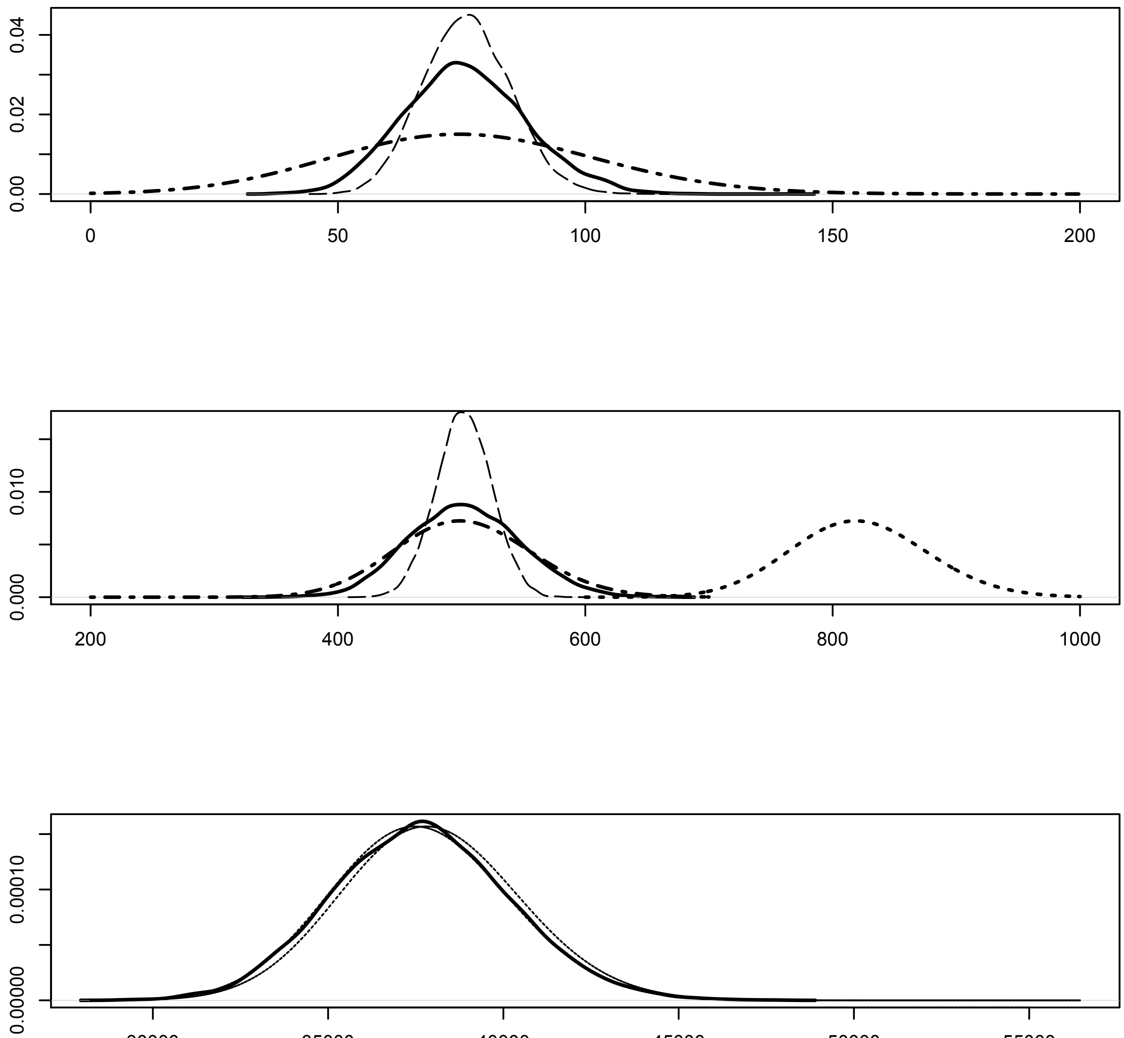


Figure 3: Distribution of $X(\tau_0)$: the estimated true distribution (plain line), asymptotic distribution (dotted line), re-centered asymptotic distribution ($\cdot - \cdot$), Poisson distribution (dashed line). Top: $\tau_0 = 5$ years, middle: $\tau_0 = 25$ years, bottom: $\tau_0 = 130$ years

4.4 Influence of $X(\tau_0)$ and estimation

We now pay attention to the importance of $X(\tau_0)$ in the estimation of ν_{acc} and ν_{cab}.

