Department Project

Behavior Modeling using Hawkes Processes

May 2017

Milliman R&D - Confidential

Pierre BOYEAU, Antoine FRANÇOIS, Olivier RIGAL et Laury ZHOU

Contents

Abstract

In this paper, we will present our approach to model insured parties' behavior using Hawkes processes. We studied a single type of financial transactions made by different groups of insured persons between two types of account.

Our first goal was to determine whether previous transactions could be modelized by an Hawkes process. If it appeared to be true, our second goal was to predict future transactions thanks to this modelization.

Therefore, this paper details the mathematical knowledge we used, the implementation of our model and the different tests we made to analyse the veracity of our model and predict future transactions.

Part I
Theoretical basis

In this part, we first introduce the Hawkes Processes theory, which is the framework of this study, and the associated notations. Then, we calculate the likelihoods which are to be computed in the algorithm for the parameter estimation.

1 Framework

We consider d groups of contracts; these may be closed to new entrants or open to new business. We consider a time origin at zero; from this time switch events occur in each group $i \in \{1, ..., d\}$ at random times denoted $(T_n^i)_{n\geq 1}$. This sequence defines a counting process $(N_t^i)_{t\geq 0}$ as

$$N_t^i = \sum_{n\geq 1} \mathbf{1}_{T_n^i \leq t}.$$

Therefore, N_t^i is the number of switches which occurred for group i in the time interval $[0, t]$. The intensity process of $(N_t^i)_{t\geq 0}$ is denoted $(\lambda_i(t))_{t\geq 0}$.

2 Model

2.1 Hawkes Processes theory

We propose a Hawkes process to model the self-excitation in each group as well the mutually excitation between groups; it is specified as follows:

- For each $i \in \{1, ..., d\}$, the base intensity is a deterministic, continuous and non-negative map $t \mapsto \mu_i(t)$,
- For each $(i, j) \in \{1, ..., d\}^2$, self and mutually-exciting maps $t \mapsto \phi_{i,j}(t)$ are introduced and also assumed to be deterministic, continuous and non-negative,
- For each $i \in \{1, ..., d\}$, the intensity process of $N_\cdot^i$ is specified as follows:

$$\lambda_i(t) = \mu_i(t) + \sum_{j=1}^{d} \sum_{T_n^j < t} \phi_{i,j}(t - T_n^j) = \mu_i(t) + \sum_{j=1}^{d} \int_0^t \phi_{i,j}(t - s) \mathrm{d}N_s^j \qquad (1)$$

In this model, the maps $\phi_{i,i}$ quantify the self excitation in the group i, whereas for $i \neq j$, the map $\phi_{i,j}$ quantifies the contagion **in** group i **caused by** an event in group j. Note here that each intensity process λ_i is adapted to the canonical filtration associated with the whole vector of processes $(N_\cdot^j)_{j\in\{1,...,d\}}$; this way the behavior of a given group may (generally) depend on that of the others.

2.2 Interpretation

We can give a qualitative interpretation of the $\phi_{i,j}$ functions. $||\phi_{i,j}|| = \alpha_{i,j}$ is the number of events of group i created a time $t+1$ by an event of group j that appears at time t. Thus, the total number of events in i created by an initial event in j is :

$$c_{i,j} = ||\phi_{i,j}|| + ||\phi_{i,j}||^2 + \cdots = \frac{||\phi_{i,j}||}{1 - ||\phi_{i,j}||} = \frac{\alpha_{i,j}}{1 - \alpha_{i,j}}$$

$c_{i,j}$ (and thus $\alpha_{i,j}$) is a fundamental parameter because it can be interpreted as a endogeneity factor.
Later on, knowing $\alpha_{i,j}$ and $\mu_{i,j}$ will allow us to simulate or predict the future intensity of an Hawkes process.

3 Likelihood

We aim at determining the parameters of the model and use an algorithm that maximizes the likelihood of the model to achieve it. Therefore, knowing how to compute this quantity is mandatory.

3.1 Observables

Let us consider that the processes are observed on the interval $[0, t]$, and denote $(t_n^k)_{1 \leq n \leq m_k}$ the times of event observed for each group k.

