



**HAL**  
open science

# A Recommendation System For Insurance Built With A Multivariate Hawkes Process Based On Customers' Life Events

Laurent Lesage, Madalina Deaconu, Antoine Lejay, Jorge Augusto Meira, Geoffrey Nichil, Radu State

► **To cite this version:**

Laurent Lesage, Madalina Deaconu, Antoine Lejay, Jorge Augusto Meira, Geoffrey Nichil, et al.. A Recommendation System For Insurance Built With A Multivariate Hawkes Process Based On Customers' Life Events. 2021. hal-03483812

**HAL Id: hal-03483812**

**<https://hal.science/hal-03483812>**

Preprint submitted on 16 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Recommendation System For Insurance Built With A Multivariate Hawkes Process Based On Customers' Life Events

Laurent Lesage<sup>†‡</sup>, Madalina Deaconu<sup>†</sup>, Antoine Lejay<sup>†</sup>,  
Jorge Augusto Meira<sup>‡</sup>, Geoffrey Nichil<sup>\*</sup>, Radu State<sup>‡</sup>

## Abstract

We construct a recommendation system for insurance with a Multivariate Hawkes Process, in order to take into account the influence of life events (e.g. marriage, birth, change of job) on the insurance covering selection from customers. This Multivariate Hawkes Process includes several specific features aiming to compute relevant recommendations to customers from a Luxembourgish insurance company. Some of these features are intent to propose a personalized background intensity for each customer thanks to a Machine Learning model, to use triggering functions suited for insurance data or to overcome flaws in real-world data by adding a specific penalization term in the objective function. Our recommendation system has been back-tested over a full year. Observations from model parameters and results from this back-test show that taking into account life events by a Multivariate Hawkes Process allows us to improve significantly the accuracy of recommendations.

**Key words:** Multivariate Hawkes Process, Recommendation System, Life Events, Insurance

## 1 Summary

### 1.1 Introduction and purpose

In [12], the authors developed a recommendation system for car insurance, aiming to detect the customers who are the most likely to accept a new insurance cover, in order to help agents to provide the best advice to their customers. The purpose of the present work is to propose an improvement of this recommendation system, by analysing and handling the influence of life events. More precisely, by observing the behaviour and the evolution of a customer in the insurance context, customers seem to modify their insurance cover when a significant event happens in their life. For instance, on the first version of our recommendation system [12], we observed that a vehicle change has an strong influence. It is thus natural to propose and study how other life events could have an influence on customer choices on insurance cover.

Our aim is to estimate the probability of a customer to add an insurance guarantee/product in real time, updated at any occurrence of a life event. Therefore, the recommendation system that we construct would be able to suggest the additional guarantee or product at the opportune moment. In the first version of the recommendation system that we introduced in [12], we propose an *a priori* estimation of the probability to add a cover. In this paper, we add an *a posteriori* estimation which allows us to update the recommendation each time a life event occurs. This reasoning is similar to the calculation of insurance premiums in [14], based on an *a posteriori* vision which updates premiums as a function of the occurrence of claims. More precisely:

$$\mathbb{P}(\text{customer adds an insurance cover}|\text{characteristics, life events}) = \underbrace{f_1(\text{characteristics})}_{\text{a priori vision}} + \underbrace{f_2(\text{life events})}_{\text{a posteriori vision}}.$$

To do so, we propose to build a recommendation system thanks to a Multivariate Hawkes Process, where each process represents one type of life event. A Multivariate Hawkes Process allows us to model the "mutually exciting effects" between all life events, that is to say the occurrence of a life event could trigger:

- Another life event, e.g. a change of job could trigger a move since the customer works in another place;

---

<sup>†</sup>University of Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France

<sup>‡</sup>University of Luxembourg, 29 Avenue JF Kennedy, JFK Building, L-1855 Luxembourg

<sup>\*</sup>Foyer Assurances, 12 Rue Léon Laval, L-3372 Leudelange

- A subscription to another insurance cover, e.g. a vehicle change could lead to a subscription to a new guarantee in the customer’s car insurance policy since the customer would protect more his new vehicle.

We developed an approach suited for the real data provided by Foyer Assurances, which performs better than other alternatives Multivariate Hawkes Processes models from literature.

## 1.2 Related work

**Recommendation systems with Hawkes Processes.** Several models based on Hawkes Processes were developed to build a recommendation system. The RecSys Challenge 2016 aimed to create a recommendation system to suggest the most suited job to users: authors in [15] proposed an approach with a Hawkes Process. Temporal user-item interactions are modelled with a Multivariate Hawkes Process in [16] and are applied to Internet Protocol television (IPTV), Yelp and Reddit data. In [17], authors construct a recommendation system on music songs and shopping items in order to estimate not only the right moment to make a recommendation, but also to predict the next returning time to the item. A Point Of Interest recommendation system based on geographic spatial data is also proposed in [18]. All these approaches are powerful and very interesting but not useful for our problem as most of these papers use data quite different from our purpose, in terms of dimensions, frequency, etc.

**Multivariate Hawkes Processes estimation.** In the numerous papers about Hawkes Processes written over the last ten years, a large variety of methods to estimate a Multivariate Hawkes Process are proposed, in order to take into account data characteristics or to improve algorithms efficiency. In [1], triggering matrix and background intensity are estimated by iteration on the complete penalized log-likelihood with an Expectation-Maximization algorithm (used in [13]), named MMEL. Triggering functions are estimated non-parametrically by solving Ordinary Differential Equations. ADM4 algorithm presented in [3] also proposes this maximum penalized likelihood estimation, but assumes that triggering functions have an exponential kernel. Authors in [2] adapt the notion of Granger causality to triggering functions to Multivariate Hawkes Processes, which are supposed to be a linear combination of basis functions. In [4], authors estimate triggering functions by solving Wiener-Hopf system, by first replacing expectations with empirical averages, then using the Nystrom method for discretizing the integrals. Estimation of background intensity is calculated simply by dividing the number of events observed by the time realization. In [5], authors estimate the integrals of triggering functions thanks to the notion of cumulants.

## 1.3 Contributions

The main contributions of the present work are listed below:

1. To the best of our knowledge, there exists no paper about Hawkes Processes adapted to recommendation systems for insurance, or a work integrating life events data for an insurance subject;
2. For most of approaches presented in the related works, estimated background intensities are identical for every user. Here, we propose to integrate a personalized background intensity for each Foyer customer, based on either Machine Learning or relevant statistical analysis;
3. We propose triggering functions which are adapted to the insurance data to our disposal and which are different from classic kernels;
4. The objective function has been penalized in a new way, in order to improve the estimated Hawkes Process robustness and compensate some defaults on real data.

## 1.4 Plan

The remainder of this paper is organized as follows. The second section introduces every relevant notation and definition, including the definition of the Multivariate Hawkes Process. The third section details the specificity of our approach: the parameters to estimate, the specific objective function and the estimation algorithm. The fourth section presents the results of our approach on real data provided by Foyer Assurances.

# 2 Multivariate Hawkes Process definition

In this section, we introduce the notations and definitions needed for the construction of our Multivariate Hawkes Process. First, we list every life event that will be considered in the adapted study.

**Definition 1. (List of life events).**

We consider  $m = 6$  types of events. Each number of event will be used for their corresponding intensity, or any mathematical object referring to a specific event. For instance,  $\lambda_1^*(\cdot)$  represents the intensity of the event number 1.

