The Density of The Duration Until The Next Event of A Hawkes Process

[1]

November 11, 2015 Stephen Crowley

Table of contents

1. Hawkes Processes

1.1. The (Standard) Exponential Hawkes Process of Arbitrary Order.

A uni-variate linear self-exciting counting process N_t is one that can be expressed as

$$\begin{aligned}\lambda(t) &= \mu(t)\kappa + \int_{-\infty}^{t} \nu(t-s)\mathrm{d}N_s \\ &= \mu(t)\kappa + \sum_{t_k<t} \nu(t-t_k)\end{aligned} \tag{1}$$

where $\mu(t)$ is some baseline which factors in sources of non-stationarity, see (150). [11][5][10][4][6, 11.3] Here, $\nu: \mathbb{R}_+ \to \mathbb{R}_+$ is a kernel function which expresses the positive influence of past events t_i on the current value of the intensity process, and κ is a scaling factor for the baseline intensity $\mu(t)$. For comparison with the multivariate case see Equation (78). The Hawkes process of order P is a defined by the exponential kernel

$$\nu(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \tag{2}$$

The intensity of the exponential Hawkes process is written as

$$\begin{aligned}\lambda(t) &= \mu(t)\kappa + \int_0^t \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-s)} \mathrm{d}N_s \\ &= \mu(t)\kappa + \sum_{j=1}^{P} \sum_{k=0}^{\check{N}_t} \alpha_j e^{-\beta_j (t-t_k)} \\ &= \mu(t)\kappa + \sum_{j=1}^{P} \alpha_j \sum_{k=0}^{\check{N}_t} e^{-\beta_j (t-t_k)} \\ &= \mu(t)\kappa + \sum_{j=1}^{P} \alpha_j B_j(N_t)\end{aligned} \tag{3}$$

1. stephenc@nym.hush.com

where $B_j(i)$ is given recursively by

$$\begin{aligned}
B_j(i) &= \sum_{k=1}^{i-1} e^{-\beta_j(t_i-t_k)} \\
&= e^{-\beta_j(t_i-t_{i-1})}\sum_{k=1}^{i-1} e^{-\beta_j(t_{i-1}-t_k)} \\
&= e^{-\beta_j(t_i-t_{i-1})}\left(1+\sum_{k=1}^{i-2} e^{-\beta_j(t_{i-1}-t_k)}\right) \\
&= e^{-\beta_j(t_i-t_{i-1})}(1+B_j(i-1))
\end{aligned} \tag{4}$$

since $e^{-\beta_j(t_{i-1}-t_{i-1})} = e^{-\beta_j 0} = e^{-0} = 1$. A uni-variate Hawkes process is stationary if the branching ratio is less than one.

$$\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} < 1 \tag{5}$$

If a Hawkes process is stationary then the unconditional mean is

$$\begin{aligned}
\mu = E[\lambda(t)] &= \frac{E[\mu(t)]}{1-E[\nu(t)]} \\
&= \frac{E[\mu(t)]}{1-\int_0^\infty \sum_{j=1}^{P} \alpha_j e^{-\beta_j t}\mathrm{d}t} \\
&= \frac{E[\mu(t)]}{1-\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}}
\end{aligned} \tag{6}$$

where $E(\cdot)$ is the Lebesgue integral over the positive real axis. For consecutive events, the dual-predictable projection is expressed (10)

$$\begin{aligned}
\Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \lambda(t)\mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \left(\nu(t) + \sum_{j=1}^{P} \alpha_j B_j(N_t)\right)\mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \nu(s)\mathrm{d}s + \int_{t_{i-1}}^{t_i} \sum_{j=1}^{P} \alpha_j \sum_{k=0}^{i-1} e^{-\beta_j(t-t_k)}\mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \nu(s)\mathrm{d}s + \sum_{j=1}^{P} \alpha_j \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} e^{-\beta_j(t-t_k)}\mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \nu(s)\mathrm{d}s + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} \nu(t-t_k)\mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \nu(s)\mathrm{d}s + \sum_{k=0}^{i-1} \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(e^{-\beta_j(t_{i-1}-t_k)} - e^{-\beta_j(t_i-t_k)}) \\
&= \int_{t_{i-1}}^{t_i} \nu(s)\mathrm{d}s + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(1-e^{-\beta_j \Delta t_i})A_j(i-1)
\end{aligned} \tag{7}$$