3.2 Likelihood for the single group case

Let us start with the single group case; we omit the group index for simplicity of notation. Note that to calibrate such model on each single group amounts to specify $\phi_{i,j} \equiv 0$ for $i \neq j$ in Equation (1).
We introduce the notation $\mathcal{H}_n = \{T_n = t_n, ..., T_1 = t_1\}$ the information on the first n times of event, and add the conventions $\mathcal{H}_0 = \emptyset$ and $T_0 = 0$. Having observed the times $(t_n)_{1 \leq n \leq m}$, the likelihood writes (by abuse of notation we keep the T_n)

$$\begin{aligned}
\mathcal{L}(\mu, \phi) &= \mathbb{P}(\forall 1 \leq n \leq m, T_n = t_n, \text{ and } T_{m+1} > t) \\
&= \mathbb{P}(T_{m+1} > t \mid \mathcal{H}_m) \prod_{n=1}^{m} \mathbb{P}(T_n = t_n \mid \mathcal{H}_{n-1}) \\
&= \exp\left(- \int_{T_m}^{t} \lambda(s) \mathrm{d}s \right) \prod_{n=1}^{m} \exp\left(- \int_{T_{n-1}}^{T_n} \lambda(s) \mathrm{d}s \right) \lambda(T_n) \\
&= \exp\left(- \int_{0}^{t} \lambda(s) \mathrm{d}s \right) \prod_{n=1}^{m} \lambda(T_n)
\end{aligned}$$

Then the log-likelihood writes

$$\log \mathcal{L}(\mu, \phi) = - \int_0^t \lambda(s) \mathrm{d}s + \sum_{n=1}^{m} \log \lambda(T_n) = - \int_0^t \lambda(s) \mathrm{d}s + \int_0^t \log \lambda(s) \mathrm{d}N_s. \quad (2)$$

This is the standard log-likelihood for any counting process N with intensity λ; it now remains to further specify it in the context of the model introduced in Equation (1), as

$$\log \mathcal{L}(\mu, \phi) = -\int_0^t \mu(s)\mathrm{d}s - \int_0^t \sum_{T_n<s} \phi(s-T_n)\mathrm{d}s + \sum_{n=1}^{m} \log\left(\mu(T_n) + \sum_{k=1}^{n-1} \phi(T_n - T_k)\right) \tag{3}$$

If we further specify the self-exciting map as $\phi(t) = \alpha \exp(-\beta t)$, with non-negative α and β, we obtain

$$\begin{aligned} \log \mathcal{L}(\mu, \alpha, \beta) = &-\int_0^t \mu(s)\mathrm{d}s - \alpha \int_0^t \sum_{T_n<s} \exp\left(-\beta(s - T_n)\right) \mathrm{d}s \\ &+ \sum_{n=1}^{m} \log\left(\mu(T_n) + \alpha \sum_{k=1}^{n-1} \exp\left(-\beta(T_n - T_k)\right)\right) \end{aligned} \tag{4}$$

Note 1: one has to keep in mind that the log-likelihood is not concave in β when choosing the optimization algorithm

Note 2: one can take advantage of the fact that the quantity $g(t) = \sum_{T_k<t} \exp(-\beta(t - T_k))$ can be recursively computed since

$$g(T_{n+1}) = \exp\left(-\beta(T_{n+1} - T_n)\right)\left(1 + g(T_n)\right)$$

3.3 Likelihood for the mutually-exciting model

We first compute the likelihood using a method similar to the one above. We denote $\{T_n^i = t_n^i, \forall 1 \leq i \leq d, \forall 1 \leq n \leq m_i\}$ by $\mathcal{H}$:

$$\mathcal{L}(\mu, \phi) = \mathbb{P}(\forall 1 \leq i \leq d, \forall 1 \leq n \leq m_i, T_n^i = t_n^i, \text{ and } T_{m+1}^i > t)$$

Therefore :

$$\log \mathcal{L}(\mu, \phi) = \sum_{i=1}^{d} \log \mathcal{L}^i(\mu, \phi) \tag{5}$$

