

EURIA  
Euro-Institut d'Actuariat



MASTER 1 ACTUARIAT

BUREAU D'ÉTUDE

---

# Modélisation du risque terroriste par les processus de Hawkes

---



Yannick Bessy-Roland  
Alice Launay  
Promotion 2019

Tuteurs académiques : Rainer Buckdahn  
Anthony Nahelou  
Tuteur professionnel : Antoine Quantin

2017-2018

# Table des matières

<b>Remerciements</b>	<b>4</b>
<b>Introduction</b>	<b>5</b>
<b>I Présentation théorique des processus de Hawkes</b>	<b>6</b>
<b>1 Définition d'un processus de Hawkes à une dimension</b>	<b>7</b>
1.1 Définitions . . . . .	7
1.1.1 Processus ponctuel . . . . .	7
1.1.2 Processus de comptage . . . . .	7
1.1.3 Processus d'intensité . . . . .	8
1.1.4 Processus linéaire auto-excitant . . . . .	8
1.2 Processus de Hawkes . . . . .	8
1.3 Hypothèse de stationnarité . . . . .	10
<b>2 Interprétation et applications d'un tel modèle</b>	<b>12</b>
2.1 Interprétation de la forme du modèle . . . . .	12
2.2 Applications au risque terroriste . . . . .	13
<b>3 Simulation d'un processus</b>	<b>14</b>
3.1 Algorithme de simulation d'Ogata . . . . .	15
3.2 Interprétation des paramètres . . . . .	16
<b>4 Estimation des paramètres</b>	<b>23</b>
4.1 Méthode du maximum de vraisemblance . . . . .	23
4.2 Estimation des paramètres d'un processus simulé avec l'algorithme d'Ogata . . . . .	24

<b>II</b>	<b>Application sur la Global Terrorism Database : Étude du nombre d'attentats</b>	<b>25</b>
1	Présentation de la base de données	26
2	Premiers calibrages sur la période globale	27
2.1	Attaques regroupées . . . . .	27
2.2	Attaques distinguées . . . . .	28
3	Calibrages sur différents découpages de temps	30
3.1	Méthode de calibration . . . . .	30
3.2	Détermination du nombre optimal de périodes . . . . .	31
3.3	Expansion du modèle . . . . .	32
3.3.1	Échelle mensuelle . . . . .	32
3.3.2	Échelle journalière . . . . .	33
4	Tentatives de prédictions avec le modèle de Hawkes	35
4.1	Prédictions avec un calibrage sur tout l'historique disponible . .	36
4.2	Prédictions avec un calibrage sur un historique proche de la pré- diction . . . . .	37
4.3	Prolongement du processus réel . . . . .	40
<b>III</b>	<b>Etude des coûts</b>	<b>42</b>
1	Détermination de la loi des montants	43
1.1	Loi discrète . . . . .	43
1.2	Loi continue . . . . .	44
1.3	Intérêts et désavantages des deux modèles retenus . . . . .	46
2	Identité de Wald	47
3	Étude de la ruine	48
	Conclusion	51
	Annexes	54
A	Comparaison des méthodes de prédiction	54



# Remerciements

Nous souhaitons avant tout remercier **M. Rainer Buckdahn** (enseignant-chercheur à l'UBO de Brest) notre tuteur universitaire pour nous avoir suivis et guidés au long du projet, notamment pour sa rigueur sur les aspects théoriques.

Nous remercions également **M. Antoine Quantin** (actuaire à la CCR) notre tuteur professionnel, avant tout pour avoir proposé ce sujet. Ce fût un véritable plaisir d'explorer ces processus assez peu "standards". Mais également d'avoir pris le temps de suivre notre travail, et de nous avoir apporté son point de vue plus pratique.

# Introduction

Ce document présente le bureau d'étude que nous avons réalisé tout au long de notre année de Master 1 à l'EURIA.

Ce bureau d'étude vient compléter celui réalisé l'année précédente par un groupe de Master 1 de l'EURIA dont le sujet était l'étude des risques liés au terrorisme [1].

Le présent travail, plus spécifique, a pour objectif d'étudier un type de processus de comptage nommé processus de Hawkes, dans le cadre de la modélisation de la survenance des attaques terroristes, afin de palier aux limites des modèles proposés dans le bureau d'étude précédent (notamment un nombre d'attentats prédit pouvant diverger vers l'infini).

Dans la littérature nous avons trouvé des documents traitant des processus de Hawkes, le plus souvent sur la partie théorique, et assez peu d'applications réelles et de résultats, que cela soit pour le terrorisme ou pour d'autres domaines pouvant faire intervenir ce type de processus (neurosciences, génomique, sismologie, finance, dynamisme des populations...).

Nous nous sommes penchés sur différentes méthodes pour effectuer des prédictions dans le cadre du terrorisme à l'aide de ces processus.

Première partie

Présentation théorique des  
processus de Hawkes

# Chapitre 1

## Définition d'un processus de Hawkes à une dimension

Les processus de Hawkes constituent une famille de processus ponctuels introduite dans les années 70. Pour notre étude nous nous intéresserons au processus de Hawkes à une dimension.

Afin de comprendre la définition d'un processus de Hawkes, nous allons dans un premier temps définir les outils mathématiques dont nous aurons besoin ainsi que leurs notations.

### 1.1 Définitions

#### 1.1.1 Processus ponctuel

**Définition** Soit  $(\Omega, F, P)$  un espace probabilisé complet. Soit  $(t_i)_{i \in \mathbb{N}^*}$  une suite strictement croissante de variables aléatoires positives ou nulles. On appelle  $(t_i)_{i \in \mathbb{N}^*}$  un processus ponctuel sur  $\mathbb{R}^+$ .

#### 1.1.2 Processus de comptage

**Définition** Soit  $(t_i)_{i \in \mathbb{N}^*}$  un processus ponctuel sur  $\mathbb{R}^+$ . On appelle processus de comptage le processus  $N$  continu à droite tel que

$$N(t) = \sum_{i \in \mathbb{N}^*} 1_{[t_i, +\infty[}(t) \quad , t \geq 0$$

### 1.1.3 Processus d'intensité

**Définition** Soit  $N$  un processus de comptage adapté à la filtration  $(\mathcal{F}_t)_{t \geq 0}$ , le processus d'intensité continu à gauche est défini par sa fonction d'intensité  $\lambda$  telle que

$$\begin{aligned}\lambda(t) &= \lim_{h \rightarrow 0} E \left[ \frac{N(t+h) - N(t)}{h} \middle| \mathcal{F}_t \right] \\ &= \lim_{h \rightarrow 0} \frac{1}{h} P [N(t+h) - N(t) > 0 | \mathcal{F}_t] \quad , t \geq 0.\end{aligned}$$

L'intensité  $\lambda(t)$  du processus représente donc la probabilité qu'il y ait un saut entre le temps  $t$  et le temps  $t+h$  lorsque  $h$  tend vers 0, connaissant le nombre d'évènements déjà survenus avant  $t$  (attention : ce n'est pas une probabilité, elle peut prendre des valeurs supérieures à 1). En d'autres termes, cela représente le **risque qu'un évènement survienne au temps  $t$** . Ici, la notion d'intensité n'a pas de rapport avec la gravité d'un attentat.

Dans la suite nous utiliserons la filtration  $(\mathcal{F}_t)_{t \geq 0}$  telle que

$$\mathcal{F}_t := (\cap_{s > t} \sigma(N(u) | u \leq s)) \vee \mathcal{N}_p$$

avec  $\mathcal{N}_p = \{A \in \Omega | \exists B \in \mathcal{F}, P(B) = 0, A \subset B\}$ .

### 1.1.4 Processus linéaire auto-excitant

**Définition générale** Un processus linéaire auto-excitant  $N$  est un processus de comptage de fonction d'intensité  $\lambda$  de la forme

$$\begin{aligned}\lambda(t) &= \lambda_0(t) + \int_{]-\infty, t[} \nu(t-s) dN_s \quad \text{où } \begin{cases} \lambda_0 : \mathbb{R}^+ \longrightarrow \mathbb{R}^+ \\ \nu : \mathbb{R}^+ \longrightarrow \mathbb{R}^+ \end{cases} \\ &= \lambda_0(t) + \sum_{t_i < t} \nu(t-t_i); \end{aligned}$$

$\lambda_0$  est la fonction d'intensité de base (déterministe), et  $\nu$  représente l'influence positive des évènements passés  $t_i$  sur la valeur actuelle du processus d'intensité en  $t$ .

## 1.2 Processus de Hawkes

**Définition** Un processus de Hawkes est un processus linéaire auto-excitant  $N$  dont la fonction  $\nu$  est de la forme :

$$\nu(t) = \sum_{j=1}^M \alpha_j e^{-\beta_j t} 1_{\mathbb{R}^+}(t), \quad t \geq 0.$$

où  $\alpha_j > 0, \beta_j > 0$ , pour tout  $j \in \{1, \dots, M\}$ , et  $M \in \mathbb{N}^*$ .

La fonction d'intensité devient alors

$$\lambda(t) = \lambda_0(t) + \sum_{t_i < t} \sum_{j=1}^M \alpha_j e^{-\beta_j(t-t_i)}, \quad t \geq 0.$$

**L'objet de notre étude est de déterminer le nombre d'attentats  $N(t)$  au cours d'une période  $[0, t]$ , où  $N(t)$  est un processus de Hawkes d'intensité  $\lambda(t)$ ,  $t \geq 0$ .**

Nous commencerons notre étude dans le cas le plus simple, c'est-à-dire  $M = 1$ , nous noterons  $\alpha = \alpha_1, \beta = \beta_1$  et nous supposerons dans la suite que  $\lambda_0$  est constant.

### Exemple

Voici ci-dessous une simulation de l'intensité  $\lambda$  d'un processus de Hawkes N, sur une période  $[0, 10]$  et de paramètres  $\lambda_0 = 1.2, \alpha = 0.6, \beta = 0.8$ . Chaque pic d'intensité se forme en un temps de survénance  $t_i$ , puis décroît de manière exponentielle dans le temps. Ici, nous pouvons observer un nombre de survénances simulées  $N(10) = 6$ .

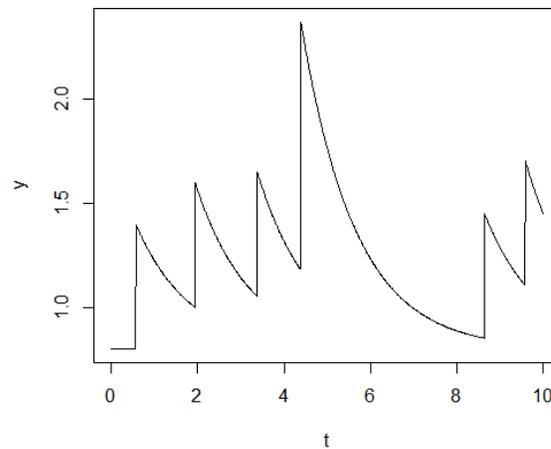


FIGURE 1.1 – Intensité d'un processus de Hawkes 1D en fonction du temps

### 1.3 Hypothèse de stationnarité

Pour éviter de modéliser un risque qui parte à l'infini, et ainsi rester dans un cadre réaliste, nous préférons supposer que l'intensité est stationnaire. Dans ce cas, nous avons  $E[\lambda(t)] = \mu$  (où  $\mu$  est une constante).

$$\begin{aligned}\mu &= E[\lambda(t)] = E \left[ \lambda_0 + \int_{]-\infty, t[} \nu(t-s) dN_s \right] \\ &= \lambda_0 + E \left[ \int_{-\infty}^t \nu(t-s) \lambda(s) ds \right] \\ &= \lambda_0 + \int_{-\infty}^t \nu(t-s) \mu ds \\ &= \lambda_0 + \mu \int_0^{\infty} \nu(v) dv\end{aligned}$$

Ce qui nous donne

$$\begin{aligned}\mu &= \frac{\lambda_0}{1 - \int_0^{\infty} \nu(v) dv} \\ &= \frac{\lambda_0}{1 - \frac{\alpha}{\beta}} \quad \text{pour } M = 1.\end{aligned}$$

Une condition nécessaire à la stationnarité de l'intensité :

$$\sum_{j=1}^M \frac{\alpha_j}{\beta_j} < 1; \quad \text{c'est-à-dire } \alpha < \beta \quad \text{pour } M = 1.$$

Cette hypothèse semble être justifiée car la condition nécessaire à la stationnarité est vérifiée sur les données réelles. En effet, nous avons estimé les paramètres d'un processus de Hawkes (estimation sans imposer de contraintes) sur l'ensemble des données françaises qui nous sont disponibles, et nous observons que  $\alpha = 40.94 < 72.95 = \beta$ .

Sous cette hypothèse, nous pouvons calculer explicitement le **nombre moyen théorique** de survenances sur une période donnée, dans le but final de le comparer au nombre réel d'attentats survenus au cours de cette même période. Le calcul est le suivant :

Notons  $M_t = N(t) - \int_0^t \lambda(s) ds$ ,  $t \geq 0$ . Le processus  $M = (M_t)_{t \geq 0}$  est une  $(F, P)$  martingale. Nous avons alors :

$$E[M_t] = E[M_0] = 0, \quad t \geq 0$$

Il en découle

$$E[N(t)] = E \left[ \int_0^t \lambda(s) ds \right] = \int_0^t E[\lambda(s)] ds, \quad t \geq 0$$

Ce qui implique

$$E[N(t)] = t * \frac{\lambda_0}{1 - \frac{\alpha}{\beta}}, \quad t \geq 0.$$

## Chapitre 2

# Interprétation et applications d'un tel modèle

### 2.1 Interprétation de la forme du modèle

A l'origine, les processus de Hawkes ont été introduits dans le but de modéliser les répliques sismiques. Dans le modèle, les  $t_i$  sont les temps de survenance de chaque événement  $i$  (ici de chaque séisme).

Une particularité des processus de Hawkes est la non-indépendance entre les événements. Plus précisément, l'intensité en un temps  $t$  dépend des événements déjà survenus. La fonction  $\nu$  (présente dans l'écriture de l'intensité  $\lambda(t) = \lambda_0(t) + \sum_{t_i < t} \nu(t - t_i)$ ) caractérise l'effet auto-excitant. En effet c'est une fonction à valeurs positives qui dépend des écarts entre les temps de survenance (précédant  $t$ ) et le temps actuel  $t$ . Dans le cadre d'un processus de Hawkes, le choix d'une fonction exponentielle négative ( $\nu(t - t_i) = \sum_{j=1}^M \alpha_j e^{-\beta_j(t - t_i)} \mathbf{1}_{\mathbb{R}_+}(t - t_i)$ ) caractérise la manière dont l'intensité décroît après chaque nouvel événement.

On comprend alors le choix d'une telle forme pour modéliser les tremblements de terre qui sont des événements que l'on peut supposer dépendants dans le temps : la probabilité qu'il y ait un séisme augmente si il y en a eu un récemment, mais elle diminue lorsque l'on s'éloigne dans le temps du dernier événement survenu.