compared with the multivariate compensator in Equation (86)where there is the recursion

$$\begin{aligned}A_j(i) &= \sum_{t_k \leqslant t_i} e^{-\beta_j(t_i - t_k)} \\ &= \sum_{k=0}^{i-1} e^{-\beta_j(t_i - t_k)} \\ &= 1 + e^{-\beta_j \Delta t_i} A_j(i-1)\end{aligned} \tag{8}$$

with $A_j(0) = 0$ since the integral of the exponential kernel (16) is

$$\begin{aligned}\int_{t_{i-1}}^{t_i} \nu(t)\mathrm{d}t &= \int_{t_{i-1}}^{t_i} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t - t_k)}\mathrm{d}t \\ &= \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} (e^{-\beta_j t_i} - e^{-\beta_j t_{i-1}})\end{aligned} \tag{9}$$

If $\lambda_0(t)$ does not vary with time, that is, $\lambda_0(t) = \lambda_0$ then (20) simplifies to

$$\begin{aligned}\Lambda(t_{i-1}, t_i) &= (t_i - t_{i-1})\lambda_0 + \sum_{k=0}^{i-1} \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} (e^{-\beta_j(t_{i-1} - t_k)} - e^{-\beta_j(t_i - t_k)}) \\ &= (t_i - t_{i-1})\lambda_0 + \sum_{k=0}^{i-1} \int_{t_{i-1} - t_k}^{t_i - t_k} \nu(t)\mathrm{d}t \\ &= (t_i - t_{i-1})\lambda_0 + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_i - t_{i-1})}) A_j(i-1)\end{aligned} \tag{10}$$

Similarly, another parametrization is given by

$$\begin{aligned}\Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \kappa \lambda_0(s)\mathrm{d}s + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_i - t_{i-1})}) A_j(i-1) \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_i - t_{i-1})}) A_j(i-1) \\ &= \kappa \Lambda_0(t_{i-1}, t_i) + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_i - t_{i-1})}) A_j(i-1)\end{aligned} \tag{11}$$

where κ scales the predetermined baseline intensity $\lambda_0(s)$. In this parametrization the intensity is also scaled by κ

$$\lambda(t) = \kappa\, \lambda_0(t) + \sum_{j=1}^{P} \alpha_j\, B_j(N_t) \tag{12}$$

this allows to precompute the deterministic part of the compensator $\Lambda_0(t_{i-1}, t_i) = \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s$.

1.1.1. Maximum Likelihood Estimation.

The log-likelihood of a simple point process is written as

$$\begin{aligned}\ln\mathcal{L}(N(t)_{t \in [0,T]}) &= \int_0^T (1 - \lambda(s))\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s \\ &= T - \int_0^T \lambda(s)\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s\end{aligned} \tag{13}$$

which in the case of the Hawkes model of order P can be explicitly written [8] as

$$\begin{aligned}
\ln\mathcal{L}(\{t_i\}_{i=1\ldots n}) =& T-\Lambda(0,T)+\sum_{i=1}^{n}\ln\lambda(t_i)\\
=& T+\sum_{i=1}^{n}(\ln\lambda(t_i)-\Lambda(t_{i-1},t_i))\\
=& T-\Lambda(0,T)+\sum_{i=1}^{n}\ln\lambda(t_i)\\
=& T-\Lambda(0,T)+\sum_{i=1}^{n}\ln\left(\kappa\lambda_0(t_i)+\sum_{j=1}^{P}\sum_{k=1}^{i-1}\alpha_j e^{-\beta_j(t_i-t_k)}\right)\\
=& T-\Lambda(0,T)+\sum_{i=1}^{n}\ln\left(\kappa\lambda_0(t_i)+\sum_{j=1}^{P}\alpha_j B_j(i)\right)\\
=& T-\int_0^T\kappa\lambda_0(s)\mathrm{d}s-\sum_{i=1}^{n}\sum_{j=1}^{P}\frac{\alpha_j}{\beta_j}(1-e^{-\beta_j(t_n-t_i)})\\
&+\sum_{i=1}^{n}\ln\left(\kappa\lambda_0(t_i)+\sum_{j=1}^{P}\alpha_j B_j(i)\right)
\end{aligned} \tag{14}$$

where $T=t_n$ and $B_j(i)$ [7] is defined by (4) If the baseline intensity is constant $\lambda_0(t)=1$ then the log-likelihood can be written

$$\begin{aligned}
\ln\mathcal{L}(\{t_1,\ldots,t_n\}) =& T-\kappa T-\sum_{i=1}^{n}\sum_{j=1}^{P}\frac{\alpha_j}{\beta_j}(1-e^{-\beta_j(T-t_i)})\\
&+\sum_{i=1}^{n}\ln\left(\kappa+\sum_{j=1}^{P}\alpha_j B_j(i)\right)
\end{aligned} \tag{15}$$