Using the same method as above, we find by having observed $((t_n^i)_{1\leq n\leq m_i})_{1\leq i\leq d}$ the associated log-likelihood :

$$\begin{aligned} \log \mathcal{L}(\mu, \phi) &= -\sum_{i=1}^{d} \int_0^t \lambda_i(s)\mathrm{d}s + \sum_{i=1}^{d}\sum_{n=1}^{m_i} \log \lambda_i(t_n^i) \\ &= -\sum_{i=1}^{d} \int_0^t \left(\mu_i(s) + \sum_{j=1}^{d} \int_0^s \phi_{i,j}(s-u)\mathrm{d}N_u^j\right)\mathrm{d}s \\ &+ \sum_{i=1}^{d}\sum_{n=1}^{m_i} \log\left(\mu_i(t_n^i) + \sum_{j=1}^{d} \int_0^{t_n^i} \phi_{i,j}(t_n^i - s)\mathrm{d}N_s^j\right) \end{aligned} \tag{6}$$

We then specify the self and mutually-exciting maps : $\phi_{i,j}(t) = \alpha_{i,j} \exp(-\beta t)$.

Note that the speed β at which interaction kernel decays is the same whatever the groups considered and $\alpha_{i,j}$ quantifies the impact of group j on group i.
This exponential specification leads to the following log-likelihood:

$$\begin{aligned}\log \mathcal{L}(\mu,\alpha,\beta) = &-\sum_{i=1}^{d}\int_0^t \left(\mu_i(s) + \sum_{j=1}^{d}\int_0^s \alpha_{i,j}\exp\left(-\beta(s-u)\right)\mathrm{d}N_u^j\right)\mathrm{d}s \\ &+\sum_{i=1}^{d}\sum_{n=1}^{m_i}\log\left(\mu_i(t_n^i) + \sum_{j=1}^{d}\int_0^{t_n^i}\alpha_{i,j}\exp\left(-\beta(t_n^i-s)\right)\mathrm{d}N_s^j\right)\end{aligned} \tag{7}$$

We define for all $j \in \{1,...,d\}$:

$$g_j(t) = \sum_{t_{n'}^j \le t}\exp\left(-\beta(t-t_{n'}^j)\right) = \int_0^t \exp\left(-\beta(t-s)\right)\mathrm{d}N_s^j$$

$$G_j(t) = \int_0^t g_j(s)\mathrm{d}s$$

This enables us to rewrite the log-likelihood this way :

$$\begin{aligned}\log \mathcal{L}(\mu,\alpha,\beta) &= -\sum_{i=1}^{d}\left[\int_0^t \mu_i(s)\mathrm{d}s + \sum_{j=1}^{d}\alpha_{i,j}\int_0^t g_j(s)\mathrm{d}s\right] \\ &\quad+\sum_{i=1}^{d}\sum_{n=1}^{m_i}\log\left(\mu_i(t_n^i) + \sum_{j=1}^{d}\int_0^{t_n^i}\alpha_{i,j}\exp\left(-\beta(t_n^i-s)\right)\mathrm{d}N_s^j\right) \\ \log \mathcal{L}(\mu,\alpha,\beta) &= -\sum_{i=1}^{d}\left[\int_0^t \mu_i(s)\mathrm{d}s + \sum_{j=1}^{d}\alpha_{i,j}G_j(t)\right] \\ &\quad+\sum_{i=1}^{d}\sum_{n=1}^{m_i}\log\left(\mu_i(t_n^i) + \sum_{j=1}^{d}\alpha_{i,j}g_j(t_n^i)\right) \\ \log \mathcal{L}(\mu,\alpha,\beta) &= -\sum_{i=1}^{d}\left[\sum_{n=1}^{m_i}\int_{t_{n-1}^i}^{t_n^i}\mu_i(s)\mathrm{d}s + \sum_{j=1}^{d}\alpha_{i,j}\sum_{n=1}^{m_i}\left(G_j(t_n^i) - G_j(t_{n-1}^i)\right)\right] \\ &\quad+\sum_{i=1}^{d}\left[\int_{t_{m_i}^i}^{t}\mu_i(s)\mathrm{d}s + \sum_{j=1}^{d}\alpha_{i,j}\left(G_j(t) - G_j(t_{m_i}^i)\right)\right] \\ &\quad+\sum_{i=1}^{d}\sum_{n=1}^{m_i}\log\left(\mu_i(t_n^i) + \sum_{j=1}^{d}\alpha_{i,j}g_j(t_n^i)\right) \\ \log \mathcal{L}(\mu,\alpha,\beta) &= -\sum_{i=1}^{d}\sum_{n=1}^{m_i}\left[\int_{t_{n-1}^i}^{t_n^i}\mu_i(s)\mathrm{d}s + \sum_{j=1}^{d}\alpha_{i,j}\left(G_j(t_n^i) - G_j(t_{n-1}^i)\right)\right] \\ &\quad+\sum_{i=1}^{d}\sum_{n=1}^{m_i}\left[\log\left(\mu_i(t_n^i) + \sum_{j=1}^{d}\alpha_{i,j}g_j(t_n^i)\right)\right] \\ &\quad+\sum_{i=1}^{d}\left[\int_{t_{m_i}^i}^{t}\mu_i(s)\mathrm{d}s + \sum_{j=1}^{d}\alpha_{i,j}\left(G_j(t) - G_j(t_{m_i}^i)\right)\right]\end{aligned}$$