Les processus de Hawkes sont également utilisés dans d'autres applications telles que les neurosciences, les réseaux sociaux et la finance statistique.

## 2.2 Applications au risque terroriste

### **Pourquoi choisir les processus de Hawkes pour modéliser le risque terroriste ?**

Les propriétés citées précédemment sont intéressantes si l'on souhaite modéliser des temps de survenance correspondants aux actes terroristes.

En effet, en partant de l'hypothèse qu'il n'est pas invraisemblable que les actes terroristes surviennent par "vague", de la même façon que les répliques sismiques, nous aurions alors bien un risque qui monte en pic lorsqu'une attaque vient de survenir, et un risque qui décroît à mesure que l'on s'en éloigne dans le temps.

Les propriétés des processus de Hawkes permettraient alors de prendre en compte ces tendances dans le terrorisme.

Plusieurs travaux ont déjà été faits sur ce sujet, parmi lesquels [1] et [2].

## Chapitre 3

# Simulation d'un processus

Notre objectif est d'étudier l'ajustement d'un processus de Hawkes, dans le but de modéliser et prédire le nombre d'attaques terroristes susceptibles de survenir sur une période donnée.

Notre première idée a été de calibrer les paramètres d'un processus de Hawkes simple ( $M = 1$ ) à une dimension sur les temps de survenance des actes terroristes passés ; dans le but final de comparer le nombre de survenances simulées avec notre modèle et le nombre d'attentats réel donné par la GTD<sup>1</sup>. Cela nécessite d'abord d'être capable d'estimer les paramètres d'un processus de Hawkes si nous fournissons des temps de survenance ; et ensuite d'être capable de générer une simulation d'un processus de Hawkes sous condition de fournir les paramètres.

Nous avons donc commencé par programmer une fonction sous R permettant de simuler des temps de survenance suivants un processus de Hawkes (la fonction prenant en argument les paramètres du processus à simuler).

Ensuite nous avons créé une seconde fonction renvoyant une estimation des paramètres d'un processus de Hawkes (cette fonction prenant en argument des temps de survenance).

Les sections suivantes (3.1 et 3.2) concernent ces deux travaux. Nous rappelons la forme de l'intensité d'un processus de Hawkes simple ( $M=1$ ) :

$$\lambda(t) = \lambda_0 + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}, \quad t \geq 0.$$

---

1. GTD : Global Terrorism Database. C'est la base de données que nous utiliserons dans la partie II.

### 3.1 Algorithme de simulation d'Ogata

L'algorithme d'Ogata (1981) est un algorithme qui génère des temps de survenance suivants un processus de Hawkes sur un intervalle de temps  $[0, T]$  sous réserve de fournir les paramètres  $\lambda_0$ ,  $\alpha$ , et  $\beta$ ; ainsi que la fonction d'intensité  $\lambda$  qui dépend de ces trois paramètres.

---

**Algorithm 1** Initialisation

---

**Require:**  $\lambda_0$   $\alpha$   $\beta$   $T$   
 $\lambda^* \leftarrow \lambda_0$ ,  $n \leftarrow 1$   
**First event :** Generate  $U \rightsquigarrow U_{[0,1]}$  and set  $s \leftarrow -\frac{1}{\lambda^*} \ln(U)$   
**if**  $s \leq T$  **then**  
     $t_1 \leftarrow s$   
**else**  
    go to the First event  
**end if**

---

Cette étape d'initialisation crée le premier temps de survenance  $t_1$ , qui est ici une réalisation d'une loi exponentielle de paramètre  $\lambda_0$ .

---

**Algorithm 2** Routine générale

---

**while**  $t_n \leq T$  **do**  
    Set  $n \leftarrow n + 1$   
    **a)**  $\lambda^* \leftarrow \lambda(t_{n-1}) + \alpha$   
    *à chaque étape dans la boucle, l'algorithme modifie le paramètre  $\lambda^*$  de la loi exponentielle en rajoutant un saut  $\alpha$  à la valeur de l'intensité précédente.*  
    **b) New event :** Generate  $U \rightsquigarrow U_{[0,1]}$  and set  $s \leftarrow s - \frac{1}{\lambda^*} \ln(U)$   
    **if**  $s \leq T$  **then**  
         $t_1 \leftarrow s$   
    **else**  
        go to the last step  
    **end if**  
    **c) Rejection test :** Generate  $D \rightsquigarrow U_{[0,1]}$   
    **if**  $D \leq \frac{\lambda(s)}{\lambda^*}$  **then**  
         $t_n \leftarrow s$  and go through the general routine again  
    **else**  
         $\lambda^* \leftarrow \lambda(s)$  and try a new date at step (b) of the general routine  
    **end if**  
**end while**

---

La routine générale de l'algorithme génère les autres temps de survenances à l'aide d'un test de rejet.

Le principe général de l'algorithme est le suivant :

- **Initialisation** : l'algorithme génère le premier temps de survenance  $s$ , de loi exponentielle  $\lambda_0$  (intensité initiale), il ne le retient que si ce temps est inférieur à la durée limite  $T$ .
- **Routine générale** : l'algorithme considère l'intensité au moment du dernier pic ( $\lambda^* \leftarrow \lambda(t_{n-1}) + \alpha$ ) et génère la durée d'attente avant la prochaine survenance, de loi exponentielle de paramètre  $\lambda^*$ . Le prochain temps de survenance potentiel est donc le dernier temps de survenance auquel on ajoute cette durée (toujours sous la contrainte d'un temps inférieur à  $T$ , si le temps est supérieur, on génère une nouvelle variable aléatoire de loi exponentielle jusqu'à obtenir un temps inférieur à  $T$ ). L'algorithme effectue ensuite un test de rejet, si la condition est vérifiée, on "accepte" le nouveau temps de survenance et on recommence la boucle. Si la condition n'est pas vérifiée, on n'enregistre pas le temps de survenance, mais  $\lambda^*$  prend la valeur de l'intensité en ce temps, et on revient à l'étape où l'on génère un nouveau temps de survenance.

## 3.2 Interprétation des paramètres

Afin d'étudier l'impact des paramètres sur l'intensité, nous avons simulé plusieurs processus en faisant varier un paramètre (en gardant les deux autres constants).

- **Influence du paramètre  $\lambda_0$**

$$\lambda(t) = \lambda_0 + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}, \quad t \geq 0$$

Le paramètre  $\lambda_0$  représente l'intensité de base du processus. Il ne dépend pas des événements passés. Le nombre d'événements (les temps de survenance) ainsi que la valeur moyenne des pics d'intensité augmentent avec le risque de base. C'est bien ce que nous observons sur les graphiques suivants qui illustrent l'évolution de l'intensité en fonction du temps lorsque  $\lambda_0$  augmente (avec  $\alpha = 0.6$  et  $\beta = 0.8$ ).

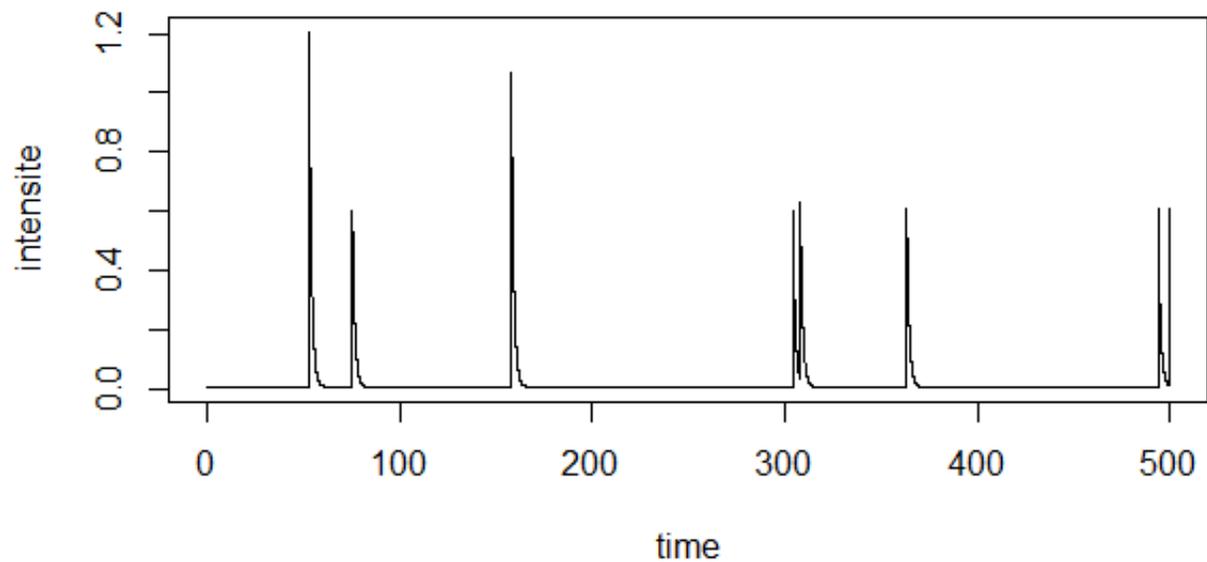


FIGURE 3.1 -  $\lambda_0 = 0.01$

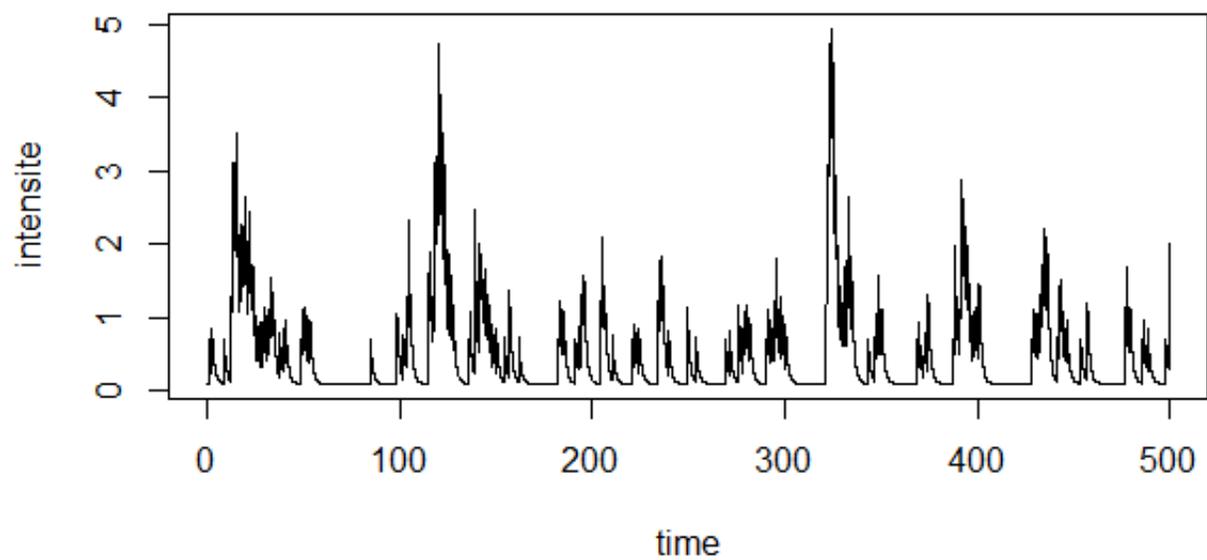


FIGURE 3.2 -  $\lambda_0 = 0.1$

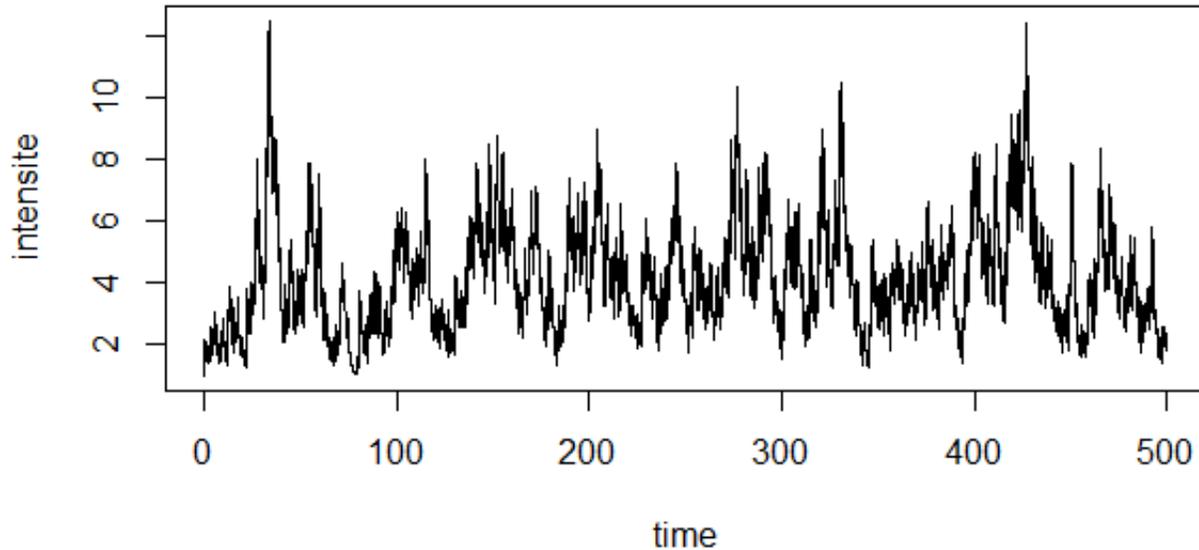


FIGURE 3.3 –  $\lambda_0 = 1$

- Influence du paramètre  $\alpha$

$$\lambda(t) = \lambda_0 + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}, \quad t \geq 0$$

Le paramètre  $\alpha$  est le coefficient multiplicateur devant la somme qui augmente l'intensité à chaque survenance. Nous pouvons donc nous attendre à ce qu'une augmentation de ce paramètre augmente la valeur moyenne de l'intensité aux moments de chaque survenance (lorsque  $t = t_j$ , où le j-ième événement est le dernier survenu, l'intensité augmente de  $\alpha$  car l'exponentielle concernée vaut alors 1). C'est bien le cas avec nos simulations présentées ci-dessous : plus  $\alpha$  est grand, plus les pics d'intensité aux moments des survenances ont une intensité élevée. (La constante tracée en rouge est la moyenne  $\mu$  calculée précédemment dans l'hypothèse de stationnarité, ici nous avons fixé  $\lambda_0 = 0.1$  et  $\beta = 1.2$ ).