Note that it was necessary to shift each t_i by t_1 so that $t_1=0$ and $T=t_n$. Also note that T is just an additive constant which does not vary with the parameters so for the purposes of estimation can be removed from the equation.

1.1.2. The case when P=1 and the Lambert W Function: Transcendental Recursion. The Hawkes process of order $P=1$ is a defined by the exponential kernel

$$\nu(t)=\alpha e^{-\beta t} \tag{16}$$

The intensity of the exponential Hawkes process is written as

$$\begin{aligned}
\lambda(t) =& \lambda_0(t)\kappa+\int_0^t\alpha e^{-\beta(t-s)}\mathrm{d}N_s\\
=& \lambda_0(t)\kappa+\sum_{k=1}^{\check{N}_t}\alpha e^{-\beta(t-t_k)}\\
=& \lambda_0(t)\kappa+\sum_{k=1}^{\check{N}_t}\alpha e^{-\beta(t-t_k)}\\
=& \lambda_0(t)\kappa+\alpha\sum_{k=1}^{\check{N}_t}e^{-\beta(t-t_k)}\\
=& \lambda_0(t)\kappa+\alpha B(N_t)
\end{aligned} \tag{17}$$

where $B(t)$ is given by equation (4). A uni-variate Hawkes process is stationary if the branching ratio is less than one.

$$\frac{\alpha}{\beta}<1 \tag{18}$$

If a Hawkes process is stationary then the unconditional mean is

$$
\begin{aligned}
\mu = E[\lambda(t)] &= \frac{\lambda_0}{1 - \int_0^\infty \nu(t)\mathrm{d}t} \\
&= \frac{\lambda_0}{1 - \int_0^\infty \alpha\, e^{-\beta\, t}\mathrm{d}t} \\
&= \frac{\lambda_0}{1 - \frac{\alpha}{\beta}}
\end{aligned} \tag{19}
$$

For consecutive events, let $\Delta t_i = t_i - t_{i-1}$, then the compensator, also known as the dual predictable projection, is (10)

$$
\begin{aligned}
\Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \lambda(t)\mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} (\lambda_0(t) + \alpha\, B_j(N_t))\mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \int_{t_{i-1}}^{t_i} \alpha \sum_{k=1}^{i-1} e^{-\beta(t-t_k)}\mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \alpha \sum_{k=1}^{i-1} \int_{t_{i-1}}^{t_i} e^{-\beta(t-t_k)}\mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{k=1}^{i-1} \int_{t_{i-1}}^{t_i} \nu(t-t_k)\mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{k=1}^{i-1} \frac{\alpha}{\beta}(e^{-\beta(t_{i-1}-t_k)} - e^{-\beta(t_i-t_k)}) \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \frac{\alpha}{\beta}(1 - e^{-\beta\Delta t_i})A_j(i-1)
\end{aligned} \tag{20}
$$

compared with the multivariate compensator in Equation (86) where there is the recursion

$$
\begin{aligned}
A_j(i) &= \sum_{t_k \leqslant t_i} e^{-\beta(t_i-t_k)} \\
&= \sum_{k=0}^{i-1} e^{-\beta(t_i-t_k)} \\
&= 1 + e^{-\beta\Delta t_i} A_j(i-1)
\end{aligned} \tag{21}
$$

with $A_j(0) = 0$ since the integral of the exponential kernel (16) is

$$
\begin{aligned}
\int_{t_{i-1}}^{t_i} \nu(t)\mathrm{d}t &= \int_{t_{i-1}}^{t_i} \alpha\, e^{-\beta\,(t-t_k)}\mathrm{d}t \\
&= \frac{\alpha}{\beta}\,(e^{-\beta\, t_i} - e^{-\beta t_{i-1}})
\end{aligned} \tag{22}
$$