When $X(\tau_0)$ is unknown we can consider several solutions: fixing it at an arbitrarly chosen value, estimating it using a uniform prior distribution and using the asymptotic distribution as a prior. Using the following parameter value

$$\nu_{acc} = 1.10^{-5}, \quad \nu_{cab} = 4.10^{-6}, \qquad d = 4000$$
$$X(0) = 0, \qquad \tau_0 = 40 \text{ years}, \quad \tau = 15 \text{ years}$$

we simulate 100 trajectories starting at $X(0) = 0$ during a time interval of length $\tau + \tau_0$. For these datasets, the number of accessories at the beginning of the study –namely $x(\tau_0)$– varies from 432 to 650. The number of observations during the study interval$[\tau_0, \tau + \tau_0]$ namely $N(\tau)$ varies between 199 and 308 with a mean value equal to 254.08. Now, for each dataset we estimate the parameters ν_{acc} and ν_{cab} Note that in order to separate the sources of imprecision, in this study, we suppose that $Z_1, \ldots, Z_{N(\tau)}$ are observed.

• *Scenario 0* : Scenario 0 will refer to the case where $X(\tau_0)$ is known (and so set to the true value $x(\tau_0)$) and will be our reference. In that case, the model is conjugate and the posterior distributions on ν_{acc} and ν_{cab} have been given by equations (9).

• *Scenario 1*: The naive solution is to set $X(\tau_0)$ to its value at the instant of installation of the line ($X(0)$), 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 $x(\tau_0)/2$. The posteriors of ν_{acc} and ν_{cab} are the same as in scenario 0.

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

- *Scenario 3* : we consider the re-centered asymptotic distribution π_∞^R on $X(\tau_0)$.

Results On Figure 4 we plot the posterior densities of ν_{acc} (upper) and ν_{cab} (bottom) for one arbitrarily chosen data set. As expected, $X(\tau_0)$ does not influence the posterior distribution of ν_{cab} and the posterior densities corresponding to the 4 scenario nearly overlap. On the contrary the posterior densities for ν_{acc} clearly depend on $X(\tau_0)$. If $X(\tau_0)$ is under-evaluated (scenario 1), the posterior density of ν_{acc} (dashed line) is shifted to the right: this phenomenon was clearly expected from equation (5). When a prior on $X(\tau_0)$ is considered, the re-centered 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_{acc} is clearly narrower and closer to the reference posterior distribution (scenario 0) when π_∞^R is used.

Note that the implementation of the Metropolis Hastings algorithm for $X(\tau_0)$ requires the computation of $\mathcal{L}(X(\tau_0|\theta)$. When this distribution is the asymptotic approximation, $\mathcal{L}(X(\tau_0)|\theta)$ is equal to:

$$\mathcal{L}(X(\tau_0)|\theta) = e^{-\tau_0\nu_{acc}} \sum_{k=0}^{\infty} e^{-\rho} \frac{\rho^k}{k!} f_{\Gamma(k+\rho+X(0),1)}(e^{-\tau_0\nu_{acc}}(X(\tau_0) + 2\rho))$$

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 define a truncation $t(\theta)$ using the following criteria

$$t(\theta) = \max\left\{1000, \inf\left\{k \geq \rho | e^{-\rho}\frac{\rho^k}{k!} < 10^{-324}\right\}\right\}.$$

We compute for each dataset and each scenario the posterior means and variances of ν_{acc} and ν_{cab} which we denote by denote by $E[\nu_{acc}^{\ell,s}|\mathbf{Y}^\ell]$, $E[\nu_{cab}^{\ell,s}|\mathbf{Y}^\ell]$, $Var[\nu_{acc}^{\ell,s}|\mathbf{Y}^\ell]$ and $Var[\nu_{cab}^{\ell,s}|\mathbf{Y}^\ell]$ respectively, with $s \in \{0,\ldots,3\}$ and $\ell \in \{1,\ldots,100\}$ index the scenario and the dataset respectively. In figure 5, a summary of these quantities is presented in the form of their densities.