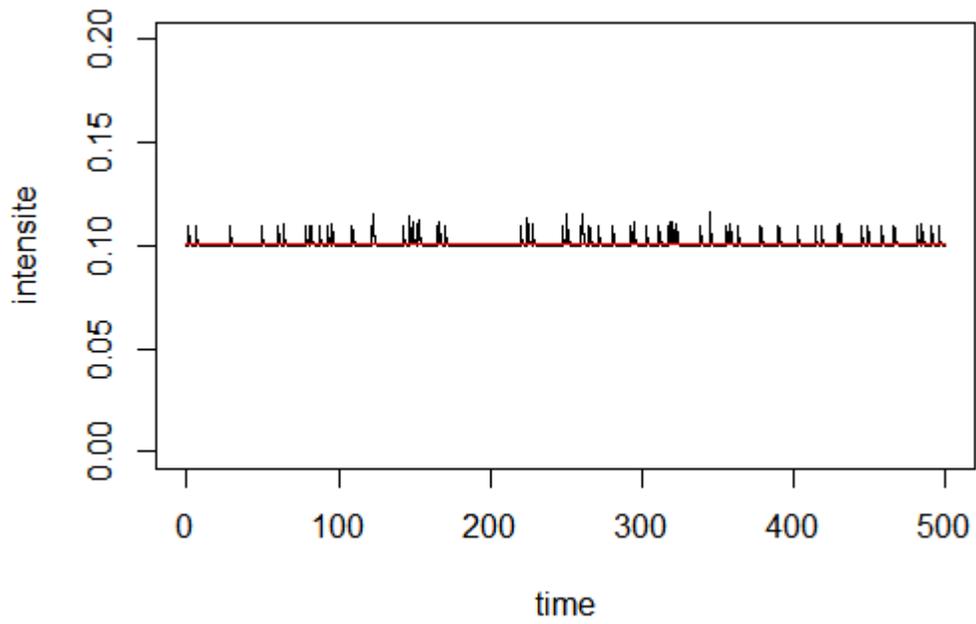


FIGURE 3.4 -  $\alpha = 0.01$

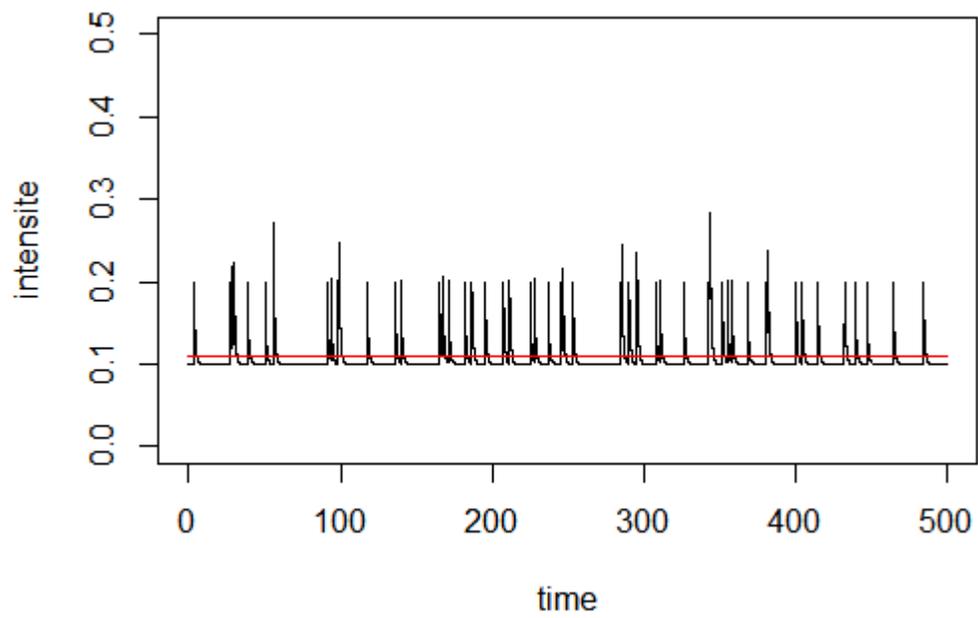


FIGURE 3.5 -  $\alpha = 0.1$

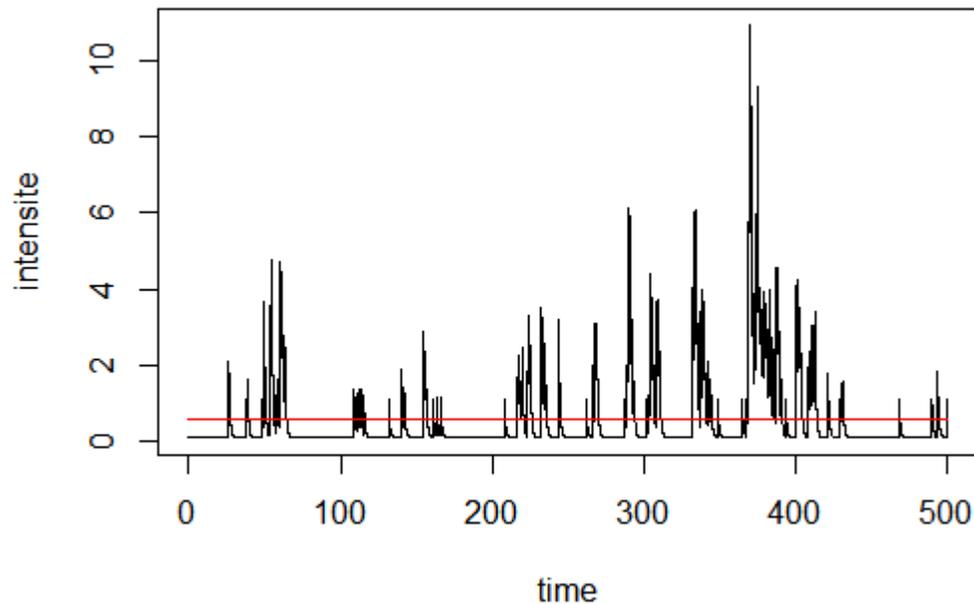


FIGURE 3.6 –  $\alpha = 1$

• Influence du paramètre  $\beta$

$$\lambda(t) = \lambda_0 + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}, \quad t \geq 0$$

Le paramètre  $\beta$  se présente dans l'exponentielle décroissante comme coefficient multiplicateur devant les écarts entre le temps actuel et les temps de survenance passés. Il modifie la vitesse de décroissance de l'exponentielle. Plus  $\beta$  est grand, plus l'intensité décroît rapidement. Les simulations présentées sur les graphiques suivants montrent cette différence de vitesse lorsque  $\beta$  varie. (La constante tracée en rouge est la moyenne  $\mu$  calculée précédemment dans l'hypothèse de stationnarité).

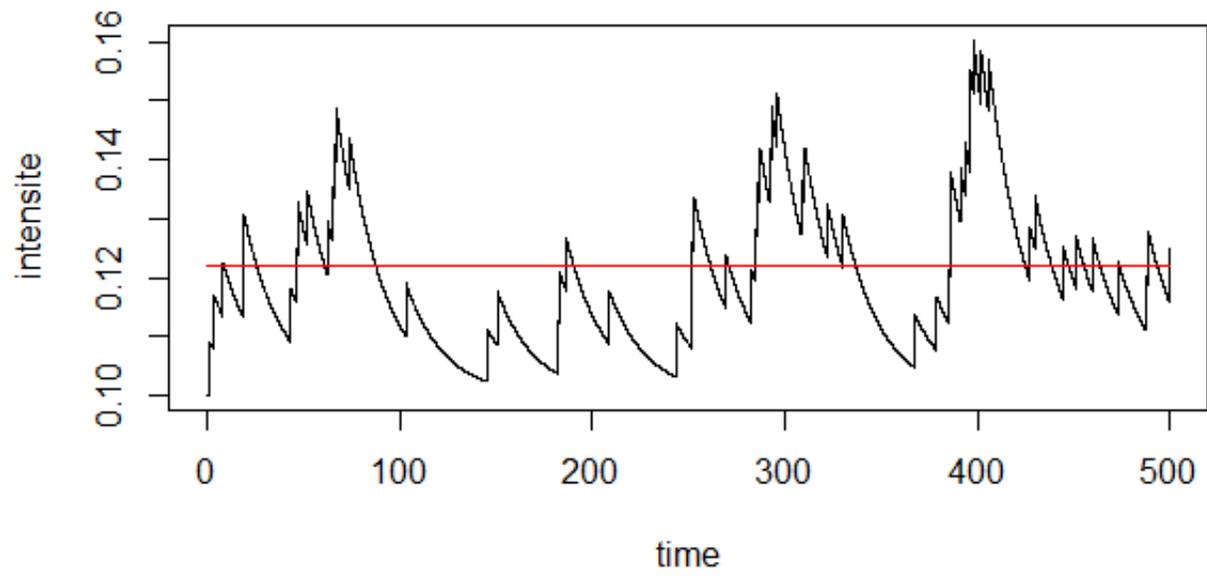


FIGURE 3.7 -  $\beta = 0.05$

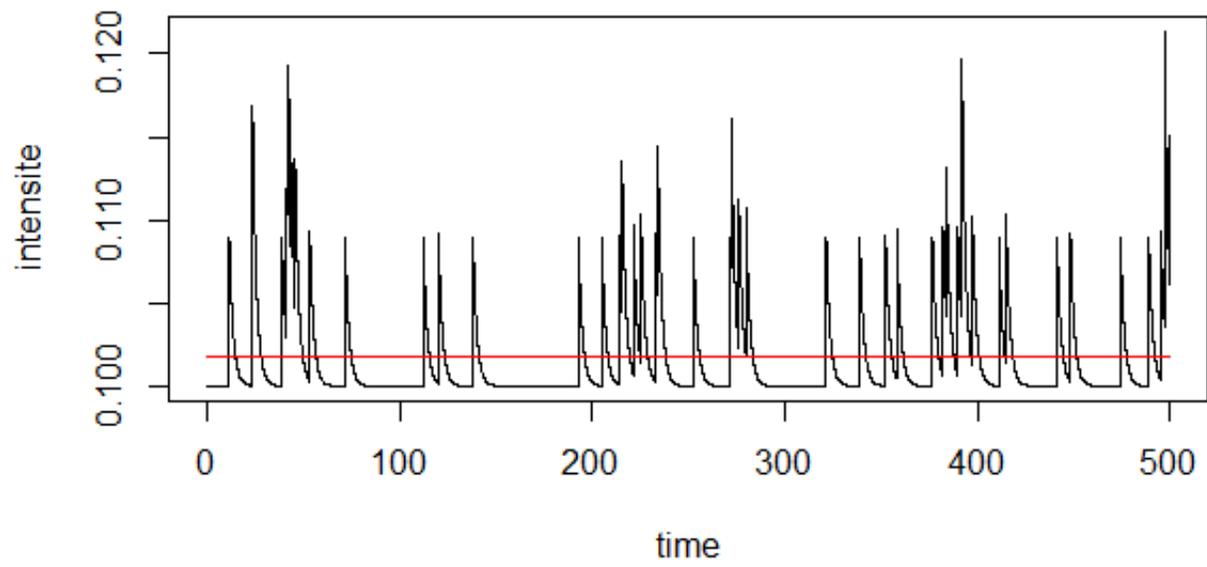


FIGURE 3.8 -  $\beta = 0.5$

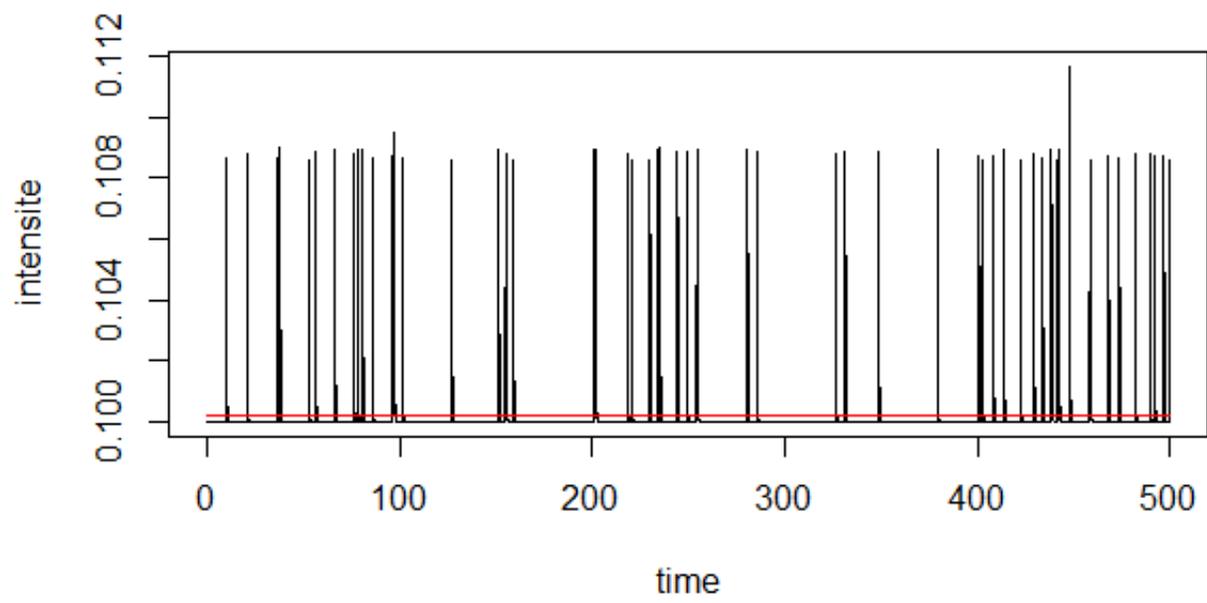


FIGURE 3.9 -  $\beta = 5$

## Chapitre 4

# Estimation des paramètres

### 4.1 Méthode du maximum de vraisemblance

Pour estimer les paramètres du processus, nous utilisons la méthode du maximum de vraisemblance.

#### Log-vraisemblance d'un processus de Hawkes à 1 dimension

$$\begin{aligned} & \ln(L\{t_i\}_{i=1,\dots,n}) \\ &= t_n - \int_0^{t_n} \lambda_0(s) ds - \sum_{i=1}^n \sum_{j=1}^M \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_n - t_i)}) + \sum_{i=1}^n \ln(\lambda(t_i)). \end{aligned}$$

En simplifiant avec nos hypothèses, la log-vraisemblance devient

$$\begin{aligned} & \ln(L\{t_i\}_{i=1,\dots,n}) \\ &= t_n - t_n \lambda_0 - \sum_{i=1}^n \frac{\alpha}{\beta} (1 - e^{-\beta(t_n - t_i)}) + \sum_{i=1}^n \ln \left( \lambda_0 + \sum_{t_k < t_i} \alpha e^{-\beta(t_i - t_k)} \right). \end{aligned}$$

Après avoir implémenté cette fonction de la log-vraisemblance sous R, nous avons utilisé la fonction "ConstrOptim" afin de la maximiser. ConstrOptim est une fonction d'optimisation sous contraintes linéaires, basée sur l'algorithme de Nelder Mead. Notons ici que le fait d'utiliser une fonction d'optimisation et non une formule explicite, induit de l'incertitude au niveau des paramètres.

Pour une raison de complexité (vitesse d'exécution de notre fonction) nous utiliserons pour notre troisième partie (application avec la GTD), la fonction *likelihoodHawkes* du package "hawkes" déjà implémentée sous R, qui renvoie la fonction du log-vraisemblance explicitée ci-dessus (plus rapide que la notre).