- **Event number 1:** Subscription to a new insurance cover, for which the intensity would represent the probability for the customer to accept a recommendation;
- **Event number 2:** Change of vehicle;
- **Event number 3:** Modification of household composition (including weddings);
- **Event number 4:** Births;
- **Event number 5:** Change of job;
- **Event number 6:** Move.

Then, we introduce some notations which are necessary to define the Multivariate Hawkes Process.

**Notation 1.** We denote:

1.  $U$  the number of customers in the under study data, on which we develop the procedure to estimate the probability to add a guarantee;
2.  $\mathcal{U} = \{1, \dots, U\}$  the set of customers;
3.  $\mathbf{x}_u$  a vector of characteristics describing the profile of the customer  $u \in \mathcal{U}$ , which contains information about:
  - His profile: gender, age, location, profession...
  - His current insurance cover: how many products/guarantees he subscribed to, premiums, data about his agent...
4.  $T_u$  the time length we observe life events for customer  $u$ ;
5.  $\mathcal{H}_u(t)$  the list of life events which occurred to customer  $u$  until time  $t$ ;
6.  $N^u(t)$  the counting process which corresponds to the total number of events observed for customer  $u$  from time 0 to time  $t$ ;
7.  $N_i^u(t)$  the counting process which corresponds to the total number of events of type  $i$  observed for customer  $u$  from time 0 to time  $t$ . We have  $N^u(t) = \sum_{i=1}^m N_i^u(t)$ ;
8.  $s_u = \{(e_k^u, t_k^u)_{k=1}^{N^u}\}$  the sequence of events observed for customer  $u$ :  $e_k^u$  is the number of the  $k^{th}$  event and  $t_k^u$  the corresponding time.

Before defining the Multivariate Hawkes Process, we introduce the univariate version of Hawkes Processes (see also [13] for a background about point processes). A Hawkes Process is a counting process denoted  $N(\cdot)$  who has a so-called "self-exciting" property, which means that each occurrence increases the probability to observe another event a short time after the first occurrence. It is characterized by its conditional intensity function. Each occurrence increases the conditional intensity according to the triggering function.

**Definition 2. (Hawkes process).** Let us consider  $\lambda > 0$ , the so-called *background intensity*, and  $\mu : [0, +\infty[ \rightarrow [0, +\infty[$ , the so-called *triggering function*. We denote  $\{t_1, \dots, t_{N(t)}\}$  the sequence of past occurrences until time  $t$ . A point process is a Hawkes process if its conditional intensity function is of the form:

$$\lambda^*(t) = \lambda + \int_0^t \mu(t-v) dN(v) = \lambda + \sum_{k=1}^{N(t)} \mu(t-t_k). \quad (2.1)$$

For our purpose, we use a multivariate Hawkes Process, which is an extension of the univariate case. The initial objective is to model the influence of life events on the propensity of the customer to modify his insurance cover. To do so, we build a Multivariate Hawkes Process which will be defined by  $m$  conditional intensity functions, one per type of event, instead of one as in the univariate case. For each event, there exists not only a "self-exciting effect" (i.e. an occurrence of event  $i$  could influence the intensity of event  $i$ ), but also a "mutual exciting effect" (i.e. an occurrence of event  $i$  could influence the intensity of any event  $j, j \in \{1, \dots, m\}$ ). For each customer  $u, u \in \mathcal{U}$ , we consider the counting processes  $\{N_i^u(\cdot), i \in \{1, \dots, m\}\}$ , where  $N_i^u(\cdot)$  refers to the type of event number  $i$  and where each type of event  $i$  could influence any other event  $j, j \in \{1, \dots, m\}$ , including  $j = i$ .

**Definition 3. (Multivariate Hawkes processes).** We consider a customer  $u \in \mathcal{U}$ . For each type of event  $i \in \{1, \dots, m\}$ , we consider:

- $\lambda_i(\mathbf{x}_u) > 0$  the *background intensity* (we denote  $\boldsymbol{\lambda}(\mathbf{x}_u) = [\lambda_i(\mathbf{x}_u)]_{i \in \{1, \dots, m\}}$ );
- $\mu_{i,j} : [0, +\infty[ \rightarrow [0, +\infty[, j \in \{1, \dots, m\}$ , the *triggering function* that models the effect of event  $j$  on event  $i$ .

The set of counting processes  $\{N_1^u(\cdot), \dots, N_m^u(\cdot)\}$  is a multivariate Hawkes process if the conditional intensity function of each process  $N_i^u(\cdot)$  is of the form:

$$\lambda_i^{u,*}(t) = \lambda_i(\mathbf{x}_u) + \sum_{j=1}^m \int_0^t \mu_{i,j}(t-v) dN_j^u(v) = \lambda_i(\mathbf{x}_u) + \sum_{j=1}^m \sum_{k=1}^{N_j^u(t)} \mu_{i,j}(t-t_j^k). \quad (2.2)$$

Therefore,  $\lambda_i^{u,*}(t)$  represents the propensity for customer  $u$  to add an insurance cover at time  $t$ ; this quantity will be integrated to the recommendation system. The probability that the customer would add a guarantee/product is influenced by  $\lambda_1(\mathbf{x}_u)$ , which models the impact of the characteristics of the customer (*a priori* vision), and by the sum of  $\mu_{1,j}(t-t_j^k)$ , which models the impact of the  $k^{th}$  event of type  $j$ , occurred at time  $t_j^k$  (*a posteriori* vision).

## 3 The proposed approach

### 3.1 Model

#### 3.1.1 Background intensity

For each event  $i$ , we would like to estimate the background intensity of customer  $u$ ,  $\lambda_{u,i}(\mathbf{x}_u)$ , where  $\mathbf{x}_u$  is a set of features describing the customer  $u$ . In our approach, we model for the event  $i$  the background intensity of customer  $u$ , described by  $\mathbf{x}_u$ , as follows:

$$\lambda_i(\mathbf{x}_u) = \mu_i f_i(\mathbf{x}_u), \quad (3.1)$$

where:

- $\mu_i$  is a proportional factor, common for each customer  $u$ . We denote  $\boldsymbol{\mu} = (\mu_i)_{i \in \{1, \dots, m\}}$ ;
- $f_i$  is a function which, to each data  $\mathbf{x}_u$ , associates  $f_i(\mathbf{x}_u) \in \mathbb{R}$  which reflects the propensity for an event  $i$  to occur for a customer  $u$ , excluding the triggering of other events.  $f_i$  are estimated as follows:
  - **Event number 1:**  $f_1(\mathbf{x}_u)$  has been estimated in [12], where a probability is built to add an additional insurance cover for a customer in function of characteristics that are included in  $\mathbf{x}_u$ ;
  - **Event number 2:**  $f_2(\mathbf{x}_u)$  is estimated by an Extreme Gradient Boosting algorithm, which learned the vehicle change frequency over customers who subscribed to a car insurance for five years at least;
  - **Event number 3:**  $f_3(\mathbf{x}_u) = 1$  for all customer  $u \in \mathcal{U}$ . A modification of household composition includes many types of events (e.g. flatsharing, civil union, weddings, step-families), implies complex interactions and could occur at any moment of life. That's why we assume that background intensity is equal for each customer;
  - **Event number 4:**  $f_4(\mathbf{x}_u)$  equals the birthrate of the age group the customer  $u$  belongs to. Birthrates were extracted from the national statistics portal of Luxembourg [8];

- **Event number 5:**  $f_5(\mathbf{x}_u)$  is built from a survey of the French Institute of Statistics (see [11]).  $f_5(\mathbf{x}_u)$  equals the rate of people who changed of job over five years by age group;
- **Event number 6:**  $f_6(\mathbf{x}_u) = 1$  for all customer  $u \in \mathcal{U}$ . Several studies indeed show that moving out is almost always linked to a life event: 70% are due to family events (births, weddings...) and 30% to professional reasons according to the French Institute of Statistics (see [19]). Thus, we assume that the background intensity is the same for every customer and the event is only influenced by the occurrence of life events.