If $\lambda_0(t)$ does not vary with time, that is, let $\lambda_0(t) = \lambda_0$ and $\Delta t_i = t_i - t_{i-1}$ then (20) simplifies to

$$
\begin{aligned}
\Lambda(t_{i-1}, t_i) &= \Delta t_i\, \lambda_0 + \sum_{k=1}^{i-1} \frac{\alpha}{\beta}(e^{-\beta(t_{i-1}-t_k)} - e^{-\beta(t_i-t_k)}) \\
&= \Delta t_i \lambda_0 + \sum_{k=1}^{i-1} \int_{t_{i-1}-t_k}^{t_i-t_k} \nu(t)\mathrm{d}t \\
&= \Delta t_i \lambda_0 + \frac{\alpha}{\beta}(1 - e^{-\beta\Delta t_i})A_j(i-1)
\end{aligned} \tag{23}
$$

Similarly, another parametrization is given by

$$\begin{aligned}\Lambda(t_{i-1},t_i) &= \int_{t_{i-1}}^{t_i} \kappa\lambda_0(s)\mathrm{d}s + \frac{\alpha}{\beta}(1-e^{-\beta\Delta t_i})A_j(i-1)\\ &= \kappa\int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \frac{\alpha}{\beta}(1-e^{-\beta\Delta t_i})A_j(i-1)\\ &= \kappa\Lambda_0(t_{i-1},t_i) + \frac{\alpha}{\beta}(1-e^{-\beta\Delta t_i})A_j(i-1)\end{aligned} \tag{24}$$

where κ scales the baseline intensity $\lambda_0(s)$. In this parametrization the intensity is also scaled by κ

$$\lambda(t) = \kappa\lambda_0(t) + \alpha B_j(N_t) \tag{25}$$

this allows the baseline(apriori) part of the compensator be defined by $\Lambda_0(t_{i-1},t_i) = \int_{t_{i-1}}^{t_i}\lambda_0(s)\mathrm{d}s$.

1.1.3. Maximum Likelihood Estimation.

The log-likelihood of a simple point process is written as

$$\begin{aligned}\ln\mathcal{L}(N(t)_{t\in[0,T]}) &= \int_0^T (1-\lambda(s))\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s\\ &= T - \int_0^T \lambda(s)\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s\end{aligned} \tag{26}$$

which in the case of the Hawkes model of order P can be explicitly written [8] as

$$\begin{aligned}\ln\mathcal{L}(\{t_i\}_{i=1\ldots n}) &= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\lambda(t_i)\\ &= T + \sum_{i=1}^{n} (\ln\lambda(t_i) - \Lambda(t_{i-1},t_i))\\ &= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\lambda(t_i)\\ &= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\left(\kappa\lambda_0(t_i) + \sum_{j=1}^{P}\sum_{k=1}^{i-1} \alpha_j e^{-\beta_j(t_i-t_k)}\right)\\ &= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\left(\kappa\lambda_0(t_i) + \sum_{j=1}^{P} \alpha_j R_j(i)\right)\\ &= T - \int_0^T \kappa\lambda_0(s)\mathrm{d}s - \sum_{i=1}^{n}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(1-e^{-\beta_j(t_n-t_i)})\\ &+ \sum_{i=1}^{n} \ln\left(\kappa\lambda_0(t_i) + \sum_{j=1}^{P} \alpha_j R_j(i)\right)\end{aligned} \tag{27}$$

where $T = t_n$ and we have the recursion[7]

$$R_j(i) = \sum_{k=1}^{i-1} e^{-\beta_j(t_i-t_k)} = e^{-\beta_j(t_i-t_{i-1})}(1+R_j(i-1)) \tag{28}$$

If we have constant baseline intensity $\lambda_0(t) = 1$ then the log-likelihood can be written

$$\begin{aligned}\ln\mathcal{L}(\{t_1,\ldots,t_n\}) &= T - \kappa T - \sum_{i=1}^{n=N_t}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(1-e^{-\beta_j(T-t_i)})\\ &+ \sum_{i=1}^{n} \ln\left(\kappa + \sum_{j=1}^{P} \alpha_j R_j(i)\right)\end{aligned} \tag{29}$$

Note that it was necessary to shift each t_i by t_1 so that $t_1 = 0$ and $T = t_n$. Also note that T is just an additive constant which does not vary with the parameters so for the purposes of estimation can be removed from the equation.