As already remarked, the posterior variance and expectation of ν_{cab} 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_{acc} (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_{acc} (figure 5, top right, density in $\cdot - \cdot$ line) whereas the use of the re-centered asymptotic prior distribution on $X(\tau_0)$ leads to a much more sensible posterior variance (figure 5, top right, density in dotted line).

4.5 Estimation on pseudo-real data set

For confidentiality reasons, no real data set was provided. To illustrate the performance of our prior distribution and our estimation method, we propose to consider a pseudo-real data set, that is to say a dataset simulated with realistic parameters (length of study, length of the network, breakdown rates, etc)

We consider a $d = 40000$ meters network, $\tau_0 = 25$ years old at the beginning of the study.

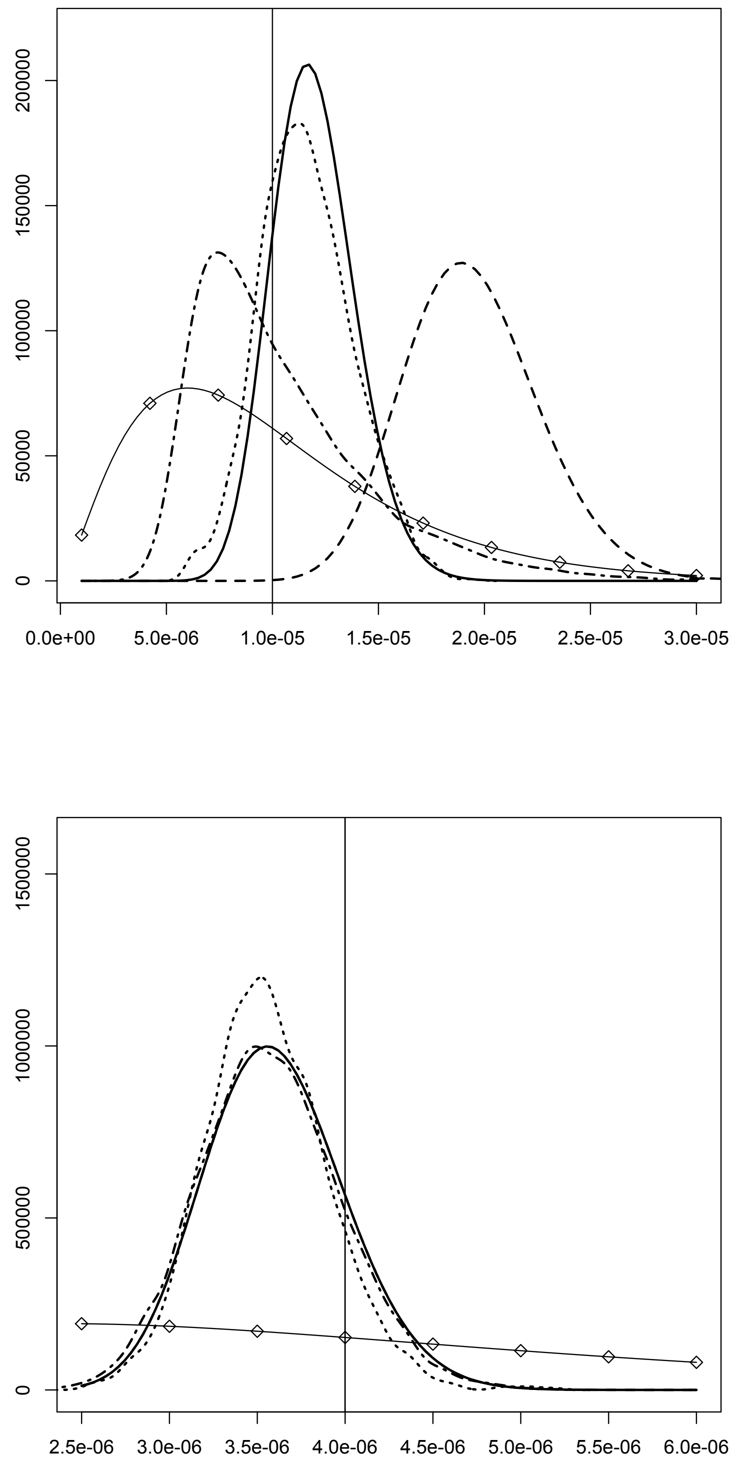


Figure 4: Influence of the non-observation of $X(\tau_0)$ on the posterior distributions of ν_{acc} (upper figure) and ν_{cab} (bottom figure) for one data set: prior distribution (plain line with diamonds), posterior distribution with the true $X(\tau_0)$ (Scenario 0) (plain line), posterior distribution 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).