### 3.1.2 Triggering functions

We assume that the triggering functions are a linear combination of basis functions.

**Notation 2.** We denote:

- $D$  the number of basis functions;
- $g_d$  the  $d^{th}$  basis function,  $d \in \{1, \dots, D\}$ ;
- $\mathbf{A}_d = (a_{i,j}^d)_{i,j=1,\dots,m}$  the  $d^{th}$  infectivity matrix, where  $a_{i,j}^d$  captures the impact of event  $j$  on event  $i$  relatively to the  $d^{th}$  basis function;
- $\mathbf{A} = \{\mathbf{A}_d\}_{d \in \{1, \dots, D\}}$  the set of the  $D$  infectivity matrix.

Therefore, the triggering function  $\mu_{i,j}$  that models the effect of event  $j$  on event  $i$  follows the equation below:

$$\mu_{i,j}(t) = \sum_{d=1}^D a_{i,j}^d g_d(t). \quad (3.2)$$

We propose the following basis functions for our data:

$$g_d(t) = \begin{cases} \frac{1}{\sqrt{2\pi}\omega} \exp(-\frac{t^2}{2\omega^2}) & \text{(Gaussian density) if } d = 1; \\ \frac{t^{k-1}}{(d\omega)^k \Gamma(k)} \exp(-\frac{t}{d\omega}) & \text{(Gamma density) otherwise,} \end{cases} \quad (3.3)$$

where  $\omega$  and  $k$  are respectively scale and shape parameters for the Gamma density. The first basis function is inspired from [2] to take into account immediate triggering, while we sketched the main interests of Gamma density as triggering function for insurance data in [13].

## 3.2 Objective function

We denote  $\Theta$  the parameters of the model to estimate:  $\Theta = (\boldsymbol{\mu}, \mathbf{A})$ .

The objective function of the problem is the sum of the Negative Log-Likelihood and regularizers described below:

- **Negative Log-Likelihood:** denoted  $\text{NLL}(\Theta)$ , it equals:

$$\text{NLL}(\Theta) = \sum_{u=1}^U \left\{ \sum_{i=1}^m \left( T_u \lambda_i(\mathbf{x}_u) + \sum_{k=1}^{N^u} \sum_{d=1}^D a_{i,e_k^u}^d G_d(T_u - t_k^u) \right) - \sum_{k=1}^{N^u} \log \left( \lambda_{e_k^u}(\mathbf{x}_u) + \sum_{j=1}^{k-1} \sum_{d=1}^D a_{e_k^u, e_j^u}^d g_d(t_k^u - t_j^u) \right) \right\}, \quad (3.4)$$

where  $G_d(T_u - t_k^u) = \int_0^{T_u - t_k^u} g_d(t) dt$ ;

- **Sparsity:** we expect that for couples of events which have no influence on each other, the infectivity matrix has zeros for the corresponding coefficients. That is why we add a sparsity constraint, denoted  $R_1$ :

$$R_1(\Theta) = \alpha_1 \left( \sum_{d=1}^D \sum_{i=1}^m \sum_{j=1}^m \frac{(a_{i,j}^d)^2}{2} \right), \quad (3.5)$$

where  $\alpha_1$  is a parameter to control the influence of this regularization;

- **Minimizing the influence of overlapping events:** for some events, the date of occurrence that is in the data-set may be delayed of several days, because of administrative issues. Therefore, we could believe that an event 1 is triggered by an event 2, while event 2 actually occurred after event 1. Therefore we propose to penalize the squared sum of coefficients  $a_{i,j}^d$  and  $a_{j,i}^d$ , in order to correct the influence of overlapping events:

$$R_2(\Theta) = \alpha_2 \left( \sum_{d=1}^D \sum_{i=1}^m \sum_{j=1}^m \frac{(a_{i,j}^d + a_{j,i}^d)^2}{2} \right). \quad (3.6)$$

Therefore, the objective function denoted  $F$  is:

$$\begin{aligned} F(\Theta) &= \text{NLL}(\Theta) + R_1(\Theta) + R_2(\Theta) \\ &= \sum_{u=1}^U \left\{ \sum_{i=1}^m \left( T_u \lambda_i(\mathbf{x}_u) + \sum_{k=1}^{N^u} \sum_{d=1}^D a_{i,e_k^u}^d G_d(T_u - t_k^u) \right) - \sum_{k=1}^{N^u} \log \left( \lambda_{e_k^u}(\mathbf{x}_u) + \sum_{j=1}^{k-1} \sum_{d=1}^D a_{e_k^u, e_j^u}^d g_d(t_k^u - t_j^u) \right) \right\} \\ &+ \alpha_1 \left( \sum_{d=1}^D \sum_{i=1}^m \sum_{j=1}^m \frac{(a_{i,j}^d)^2}{2} \right) + \alpha_2 \left( \sum_{d=1}^D \sum_{i=1}^m \sum_{j=1}^m \frac{(a_{i,j}^d + a_{j,i}^d)^2}{2} \right), \end{aligned}$$

and the learning problem is:

$$\min_{\Theta} F(\Theta). \quad (3.7)$$

### 3.3 Parameters estimation

In order to estimate parameters, we use an Expectation-Maximization (EM) algorithm. It consists in introducing the notion of complete likelihood, which assumes that we could observe, for each event, which previous event has triggered this event (see [13]).

To do so, we construct a surrogate objective function (expectation step), whose iterative minimization allows initial objective function to decrease monotonically (maximization step). We denote  $\hat{\Theta}^{(l)}$  the estimation of parameters at iteration  $l$ :  $\hat{\Theta}^{(l)} = (\hat{\boldsymbol{\mu}}^{(l)}, \hat{\mathbf{A}}^{(l)})$ . The surrogate objective function denoted  $F_s(\Theta, \hat{\Theta}^{(l)})$  equals:

$$F_s(\Theta, \hat{\Theta}^{(l)}) = -Q(\Theta, \hat{\Theta}^{(l)}) + R_1(\Theta) + \hat{R}_2^{(l)}(\Theta, \hat{\Theta}^{(l)}), \quad (3.8)$$

where:

1.

$$\begin{aligned} Q(\Theta, \hat{\Theta}^{(l)}) &= \sum_{u=1}^U \left\{ - \sum_{i=1}^m \left( T_u \lambda_i(\mathbf{x}_u) + \sum_{k=1}^{N^u(T_u)} \sum_{d=1}^D a_{i,e_k^u}^d G_d(T_u - t_k^u) \right) \right. \\ &+ \left. \sum_{k=1}^{N^u(T_u)} p_{e_k^u, e_k^u}^{(l)} \log \left( \frac{\lambda_{e_k^u}(\mathbf{x}_u)}{p_{e_k^u, e_k^u}^{(l)}} \right) + \sum_{j=1}^{k-1} \sum_{d=1}^D p_{e_k^u, e_j^u}^{(l)} \log \left( \frac{a_{e_k^u, e_j^u}^d g_d(t_k^u - t_j^u)}{p_{e_k^u, e_j^u}^{(l)}} \right) \right\}; \end{aligned} \quad (3.9)$$