We remark that g_j and G_j can be easily computed by using the following recursive relations :

$$g_j(t^i_{n+1}) = g_j(t^i_n)\exp\left(-\beta(t^i_{n+1} - t^i_n)\right) + \sum_{t^i_n \le t^j_{n'} < t^i_{n+1}} \exp\left(-\beta(t^i_{n+1} - t^j_{n'})\right)$$

$$G_j(t^i_{n+1}) = G_j(t^i_n) + \int_{t^i_n}^{t^i_{n+1}} g_j(s)ds$$

We can observe that for each $s \in [t^i_n, t^i_{n+1}]$, $g_j(s)$ can be written as:

$$\sum_{t^i_{n'} \le t^i_n} e^{-\beta(s - t^i_{n'})} + \sum_{t^i_n < t^j_k < t^i_{n+1}} e^{-\beta(s - t^j_k)}$$

$$g_j(s) = g_j(t^i_n)\exp(-\beta(s - t^i_n)) + \sum_{t^i_n \le t^j_k < s} \exp(-\beta(s - t^j_k))$$

Thus:

$$G_j(t^i_{n+1}) = G_j(t^i_n) + \int_{t^i_n}^{t^i_{n+1}} \sum_{t^i_{n'} \le t^i_n} e^{-\beta(t^i_n - t^i_{n'})} e^{-\beta(s - t^i_n)} ds + \int_{t^i_n}^{t^i_{n+1}} \sum_{t^i_n \le t^j_k \le t^i_{n+1}} e^{-\beta(s - t^j_k)} ds$$

$$G_j(t^i_{n+1}) = G_j(t^i_n) + g_j(t^i_n)\frac{1 - e^{-\beta(t^i_{n+1} - t^i_n)}}{\beta} + \sum_{t^i_n \le t^j_k \le t^i_{n+1}} \frac{1 - e^{-\beta(t^i_{n+1} - t^j_k)}}{\beta}$$

Part II
Parameters estimation

Once we have explicited the function $(\mu, \alpha, \beta) \rightarrow \mathcal{L}(\mu, \alpha, \beta)$, we want to find the most likely parameters linked to our observations.
To do so, we are going to minimize efficiently this function in order to determine these parameters.

1 Parameters interpretation

In order to interpret studied datasets, we will mainly focus on α. For $i \neq j$, $\alpha_{i,j}$ represents the influence of group j on group i, and $\alpha_{i,i}$ is an indicator of group i auto-excitation.

Thus, this parameter can help us answer the questions:

- Which are the "leaders" groups, those that are influenced?
- Are the observed phenomena exogenous?