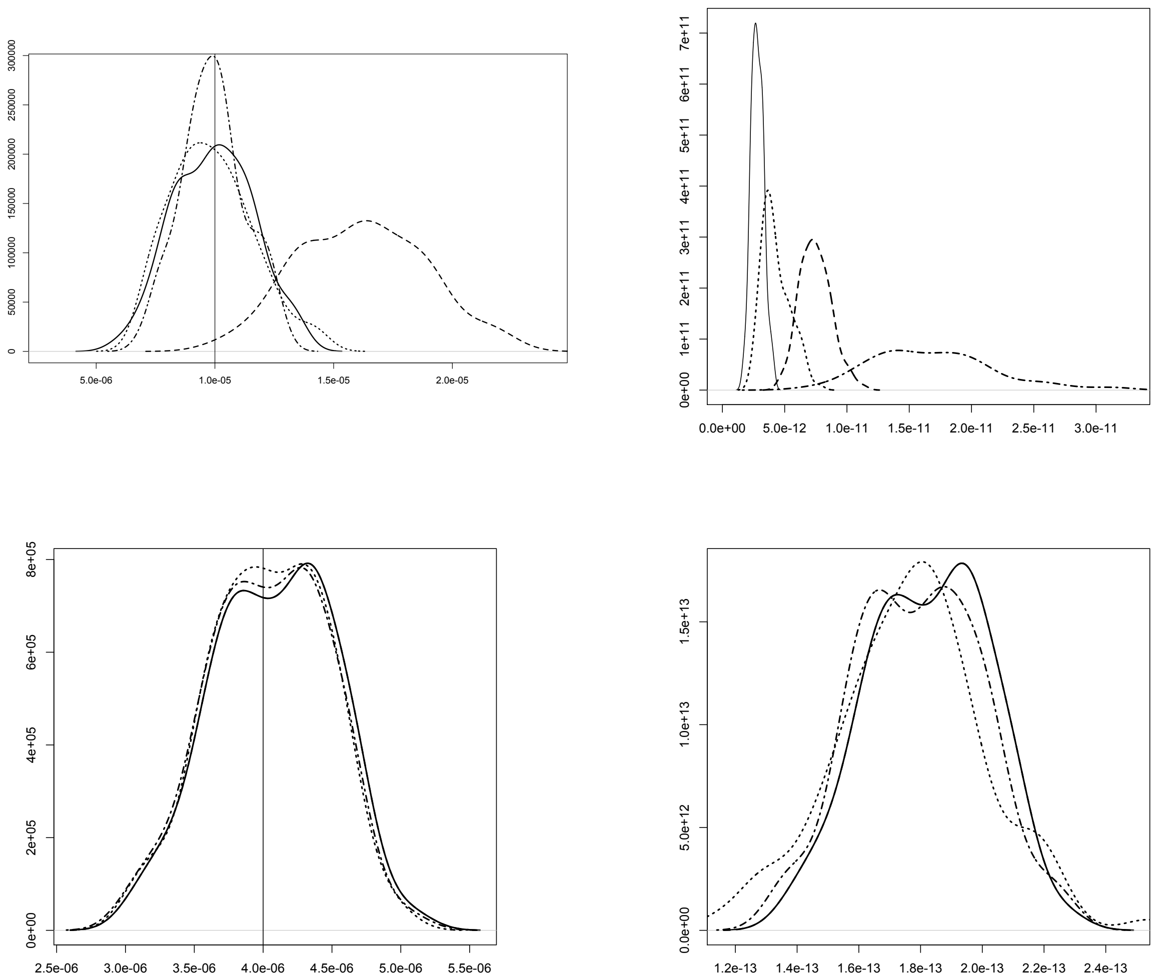


Figure 5: Influence of the non-observation of $X(\tau_0)$ on the posterior expectation (left) and variance (right) of ν_{acc} (upper) and ν_{cab} (bottom) for the 100 datasets: estimated density with the true $X(\tau_0)$ (Scenario 0) (plain line), with under-evaluated $X(\tau_0)$ (Scenario 1) (dashed 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).

The study lasts $\tau = 4$ years and half of the breakdowns types are reported whereas the other ones are unknown. We observe $N(\tau)$ breakdowns. $X(0)$ is fixed at 10 and the state of the network at the beginning of the study $-X(\tau_0)-$ is unobserved.

We use the re-centerered asymptotic prior distribution π_∞^R for $X(\tau_0)$. We sample the posterior distribution using the Gibbs algorithm described in Section 3.2.2. The algorithm is initialized on the following value:

$$
\begin{aligned}
\nu_{cab}^{(0)} &= \frac{\frac{N(\tau)}{n_{obs}} \sum_{i=1}^{n_{obs}} 1\!\!1_{Z_{obs,i}=2}}{\tau \dot{d}}, \quad & \nu_{acc}^{(0)} &\sim \Gamma(\alpha_{acc}, \beta_{acc}) \\
\rho^{(0)} &= \nu_{cab} d / \nu_{acc}, & X^{(0)}(0) &= e^{\tau_0 \nu_{acc}^{(0)}} \left(2\rho^{(0)}(1 - e^{-\tau_0 \nu_{acc}^{(0)}}) + X(0) \right)
\end{aligned}
$$

The algorithm is implemented with 50000 iterations. A period of burn in of 10000 iterations is removed. The trajectories of the Gibbs algorithm are plotted in figure 6. The posterior distributions of $X(\tau_0)$, ν_{acc} and ν_{cab} are plotted in figure 7.

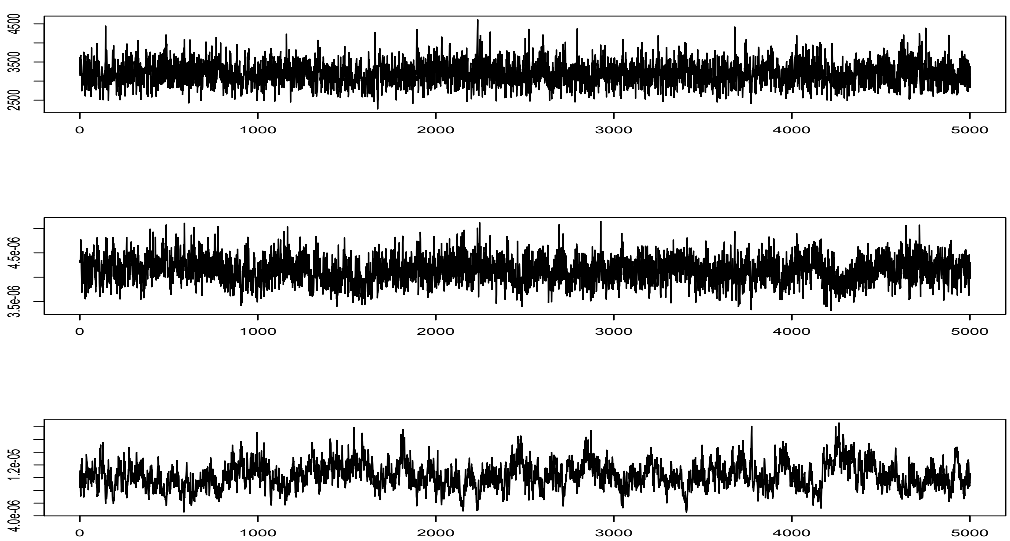


Figure 6: *Pseudo real data set*: trajectories of the MCMC algorithm. Top: $X(\tau_0)$. Middle: ν_{cab}. Bottom: ν_{acc}