2.

$$p_{e_k^u, e_k^u}^{(l)} = \frac{\hat{\boldsymbol{\mu}}_{e_k^u}^{(l)} f_{e_k^u}(\mathbf{x}_u)}{\hat{\boldsymbol{\mu}}_{e_k^u}^{(l)} f_{e_k^u}(\mathbf{x}_u) + \sum_{j=1}^{k-1} \sum_{d=1}^D \hat{a}_{e_k^u, e_j^u}^{d, (l)} g_d(t_k^u - t_j^u)}, \quad (3.10)$$

which represents the probability that event  $(e_k^u, t_k^u)$  was triggered by the background intensity of event of type  $e_k^u$ ;

3.

$$p_{e_k^u, e_j^u}^{(l)} = \frac{\sum_{d=1}^D \hat{a}_{e_k^u, e_j^u}^{d, (l)} g_d(t_k^u - t_j^u)}{\hat{\boldsymbol{\mu}}_{e_k^u}^{(l)} f_{e_k^u}(\mathbf{x}_u) + \sum_{j=1}^{k-1} \sum_{d=1}^D \hat{a}_{e_k^u, e_j^u}^{d, (l)} g_d(t_k^u - t_j^u)}, \quad (3.11)$$

which represents the probability that event  $(e_k^u, t_k^u)$  was triggered by event  $(e_j^u, t_j^u)$ ;

4.

$$\hat{R}_2^{(l)}(\Theta, \hat{\Theta}^{(l)}) = \alpha_2 \left( \sum_{d=1}^D \sum_{i=1}^m \sum_{j=1}^m \frac{(a_{i,j}^d + \hat{a}_{j,i}^{d,(l)})^2}{2} \right). \quad (3.12)$$

We update iteratively the parameters estimation by calculating:

$$\Theta^{(l+1)} = \underset{\Theta}{\operatorname{argmax}} F_s(\Theta, \hat{\Theta}^{(l)}). \quad (3.13)$$

We solve  $\frac{\partial F}{\partial \Theta} = 0$  (see Appendix) and we obtain, for all  $d \in \{1, \dots, D\}, k \in \{1, \dots, m\}, (i, j) \in \{1, \dots, m\}^2$ :

$$\hat{\mu}_k^{(l+1)} = \frac{\sum_{u=1}^U \sum_{e_i^u = k} p_{e_i^u, e_i^u}^{(l)}}{\sum_{u=1}^U T_u}; \quad (3.14)$$

$$\hat{a}_{i,j}^{d,(l+1)} = \frac{-(A_{i,j} + \alpha_2 \hat{a}_{j,i}^{d,(l)}) + \sqrt{(A_{i,j} + \alpha_2 \hat{a}_{j,i}^{d,(l)})^2 - 4(\alpha_1 + \alpha_2) B_{i,j}^{(l)}}}{2(\alpha_1 + \alpha_2)}, \quad (3.15)$$

where

$$A_{i,j} = - \sum_{u=1}^U \sum_{e_k^u = j} G_d(T_u - t_k^u), \quad (3.16)$$

$$B_{i,j}^{(l)} = \sum_{u=1}^U \sum_{e_k^u = i} \sum_{e_m^u = j} p_{e_k^u, e_m^u}^{(l)}. \quad (3.17)$$

Optimizing  $F_s$  this way allows us to decrease the Negative Log-Likelihood monotonically. Indeed, from the two following inequalities (see a proof from Appendix in [1]):

- $-Q(\Theta, \hat{\Theta}^{(l)}) \geq \text{NLL}(\Theta)$ ;
- $-Q(\hat{\Theta}^{(l)}, \hat{\Theta}^{(l)}) = \text{NLL}(\hat{\Theta}^{(l)})$ ,

we could deduce that  $\text{NLL}(\hat{\Theta}^{(l-1)}) = -Q(\hat{\Theta}^{(l-1)}, \hat{\Theta}^{(l-1)}) \geq -Q(\hat{\Theta}^{(l)}, \hat{\Theta}^{(l-1)}) \geq \text{NLL}(\hat{\Theta}^{(l)})$ . We update iteratively  $\hat{\Theta}^{(l)} = (\hat{\boldsymbol{\mu}}^{(l)}, \hat{\mathbf{A}}^{(l)})$  while the relative difference between  $\text{NLL}(\hat{\Theta}^{(l-1)})$  and  $\text{NLL}(\hat{\Theta}^{(l)})$ , denoted:

$$\text{Err}(l) = \left| \frac{\text{NLL}(\hat{\Theta}^{(l)})}{\text{NLL}(\hat{\Theta}^{(l-1)})} - 1 \right|, \quad (3.18)$$

is above  $\varepsilon$  and the number of iterations is below  $N_{iter}$ . This methodology is summarized in Algorithm 1.

---

**Algorithm 1:** Parameters estimation

---

**Result:** Estimation of  $\Theta = (\boldsymbol{\mu}, \mathbf{A})$

Initialize  $l = 0$  and  $(\hat{\boldsymbol{\mu}}^{(l)}, \hat{\mathbf{A}}^{(l)})$ ;

$l = l + 1$ ;

Update  $(\hat{\boldsymbol{\mu}}^{(l)}, \hat{\mathbf{A}}^{(l)})$  according to Equations (3.14) and (3.15);

**while**  $\text{Err}(l) \geq \varepsilon$  and  $l \leq N_{iter}$  **do**

$l = l + 1$ ;

    Update  $(\hat{\boldsymbol{\mu}}^{(l)}, \hat{\mathbf{A}}^{(l)})$  according to Equations (3.14) and (3.15);

**end**

---

## 4 Results

In this section, we focus on the results of our approach on real data provided by Foyer Assurances. First, we introduce data and hyperparameters used for estimation. Then, we present the parameters estimation. We also study the impact of regularization on parameters estimation, to conclude by showing that the proposed model overperforms existing approaches.

## 4.1 Data and hyperparameters

This work uses data provided by Foyer Assurances. We extracted sequences of life events from Foyer customers observed between  $T_1 = 2015$  and  $T_2 = 2018$  included. Each row from the final data-set represents one event. The columns are:

- Customer's ID number  $u \in \mathcal{U}$ ;
- Time length we observe life events for customer  $u$ ,  $T_u$ ;
- Type of event ( $e_k^u$  denotes the  $k^{\text{th}}$  event that occurred to customer  $u$ );
- Time of occurrence ( $t_k^u$  denotes the time when the  $k^{\text{th}}$  event occurred to customer  $u$ ).

Some descriptive statistics about this data-set are listed below:

- There are around  $U = 63.000$  customers;
- The number of rows is  $\sum_{u=1}^U N^u(T_2) = 149.000$  approximately, therefore there are around 2.4 events per customer on average;
- We also present the distributions of events in Table 3 below.

For hyperparameters introduced in the previous section we set the following values:

- $N_{iter} = 100$ ;
- $\varepsilon = 0.005$ ;
- $\omega = 0.05$ , fixed from the rule of thumb detailed in [2];
- $k = 2$ ;
- $D = 6$ ;
- $\alpha_1 = 10^4$ ;
- $\alpha_2 = 10^6$ .