## 4.2 Estimation des paramètres d'un processus simulé avec l'algorithme d'Ogata

Une méthode simple pour vérifier que notre estimation est efficace, est la suivante :

- Nous simulons  $n$  fois un processus de Hawkes sur une période  $T$  avec l'algorithme d'Ogata, dont les paramètres sont bien connus (pour ce test nous prenons  $\lambda_0 = 1.2$ ,  $\alpha = 0.6$  et  $\beta = 0.8$ ).
- Nous estimons les trois paramètres pour chaque processus simulé. Puis nous calculons la moyenne de chaque paramètre.
- Ces paramètres estimés sont censés converger vers les paramètres utilisés pour la simulation des processus.

Voici ci-dessous la moyenne des paramètres estimés en fonction de  $n$  et  $T$  (nous n'avons pas calculé la moyenne pour  $T=500$  et  $n=200$  étant donné que le temps de calcul pour  $T=500$  et  $n=100$  était d'environ 4 heures).

	<b>n = 10</b>	<b>n = 100</b>	<b>n = 200</b>	<b>n = 10</b>	<b>n = 100</b>	<b>n = 200</b>
<b>T = 50</b>	1.530999	1.533462	1.526972	0.5227487	0.5473307	0.583421
<b>T = 100</b>	1.410459	1.388181	1.384644	0.5135027	0.6097493	0.5752622
<b>T = 200</b>	1.593578	1.355312	1.330897	0.5869971	0.6091595	0.5986176
<b>T = 500</b>	1.220407	1.2360074		0.5984739	0.5976343	
	<b>Lamba0</b>			<b>Alpha</b>		
	<b>n = 10</b>	<b>n = 100</b>	<b>n = 200</b>			
<b>T = 50</b>	0.7393073	0.8620208	1.023031			
<b>T = 100</b>	0.7254344	0.8804755	0.8343176			
<b>T = 200</b>	0.8668946	0.8617870	0.8363573			
<b>T = 500</b>	0.7919300	0.8086969				
	<b>Beta</b>					

Sans surprise, nous observons que plus nous augmentons le nombre de simulations  $n$ , meilleure est la convergence. Nous observons de plus que la convergence est meilleure lorsque nous augmentons la période  $T$ .

Ces résultats nous montrent que ce n'est pas tant le nombre de simulations qui est important, mais surtout la période sur laquelle se fait l'étude.

## Deuxième partie

# Application sur la Global Terrorism Database : Étude du nombre d'attentats

# Chapitre 1

## Présentation de la base de données

Pour appliquer notre modèle, nous avons utilisé la GTD (Global Terrorism Database [2]). C'est une des bases de données accessible au public la plus complète sur le recensement des attaques terroristes. Elle a été réalisée et est mise à jour chaque année par le centre de recherche START (National Consortium for Study of Terrorism and Responses to Terrorism). Cette base recense 156 772 actes terroristes survenus entre 1970 et 2015 dans le monde, expliqués par 137 variables.

Dans cette partie II du rapport, nous nous intéressons à la modélisation du nombre d'événements. C'est pourquoi nous nous restreignons pour l'instant aux variables suivantes :

- **iyear** : année de l'attaque.
- **imonth** : mois de l'attaque.
- **iday** : jour de l'attaque.
- **country\_txt** : pays où a eu lieu l'attaque.

Précisons que l'année 1993 est absente de la base de données. Cependant, ceci n'est pas gênant pour les calibrages qui sont effectués dans le seul but de voir que le processus s'adapte aux données que nous lui donnons.

En revanche, afin d'être sûrs de ne pas fausser nos prédictions, nous ne considérerons que les attaques survenues après 1994 dans le chapitre 4 de cette deuxième partie.

## Chapitre 2

# Premiers calibrages sur la période globale

Dans un premier temps nous avons décidé de nous concentrer uniquement sur la **France**. Nous avons choisi de commencer avec une échelle de temps **mensuelle**, c'est-à-dire que nous avons exprimé les temps de survenance d'attaques terroristes en mois (avec la convention 1 mois = 31 jours). Ainsi les temps de survenance s'étendaient sur une période  $T$  d'environ 500 unités de temps. Nous avons vu précédemment que pour une période de temps assez grande, une unique estimation ponctuelle suffisait à obtenir des paramètres assez précis. Ce qui explique ce choix d'échelon de temps.

Dans notre base de données, les temps de survenance de chaque attentat sont caractérisés par le jour, le mois et l'année de leur survenance. Il arrive souvent que plusieurs attaques se soient passées le même jour. Sachant que tous les temps de survenance doivent être distincts pour effectuer nos études, nous faisons face à deux choix :

- Soit nous **regroupons** en une unique survenance les attaques survenues le même jour.
- Soit nous **distinguons** chaque attaque en les séparant arbitrairement de manière uniforme dans la journée.

### 2.1 Attaques regroupées

Notre premier calibrage consiste en l'ajustement d'un processus de Hawkes sur les attaques terroristes **en regroupant** celles survenues le même jour. Autrement dit nous avons supprimé les lignes identiques (même année, même mois et même jour de survenance) pour ne garder qu'un exemplaire représentatif du

regroupement d'attaques ; nous ramenant ainsi au nombre de 1283 attaques recensées (en France). Puis nous avons effectué un unique calibrage sur les 44 années disponibles.

- **Estimation des paramètres** (calibrés sur les 44 années accessibles) :

$$\lambda_0 = 0.4705904, \alpha = 0.5931104, \beta = 0.7339127$$

- **Estimations du nombre de survenances de 1972 à 2015** (le nombre *simulé* est obtenu en simulant 10 000 fois un processus de Hawkes avec les paramètres estimés ci-dessus ; le nombre *théorique* est calculé par la formule expliquée dans la première partie de ce rapport) :

Nombre réel	Nombre simulé	Nombre théorique
1283	1281.8	1296.6

Le choix de regrouper les attentats survenus le même jour ou non dépend de ce que nous souhaitons modéliser, ici pour notre étude c'est un choix arbitraire.

Cependant il faut prendre en considération que des attentats survenus le même jour peuvent être un seul même attentat organisé, ainsi le regroupement permet de modéliser le fait qu'un attentat survienne.

En revanche si l'on souhaite modéliser les coûts, il peut être plus intéressant de prendre en compte chaque attentat survenu le même jour, chacun engendrant un coût différent.

## 2.2 Attaques distinguées

Notre second calibrage consiste à ajuster un processus de Hawkes de la même façon que le précédent calibrage, mais cette fois-ci **en distinguant** les attaques lorsqu'elles sont survenues le même jour. Nous nous retrouvons alors cette fois-ci avec un nombre de 2544 attaques recensées.

- **Estimation des paramètres**

$$\lambda_0 = 2.112516, \alpha = 40.939447, \beta = 72.947672$$

- **Estimations du nombre de survenances de 1972 à 2015**

Nombre réel	Nombre simulé	Nombre théorique
2544	2546.1	2545.8

Sur le graphique suivant, chaque palier tracé en bleu correspond au nombre d'attentats réellement survenus en France chaque année (en prenant en compte

chaque attaque). Les paliers tracés en rouge correspondent aux nombres d'attentats estimés chaque année, par une **unique** simulation d'un processus de Hawkes.

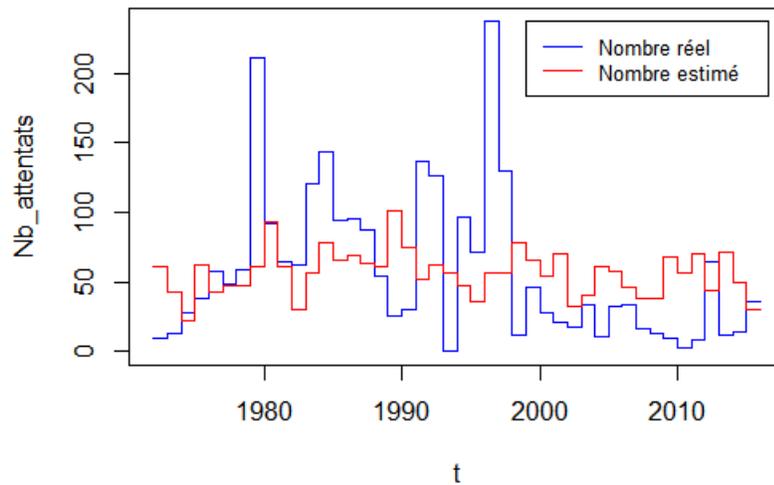


FIGURE 2.1 – Nombre d'attentats en France par an

Remarquons que ce calibrage estime correctement le nombre de survenances sur la totalité de la période, mais qu'il est incapable de décrire plus finement les variations du terrorisme. Nous allons traiter ce problème dans le chapitre suivant.

## Chapitre 3

# Calibrages sur différents découpages de temps

Dans le chapitre précédent nous avons calibré sur l'entièreté de la période étudiée, c'est-à-dire 44 ans (soit 528 mois). Cependant il semble plus réaliste de supposer que la tendance du terrorisme n'est pas la même sur les 44 années observées. Nous avons donc eu l'idée de découper la période globale en un nombre entier  $k \geq 2$  de périodes afin de faire un calibrage différent sur chacune de ces  $k$  périodes.

### 3.1 Méthode de calibration

Nous avons créé une fonction  $Calibrage(temps, p)$  sous R (où "temps" est le vecteur contenant les temps de survénances réelles, et "p" est la taille de la période, c'est-à-dire ici le nombre de mois tel que  $k = \lfloor \frac{max(temps)}{p} \rfloor$ ), nous renvoyant :

- un vecteur  $R$  contenant le nombre d'attentats réels survenus par périodes ;
- un vecteur  $S$  contenant le nombre d'attentats théorique estimé avec les paramètres calibrés sur chaque période (la fonction ajuste un nouveau modèle différent sur chaque période sur les temps de survénances réelles) ;
- un vecteur  $Diff$  contenant le pourcentage d'erreur du nombre d'attentats simulés par rapport au nombre d'attentats réellement survenus sur chaque période.

Pour estimer les paramètres sur chaque période, il faut faire attention à bien calibrer sur les temps de survénances réels d'une période de taille  $p$ , mais ramenés au temps 0. En effet, dans le cas contraire, les paramètres estimés correspondraient en fait à **un processus de Hawkes n'ayant aucune survénance avant la période considérée.**

Nous avons également construit un intervalle de confiance à 95% de niveau de confiance pour l'espérance du nombre d'attentats simulé, sur chaque période. Pour cela nous avons utilisé le théorème central limite :

$$\frac{\sum_{i=1}^n X_i - n * E[X_1]}{\sqrt{n * var(X_1)}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Z \sim \mathcal{N}(0, 1)$$

où

- $(X_i)_{1 \leq i \leq n}$  identiquement distribués et indépendants : chaque  $X_i$  représente le nombre d'attentats simulé de façon indépendante sur une même période ;
- nous effectuons  $n = 1000$  simulations sur une même période ;
- nous prenons la variance empirique pour remplacer  $var(X_1)$ .

Afin de mieux visualiser ce que renvoie cette fonction *calibrage*, voici un exemple : la commande *Calibrage(temps, 176)* calibre un processus de Hawkes sur  $k = \lfloor \frac{528.77}{176} \rfloor = 3$  périodes différentes (176 mois  $\simeq$  14.66 ans).

Les résultats que nous renvoie cette commande sont présentés dans le tableau suivant :

Périodes (en années)	<b>[1972, 1972+14.66[</b>	<b>[1986.66, 2001.32[</b>	<b>[2001.32, 2015.98[</b>
<b>R</b> (nb réel)	1083	1138	321
<b>S</b> (nb simulé)	1085.46	1140.86	321.99
<b>Diff</b> (err relative)	0.002278328	0.002515953	0.003112904
<b>IC</b> (95%)	[1081.12 ; 1090.72]	[1133.90 ; 1143.75]	[318.83 ; 322.78]

Les résultats obtenus sont excellents. Mais comme dans le cas du chapitre précédent (calibrage sur une seule période), même si le nombre total d'attentats simulés sur une période est très proche de la réalité, les survenances des attentats simulés ne sont probablement pas bien réparties au sein de chaque période (voir Figure 2.1).

## 3.2 Détermination du nombre optimal de périodes

L'idée est donc de trouver un nombre de périodes assez grand tel que le pourcentage d'erreur (Diff) reste le plus petit possible sur chaque période.

Pour trouver ce nombre optimal de périodes, il suffit de calculer le pourcentage d'erreur moyen (sur les  $k$  périodes) et de comparer ces pourcentages en faisant varier le nombre  $k$  de périodes.

Nous avons donc fait varier  $p$  de 1 à 44, et nous trouvons une erreur minimale pour  $p = 14$ , c'est-à-dire des périodes de 14 mois, soit un nombre  $k = 37$  périodes.

Nous avons tracé ci-dessous le nombre d'attentats réellement survenus (en bleu) et le nombre d'attentats simulés (en rouge) par périodes.

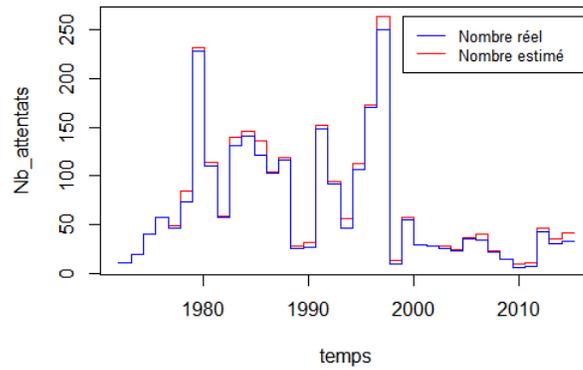


FIGURE 3.1 – Nombre d’attentats en France tous les 14 mois

Erreur absolue moyenne par période
3.926835

### 3.3 Expansion du modèle

Après avoir effectué un calibrage sur environ 2500 attaques (en France), nous avons voulu tester un ajustement sur un plus grand nombre d’attaques. **L’Irak** étant l’un des pays les plus soumis aux attaques terroristes (avec 18750 attentats recensés entre 1975 et 2015), nous avons déterminé de manière similaire au cas de la France le nombre optimal de périodes permettant de minimiser le pourcentage d’erreur.

#### 3.3.1 Échelle mensuelle

En exprimant dans un premier temps les temps de survenances réelles en mois, nous trouvons un nombre optimal de période en découpant l’intervalle global en périodes de 7 mois (soit un nombre de  $\lfloor \frac{493}{7} \rfloor = 70$  périodes).

Erreur absolue moyenne par période
6.151194

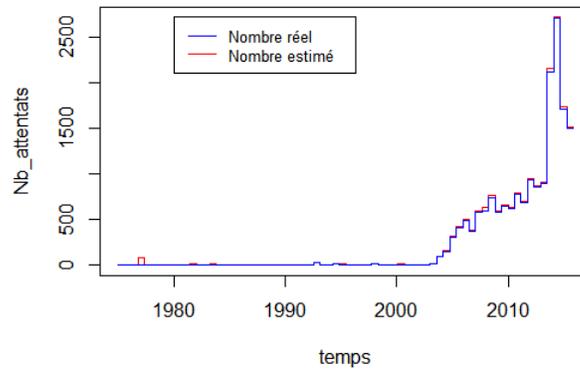


FIGURE 3.2 – Nombre d’attentats en Irak tous les 7 mois

### 3.3.2 Échelle journalière

Nous remarquons sur la Figure 6.2 qu’il y a eu peu d’attentats survenus avant environ l’année 2003, date de la disparition du régime de Saddam Hussein. Depuis cette date, le nombre d’attaques par jour a fortement augmenté (début de la guerre d’Irak). Nous avons donc exprimé les temps de survenances en jours, à partir de l’année 2003.