2 Stochastic gradient descent algorithm

In order to evaluate the different parameters, we have implemented a stochastic gradient descent algorithm. The principle of the algorithm is presented below:

Algorithm 1 Stochastic Gradient Descent

$n \leftarrow 0$
while Converged = False **do**
  $n \leftarrow n + 1$
  Choose at random (i, k) in $1, ..., D$ and $1, ..., N^i$
  Compute $s = \mu^i + \sum_{j'=1}^{D} a^{ij'} g^{ij'}(t_i^k)$
  $\mu^i \leftarrow mu^i + \gamma_n(\frac{1}{s} - (t_i^k - t_i^{k-1}))$
  for $j \leftarrow 1$ to D **do**
    $a^{ij} \leftarrow a^{ij} + \gamma_n(\frac{g^{ij}(t_i^k)}{s} - (G^{ij}(t_i^k) - G^{ij}(t_i^{k-1}))$
  end for
end while

It is an iterative algorithm. At each step, we change the parameters following one direction of the gradient. This method has been selected over other techniques (ex : normal gradient descend for instance) because it can be applied to a large amount of data.
γ_n represents the size of each step. For this function, we have tested several possibilities like a constant or $\frac{1}{n}$. However, we observed that the results were much better with $\gamma_n = \sqrt{\frac{\eta}{n}}$
For the test of convergence, we have tried 2 possibilities:

- The first one was to consider the norm of the gradient. We stopped the algorithm once this norm was smaller than a given ϵ.
- The second one was to do N iterations where N is a parameter of the algorithm.

3 Test of the algorithm on simulated data

Because this algorithm might be reused by Milliman, its efficiency and results needed to be guaranteed. This is the reason why we have tested it on simulated data first. To simulate the data, we used the R package 'Hawkes' (4).
For instance, on the graph 1, we represented the evolutions of the intensity of two Hawkes processes of parameters $\alpha = \begin{bmatrix} 0.1 & 0.2 \\ 0.8 & 0.8 \end{bmatrix}$ and $\mu = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$. On this graph, we can clearly see that the second process' excitations cause excitations for the first process.

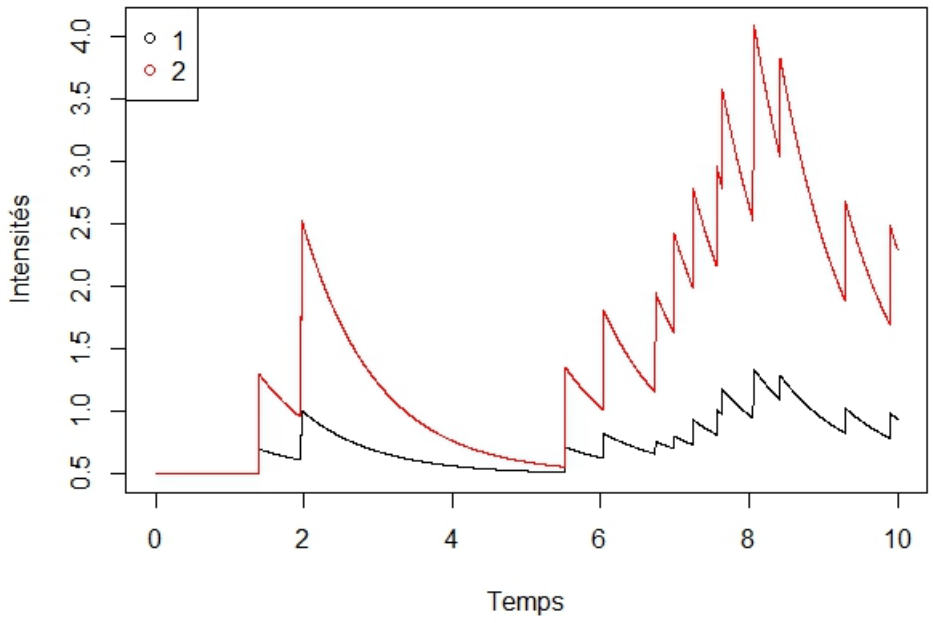

Figure 1: Simulation of interdependant Hawkes processes

The algorithm allows us to evaluate the parameters μ and α. In order to evaluate the parameter β, we execute the algorithm for several values of β and then we select the value which maximizes the likelihood.

- Thus, for several configurations such as this one, we have confronted our results with the expected ones. It has allowed us to confirm the accuracy of our algorithm.