To set  $\alpha_1$  and  $\alpha_2$ , we performed a grid-search over a parameter grid based on NLL value, where  $(\alpha_1, \alpha_2) \in [10^0, 10^8]^2$ .

In Figure 1 we plot the  $D$  basis functions  $g_d, d \in \{1, \dots, D\}$ .

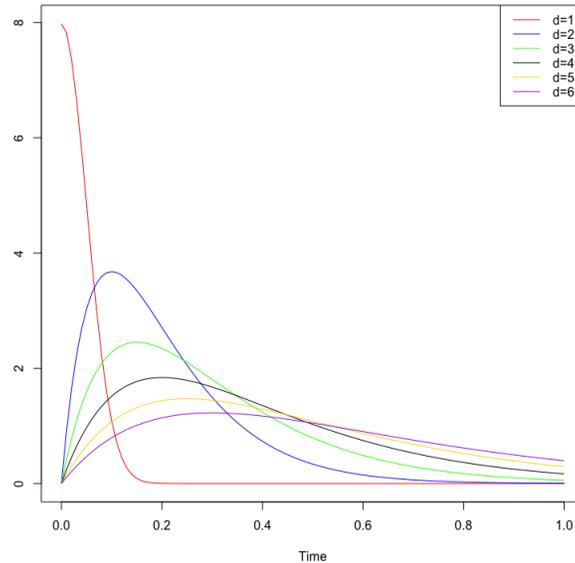


Figure 1: Plot of basis functions  $g_d, d \in \{1, \dots, D\}$

## 4.2 Parameters estimation

The parameters estimation leads to the following results:

- $\hat{\boldsymbol{\mu}} = (\hat{\mu}_i)_{i \in \{1, \dots, m\}} = [2.74 \times 10^{-2}, 6.76 \times 10^{-1}, 8.65 \times 10^{-2}, 2.11 \times 10^{-4}, 1.52 \times 10^{-3}, 8.24 \times 10^{-2}]$ ;
- $\hat{\mathbf{A}}_d = (\hat{a}_{i,j}^d)_{i,j=1, \dots, m}$ ,  $\hat{\mathbf{A}} = \{\hat{\mathbf{A}}_d\}_{d \in \{1, \dots, D\}}$ : in Table 1 we compute  $\hat{\mathbf{A}}^{sum}$ , estimation of  $\mathbf{A}^{sum} = \sum_{d=1}^D \mathbf{A}_d = (a_{i,j}^{sum})_{i,j=1, \dots, m}$ , the sum of the  $D$  infectivity matrix.

Triggering \ Triggered	Insurance cover modif.	Change of vehicle	Household composition	Birth	Change of job	Move
Insurance cover modif.	$1.46 \times 10^{-2}$	$4.45 \times 10^{-2}$	$8.29 \times 10^{-4}$	$3.42 \times 10^{-6}$	$2.28 \times 10^{-3}$	$1.65 \times 10^{-3}$
Change of vehicle	$1.74 \times 10^{-15}$	$6.38 \times 10^{-15}$	$1.50 \times 10^{-5}$	$4.82 \times 10^{-4}$	$1.14 \times 10^{-3}$	$7.34 \times 10^{-8}$
Household composition	$2.52 \times 10^{-3}$	$8.64 \times 10^{-3}$	$6.27 \times 10^{-23}$	$5.49 \times 10^{-3}$	$1.02 \times 10^{-2}$	$3.07 \times 10^{-2}$
Birth	$9.64 \times 10^{-5}$	$6.87 \times 10^{-5}$	$3.14 \times 10^{-5}$	$3.92 \times 10^{-15}$	$1.56 \times 10^{-3}$	$3.61 \times 10^{-3}$
Change of job	$1.02 \times 10^{-3}$	$2.07 \times 10^{-4}$	$5.66 \times 10^{-4}$	$9.78 \times 10^{-4}$	$1.29 \times 10^{-2}$	$2.02 \times 10^{-3}$
Move	$9.75 \times 10^{-4}$	$6.58 \times 10^{-7}$	$4.23 \times 10^{-3}$	$4.96 \times 10^{-4}$	$2.86 \times 10^{-3}$	$1.80 \times 10^{-18}$

Table 1: Estimation of the sum of the  $D$  infectivity matrix  $\mathbf{A}^{sum}$

We propose below an interpretation of this parameters estimation.

**Highest interactions between events.** The matrix  $\mathbf{A}^{sum}$  models the way events trigger each other. The higher is  $a_{i,j}^{sum}$ , the higher is the increasing of the intensity of event  $i$  when  $j$  occurred. We sort in Table 2 the interactions Event  $i \rightarrow$  Event  $j$  by decreasing coefficients  $a_{i,j}^{sum}$ .

Event triggering	Event triggered	Coefficient
Change of vehicle	Insurance cover modification	$4.45 \times 10^{-2}$
Move	Household composition	$3.07 \times 10^{-2}$
Insurance cover modification	Insurance cover modification	$1.46 \times 10^{-2}$
Change of job	Change of job	$1.29 \times 10^{-2}$
Change of job	Household composition	$1.02 \times 10^{-2}$
Change of vehicle	Household composition	$8.64 \times 10^{-3}$
Birth	Household composition	$5.49 \times 10^{-3}$
Household composition	Move	$4.23 \times 10^{-3}$
Move	Birth	$3.61 \times 10^{-3}$
Change of job	Move	$2.86 \times 10^{-3}$

Table 2: Highest interactions between events

We notice that insurance cover modifications are likely to be triggered by a change of vehicle and by a previous modification. It means that according to our model, we should recommend a new guarantee when the customer buys a new car (a different vehicle implies new needs of insurance) or has recently modified his cover (the customer is more receptive after a first change). Other high interactions also make sense, such as the link between move and household composition.

**Expected value.** For each customer  $u$  and type of event  $i$ , assuming  $N_i^u$  is a stationary process, the expected total of events at the end of observation period is denoted  $n_i^u$ . We also denote  $\mathbf{n}^u = [n_i^u]_{i \in \{1, \dots, m\}}$ :  $\mathbf{n}^u$  is given by (see [4]):

$$\mathbf{n}^u = T_u \left( I_m - \mathbf{A}^{sum} \right)^{-1} \boldsymbol{\lambda}(\mathbf{x}_u), \quad (4.1)$$

where  $I_m$  is the identity matrix of dimensions  $m \times m$ . This first order moment allows us to check whether the order of magnitude of parameters is correct, by comparing the theoretical expected value of the total of events and the distribution observed in the data-set. We could also quantify the average part of occurrences

triggered by past events, for event  $i$  and customer  $u$ , by computing  $\frac{n_i^u - \lambda_i(\mathbf{x}_u)T_u}{n_i^u}$ . On the opposite, the quotient  $\frac{\lambda_i(\mathbf{x}_u)T_u}{n_i^u}$  represents the part of events which occur due to the "natural" tendency of the customer, modelled by the background intensity. In Table 3, we present:

- The distribution of events observed in the data-set, on average per customer:  $\frac{1}{U} \sum_{u=1}^U N_i^u(T_u)$ ;
- The average of expected value of the number of events, over every customer:  $\frac{1}{U} \sum_{u=1}^U n_i^u$ ;
- The average of background intensities multiplied by time exposure, over every customer:  $\frac{1}{U} \sum_{u=1}^U \lambda_i(\mathbf{x}_u)T_u$ ;
- The average part of occurrences triggered by past events, over every customer:  $\frac{1}{U} \sum_{u=1}^U \frac{n_i^u - \lambda_i(\mathbf{x}_u)T_u}{n_i^u}$ .