De nouveau nous avons déterminé la période optimale minimisant l’erreur à l’aide de notre fonction *Calibrage*. Nous trouvons une période de 32 jours.

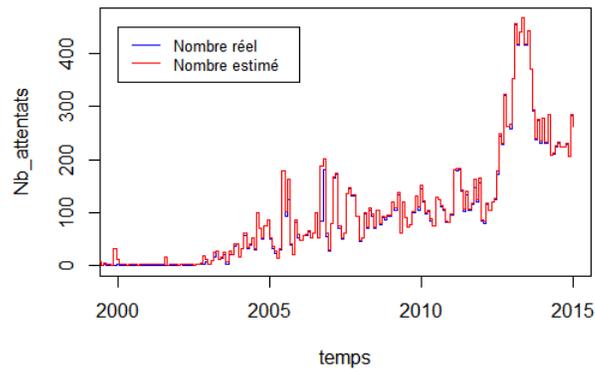


FIGURE 3.3 – Nombre d’attentats en Irak tous les 32 jours

Erreur absolue moyenne par période
3.265863

## Résultats

Que les temps de survenances soient exprimés en jours ou en mois, le modèle de Hawkes se calibre très bien sur les données réelles, dans le sens où il est capable d'expliquer le nombre d'attentats sur une période.

Cela nous conforte dans l'idée que les hypothèses du modèle de Hawkes précédemment évoquées (dépendances des événements, etc...) semblent être adaptées dans le cadre du risque terroriste.

## Chapitre 4

# Tentatives de prédictions avec le modèle de Hawkes

Toujours en se basant sur les données recensées dans la Global Terrorism Database, nous tentons maintenant des prédictions.

Nous avons utilisé une méthode assez "naturelle" qui est de calibrer sur une période choisie, évidemment précédente à celle que nous souhaitons prédire. Puis nous simulons un grand nombre de fois un processus de Hawkes avec les paramètres estimés, sur une période équivalente à celle que l'on souhaite prédire. Finalement nous retenons un nombre moyen d'évènement. Nous considérons également le nombre moyen théorique.

Cependant deux choix s'offrent à nous :

- Le premier est de prédire en considérant que nous démarrons une nouvelle période, c'est à dire que nous comptons le nombre d'attentats survenus entre le temps zero et le temps final. Dans la suite de ce chapitre, nous appellerons ce nombre "**Nb simulé 1**".
- Le second choix est de considérer que la période de prédiction est le prolongement de la période de calibrage. Dans ce cas, pour effectuer une prédiction, nous devons simuler des processus de longueur "période de calibrage"+"période de prédiction" et retenir le nombre d'attentats survenus sur la période de prédiction. Nous appellerons ce nombre "**Nb simulé 2**".

## 4.1 Prédications avec un calibrage sur tout l'historique disponible

Notre première tentative a été de prédire le nombre d'attaques en 2015 avec un calibrage sur tout l'historique précédant 2015, donc de 1994 à 2014 (ici nous considérons un seul attentat par jour). "Nb simulé 1" correspond à la prédiction depuis un temps zero. "Nb simulé 2" correspond à la prédiction dans la continuité de la période de calibrage.

- **Estimation des paramètres** (calibrés sur les temps de 1994 à 2014) :

$$\lambda_0 = 0.4873506, \alpha = 0.8691401, \beta = 1.1411587$$

- **Prédiction** pour l'année 2015 avec les paramètres estimés entre 1994 et 2014 (moyenne sur 10 000 simulations) :

Année prédite	Nb réel	Nb simulé 1	Nb simulé 2	Nb théorique
2015	19	19.03	25.92	24.53

Nous avons également suivi la même méthode cette fois-ci en prenant en compte les attentats survenus le même jour :

- **Estimation des paramètres** (calibrés sur les temps de 1994 à 2014) :

$$\lambda_0 = 1.545339, \alpha = 27.991541, \beta = 48.925394$$

- **Prédiction** pour l'année 2015 avec les paramètres estimés entre 1994 et 2014 (moyenne sur 10 000 simulations) :

Année prédite	Nb réel	Nb simulé 1	Nb simulée 2	Nb théorique
2015	36	43.26	46.09	43.34

Nous avons ensuite tenté quelques autres prédictions (avec la première méthode qui semble être la plus probante) pour cette fois-ci les années : 2014, 2013, 2012 et 2011. Avec un calibrage sur tout l'historique précédant chaque année prédite.

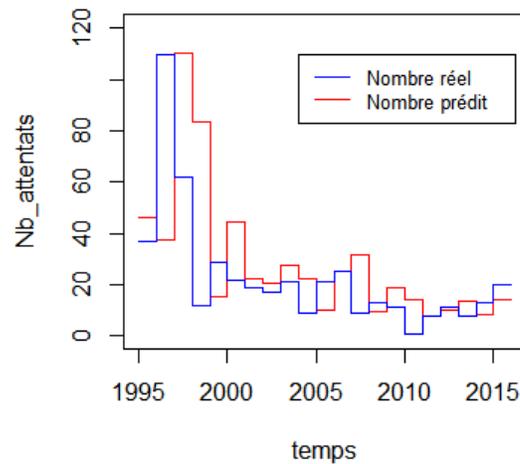
Année prédite	Nb réel	Nb simulé 1	Nb simulé 2	Nb théorique
2014	12	19.41	25.56	25.30
2013	11	19.61	25.62	25.64
2012	11	19.89	26.76	26.45
2011	8	20.69	27.46	27.52

Les nombres prédits sont sensiblement les mêmes à chaque fois, sans grande surprise, car nous calibrons toujours sur le même historique à une année près, les paramètres sont donc très proches ainsi que le nombre d'attaques prédit. Ces résultats suggèrent de calibrer sur des périodes plus proches de la période à prédire, afin de ne prendre en compte que la tendance "récente" du terrorisme, en effet nous pouvons penser que la tendance du terrorisme dans les années 90 n'est pas la même qu'en 2015. Nous recherchons donc une période plus courte, sur laquelle la tendance serait identique, ainsi que les paramètres.

## 4.2 Prédications avec un calibrage sur un historique proche de la prédiction

Nous avons par la suite tenté différents calibrages en faisant varier la période de calibrage ainsi que la période de prédiction. Pour comparer les différentes méthodes nous avons considéré les écarts relatifs absolus ainsi que les écarts absolus à la réalité. Voici quelques résultats (en utilisant le nombre "simulé 2") :

- **Prédiction sur une période de 12 mois, avec un calibrage sur la période de 12 mois qui la précède :**



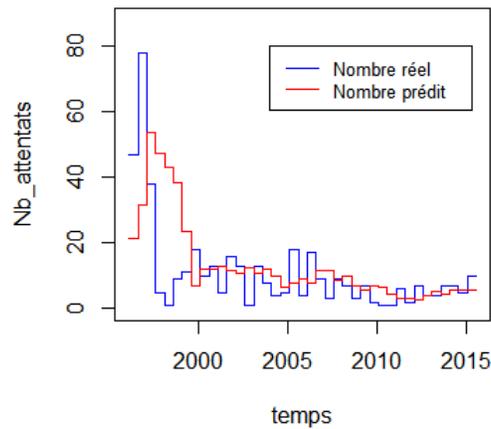
Erreur absolue moyenne par période
16.17

On peut observer que les prédictions correspondent seulement au même schéma que la réalité mais avec une période de retard, en réalité elles ne sont pas très intéressantes car calibrer sur une période de durée équivalente à celle que l'on prédit revient à considérer que sur la prochaine période il y aura autant d'attentats que sur la période en cours.

Nous avons alors recherché une taille de période de calibrage "c" et une taille de période de prédiction "p" de façon à minimiser l'erreur, et telle que  $c > p$ .

Voici un exemple parmi les meilleurs :

- **Prédiction sur une période de 6 mois, avec un calibrage sur la période de 25 mois qui la précède :**



Erreur absolue moyenne par période
2.45

Les résultats sont moins probants que les calibrages. Cependant sur le graphique ci-dessus nous pouvons voir qu'il est possible de prédire la tendance dans le nombre d'attaques au cours du temps. C'est un résultat plutôt intéressant car il vient palier les limites des études réalisées dans le cadre du même bureau d'étude l'an dernier.

En effet les modèles proposés permettaient également de répliquer la tendance dans le nombre d'attaques, mais ils pouvaient prédire un nombre infini d'attentats.

• **Prédiction sur un mois, en calibrant sur différentes durées**

Ci-dessous, la moyenne (sur 10 000 simulations) des écarts entre nombre prédit et nombre réel sur chaque période avec différentes méthodes :

- **Ecart 1** : C'est l'écart absolu moyen en utilisant la première méthode (simulation entre le temps 0 et le temps final). C'est une moyenne sur toutes les périodes simulées.
- **Ecart 2** : C'est l'écart absolu moyen en utilisant la seconde méthode (période de prédiction = prolongement de la période de calibrage). C'est une moyenne sur toutes les périodes simulées.
- **Ecart 3** : C'est l'écart absolu moyen en utilisant le nombre moyen théorique. C'est une moyenne sur toutes les périodes simulées.
- **Nb de périodes** : C'est le nombre  $k$  de prédictions que l'on effectue sur les 20 années disponibles (de 1994 à 2014) ; c'est-à-dire  $k = \lfloor \frac{T}{p} \rfloor$  où  $T$  est la durée de la période globale et  $p$  est la durée de la période de prédiction.

Durée de calibrage	Nb de périodes	Ecart 1	Ecart 2	Ecart 3
4 mois	260	2.10	2.28	3.58
8 mois	256	2.05	1.73	1.81
12 mois	252	2.10	1.72	1.75
16 mois	248	2.11	1.77	1.79

## Résultats

Nous pouvons déduire de ces résultats quelques corrections quant à la façon de prédire.

Tout d'abord, il paraît surprenant que le nombre d'attentats simulé avec la méthode 1 soit aussi éloigné du nombre théorique. Les écarts entre la méthode 1 et la méthode 2 laissent penser qu'il y a une tendance, et donc que le processus n'est pas stationnaire. C'est pour cela que nous avons distingué les méthodes. Cependant, le nombre obtenu avec la méthode 2 est étonnamment toujours proche du nombre théorique. Nous nous sommes rendus compte qu'il n'y a en fait pas de tendance. En effet :

Nous savons que :

$$E[N(t)] = t \cdot \frac{\lambda_0}{1 - \frac{\alpha}{\beta}}, \quad t \geq 0.$$

donc :

$$E[N(t_2) - N(t_1)] = (t_2 - t_1) \cdot \frac{\lambda_0}{1 - \frac{\alpha}{\beta}} = E[N(t_2 - t_1)] = E[N(t_2 + h) - N(t_1 + h)]$$

Pour tous  $t_1 \leq t_2$ ,  $t_1, t_2, h \geq 0$ .

Nous pouvons en déduire qu'il n'y a pas d'intérêt à considérer des périodes décalées dans le temps comme avec les méthodes 1 et 2. Nous en déduisons également que les méthodes 1 et 2 devraient donner un nombre proche du nombre théorique, or cela ne semble fonctionner que pour la méthode 2. Nous en déduisons que le processus n'est pas représentatif de ses paramètres pendant son "démarrage", peut être car l'intensité de base est fixée à une valeur qui n'est pas la valeur moyenne de l'intensité.

Finalement la méthode la plus sensée parmi ces trois méthodes est de considérer le **nombre moyen théorique**.

### 4.3 Prolongement du processus réel

Notre dernier modèle a pour but de simuler un processus de Hawkes qui serait le prolongement du processus réel ; c'est-à-dire forcer le processus à posséder les temps réels de survenances, puis le prolonger.

Notre idée est de modifier l'algorithme d'Ogata vu précédemment, qui simule des temps de survenance d'un processus de Hawkes, en partant d'une intensité  $\lambda_0$  puis en faisant évoluer cette intensité en fonction des temps générés. Nous avons donc modifié la partie "initialisation" de l'algorithme, en prenant comme intensité de base, l'intensité qu'aurait eu le processus s'il avait généré auparavant tous les temps réels, et en considérant un vecteur de temps déjà généré (les temps de survenance réels).

L'algorithme renvoie ensuite le vecteur de temps complet. Il ne reste plus qu'à considérer les temps de survenances au-delà du dernier temps de survenance historique pour voir les temps générés.

Voici quelques résultats à titre de comparaison :

Durée de calibrage	Nb de périodes	Ecart 1	Ecart 2	Ecart 3	<b>Ecart 4</b>
4 mois	260	2.10	2.28	3.58	<b>1.82</b>
8 mois	256	2.05	1.73	1.81	<b>1.73</b>
12 mois	252	2.10	1.72	1.75	<b>1.69</b>
16 mois	248	2.11	1.77	1.79	<b>1.67</b>

Les écarts ("*Ecart 4*") sont légèrement meilleurs.

En Annexe A, nous laissons plus de résultats de différents calibrages/prédictions (on affiche l'erreur absolue moyenne par période). Les colonnes en bleu foncé contiennent les résultats des estimations par la moyenne théorique. Celles en bleu clair contiennent les résultats obtenus par la méthode qui prolonge le processus réel.

Nous pouvons constater que pour une période de prédiction fixée, l'erreur diminue au fur et à mesure que la période de calibrage augmente, mais la tendance

s'inverse/stagne passé une certaine période de calibrage. Nous avons surligné en jaune les erreurs parmi les plus faibles au delà desquelles la tendance semble s'inverser/stagner.

Cela laisse donc supposer qu'il n'y a pas d'intérêt à considérer le plus d'historique possible.

Troisième partie

Etude des coûts

# Chapitre 1

## Détermination de la loi des montants

Nous nous intéressons maintenant aux coûts des actes terroristes.

Nous utilisons toujours les données présentes dans la GTD. Cependant, il y a peu d'informations disponibles sur le coût des attentats : seulement 300 attentats sont accompagnés d'un montant, sur 2400 attentats ayant "réussi" en France. Par manque de temps pour rechercher plus de données sur les montants, nous nous sommes restreints aux deux variables suivantes :

- **propvalue** : coût des dommages matériels (en dollars).
- **success** : 1 si l'attaque a été un succès pour les terroristes, 0 sinon.

### 1.1 Loi discrète

Nous avons commencé par considérer tous les coûts disponibles depuis 1972 en France. Ces derniers sont recensés par des montants variants de 150 à 10 millions de dollars, et ne prennent que quelques valeurs différentes.

Bien que le coût des dommages doit être modélisé à l'aide d'une variable aléatoire continue, les données disponibles dans la GTD étant sous forme de classe, il paraît plus simple dans un premier temps de modéliser cette variable à l'aide d'une variable aléatoire discrète.