- We also checked that the SGD was stable, i.e. that if we changed the initial values of α and μ, or if we sensibly altered the form of γ_n, the obtained parameters' value didn't change much.
- To be assured that the algorithm converges, we plot the evolution of the log-likelihood with respect to the number of iterations. This is another proof of the efficiency of our algorithm.

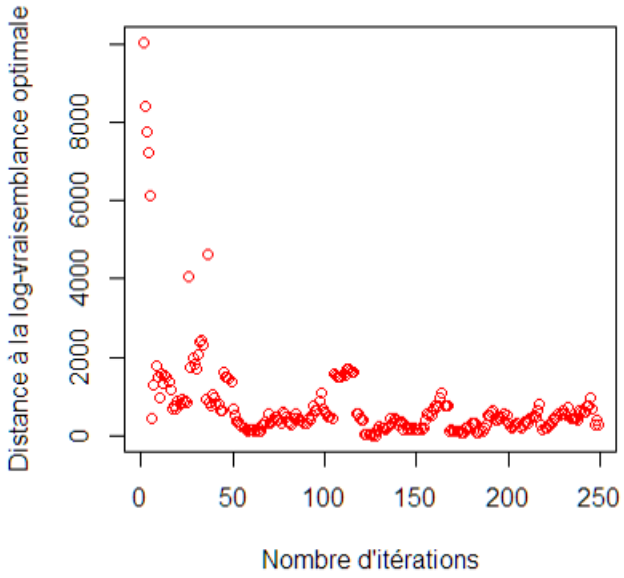

Figure 2: Evolution of log-likelihood

Part III
Results analysis on an actuarial database

In the last step of our approach, we decided to test our algorithm on Milliman's database. In this part, we will, first, explain the content of the database. Then, we will explicit the meaning of the several results obtained from the algorithm. And finally, we will show how, from the obtained results, we were able to predict future behavior for the different groups person insured.

1 Content of the database

The database given by Milliman include the operations made by four different groups during a time period of eleven years. As it is confidential information, we did not have access to the detailed description of each group which will be named "a", "b", "c" and "d".

However, we can interpret the differences between each group as a different purchasing medium. To give a practical example, we consider a insurance company who wants to sell a new financial product. To do so, the insurance company can use different distribution channel :

- itself
- a partner bank
- a insurance broker

Therefore, the different groups would represent the groups of person buying this financial product through the different distribution channels.

Before using our algorithm on the database, we plotted the data for each group, so that we could see, roughly, if the data followed an Hawkes process. To plot the data, we decided to plot the number of operations with a step of eighty days.
The following graphs show the shape of the different groups of the database.

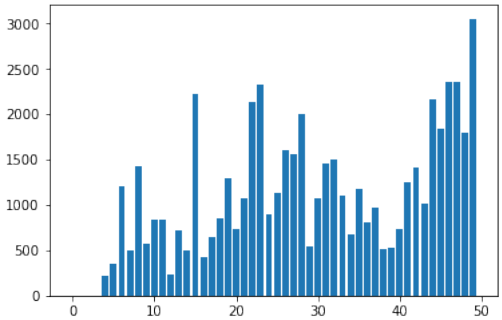

Figure 3: Group "a"

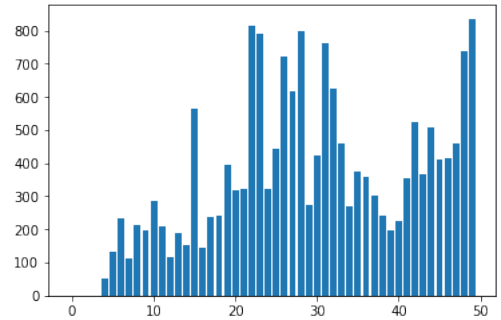

Figure 4: Group "b"