Type of event	Observed	Expected value	Background int.	% from past events
Insurance cover modif.	$9.70 \times 10^{-2}$	$9.16 \times 10^{-2}$	$1.50 \times 10^{-2}$	83.6%
Change of vehicle	1.22	1.21	1.21	0.01%
Household composition	$4.18 \times 10^{-1}$	$4.35 \times 10^{-1}$	$4.11 \times 10^{-1}$	5.66%
Birth	$5.64 \times 10^{-2}$	$5.36 \times 10^{-2}$	$5.18 \times 10^{-2}$	3.29%
Change of job	$1.63 \times 10^{-1}$	$1.52 \times 10^{-1}$	$1.49 \times 10^{-1}$	2.22%
Move	$4.32 \times 10^{-1}$	$3.93 \times 10^{-1}$	$3.91 \times 10^{-1}$	0.60%

Table 3: Comparison between the distribution of events observed in the data-set and the expected total of events

First, we observe that the order of magnitude of parameters is correct, by comparing actual average of events with their expected value. Estimation on event number 1 seems more accurate than other events, which could be explained by the fact that change of vehicle represents half of total events. We also notice that according to our model, 83% of insurance cover modifications are triggered by other life events. This observation justifies our model a posteriori, since the objective is to take into account the influence of life events on this type of event. We could also observe that on the opposite the changes of vehicle are almost never triggered by other events. This seems to be in compliance with Luxembourgish customers behavior: they buy a new vehicle every 3-4 years without waiting for a specific event.

**Prediction by simulation.** The main interest of building our recommendation system with a Multivariate Hawkes Process is that the recommendations are updated with respect to events occurrences, which allows us to have the best timing to suggest insurance covers to customers. The recommendation system is not based on expected values provided by the model, but on instantaneous intensities updated by events occurrences. However, it is useful to know the distribution of the total of events for each customer for other purposes than marketing (for instance, forecasting the workload of company's staff, which is proportional to the total of events to manage). We evaluate the distribution numerically by simulating  $n$  trajectories and calculating empirical distribution.

In order to simulate the estimated Multivariate Hawkes Process for customer  $u$  from time 0 to  $T_u$ , we use the Ogata's thinning algorithm (see [20]), written in Algorithm 2 in Appendix.

We plot in Figure 2 the simulated distribution of events over 10 000 simulations for a 45-year-old Luxembourgish customer with parameters:  $T_u = 5$ ,  $f_1(\mathbf{x}_u) = 0.12$ ,  $f_2(\mathbf{x}_u) = 0.32$ ,  $f_4(\mathbf{x}_u) = 47.97$ ,  $f_5(\mathbf{x}_u) = 16$ .

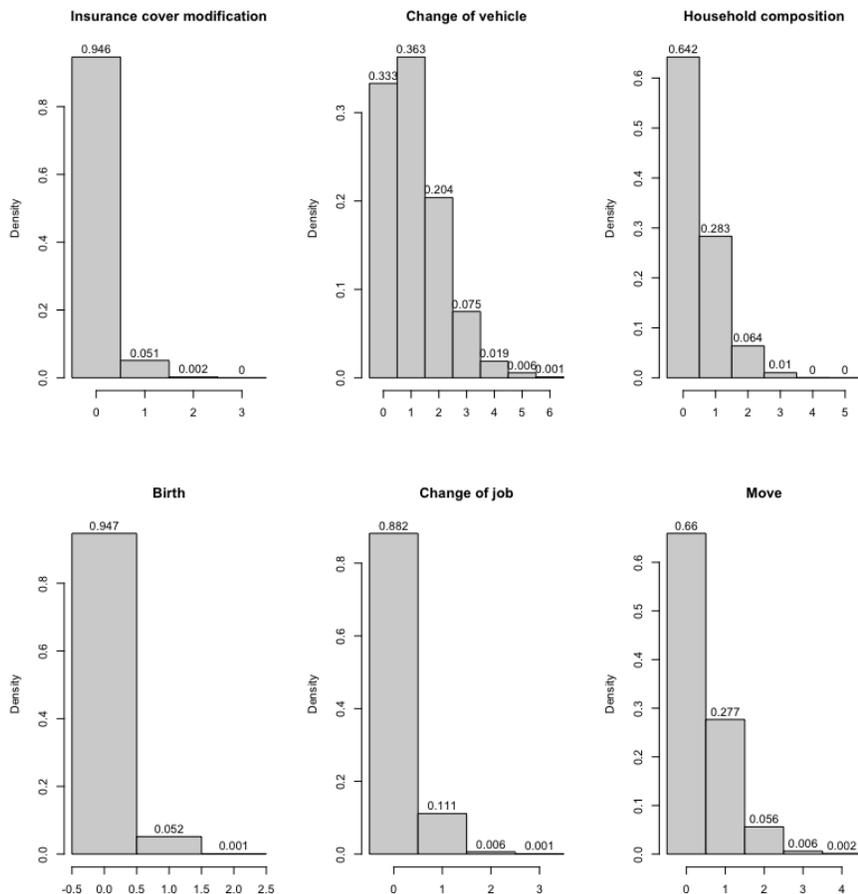


Figure 2: Distribution of events over 10 000 simulations with parameters:  $T_u = 5$ ,  $f_1(\mathbf{x}_u) = 0.12$ ,  $f_2(\mathbf{x}_u) = 0.32$ ,  $f_4(\mathbf{x}_u) = 47.97$ ,  $f_5(\mathbf{x}_u) = 16$ .

For instance, we could observe that the simulated customer has a 5% probability for modifying his insurance cover at least one time over the next 5 years. We also notice that other events distributions make sense given customer’s age and situation.

### 4.3 Influence of regularization

As a reminder, regularization aims to correct the main default observed on the data-set, which is the inaccuracy of times of occurrence. Some events might happen before some others according to the data-set, while it is the opposite in real life. The consequence of this inaccuracy may be that these misreported dates could reinforce the wrong coefficients from triggering matrix. The objective of this subsection is to show that regularization allows us to fix this problem.

To do so, we compare the list of higher interactions without and with regularization. Especially, we check whether couples of events (Event 1, Event 2) which appear on top of the list with both causations Event 1  $\rightarrow$  Event 2 and Event 2  $\rightarrow$  Event 1 without regularization appear with only one causation with regularization. Table 4 compares higher interactions (as defined in Section 4.2) in both cases by highlighting the various couples of events.