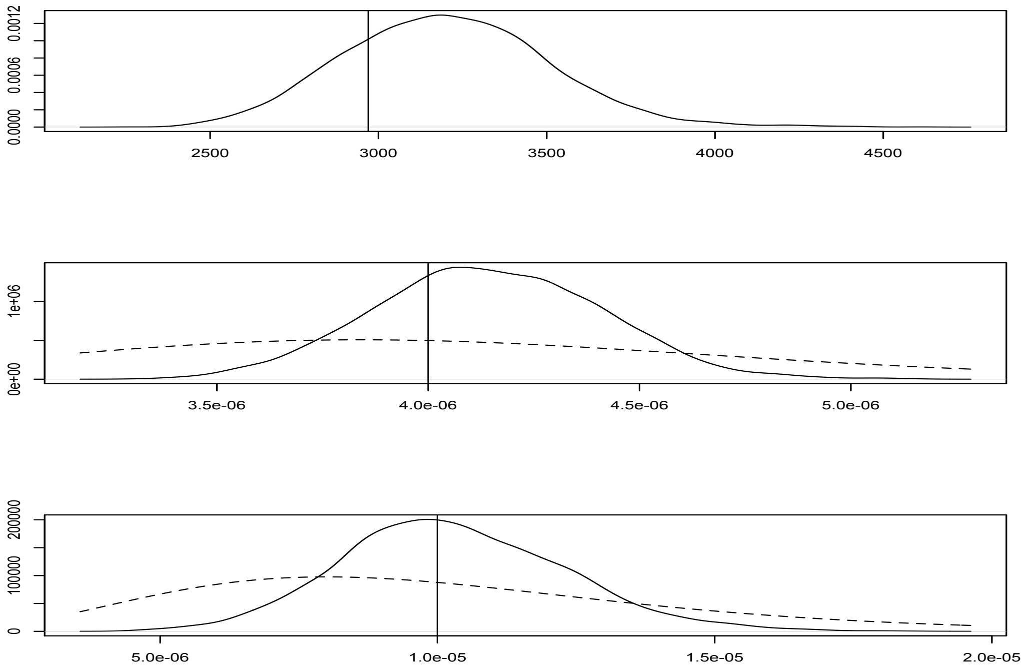


Figure 7: *Pseudo real data set*: posterior distribution (plain) line and prior distribution in dashed line. Top: $X(\tau_0)$. Middle: ν_{cab}. Bottom: ν_{acc}

5 Discussion and possible extensions of model (1)

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

First, in that model, we consider 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 report with error so that each true type of event $Z_j^t = Z_j^r$ with an unknown probability p. No major difficulty would come from that extension of the model.

An other 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.

When the covariates are allowed to vary with time, things become more complicated. An interesting extension is to consider ν_l, $l \geq 1$ as a function of t. For instance, in the example of the electrical network, described in Section 1, the failure rate of the electrical cable could increase smoothly with time. The likelihood still has a relatively simple expression in the form :

$$\mathcal{L}(X(\tau_0), \{N_{0:K}(t), \tau_0 \leq t \leq \tau_0 + \tau\}; \theta) = \mathcal{L}(X(\tau_0); \theta) \prod_{k=0}^{K} \prod_{j=1}^{n} [\nu_k(T_j)]^{1\!\!1_{Z_j=k}} \prod_{j=1}^{N(\tau)} (\nu_0 X(T_{j-1}))^{1\!\!1_{Z_j=0}}$$
$$\times \exp\left[-\nu_0 \sum_{j=1}^{N(\tau)+1} (T_j - T_{j-1}) X(T_{j-1}) - \tilde{\mu}(\tau) \right]$$