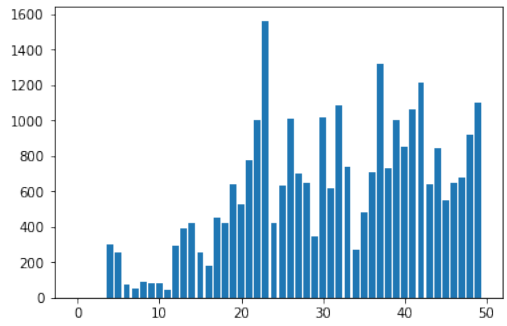

Figure 5: Group "c"

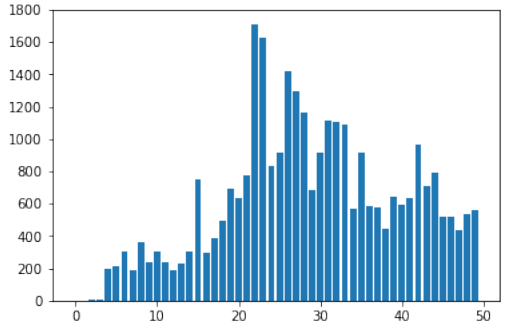

Figure 6: Group "d"

On these graphs, we can see that for each group, there are peaks of operations. Furthermore, we can observe that, at certain times, there are peaks of the same magnitude for each group. Consequently, these graphs strengthen our approach of using Hawkes processes to simulate insured persons' behavior.

2 Estimation of Hawkes Processes on the database

Thanks to our algorithm, we were able to estimate the coefficient of an Hawkes process for each pair of groups of the database.
On the following charts, we can see the obtained results which contains the matrix of coefficient α and the vector of coefficients μ .

α	0,803870341	0,477588482
	0,129630271	0,479297161
μ	0,96164748	0,355941418

Figure 7: Group "a" on "b"

α	0,938661783	0,164831226
	0,130191719	0,684421461
μ	0,722930309	0,365453626

Figure 8: Group "a" on "c"

α	0,791662874	0,284513677
	0,177365306	0,591357309
μ	0,714744436	0,467739993

Figure 9: Group "a" on "d"

α	0,810878558	0,77723232
	1,16293784	1,561918209
μ	0,992732965	1,045082092

Figure 10: Group "b" on "c"

α	0,57093589	0,246501053
	0,360192776	0,699638022
μ	0,235656909	0,537798004

Figure 11: Group "b" on "d"

α	0,689136867	0,171731892
	0,229801674	0,692029909
μ	0,701974829	0,95732320

Figure 12: Group "c" on "d"

We can give an example of interpretation that can be made on these coefficients. For instance, if we study the chart representing the groups "a" and "b", we can observe that "a" has a lot of endogeneity ($\alpha_{aa} = 0.803$) whereas "b" has a small coefficient of endogeneity ($\alpha_{bb} = 0.479$). Moreover, we can observe that "b" has more influence on "a" ($\alpha_{ab} = 0.477$) and that "a" has very little influence on "b" ($\alpha_{ba} = 0.129$).

The first observation we can mention is that for a chosen group, its coefficients differ when we change the group it is compared to. This shows the impact each group has on the others.

Furthermore, we compared the mean on each scenarii for each type of coefficients and each group. There are three types of coefficients :

- α_{ii} represents the endogeneity within a the group i
- α_{ij} represents the influence of the group j on the group i
- α_{ji} represents the influence of the group i on the group j

The calculus gave the following results :

i	a	b	c	d
α_{ii}	0,844731666	0,620370536	0,978492179	0,661008413
α_{ij}	0,308977795	0,384454548	0,48828715	0,255786585
α_{ji}	0,145729099	0,666906366	0,39062174	0,234248874

Figure 13: Mean on each scenarii ("j") for each type of coefficients and each group

From this chart, we can make three observrations :

- The group with the most influence on the other groups is the group "b".
- The group which is the most influenced by others is the group "c".
- Each group has a lot of endogeneity and the one with the most endogeneity is the group "c".