Il nous suffit alors de calculer les fréquences d'apparition de chaque coût pour déterminer leur loi.

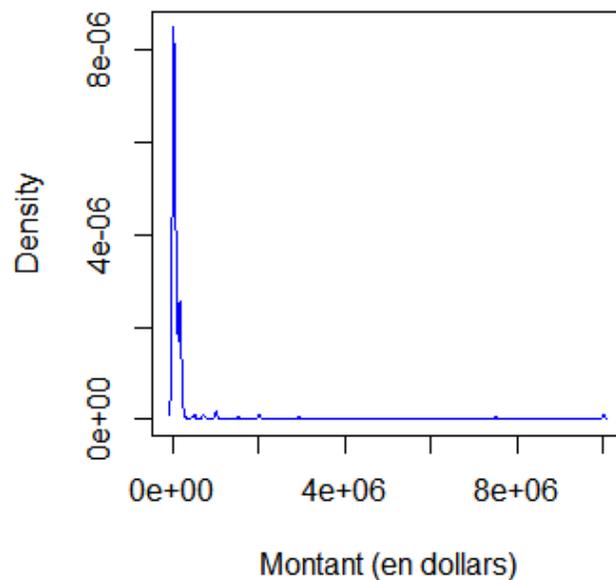
Voici ci-après les valeurs que peut prendre la variable des montants, et le nombre de fois qu'elles apparaissent.

Montants	150	200	250	500	800	1000
Fréquence	3	1	2	3	1	5
Montants	6000	7000	8000	9000	10000	15000
Fréquence	2	1	5	1	20	14
Montants	52000	60000	70000	75000	80000	90000
Fréquence	1	3	2	6	4	1
Montants	250000	300000	440000	500000	700000	750000
Fréquence	1	1	1	2	2	1
Montants	1500	2000	2500	3000	4000	5000
Fréquence	3	19	9	10	1	19
Montants	16000	20000	25000	30000	40000	50000
Fréquence	1	24	6	53	2	16
Montants	100000	120000	150000	160000	172000	200000
Fréquence	18	3	3	50	1	11
Montants	1000000	1500000	2000000	2926563	7500000	10000000
Fréquence	4	1	2	1	1	2

## 1.2 Loi continue

Par soucis de réalisme, nous avons préféré ajuster une loi continue sur les montants disponibles.

La **densité empirique** des montants recensés en France, obtenue à l'aide de la fonction *"density"* sous R, est la suivante :



En considérant la forme de cette densité empirique, nous avons tenté d'ajuster différentes lois continues à valeur dans  $\mathbb{R}^+$ , dont la densité dessine une

"cloche" au niveau des faibles valeurs et diminue fortement vers les grandes valeurs.

Selon ces caractéristiques, nous avons donc essayé d'ajuster les lois suivantes :

- Loi exponentielle
- Loi log-normale
- Loi gamma
- Loi de pareto
- Loi du  $\chi^2$
- Loi GPD

Pour l'ajustement de ces lois, nous avons estimé les paramètres (en calibrant sur tous les montants observables), à l'aide des estimateurs de maximum de vraisemblance et de la fonction *"fitdistr"* sous R du package *MASS*.

Puis nous avons comparé la performance de ces modèles, en nous basant principalement sur deux types comparaisons : celle de la densité et celle des quantiles.

La grande difficulté étant de réussir à modéliser à la fois les faibles montants (plus fréquents) et les montants très élevés (assez présents pour ne pas être négligés), nous avons également essayé d'ajuster une loi hybride, c'est-à-dire l'agrégation de deux modèles : une première loi pour modéliser les montants inférieurs à un certain seuil, une seconde loi pour le reste des montants. Cependant les résultats pour cet ajustement n'ont pas été probants.

Finalement, la loi qui s'ajuste le mieux à notre échantillon est la **loi Log-normale**.

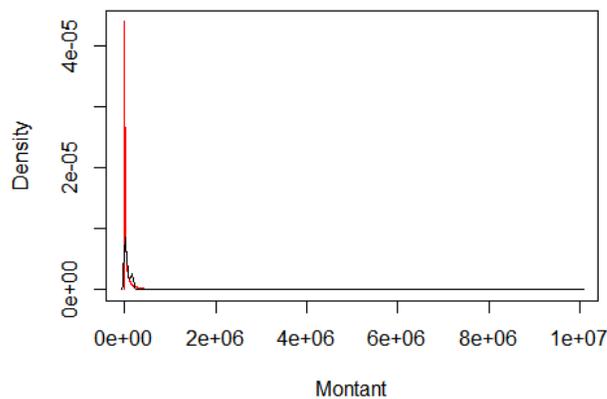


FIGURE 1.1 – Comparaison des densités (empirique en noir, théorique en rouge)

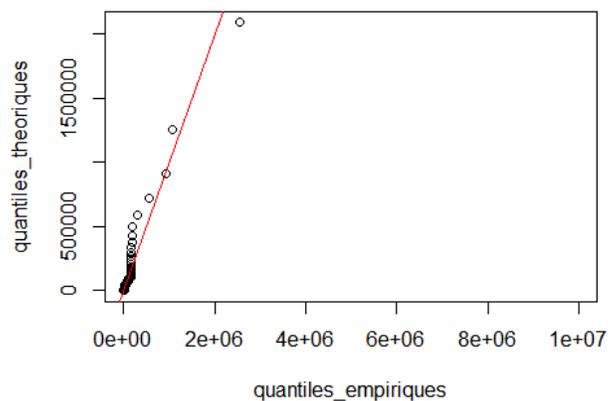


FIGURE 1.2 – Comparaison des quantiles

### 1.3 Intérêts et désavantages des deux modèles retenus

— **Modèle de coûts discret**

L'intérêt de ce modèle est de disposer d'une espérance identique à la moyenne empirique des données disponibles. Il modélise bien les coûts tels qu'ils sont recensés dans la GTD.

En revanche, il ne permet pas de simuler des coûts non présents dans les données.

— **Modèle de coûts continu**

Sous l'hypothèse que dans la réalité les montants ne sont pas aussi standardisés que dans la GTD, et suivraient plutôt une loi continue, nous avons à travers ce modèle tenté de déterminer quelle serait cette loi.

Nous avons calibré cette loi en supposant que les montants suivent cette même loi indépendamment du temps (hypothèse que l'on dispose d'un échantillon de variables aléatoires identiquement distribuées et indépendantes).

Cependant, si l'on considère d'autres périodes (par exemple sur 6 ans), une loi log-normale ne s'ajuste pas systématiquement de façon satisfaisante sur chaque groupe d'années.

## Chapitre 2

# Identité de Wald

Nous nous sommes penchés sur le coût moyen des attentats terroristes sur une période  $[0, t]$ . C'est-à-dire que nous cherchons à déterminer :  $E \left[ \sum_{k=0}^{N_t} X_k \right]$  avec  $N_t$  un processus de Hawkes et  $X_k$  le montant du k-ième attentat.

Sous les hypothèses que

- Les  $(X_k)_{k \in \mathbb{N}}$  sont iid ;
- $N_t$  et les  $(X_k)_{k \in \mathbb{N}}$  sont intégrables ;
- $N_t$  est indépendant des  $(X_k)_{k \in \mathbb{N}}$  ;

nous pouvons appliquer la formule de Wald :

$$E \left[ \sum_{k=0}^{N_t} X_k \right] = E [N_t] * E [X_1]$$

Nous connaissons l'espérance théorique de  $N_t$  et nous savons l'estimer empiriquement avec des simulations. De plus, nous avons précédemment modélisé de deux façons différentes la loi des  $(X_k)_{k \in \mathbb{N}}$ . Nous sommes donc en mesure d'estimer le coût moyen des attentats sur une période donnée.

### Application

Nous avons calibré un processus de Hawkes sur les attentats qui ont abouti en 2015, il nous servira à simuler le processus  $N_t$ . Nous avons par la suite appliqué la formule de Wald afin de déterminer le coût moyen des attentats sur une période de 12 mois, nous obtenons les résultats suivants :

Modèle coût lognormal	Modèle coût discret
5 765 489	6 714 868

## Chapitre 3

# Étude de la ruine

Étant donné que la théorie de la ruine est non triviale dans le cas où il existe une dépendance entre les temps de survie, nous avons effectué des études empiriques pour palier à ce manque de résultats.

Nous avons implémenté une fonction permettant de déterminer la probabilité de se ruiner (capital inférieur à zéro) étant donné un capital initial et un montant de prime fixé. Le modèle pour le capital est le suivant :

$$R_t = u + c * t - \sum_{k=0}^{N_t} X_k$$

Avec

- $R_t$  le capital au temps  $t$  ;
- $N_t$  un processus de Hawkes, calibré sur l'année 2015 ;
- Les  $(X_k)_{k \in \mathbb{N}}$  représentent les coûts, et sont iid ;
- $u$  le capital initial ;
- $c$  le montant de prime mensuelle reçu au cours du temps.

La fonction simule un grand nombre de trajectoires (10 000) de  $R_t$  et détermine la probabilité de tomber en ruine. En minimisant une seconde fonction renvoyant l'écart absolu entre cette première fonction et 0.05 par exemple, nous pouvons déterminer le montant de primes nécessaire pour éviter la ruine dans 95% des cas, pour une durée donnée, et un montant de capital initial donné.

Par exemple, pour un capital initial de  $u = 2000000$ , une durée de 100 ans, les résultats sont les suivants.

Prime (loi log-normale)	Prime (loi discrète)
2 442 462	3 130 820

La prime nécessaire calculée avec la loi discrète est plus élevée. Ceci est justifié par le fait que l'espérance de la loi log-normale est plus faible (155 394.4) que l'espérance de la loi discrète (180 982.5).

Avec ces montants de primes, les probabilités de ruine à horizon 100 ans sont les suivantes :

	Modèle coût log-normal	Modèle coût discret
Proba de ruine	0.037	0.051
Intervalle confiance 0.95	[0.033; 0.041]	[0.037; 0.065]

Ci-dessous nous avons tracé la probabilité de faire faillite en 100 ans en fonction de la prime mensuelle où les montants suivent une loi discrète (prime plus élevée donc moins de prise de risque), et avec un capital initial de 2 000 000 \$ (arbitraire).

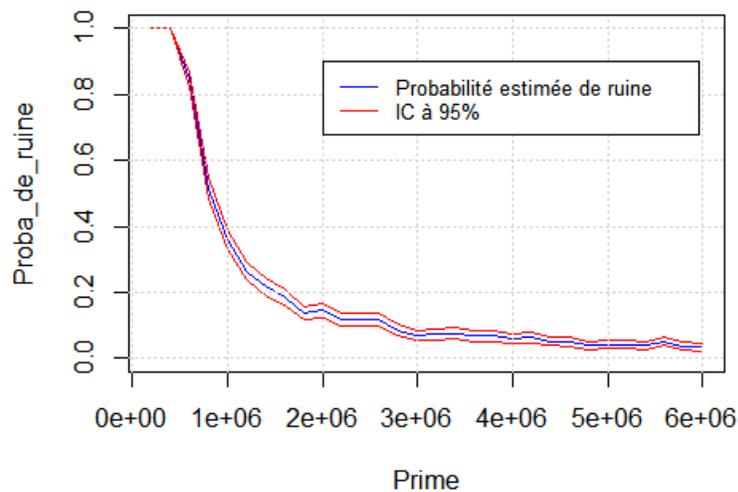


FIGURE 3.1 – Probabilité de faire faillite en 100 ans en fonction de la prime mensuelle

Nous pouvons observer que la probabilité de ruine décroît très rapidement sur la première partie de la courbe, elle décroît ensuite très lentement. Cela signifie que si l'on souhaite diminuer une probabilité de ruine déjà faible, il faut ajouter des montants de plus en plus conséquents.

Nous avons également tracé la probabilité de faire faillite en 100 ans en fonction du capital initial où les montants suivent toujours une loi discrète, et avec une prime de 3 130 820 \$ (prime minimale nécessaire pour éviter la ruine à 95% avec un capital initial de 2 000 000 \$).

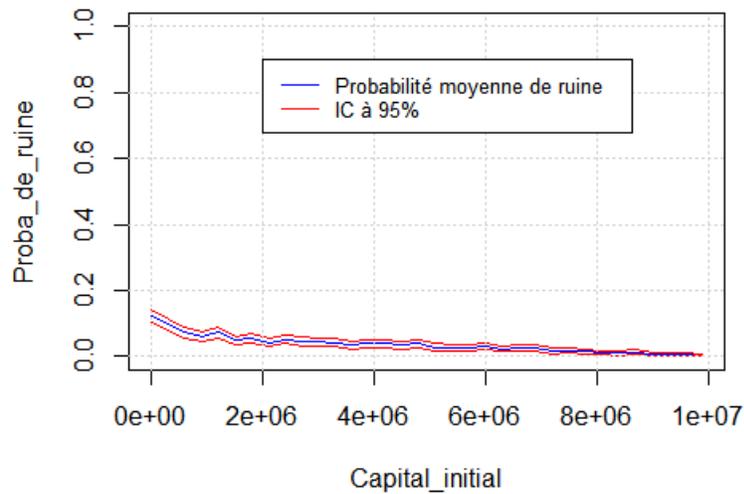


FIGURE 3.2 – Probabilité de faire faillite en 100 ans en fonction du capital initial

De la même façon que dans la seconde partie du graphique précédent, il est nécessaire d'ajouter des montants de plus en plus élevés afin de diminuer faiblement la probabilité de ruine.

# Conclusion

Les processus de Hawkes sont des processus introduits récemment, et la théorie étant complexe, il y a encore peu d'information dans la littérature à propos de leurs applications pratiques (les débuts d'applications au terrorisme sont relativement récents, les applications en finance sont plus présentes dans la littérature). L'objectif de notre bureau d'étude était d'étudier l'un de ses nombreux domaines d'application (le risque terroriste), notamment dans le but de palier aux limites du bureau d'étude de l'année dernière [1] (possibilité de prédire un nombre d'attentats qui parte à l'infini).

Après avoir vérifié que les hypothèses du modèle de Hawkes semblent être adaptées à l'étude du risque terroriste (chapitres sur le calibrage), nous avons étudié la capacité prédictive de ces processus. Nous avons retenu deux méthodes afin de prédire le nombre d'attentats sur une période donnée. La première étant d'estimer les paramètres du processus sur une période suffisamment récente puis de calculer le nombre théorique moyen. La seconde étant d'effectuer le même calibrage et de prolonger le processus réel, puis de calculer un nombre moyen de survenances à l'aide de simulations. Cette seconde méthode semble être plus précise.

Ce bureau d'étude se déroulant dans un cadre assurantiel, nous nous sommes penchés sur l'étude des coûts des dommages, ainsi que celle de la ruine d'une compagnie. Bien entendu, les résultats numériques sont en partie basés sur des hypothèses arbitraires et sont à relativiser. L'objectif était de mettre en place des fonctions relatives à l'assurance (prime, ruine) tout en y intégrant le processus de Hawkes.