1.2. An Expression for the Density of the Duration Until the Next Event When $P = 1$.

The simplest case occurs when the baseline intensity $\lambda_0(t) = 1$ is constant unity (apparently this is the 'shot noise' case) and $P = 1$ where we have

$$\lambda(\{t_i\}) = \kappa + \sum_{t_i < t} \sum_{j=1}^{1} \alpha_j \, e^{-\beta_j (t - t_i)} = \kappa + \sum_{t_i < t} \alpha \, e^{-\beta (t - t_i)} \tag{30}$$

where

$$E[\lambda(t)] = \frac{\kappa}{1 - \frac{\alpha}{\beta}} \tag{31}$$

is the expected value of the unconditional mean intensity.

$$a_n = \sum_{k=0}^{n} e^{\beta t_k} \tag{32}$$

$$b_n = \sum_{k=0}^{n} e^{\beta (t_k - t_n)} \tag{33}$$

$$c_n = \sum_{k=0}^{n} \sum_{l=0}^{n} e^{\beta (t_k + t_l - t_n)} \tag{34}$$

The expected time of the next point can be obtained by integrating out the (unit exponentiallly distributed, the fundamental normal shot noise) ε appearing in the inverse compensator

$$\Lambda^{-1}(\varepsilon; \alpha, \beta) = \begin{array}{l} \dfrac{t \, a_n}{A_1(n)} \cdot e^{-\beta t} + \\ \dfrac{a_n}{\beta \, A_1(n)} \cdot e^{-\beta t_n} W\!\left(\dfrac{\alpha}{\kappa} A_1(n) \cdot e^{\frac{\alpha b_n - \beta \varepsilon}{\kappa}} \right) + \\ \dfrac{e^{-\beta t_n}}{\kappa \, A_1(n)} \left(a_n \, \varepsilon - \dfrac{\alpha}{\beta} \, c_n \right) \end{array} \tag{35}$$

so that

$$\mathrm{E}[\Lambda^{-1}(\varepsilon; \alpha, \beta) | \mathcal{F}_t] = \int_0^{\infty} \Lambda^{-1}(\varepsilon; \alpha, \beta) \mathrm{d}\varepsilon \tag{36}$$

The recursive equations with initial conditions $b_0 = 1$ and $d_0 = e^{\beta t_0}$ are

$$\begin{aligned} a_n &= a_{n-1} e^{-\beta \Delta t_n} + 1 \\ b_n &= b_{n-1} \, e^{-\beta \Delta t_n} + 1 \\ c_n &= c_{n-1} e^{-\beta \Delta t_n} + e^{\beta t_n} + 2 \, a_{n-1} \end{aligned} \tag{37}$$

It would be nice to have expressions like this involving the Lambert W function for $P > 1$ but neither Maple nor Mathematica were able to find any solutions in terms of "known" functions for $P > 1$. It is noted that Equation (42) has the form

$$\int_0^{\infty} (p + q W(r e^{-s x + t}) + u x) e^{-x} \mathrm{d}x \tag{38}$$

which is a function of 6 variables, $\{p, q, r, s, t, u\}$, for which it would be a very nice thing to have a closed form expression, in order to avoid a recourse to numerical or Monte Carlo integration. It seems that such an expression is very likely to exist because if we drop the variable s from Equation (38) we get a closed-form expression of the form

$$\int_0^\infty (p + qW(re^{-x+t}) + ux)e^{-x}\mathrm{d}x = qW(re^t) + \frac{q}{W(re^t)} - q + u + p - \frac{q}{re^t} \tag{39}$$

We can break this problem down into a more manageable one by calculating some more integrals to see if we can find a pattern. Let us begin with the integral

$$\int_0^\infty W(e^{-sx})e^{-x}\mathrm{d}x = W(1) + \left(-\frac{1}{s}\right)^{-\frac{1}{s}} \left(\Gamma\left(\frac{1}{s}\right) - s\Gamma\left(1 + \frac{1}{s}, -\frac{W(1)}{s}\right)\right) \tag{40}$$

whose closed-form expression was found by Vladimir Reshetnikov. [24]