Without regularization			With regularization		
Triggering	Triggered	Coefficient	Triggering	Triggered	Coefficient
Move	Household composition	$5.90 \times 10^{-2}$	Change of vehicle	Insurance cover modification	$4.45 \times 10^{-2}$
Change of vehicle	Insurance cover modification	$5.38 \times 10^{-2}$	Move	Household composition	$3.07 \times 10^{-2}$
Insurance cover modification	Change of vehicle	$4.89 \times 10^{-2}$	Insurance cover modification	Insurance cover modification	$1.46 \times 10^{-2}$
Insurance cover modification	Insurance cover modification	$4.70 \times 10^{-2}$	Change of job	Change of job	$1.29 \times 10^{-2}$
Change of vehicle	Household composition	$2.39 \times 10^{-2}$	Change of job	Household composition	$1.02 \times 10^{-2}$
Change of job	Change of job	$2.35 \times 10^{-2}$	Change of vehicle	Household composition	$8.64 \times 10^{-3}$
Household composition	Move	$1.91 \times 10^{-2}$	Birth	Household composition	$5.49 \times 10^{-3}$
Move	Birth	$1.82 \times 10^{-2}$	Household composition	Move	$4.23 \times 10^{-3}$
Change of job	Household composition	$1.35 \times 10^{-2}$	Move	Birth	$3.61 \times 10^{-3}$
Move	Insurance cover modification	$8.40 \times 10^{-3}$	Change of job	Move	$2.86 \times 10^{-3}$

Table 4: Comparison of highest interactions between events without and with regularization

We notice that with regularization, only one causation from couples of events (**Move**, **Household composition**) and (**Change of vehicle**, **Insurance cover modification**) is kept as expected. The other causation has a lower coefficient, which allows us to define which event triggers the other for every couple.

#### 4.4 Backtesting over year 2019

In order to check the accuracy of our recommendation system, we perform a backtesting over year 2019 as follows:

1. We consider the Multivariate Hawkes Process learned as described previously from data observed from 2015 to 2018, whose parameters are given in Section 4.2. For each customer  $u$ , we focus on  $\lambda_1^{u,*}(t)$ , the intensity related to the event 1 (i.e. subscription to a new insurance cover), where  $t$  takes its values on year 2019;
2. Over 2019, we observe  $\lambda_1^{u,*}(t)$ , considering that every event of types 2 to 6 in historical data occurs on every customer  $u$ . We calculate recommendations by two methods:
  - We consider that any time the intensity is higher than a threshold  $\lambda_{\text{threshold}}^*$ , we should recommend to the customer the upgrade of his insurance cover. This threshold equals the quantile of the maximal intensity observed on each customer in 2019 of order  $1 - \alpha$ , where  $\alpha$  is the proportion of customers who changed their cover over year 2019:

$$\lambda_{\text{threshold}}^* = m(\{\max_t (\lambda_1^{u,*}(t))\}_{u \in U}, [\alpha U]), \quad (4.2)$$

where  $m(S, n)$  is the subset of  $S$  which contains elements of  $S$  higher than the  $n^{\text{th}}$  maximum of  $S$ ;

- We consider that we should make recommendations only based on background intensities, i.e. on customers  $u$  with  $\lambda_1(\mathbf{x}_u)$  higher than a threshold  $\lambda_{\text{threshold}}$  (in a similar way than in [12]) fixed so that the two methods propose the same total of recommendations:

$$\lambda_{\text{threshold}} = m(\{\lambda_1(\mathbf{x}_u)\}_{u \in U}, [\alpha U]); \quad (4.3)$$

3. To quantify the performances of the recommendations, we compute the confusion matrix which allows us to compare actual and predicted insurance cover subscriptions. We represent the confusion matrix for both methods respectively in Tables 5 and 6.

Actual/Predicted	Subscription	No subscription
Subscription	56 734	3401
No subscription	3401	5548

Table 5: Confusion matrix to compare actual and predicted insurance cover subscriptions over year 2019 based on full intensity

Actual/Predicted	Subscription	No subscription
Subscription	56 734	5907
No subscription	5907	3042

Table 6: Confusion matrix to compare actual and predicted insurance cover subscriptions over year 2019 based on background intensity

We could see that our Multivariate Hawkes Process allows us to compute accurate recommendations. The confusion matrix shows a sensitivity of 62% with the first method (i.e. the proportion of recommendations that are correctly identified). Moreover, the comparison with the second method shows that the influence of life events allows us to improve significantly the recommendation system accuracy. The accuracy drops to 34% without considering the impact of life events.

## 5 Appendix

### 5.1 Proof of equations (3.14) and (3.15)

We update  $\hat{\Theta}^{(l)}$  such that  $\hat{\Theta}^{(l+1)} = \underset{\Theta}{\operatorname{argmax}} F_s(\Theta, \hat{\Theta}^{(l)})$ . We solve  $\frac{\partial F_s(\Theta, \hat{\Theta}^{(l)})}{\partial \Theta} = 0$ .

$$\begin{aligned}
F_s(\Theta, \hat{\Theta}^{(l)}) &= -Q(\Theta, \hat{\Theta}^{(l)}) + R_1(\Theta) + \hat{R}_2^{(l)}(\Theta, \hat{\Theta}^{(l)}) \\
&= \sum_{u=1}^U \left\{ -\sum_{i=1}^m \left( T_u \lambda_i(\mathbf{x}_u) + \sum_{k=1}^{N^u(T_u)} \sum_{d=1}^D a_{i,e_k^u}^d G_d(T_u - t_k^u) \right) \right. \\
&\quad + \sum_{k=1}^{N^u(T_u)} p_{e_k^u, e_k^u}^{(l)} \log \left( \frac{\lambda_{e_k^u}(\mathbf{x}_u)}{p_{e_k^u, e_k^u}^{(l)}} \right) + \sum_{j=1}^{k-1} \sum_{d=1}^D p_{e_k^u, e_j^u}^{(l)} \log \left( \frac{a_{e_k^u, e_j^u}^d g_d(t_k^u - t_j^u)}{p_{e_k^u, e_j^u}^{(l)}} \right) \\
&\quad \left. + \alpha_1 \left( \sum_{d=1}^D \sum_{i=1}^m \sum_{j=1}^m \frac{(a_{i,j}^d)^2}{2} \right) + \alpha_2 \left( \sum_{d=1}^D \sum_{i=1}^m \sum_{j=1}^m \frac{(a_{i,j}^d + \hat{a}_{j,i}^{d,(l)})^2}{2} \right) \right\}.
\end{aligned}$$

For all  $k \in \{1, \dots, m\}$ ,

$$\frac{\partial F_s(\Theta, \hat{\Theta}^{(l)})}{\partial \mu_k} = -\sum_{u=1}^U T_u + \left( \sum_{u=1}^U \sum_{e_i^u = k} p_{e_i^u, e_i^u}^{(l)} \right) \frac{1}{\mu_k},$$

therefore we set  $\hat{\mu}_k^{(l+1)}$  such that  $\left( \frac{\partial F_s}{\partial \mu_k} \right)_{\mu_k = \hat{\mu}_k^{(l+1)}} = 0$ , which leads to:

$$\hat{\mu}_k^{(l+1)} = \frac{\sum_{u=1}^U \sum_{e_i^u = k} p_{e_i^u, e_i^u}^{(l)}}{\sum_{u=1}^U T_u}.$$

For all  $d \in \{1, \dots, D\}, (i, j) \in \{1, \dots, m\}^2$ ,

$$\frac{\partial F_s(\Theta, \hat{\Theta}^{(l)})}{\partial a_{i,j}^d} = -\sum_{u=1}^U \sum_{e_k^u = j} G_d(T_u - t_k^u) + \left( \sum_{u=1}^U \sum_{e_k^u = i} \sum_{e_m^u = j} p_{e_k^u, e_m^u}^{(l)} \right) \frac{1}{a_{i,j}^d} + (\alpha_1 + \alpha_2) a_{i,j}^d + \alpha_2 \hat{a}_{j,i}^{d,(l)}.$$