Par manque de données et de temps, l'étude des coûts que nous avons effectuée ne concerne que les coûts matériels. Une possible poursuite de ce travail serait d'y incorporer les coûts corporels, en estimant par exemple le nombre de victimes et le coût moyen par victime, et dans l'idéal tous les coûts dédommageables (dommages psychologiques, etc).

Pour conclure, en ce qui concerne les processus de Hawkes nous n'avons étudié qu'une petite partie de ces processus (trois paramètres,  $M=1$ ), tout d'abord par contrainte de temps, mais également par manque de moyens, au niveau algorithmique (l'algorithme d'Ogata ne permet que  $M=1$ ), et au niveau de l'optimisation (plus de paramètres nécessite une optimisation sous contraintes non linéaires sur plusieurs variables). De la même façon, il est possible de calibrer des Processus de Hawkes multidimensionnels mais cela nécessite également des contraintes sur les paramètres.

Une ouverture serait donc d'explorer les possibilités, et l'évolution de la prédiction avec plus de paramètres, et éventuellement d'utiliser des processus de Hawkes multidimensionnels afin de modéliser le terrorisme dans plusieurs pays à la fois, avec de la dépendance entre chaque pays.

Il serait également possible de prendre en compte des tendances d'actualité (et donc pas seulement le passé) en tentant d'ajuster l'intensité en fonction des informations disponibles. Par exemple modifier l'intensité de base  $\lambda_0$  du processus pour prendre en compte l'influence d'une décision politique, d'un événement particulier etc...

# Annexes

## Annexe A

# Comparaison des méthodes de prédiction

période de calibrage en mois	1 mois prédit		2 mois prédits		3 mois prédits		4 mois prédits	
12	1,75	1,69	3,08	3,28	4,41	4,21	5,75	5,66
18	1,76	1,62	3,05	3,11	4,38	3,89	5,45	4,86
24	1,78	1,62	3,19	3,11	4,59	3,84	5,78	5,03
30	1,75	1,57	3,18	3,04	4,64	3,84	5,81	4,69
36	1,64	1,45	2,95	2,74	4,13	3,50	4,93	4,43
42	1,60	1,43	2,78	2,63	3,99	3,35	4,81	3,71
48	1,50	1,43	2,60	2,61	3,74	3,47	4,46	4,35
54	1,44	1,45	2,47	2,65	3,49	3,48	4,02	3,51
60	1,39	1,44	2,36	2,52	3,35	3,45	3,89	4,00
66	1,36	1,44	2,32	2,53	3,26	3,47	3,73	3,67
72	1,33	1,38	2,22	2,40	3,25	3,14	3,84	3,58
78	1,29	1,39	2,15	2,43	3,11	3,17	3,74	3,93
84	1,30	1,35	2,23	2,47	3,19	3,11	3,89	3,78
90	1,28	1,35	2,20	2,44	3,15	3,13	3,79	4,05
96	1,30	1,34	2,27	2,47	3,27	3,18	3,90	3,91
102	1,34	1,34	2,34	2,48	3,39	3,10	3,99	3,82
108	1,30	1,38	2,20	2,50	3,16	3,07	3,69	3,75
114	1,30	1,34	2,17	2,54	3,17	3,14	3,62	4,09
120	1,30	1,34	2,17	2,53	3,21	3,06	3,72	3,65
126	1,28	1,34	2,13	2,56	3,12	3,16	3,53	4,29
132	1,29	1,36	2,15	2,61	3,12	3,30	3,55	3,83
138	1,30	1,28	2,24	2,54	3,16	3,09	3,58	4,32
144	1,28	1,27	2,18	2,39	3,01	2,95	3,68	3,44
150	1,28	1,20	2,32	2,42	3,14	2,84	3,89	4,42
156	1,31	1,19	2,36	2,43	3,24	2,89	4,14	3,39

période de calibrage en mois	5 mois prédits		6 mois prédits		7 mois prédits		8 mois prédits	
12	7,74	7,31	8,09	8,11	8,99	8,27	9,13	8,62
18	6,67	5,87	8,36	7,69	9,29	9,13	9,92	9,04
24	7,32	7,02	8,86	7,44	10,11	9,51	11,17	9,86
30	7,21	5,54	8,66	6,80	8,95	7,43	11,37	8,61
36	5,87	5,10	7,60	6,66	7,90	7,35	8,80	7,99
42	5,82	4,83	7,45	6,14	7,97	6,59	9,44	6,91
48	5,48	4,15	7,07	6,02	7,51	5,33	8,25	7,13
54	5,15	5,22	6,44	6,36	6,97	5,36	7,39	5,90
60	4,77	4,19	5,97	5,67	6,30	5,87	7,00	5,99
66	4,38	4,32	5,82	6,12	5,99	5,96	7,09	6,62
72	4,55	4,51	5,86	5,59	6,59	5,75	7,20	5,76
78	4,69	4,72	5,67	5,60	5,94	5,49	6,79	6,80
84	4,99	5,24	5,89	5,41	6,40	6,11	7,09	5,75
90	4,81	4,82	5,79	5,88	6,13	6,40	7,19	7,23
96	4,27	3,96	5,81	5,99	6,38	6,35	7,00	6,53
102	4,52	4,41	5,92	5,45	6,34	5,56	7,01	6,30
108	4,33	4,18	5,26	5,33	5,49	5,42	6,40	5,78
114	4,60	4,86	5,24	5,08	6,03	5,11	6,89	7,67
120	4,54	4,43	5,22	4,86	5,71	5,22	6,55	5,86
126	4,03	4,66	4,97	5,03	5,75	6,39	6,09	6,08
132	3,95	4,86	4,85	5,44	5,77	6,90	6,60	6,49
138	4,29	3,68	5,46	4,93	5,98	5,55	7,25	8,61
144	4,76	5,46	5,25	5,02	5,92	5,67	6,84	5,93
150	4,96	5,06	5,84	4,81	6,50	6,21	7,59	7,27
156	4,93	4,89	5,97	4,69	6,95	6,18	8,26	6,70

période de calibrage en mois	9 mois prédits		10 mois prédits		11 mois prédits		12 mois prédits	
12	13,96	12,71	14,09	14,21	12,97	12,67	16,17	15,14
18	11,58	10,28	12,61	10,09	15,23	13,35	16,82	14,99
24	12,71	10,83	12,55	9,11	13,76	10,90	14,50	11,39
30	12,21	9,35	13,72	10,97	14,89	10,85	16,17	11,90
36	9,32	8,03	9,48	8,52	11,70	9,32	12,23	11,11
42	10,09	7,39	11,91	8,55	12,65	8,80	14,36	9,23
48	9,44	6,68	9,85	6,57	11,62	9,63	12,71	10,35
54	8,47	7,16	9,22	7,83	10,33	8,11	11,08	8,25
60	8,15	6,16	8,72	6,70	9,78	9,45	11,39	10,53
66	7,80	7,39	8,03	6,89	9,27	8,55	10,04	8,45
72	8,00	6,56	8,75	7,48	9,80	8,63	10,97	8,47
78	7,89	6,75	8,58	6,70	9,83	8,29	10,36	9,02
84	8,06	7,83	8,90	8,29	9,93	8,67	10,92	7,86
90	7,76	6,80	8,99	8,26	9,78	8,47	9,55	9,99
96	7,98	7,02	8,20	7,46	9,47	7,50	10,70	7,21
102	7,49	6,33	8,85	7,08	9,45	7,40	10,25	9,21
108	7,37	6,84	8,04	7,32	8,75	8,31	9,81	7,87
114	7,50	6,13	7,85	6,91	9,30	8,19	9,96	8,29
120	7,21	6,44	8,46	6,50	8,95	7,65	10,16	6,58
126	7,03	5,69	8,18	10,01	8,86	7,97	9,71	10,18
132	7,35	7,12	8,05	7,27	8,49	8,65	9,79	7,27
138	7,86	7,63	9,10	7,98	10,00	9,49	10,97	10,41
144	7,90	7,66	8,11	8,74	9,90	9,93	10,62	8,77
150	9,05	7,45	10,07	7,96	10,94	10,18	12,01	9,95
156	8,63	7,17	10,29	9,07	11,15	7,60	12,22	8,34

# Annexe B

## Codes R

```
library(hawkes)

##### ETUDE DES SURVENANCES D'ATTENTATS #####

#####
### Simulation d'un processus de Hawkes ###
#####

# Fonction d'intensite lambda d'un processus de Hawkes #
intensity = function(lambda0,alpha,betha,t,temps){
  a=temps[temps<t]
  return(lambda0 + sum(alpha*exp(-betha*(t-a))))
}

# Temps de survenance d'un processus de Hawkes generes par l'algorithme d'Ogata #
processusHawkes=function(lambda0,alpha,betha,T){
  temps = NULL #processus ponctuel (temps de survenances)
  n = 1

  #Initialisation

  lambda = lambda0
  s = -(1/lambda)*log(runif(1,0,1))
  while (s>T) {
    s = -(1/lambda)*log(runif(1,0,1))
  }
  temps[n] = s

  #Generale routine
```

```

n=n+1
while(temps[n-1]<T-0.5){

  a=FALSE
  lambdaetoile = intensity(lambda0,alpha,betha,temps[n-1],temps) + alpha
  s0=s
  s = s-(1/lambdaetoile)*log(runif(1,0,1))

  while(s>=T){
    s = s0-(1/lambdaetoile)*log(runif(1,0,1))
  }

  D=runif(1,0,1)
  if (D<=intensity(lambda0,alpha,betha,s,temps)/lambdaetoile){
    temps[n]=s
    n=n+1
    a=TRUE
  }

  while((D>intensity(lambda0,alpha,betha,s,temps)/lambdaetoile)&(a==FALSE)){
    lambdaetoile = intensity(lambda0,alpha,betha,s,temps)
    s0=s
    s = s-(1/lambdaetoile)*log(runif(1,0,1))
    while(s>=T){
      s = s0-(1/lambdaetoile)*log(runif(1,0,1))
    }
    D=runif(1,0,1)
    if (D<=intensity(lambda0,alpha,betha,s,temps)/lambdaetoile){
      temps[n]=s
      n=n+1
      a=TRUE
    }
  }

}
return(temps)
}

# Utilisation du package Hawkes #
processusHawkes2=function(lambda0,alpha,beta,T){
  return(unlist(simulateHawkes(lambda0,alpha,beta,T)))
}

```

```

#####
## Estimation des parametres d'un processus de Hawkes ##

```

```
#####

estimation_parametres=function(temps){
  lnL2=function(x){
    n=length(temps)
    logint=NULL
    lambda0 = x[1]
    alpha = x[2]
    betha = x[3]
    lambda = temps[n]*lambda0 + sum(1-exp(-betha*(temps[n]-temps)))*alpha/betha
    for(i in 1:n){logint[i]=log(intensity(lambda0,alpha,betha,temps[i],temps))}
    res = temps[n] - lambda + sum(logint)
    return(-res)
  }

  #fonction de minimisation par methode de Nelder mead
  c=constrOptim(theta = c(0.1,0.1,0.2),f = lnL2,grad = NULL,
               ui = cbind(c(1,0,0,0),c(0,1,0,-1),c(0,0,1,1)),ci = c(0,0,0,0))

  return(c$par)
}

# Utilisation du package Hawkes #
estimation_parametres2=function(temps){
  lnL=function(x){
    return(likelihoodHawkes(x[1],x[2],x[3],temps))
  }
  c=constrOptim(theta = c(0.1,0.01,30),f = lnL,grad = NULL,
               ui = cbind(c(1,0,0,0),c(0,1,0,-1),c(0,0,1,1)),ci = c(0,0,0,0.1))
  return(c$par)
}

#####
### Importation et transformation des donnees ###
#####

GTD=GTDFrance #contient les annees dans la premiere colonne,
              #les mois dans la deuxieme colonne et les jours dans la troisieme colonne

# Creation d'un vecteur contenant les temps de survenances reels exprimes en mois
temps_de_survenances_mois=function(GTD){
  n=length(GTD$year)
  tps_reel_m=NULL
  tps_reel_m[1]=(GTD$year[1]-min(GTD$year))*12+GTD$imonth[1]+GTD$iday[1]/31
  i=2

```

```

while (i <= n){
a=GTD$iyear[i]-min(GTD$iyear)
if(sum(GTD[i,]==GTD[i-1,])==3){
h=2
while((sum(GTD[i+h-1,]==GTD[i,])==3)&(i+h-1<=n)){
h=h+1 #h est le nombre d'attentat survenu le meme jour
}
kmax=h-1
for(k in 1:kmax){print(i)
tps_reel_m[i]=a*12+GTD$imonth[i]+GTD$iday[i]/31+k/(h*31)
i=i+1
}
}
else{print(i)
tps_reel_m[i]=a*12+GTD$imonth[i]+GTD$iday[i]/31
i=i+1
}
}
return(tps_reel_m)
}

# Creation d'un vecteur contenant les temps de survenances reels exprimes en jours
temps_de_survenances_jours=function(GTD){
n=length(GTD$iyear)
tps_reel_j=NULL
tps_reel_j[1]=(GTD$iyear[1]-min(GTD$iyear))*12*31+(GTD$imonth[1]-1)*31+GTD$iday[1]
i=2
while (i <= n){
a=GTD$iyear[i]-min(GTD$iyear)
m=GTD$imonth[i]-1
if(sum(GTD[i,]==GTD[i-1,])==3){
h=2
while((sum(GTD[i+h-1,]==GTD[i,])==3)&(i+h-1<=n)){
h=h+1 #h est le nombre d'attentat survenu le meme jour
}
kmax=h-1
for(k in 1:kmax){print(i)
tps_reel_j[i]=a*12*31+m*31+GTD$iday[i]+k/h
i=i+1
}
}
else{print(i)
tps_reel_j[i]=a*12*31+m*31+GTD$iday[i]
i=i+1
}
}
}

```

```

    return(tps_reel_j)
}

tps_reel_m=temps_de_survenances_mois(GTD)

#####
##### CALIBRAGE #####
#####

### sur la periode globale ###

param=estimation_parametres2(tps_reel_m)

simu=NULL
for(i in 1:10000){
  simu[i]=length(processusHawkes2(param[1],param[2],param[3],max(tps_reel_m)))
}
nb_simu=mean(simu)
# nombre simule calibre sur les 44 annees disponibles

nb_theo=jumpMean(param[1],param[2],param[3],max(tps_reel_m))
# nombre theorique calibre sur les 44 annees disponibles

### sur plusieurs periodes ###

calibrage=function(t,periode,maxannee=max(t)){
  p=maxannee%/%periode
  nbreel=NULL
  nbestim=NULL
  para=NULL
  temps=NULL
  for (i in 0:(p-1)){
    para=rbind(para,estimation_parametres2(t[(t>=i*periode)&(t<(i+1)*periode)]-i*periode))
    #On enleve le min des ti pour faire une estimation en comment
    nbreel[i+1]=length(t[(t>=i*periode)&(t<(i+1)*periode)])
    temps[i+1]=min(GTD$iyar)+i*periode/12
  }

  for(i in 1:p){nbestim[i]=jumpMean(para[i,1],para[i,2],para[i,3],periode)}
  return(list(reel=nbreel,simu=nbestim,parametre=para,diff=abs(nbreel-nbestim),
             temps=temps))
}
#renvoie nombres d'attentats (simules et reels) et erreur absolue sur chaque periode

graphique=function(tps_reel,periode){
  S=calibrage(tps_reel,periode)$simu