3 Prediction of future number of operations on the database

Thanks to our algorithm, we can estimate the coefficients defining the evolution of the different groups. These coefficients can be used in order to predict the number of events that will happen in the future.
In order to illustrate those predictions, we are going to take the example of the groups "a" and "b". In the previous section, we evaluated the different parameters influencing the evolution of the transactions in those groups. Thanks to these parameters, we can simulate an Hawkes process that will evolve consistently with the previous evolutions of the groups "a" and "b". An example of simulation over a year is presented below:

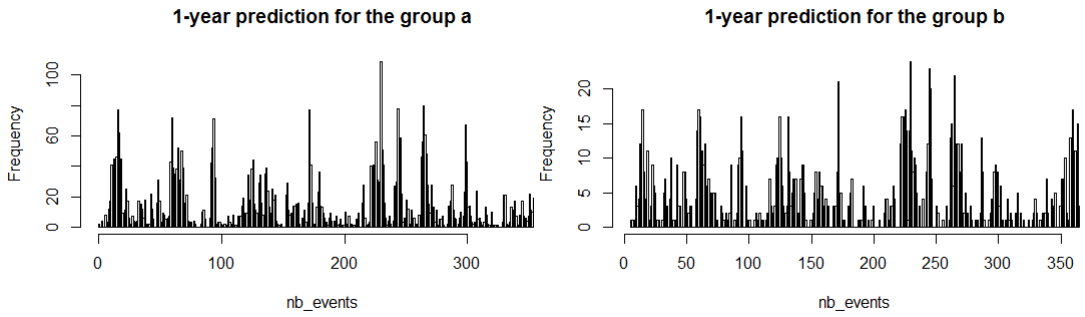

Figure 14: Simulation of the events for the group "a" and "b" over a year

In order to predict future events, we need to run numerous simulations because of their random nature. If we run 200 simulations, the histogram of the number of events in each simulation is the following:

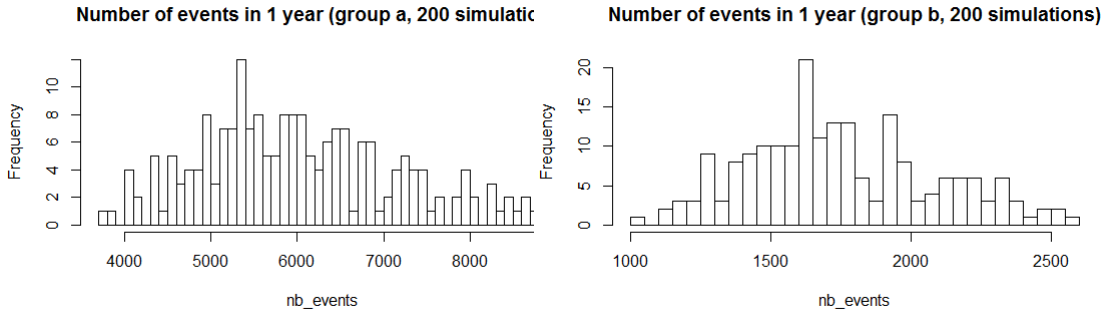

Figure 15: Histograms of the total number of events over a year for the groups "a" and "b"

For these groups, we concluded that there will be, on average, 6027 events for the group "a" and 1748 events for the group "b". However, the standard deviations are quite high, 1138 for the group "a" and 332 for the group "b". So the number of future events can't be forecast with an important accuracy with our model.

Conclusion

Thanks to a furnished mathematical documentation, we were able to successfully compute an algorithm to approximate Hawkes processes. Our algorithm answers to all the requirements of stability, convergence and accuracy. Therefore, we were able to set Hawkes processes on the actuarial database.

The results of this project offers encouraging prospects in the field of transactions' prediction. Due to a lack of time, our results are, for now, incomplete but still promising. Therefore, it may be interesting to go forward in this Hawkes modelisation in order to deliver an expertise for insurance companies.

References

[1] T. Jaisson, *Market activity and price impact throughout time scales*. CMAP, Ecole polytechnique, 2015

[2] I. Muni Toke, *An Introduction to Hawkes Processes with Applications to Finance*, cours, Centrale Paris, 2011

[3] A. Boumezoued, *Approches micro-macro des dynamiques de populations hétérogènes structurées par âge. Application aux processus auto-excitants et à la démographie*, Université Pierre et Marie Curie, 2016

[4] Package R *Hawkes*, *cran.r-project.org/web/packages/hawkes/index.html* 9