1.2.1. The Case of Any Order $P = n$.

For general values of the order P, the equation whose root is to be sought is given by the expression

$$\varphi_P(x(\varepsilon)) = \begin{array}{l} \left(\prod_{k=1}^{P} \beta_k\right)(\kappa x - (\varepsilon + \kappa T))e^{\sum_{k=1}^{P}\beta_k(x+T)} + \ldots \\ \ldots + \sum_{m=1}^{P} \left(\prod_{k=1}^{P} \left\{\begin{array}{ll} \alpha_k & m = k \\ \beta_k & m \neq k \end{array}\right.\right) \sum_{k=0}^{n} e^{\sum_{j=1}^{P}\beta_j\left(x + \left\{\begin{smallmatrix} T & j \neq m \\ t_k & j = m \end{smallmatrix}\right.\right)} - e^{\sum_{j=1}^{P}\beta_j\left(T + \left\{\begin{smallmatrix} x & j \neq m \\ t_k & j = m \end{smallmatrix}\right.\right)} \end{array} \tag{41}$$

where $T = t_n$ is arrival time of the most recent point and it is noted that the product of piece-wise functions can be written as

$$\begin{aligned} \prod_{k=1}^{P} \left\{\begin{array}{ll} \alpha_k & m = k \\ \beta_k & m \neq k \end{array}\right. &= \alpha_m \left(\prod_{k=1}^{m-1} \beta_k\right)\left(\prod_{k=m+1}^{P} \beta_k\right) \\ &= \alpha_m \prod_{\substack{k=1 \\ k \neq m}}^{P} \beta_k \end{aligned} \tag{42}$$

and the sums likewise

$$\begin{aligned} \sum_{j=1}^{P} \beta_j\left(x + \left\{\begin{array}{ll} T & j \neq m \\ t_k & j = m \end{array}\right.\right) &= \beta_m(x + t_k) + \sum_{j=1}^{m-1} \beta_j(x + T) + \sum_{j=m+1}^{P} \beta_j(x + T) \\ &= \beta_m(x + t_k) + \sum_{\substack{j=1 \\ j \neq m}}^{P} \beta_j(x + T) \\ &= \sigma_{m,k}(x, x) \end{aligned} \tag{43}$$

and

$$\begin{aligned} \sum_{j=1}^{P} \beta_j\left(T + \left\{\begin{array}{ll} x & j \neq m \\ t_k & j = m \end{array}\right.\right) &= \beta_m(T + t_k) + \sum_{j=1}^{m-1} \beta_j(x + T) + \sum_{j=m+1}^{P} \beta_j(x + T) \\ &= \beta_m(T + t_k) + \sum_{\substack{j=1 \\ j \neq m}}^{P} \beta_j(x + T) \\ &= \sigma_{m,k}(x, T) \end{aligned} \tag{44}$$

so that (41) can be rewritten as

$$\varphi_P(x(\varepsilon)) = \tau(x, \varepsilon) + \sum_{j=1}^{P} \phi_j \sum_{k=0}^{N_T} (\sigma_{j,k}(x, x) - \sigma_{j,k}(x, T)) \tag{45}$$

to be compared with the multivariate case in Equation (68), where

$$\sigma_{m,k}(x,a) = e^{(a+t_k)\beta_m + (x+T)\sum_{\substack{j=1\\ j\neq m}}^{P}\beta_j} \tag{46}$$

$$\phi_m = \alpha_m \prod_{\substack{k=1\\ k\neq m}}^{P} \beta_k = \prod_{k=1}^{P} \begin{cases} \alpha_k & k=m \\ \beta_k & k\neq m \end{cases} \tag{47}$$

$$\tau(x,\varepsilon) = ((x-T)\kappa - \varepsilon)\upsilon\eta(x) \tag{48}$$

$$\eta(x) = e^{(x+T)\sum_{k=1}^{P}\beta_k} \tag{49}$$

$$\upsilon = \prod_{k=1}^{P} \beta_k \tag{50}$$

The derivative given by

$$\varphi_P'(x(\varepsilon)) = \upsilon(\kappa\eta(x) + \tau(x,\varepsilon)) + \sum_{m=1}^{P} \phi_m \sum_{k=0}^{n} (\mu\sigma_{m,k}(x) - \mu_m\sigma_{m,k}(T)) \tag{51}$$