```

```

S=c(S,S[length(S)])
Nb_attentats=calibrage(tps_reel,periode)$reel
Nb_attentats=c(Nb_attentats,Nb_attentats[length(Nb_attentats)])
temps=calibrage(tps_reel,periode)$temps
temps=c(temps,temps[length(temps)]+periode/12)
if(max(S)>max(Nb_attentats)){
  plot(temps,S,'s',col='red',ylab='Nb_attentats')
  lines(temps,Nb_attentats,'s',col='blue')
}
else{
  plot(temps,Nb_attentats,'s',col='blue')
  lines(temps,S,'s',col='red')
}
legend(1980, max(S), legend=c("Nombre reel", "Nombre estime"),
      col=c("blue", "red"),lty=1,cex=0.8)
}
# plot du nb d'attentats simules et ceux reels

### recherche de la periode optimale ###

diffmoy=NULL
for(k in 1:528){diffmoy[k]=mean(calibrage(tps_reel_m,k)$diff)}
#on a la plus petite erreur pour une periode de 14 mois

#####
### PREDICTION ###
#####

# on ne garde que les survenances au dela de 1994
GTD1=GTD[GTD$year>1993,]
tps_reel1=temps_de_survenances_mois(GTD1)

prediction=function(t,c,p){
  pred=NULL
  i=0
  temps=NULL
  while(i*p+c+p<max(t)){
    print(i*p+c+p)
    para=EstimParaHawkes3(t[(t>=i*p)&(t<i*p+c)]-i*p)
    nbreel=length(t[(t>=i*p+c)&(t<(i*p+c+p))])
    nbsimu1=NULL
    nbsimu2=NULL
    if(length(t[(t>=i*p)&(t<i*p+c)])==0){
      nbpred1=0
      nbpred2=0
    }
  }
}

```

```

    nbpredtheo=0
  }
  else{
    for(u in 1:1000){
      print(c("u=",u))
      a=unlist(simulateHawkes(para[1],para[2],para[3],c+p))
      nbsimu2[u]=length(a[a>c]) # methode 2
      a=simulateHawkes(para[1],para[2],para[3],p)
      nbsimu1[u]=length(a) # methode 1
    }

    nbpred1=mean(nbsimu1)
    nbpred2=mean(nbsimu2)
    nbpredtheo=jumpMean(para[1],para[2],para[3],p)}

err1=abs(nbreel-nbpred1)
err2=abs(nbreel-nbpred2)
errtheo=abs(nbreel-nbpredtheo) #/nbreel #on le prend si on veut une erreur absolue
pred=cbind(pred,c(nbreel,nbpred1,nbpred2,nbpredtheo,err1,err2,errtheo))
#pred=cbind(pred, errtheo)
temps[i+1]=1994+(i*p+c)/12
i=i+1
}
temps=c(temps,temps[length(temps)]+p/12)
if(max(pred[3,])>max(pred[1,])){
  Nb_attentats=pred[3,]
  Nb_attentats=c(Nb_attentats,Nb_attentats[length(Nb_attentats)])
  a=pred[1,]
  a=c(a,a[length(a)])
  plot(temps,Nb_attentats,type='s',col='red') #les preditions sont en rouge
  lines(temps,a,'s',col='blue') #les reels sont en bleu
}
else{
  Nb_attentats=pred[1,]
  Nb_attentats=c(Nb_attentats,Nb_attentats[length(Nb_attentats)])
  a=pred[3,]
  a=c(a,a[length(a)])
  plot(temps,Nb_attentats,type='s',col='blue')
  lines(temps,a,'s',col='red')
}
return(#mean(pred))
  list(predictions=pred,moy1=mean(pred[5,]),moy2=mean(pred[6,]),
        moytheo=mean(pred[7,]))
}
# renvoie erreur moyenne theorique, erreur moyenne avec la methode 1
# et erreur moyenne avec la methode 2

```

```

ProcessusHawkesProlonge=function(lambda0,Alpha,Beta,T,temps){
  # T : duree sur laquelle on veut simuler
  # temps : vecteurs contenant les temps survenus

  T=temps[length(temps)]+T
  t=temps
  n=length(t)

  s=t[length(t)]
  #General routine
  n=n+1
  letoile=lambda(lambda0,Alpha,Beta,t,t[n-1])+Alpha
  #new event
  s0=s
  s=s-(1/letoile)*log(runif(1,0,1))
  while(s>=T){
    s=s0-1/letoile*log(runif(1,0,1))}

  D=runif(1,0,1)

  while(D>lambda(lambda0,Alpha,Beta,t,s)/letoile){

    letoile=lambda(lambda0,Alpha,Beta,t,s)
    #new event
    s0=s
    s=s-1/(letoile)*log(runif(1,0,1))
    while(s>=T){
      s=s0-1/(letoile)*log(runif(1,0,1))}

    D=runif(1,0,1)
  }

  while(D<=lambda(lambda0,Alpha,Beta,t,s)/letoile){
    t[n]=s

    n=n+1
    letoile=lambda(lambda0,Alpha,Beta,t,t[n-1])+Alpha
    #new event
    s0=s
    s=s-1/letoile*log(runif(1,0,1))
    while(s>=T){
      s=s0-1/(letoile)*log(runif(1,0,1))}
  }
}

```

```

D=runif(1,0,1)

while(D>lambda(lambda0,Alpha,Beta,t,s)/letoile){

  letoile=lambda(lambda0,Alpha,Beta,t,s)
  #new event
  s0=s
  s=s-1/(letoile)*log(runif(1,0,1))
  while(s>=T){
    s=s0-1/(letoile)*log(runif(1,0,1))}

  D=runif(1,0,1)

}
if(max(t)>=T-0.2){break()}
}

return(t)
}
# Voir algorithme d'0gata

predictionProlonge=function(t,p){
  pred=NULL

  i=1
  A=NULL
  while((i+2)*p<max(t)){
    print(i)
    para=EstimParaHawkes3(t[t<(i+1)*p])
    A=rbind(A,para)
    nbreel=length(t[(t>=(i+1)*p)&(t<(i+2)*p)])
    nbpred=NULL

    for(u in 1:1000){
      a=ProcessusHawkesProlonge(para[1],para[2],para[3],T=p,temps=t[t<(i+1)*p])
      nbpred[u]=length(a[a>(i+1)*p])
    }
    nbpred=mean(nbpred)
    i=i+1
    err=abs(nbreel-nbpred)#/nbreel on le prend si on veut une erreur absolue
    pred=cbind(pred,c(nbreel,nbpred,err))
  }
  return(#mean(pred[3,]))
  list(valeurs=pred,parametres=A)
}

```

```

}

#####
# INTERVALLE DE CONFIANCE #
#####

ICcalibrage=function(t,periode,q=0.95){
  p=max(t)/%periode
  nbestim=NULL; ICinf=NULL; ICsup=NULL
  nbreel=NULL
  for (i in 0:(p-1)){
    para=estimation_parametres2(t[(t>=i*periode)&(t<(i+1)*periode)]-i*periode)
    #On enleve le min des ti pour faire une estimation en comment
    nbreel[i+1]=length(t[(t>=i*periode)&(t<(i+1)*periode)])
    #nbreel contient le nb d'attentats reels sur chaque periode
    nb=NULL
    for(u in 1:1000){nb[u]=length(processusHawkes2(para[1],para[2],para[3],periode))}
    nbestim[i+1]=mean(nb)
    #nbestim contient le nb d'attentats estimes sur chaque periode
    ICinf[i+1]=mean(nb)-qnorm((1+q)/2)*sd(nb)/sqrt(1000)
    ICsup[i+1]=mean(nb)+qnorm((1+q)/2)*sd(nb)/sqrt(1000)
  }
  taux=mean((nbreel<=ICsup)&(nbreel>=ICinf))
  #taux de nbreel qui appartiennent a l'intervalle de confiance
  return(list(nb=rbind(nbreel,nbestim,ICinf,ICsup),succes=taux))
}
#IC pour les calibrages

ICprediction=function(t,c,p,q=0.95){
  nbreel=NULL; nbpredit=NULL; ICinf=NULL; ICsup=NULL
  i=0
  while(i*p+c+p<max(t)){
    print(i*p+c+p)
    para=estimation_parametres2(t[(t>=i*p)&(t<i*p+c)]-i*p)
    nbreel[i+1]=length(t[(t>=i*p+c)&(t<(i*p+c+p))])
    nbsimu=NULL
    for(u in 1:1000){
      pro=processusHawkes2(para[1],para[2],para[3],c+p)
      nbsimu[u]=length(pro[pro>c])
    }
    nbpredit[i+1]=mean(nbsimu)
    ICinf[i+1]=mean(nbsimu)-qnorm((1+q)/2)*sd(nbsimu)/sqrt(1000)
    ICsup[i+1]=mean(nbsimu)+qnorm((1+q)/2)*sd(nbsimu)/sqrt(1000)
  }
}

```

```

    i=i+1
  }
  return(rbind(nbreel,nbpreedit,ICinf,ICsup))
}
#IC pour les predictions

##### ETUDE DES MONTANTS #####

#####
### MODELISATION DU MONTANT DES DOMMAGES ###
##### MATERIELS EN FRANCE #####
#####

library(MASS)
montant=coutsFrance$propvalue
#contient les montants des dommages materiels en France

# Densite empirique des montants
plot(density(montant)$x,density(montant)$y,'l',col='blue',
      ylab='Density',xlab='Montant (en dollars)')

#####
### Loi continue ###
#####

# ajustement de la loi lognormale
fit1=fitdistr(montant,'lognormal')

# comparaison des densites
t=seq(0,1e7,by=1000)
plot(t,dlnorm(t,fit1$estimate[1],fit1$estimate[2]),'l',col='red',
      ylab='Density',xlab='Montant')
lines(density(montant))

# comparaison des quantiles
q=seq(0,1,by=0.01)
plot(quantile(montant,q),qlnorm(q,fit1$estimate[1],fit1$estimate[2]),
      xlab='quantiles_empiriques',ylab='quantiles_theoriques')
abline(0,1,col='red')

#####
# Proba discrete #
#####

```

```

M=as.data.frame(table(montant))
effectif=M$Freq
prob=effectif/sum(effectif)
valeurs=sort(unique(montant))
sample(valeurs,1,replace=TRUE,prob=prob) #simule un montant

#####
##### Simulation des reserves financieres #####
# d'une compagnie assurant les actes terroristes #
#####

GTD=Succes_France_distinc #on calibre uniquement sur les attentats qui ont resussis
tps_reel=temps_de_survenances_mois(GTD)

param=estimation_parametres2(tps_reel[2314:2348]-517) # on calibre sur 2015

#####
### proba de ruine en fonction de la prime ###
#####

u=2000000 # capital initial
Tmax=1200 #100 annees

probaruine=function(c){ #c : coef des primes mensuel

  nb1=NULL;nb2=NULL
  for(k in 1:1000){
    #simulation des temps de survenances
    tps_simu=processusHawkes2(param[1],param[2],param[3],Tmax)
    n=length(tps_simu)

    #simulation des montants des sinistres
    X1=sample(valeurs,n,replace=TRUE,prob=prob) #loi discrete
    S1=cumsum(X1)
    R1=u+c*tps_simu-S1
    nb1[k]=(sum(R1<0)>=1)

    X2=rlnorm(n,fit1$estimate[1],fit1$estimate[2]) #loi continue
    S2=cumsum(X2)
    R2=u+c*tps_simu-S2
    nb2[k]=(sum(R2<0)>=1)
  }
  q=0.95
  ICinf1=mean(nb1)-qnorm((1+q)/2)*sd(nb1)/sqrt(1000)
  ICsup1=mean(nb1)+qnorm((1+q)/2)*sd(nb1)/sqrt(1000)

```

```

ICinf2=mean(nb2)-qnorm((1+q)/2)*sd(nb2)/sqrt(1000)
ICsup2=mean(nb2)+qnorm((1+q)/2)*sd(nb2)/sqrt(1000)
return(list(moyennedisc=mean(nb1),ICdisc=c(ICinf1,ICsup1), # modele discret
           moyennecont=mean(nb2),ICcont=c(ICinf2,ICsup2))) # modele continu
}
#renvoie la probabilite de tomber en ruine en fonction de c des deux modeles
#(discret et continu) ainsi que leur intervalle de confiance 5%.

#####
## proba de ruine en fonction du capital initial ##
#####

c=3130820

probaruine2=function(u){ #u : capital initial

  nb1=NULL;nb2=NULL
  for(k in 1:1000){
    #simulation des temps de survenances
    tps_simu=processusHawkes2(param[1],param[2],param[3],Tmax)
    n=length(tps_simu)

    #simulation des montants des sinistres
    X1=sample(valeurs,size=n,replace=TRUE,prob=prob) #loi discrete
    S1=cumsum(X1)
    R1=u+c*tps_simu-S1
    nb1[k]=(sum(R1<0)>=1)

    X2=rlnorm(n,fit1$estimate[1],fit1$estimate[2]) #loi continue
    R2=u+c*tps_simu-cumsum(X2)
    nb2[k]=(sum(R2<0)>=1)
  }
  q=0.95
  ICinf1=mean(nb1)-qnorm((1+q)/2)*sd(nb1)/sqrt(1000)
  ICsup1=mean(nb1)+qnorm((1+q)/2)*sd(nb1)/sqrt(1000)

  ICinf2=mean(nb2)-qnorm((1+q)/2)*sd(nb2)/sqrt(1000)
  ICsup2=mean(nb2)+qnorm((1+q)/2)*sd(nb2)/sqrt(1000)
  return(list(moyennedisc=mean(nb1),ICdisc=c(ICinf1,ICsup1), # modele discret
           moyennecont=mean(nb2),ICcont=c(ICinf2,ICsup2))) # modele continu
}
#renvoie la probabilite de tomber en ruine en fonction de u des deux modeles

```

*#(discret et continu) ainsi que leur intervalle de confiance 5%.*

# Bibliographie

- [1] Lise Schryve Manon Fievez, Manon Martinez. Etude des risques liés au terrorisme. Technical report, 2016-2017.
- [2] START. Global terrorism database. <https://www.start.umd.edu/gtd/>.
- [3] Ioane Muni Toke. An introduction to Hawkes processes with applications to finance. Technical report, 2011. <http://lamp.ecp.fr/MAS/fiQuant/ioane>.
- [4] Michael D. Porter and Gentry White. Self-exciting hurdle models for terrorist activity. Technical report, 2012. <https://arxiv.org/pdf/1203.3680.pdf>.
- [5] Hussein Khraibani Zaher Khraibani. Self-exciting point process to study the evolution of the attack terrorism. *International Journal of Statistics and Applications*, 2016. <http://article.sapub.org/10.5923.j.statistics.20160606.04.html>.