We denote:

$$A_{i,j} = - \sum_{u=1}^U \sum_{e_k^u=j} G_d(T_u - t_k^u),$$

$$B_{i,j}^{(l)} = \sum_{u=1}^U \sum_{e_k^u=i} \sum_{e_m^u=j} p_{e_k^u, e_m^u}^{(l)}.$$

We set  $\hat{a}_{i,j}^{d,(l+1)}$  such that  $\left(\frac{\partial F_s}{\partial a_{i,j}^d}\right)_{a_{i,j}^d=\hat{a}_{i,j}^{d,(l+1)}} = 0$ , which leads to:

$$\begin{aligned} A_{i,j} + B_{i,j}^{(l)} \frac{1}{a_{i,j}^d} + (\alpha_1 + \alpha_2) a_{i,j}^d + \alpha_2 \hat{a}_{j,i}^{d,(l)} &= 0, \\ (\alpha_1 + \alpha_2) (a_{i,j}^d)^2 + (A_{i,j} + \alpha_2 \hat{a}_{j,i}^{d,(l)}) a_{i,j}^d + B_{i,j}^{(l)} &= 0, \end{aligned}$$

$$\hat{a}_{i,j}^{d,(l+1)} = \frac{-(A_{i,j} + \alpha_2 \hat{a}_{j,i}^{d,(l)}) + \sqrt{(A_{i,j} + \alpha_2 \hat{a}_{j,i}^{d,(l)})^2 - 4(\alpha_1 + \alpha_2) B_{i,j}^{(l)}}}{2(\alpha_1 + \alpha_2)}.$$

## 5.2 Ogata's thinning algorithm

---

**Algorithm 2:** Ogata's thinning algorithm

---

**Result:** Simulated sequence of events for customer  $u$ :  $\{(e_k^{u,sim}, t_k^{u,sim})_{k=1}^{N^{u,sim}}\}$

**1. Generate the first event:**

$k = 1$ ;

$\lambda_{max} = \sum_{i=1}^m \lambda_i^{u,*}(0)$ ;

Simulate  $V \sim \mathcal{U}[0, 1]$ ;

$s = -\frac{1}{\lambda_{max}} \ln(V)$ ;

**if**  $s > T_u$  **then**

    Go to step 3.;

**else**

    Simulate  $W \sim \mathcal{U}[0, 1]$ ;

$e_k^{u,sim} = j$  such that  $\sum_{i=1}^j \lambda_i^{u,*}(s) \leq \lambda_{max} \times W \leq \sum_{i=1}^{j+1} \lambda_i^{u,*}(s)$ ;

$t_k^{u,sim} = s$

**end**

**2. General:**

$k = k + 1$ ;

$\lambda_{max} = \sum_{i=1}^m \lambda_i^{u,*}(t_{k-1}^{u,sim}) + \sum_{i=1}^m a_{e_{k-1}^{u,sim}, i}^{sum}$ ;

Simulate  $V \sim \mathcal{U}[0, 1]$ ;

$s = s - \frac{1}{\lambda_{max}} \ln(V)$ ;

**if**  $s > T_u$  **then**

    Go to step 3.;

**else**

    Simulate  $W \sim \mathcal{U}[0, 1]$ ;

**if**  $\lambda_{max} \times W \leq \sum_{i=1}^m \lambda_i^{u,*}(s)$  **then**

$e_k^{u,sim} = j$  such that  $\sum_{i=1}^j \lambda_i^{u,*}(s) \leq \lambda_{max} \times W \leq \sum_{i=1}^{j+1} \lambda_i^{u,*}(s)$ ;

$t_k^{u,sim} = s$

**else**

$\lambda_{max} = \sum_{i=1}^m \lambda_i^{u,*}(s)$  and simulate another time  $s$ ;

**end**

**end**

**3. Output:**  $\{(e_k^{u,sim}, t_k^{u,sim})_{k=1}^{N^{u,sim}}\}$

---

**Acknowledgements**

We developed this work through a strong collaboration with Foyer Assurances, leader of individual and professional insurance in Luxembourg, which provided the domain specific knowledge and use cases. The first author also gratefully acknowledges the funding received towards our project (number 13659700) from the Luxembourgish National Research Fund (FNR).

## References

- [1] Zhou, Zha, Song, Learning Triggering Kernels for Multi-dimensional Hawkes Processes. 2013. Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA.
- [2] Xu, Farajtabar, Zha, Learning Granger Causality for Hawkes Processes. 2016. Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA.
- [3] Zhou, Zha, Song, Learning Social Infectivity in Sparse Low-rank Networks Using Multi-dimensional Hawkes Processes. 2013. Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS), Scottsdale, AZ, USA. Volume 31 of JMLR: W&CP 31.
- [4] Bacry, Muzy, First- and Second-Order Statistics Characterization of Hawkes Processes and Non-Parametric Estimation. 2015. IEEE Transactions on Information Theory, vol. 62, no. 4, pp. 2184-2202.
- [5] Achab, Emmanuel Bacry, Gaiffas, Mastromatteo and Muzy, Uncovering Causality from Multivariate Hawkes Integrated Cumulants. 2018. Journal of Machine Learning Research 18 (2018) 1-28.
- [6] Christian R. Shelton and Zhen Qin and Chandini Shetty, Hawkes Process Inference With Missing Data. 2018. AAAI Conference 2018.
- [7] Young Lee, Kar Wai Lim, Cheng Soon Ong, Hawkes Processes with Stochastic Excitations. 2016. Proceedings of The 33rd International Conference on Machine Learning, PMLR 48:79-88.
- [8] <https://statistiques.public.lu>
- [9] J.H. Friedman, Greedy function approximation: a gradient boosting machine. 2001. Annals of Statistics, pp. 1189–1232.
- [10] K.P. Burnham and D.R. Anderson, Multimodel inference: understanding AIC and BIC in Model Selection, 2004. Sociological Methods and Research, p. 261–304.
- [11] <https://dares.travail-emploi.gouv.fr/sites/default/files/pdf/2018-049v2.pdf>
- [12] Lesage, Deaconu, Lejay et al. A recommendation system for car insurance. 2020. Eur. Actuar. J. 10, pp. 377–398.
- [13] article 2
- [14] Denuit and Charpentier, Mathematiques de l'assurance non-vie ; Tome 1 : Principe fondamentaux de theorie du risque. 2004.
- [15] Wenming Xiao, Xiao Xu, Kang Liang, Junkang Mao, and Jun Wang, Job recommendation with Hawkes process: an effective solution for RecSys Challenge 2016. 2016. Proceedings of the Recommender Systems Challenge (RecSys Challenge '16). Association for Computing Machinery, New York, NY, USA, Article 11, 1–4.
- [16] J. Shang and M. Sun, Local Low-Rank Hawkes Processes for Temporal User-Item Interactions. 2018. 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 2018, pp. 427-436.
- [17] Nan Du, Yichen Wang, Niao He, Jimeng Sun, Le Song, Time-Sensitive Recommendation From Recurrent User Activities. 2015. NIPS 2015: pp. 3492-3500.
- [18] Guoming Zhang, Junshu Wang, Nan Jiang, Yehua Sheng, A Point-of-interest Recommendation Method Based on Hawkes Process. 2018. Acta Geodaetica et Cartographica Sinica, 2018, 47(9): 1261-1269.
- [19] <https://www.insee.fr/fr/statistiques/2882020>
- [20] Ogata, Yosihiko. On Lewis' simulation method for point processes. 1981. IEEE Transactions on Information Theory 27.1 (1981): 23-31.