with $\tilde{\mu}(\tau) = \int_0^\tau \sum_{j=1}^k \nu_j(t)dt$. The rest of the analysis would then follow similarly to before, writing $\nu_j(t) = \nu_j(t;\theta)$ and performing Bayesian parametric inference on ν_0 and θ. The computation of the (exact and asymptotic) distribution of $X(\tau_0)$ given $X(0)$, follows the same line as in the supplementary material. Indeed, similar computations imply that we can still write

$$\phi(s,t) = \sum_{k=0}^{\infty} e^{-\tilde{\mu}(t)} \frac{\tilde{\mu}(t)^k}{k!} J^k(s,t)$$

where

$$J(s,t) = \sum_{k=1}^{K} \frac{1}{\tilde{\mu}(t)} \int_0^t \psi(s, t-\tau)^{j_k} \nu_k(\tau) d\tau$$

with the same notation as in Appendix **??**. The exact expression of the above integral then depends on the form of $\nu_k(t)$, $k = 1, \cdots, K$. A key point in the above computations is the fact that ν_0 is constant over time. If ν_0 is not constant then $\psi(s, t-\tau)$ needs to be changed into

$$\psi(s,\tau,t) = \sum_{n=0}^{\infty} s^n Q_n(t,\tau)$$

where $Q_n(t,\tau)$ is the probability that one particle born at time τ leads to n particles at time t. This depends on the form of $\nu_0(t)$. If $\nu_j = \nu_j(t)$ for $j = 0, \ldots, K$, then the likelihood can be written as :

$$\mathcal{L}(X(\tau_0), \{N_{0:K}(t), \tau_0 \leq t \leq \tau_0 + \tau\}; \theta) = \mathcal{L}(X(\tau_0); \theta) \prod_{k=0}^{K} \prod_{j=1}^{n} [\nu_k(T_j)]^{\mathbb{1}_{Z_j = k}} \prod_{j=1}^{N(\tau)} (\nu_0(T_j) X(T_{j-1}))^{\mathbb{1}_{Z_j=0}}$$
$$\times \exp\left[- \sum_{j=1}^{N(\tau)+1} X(T_{j-1}) \int_{T_{j-1}}^{T_j} \nu_0(t) dt - \tilde{\mu}(\tau) \right]$$

A more easy way to consider an 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.

References

[1] O.O. Aalen. Nonparametric inference for a family of counting processes. *aos*, 6:701–726, 1978.

[2] Per Kragh Andersen, Ørnulf Borgan, Richard D. Gill, and Niels Keiding. *Statistical models based on counting processes.* Springer Series in Statistics. Springer-Verlag, New York, 1993.

[3] Feng Chen. Maximum local partial likelihood estimators for the counting process intensity function and its derivatives. *Satistica Sinica*, 21:107–128, 2011.

[4] L. Gamiz, K. B. Kulasekera, N. Limnios, and B.H. Lindqvist. *Applied nonparametric statistics in Reliability.* Springer Series in Reliability engineering. Springer-Verlag, New York, 2011.

[5] Gaelle Gusto and Sophie Schbath. Fado: a statistical method to detect favored or avoided distances between occurrences of motifs using the hawkesÕ model. *Statistical Applications in Genetics and Molecular Biology*, 4:1–26, 2005.

[6] Hemant Ishawaran and Lancelot James. Computational methods for multiplicative intensity models using weighted Gamma processes: proportional hazards, marked point processes, and panel count data. *jasa*, 99:175–190, 2004.

[7] Yosihiko Ogata. Seismicity analysis through point-process modelling: A review. *Pure and applied Geophysics*, 155:471–507, 1999.

[8] Patricia Reynaud-Bouret and Sophie Schbath. Adaptive estimation for Hawkes processes; application to genome analysis. *Ann. Statist.*, 38(5):2781–2822, 2010.

[9] Ron Shonkwiler. On age-dependent branching processes with immigration. *Comput. Math. Appl.*, 6(3):289–296, 1980.