where

$$\mu = \sum_{k=1}^{P} \beta_k \tag{52}$$

$$\mu_m = \sum_{\substack{j=1\\ j\neq m}}^{P} \beta_j \tag{53}$$

is needed so that the Newton sequence can be expressed as

$$\begin{aligned} x_{i+1} &= x_i - \frac{\varphi_2(x_i)}{\varphi_2'(x_i)} \\ &= x_i - \frac{\tau(x_i,\varepsilon) + \sum_{m=1}^{P}\phi_m\sum_{k=0}^{n}(\sigma_{m,k}(x_i,x_i) - \sigma_{m,k}(x_i,T))}{\upsilon(\kappa\eta(x_i) + \tau(x_i,\varepsilon)) + \sum_{m=1}^{P}\phi_m\sum_{k=0}^{n}(\mu\sigma_{m,k}(x_i) - \mu_m\sigma_{m,k}(T))} \end{aligned} \tag{54}$$

and simplified a bit(at least notationally) if we let

$$\rho(x,d) = \sum_{m=1}^{P} \phi_m \sum_{k=0}^{n} \left(\sigma_{m,k}(x) \begin{cases} 1 & d=0 \\ \mu & d=1 \end{cases} - \sigma_{m,k}(T) \begin{cases} 1 & d=0 \\ \mu_m & d=1 \end{cases} \right) \tag{55}$$

then

$$\begin{aligned} x_{i+1}(\varepsilon) &= x_i(\varepsilon) - \frac{\varphi_P(x_i(\varepsilon))}{\varphi_P'(x_i(\varepsilon))} \\ &= x_i - \frac{\tau(x_i(\varepsilon),\varepsilon) + \rho(x_i(\varepsilon),0)}{\upsilon(\kappa\eta(x_i(\varepsilon)) + \tau(x_i(\varepsilon),\varepsilon)) + \rho(x_i(\varepsilon),1)} \end{aligned} \tag{56}$$

so that

$$\Lambda_P^{-1}(\varepsilon;t_0\ldots T) = \lim_{m\to\infty} x_m(\varepsilon) \tag{57}$$

which leads to the expression for the expected arrival time of the next point

$$\int_0^\infty \Lambda_P^{-1}(\varepsilon;t_0\ldots T)e^{-\varepsilon}\mathrm{d}\varepsilon = \int_0^\infty \lim_{m\to\infty} x_m(\varepsilon)e^{-\varepsilon}\mathrm{d}\varepsilon \tag{58}$$

Fatou's lemma[9] can probably be invoked so that the order of the limit and the integral in Equation (58) can be exchanged, with perhaps the introduction of another function, which of course would greatly reduce the computational complexity of the equation. The sequence of functions is known as a Newton sequence [2, 3.3p118] There is also the limit

$$\begin{aligned} \lim_{x\to\infty} \frac{\varphi_P(x_i(\varepsilon))}{\varphi_P'(x_i(\varepsilon))} &= \lim_{x\to\infty} \frac{\tau(x_i(\varepsilon),\varepsilon) + \rho(x_i(\varepsilon),0)}{\upsilon(\kappa\eta(x_i(\varepsilon)) + \tau(x_i(\varepsilon),\varepsilon)) + \rho(x_i(\varepsilon),1)} \\ &= \frac{1}{\mu} \end{aligned} \tag{59}$$

There is more to be done here. [3] Actually, the notion of viscosity solutions and energy functional minimization in an infinite dimensional setting can be invoked to prove uniqueness and convergence (58) so that

$$x_m(\varepsilon) \xrightarrow{\Lambda_P^{-1}} x(\varepsilon) \text{ as } m \to \infty \tag{60}$$

which means that $x_m(\varepsilon)$ Λ^{-1}-converges to $x(\varepsilon)$. [6, Ch.3 Def.7 p.43]

1.3. Filtering, Prediction, Estimation, etc.

The next occurrence time of a point process, given the most recent time of occurrence of a point of a process, can be predicted by solving for the unknown time t_{n+1} when $\{t_n\}$ is a sequence of event times. Let

$$\Lambda_{\text{next}}(t_n, \delta) = \{t_{n+1}: \Lambda(t_n, t_{n+1}) = \delta\} \tag{61}$$

where

$$\Lambda(t_n, t_{n+1}) = \int_{t_n}^{t_{n+1}} \lambda(s; \mathfrak{F}_s) \mathrm{d}s \tag{62}$$

and $\mathfrak{F}_s$ is the σ-algebra filtration up to and including time s and the parameters of λ are fixed. The multivariate case is covered in Section (1.4.1). The idea is to integrate over the solution of Equation (61) with all possible values of ε, distributed according to the unit exponential distribution. The reason for the plural form, time(s), rather than the singular form, time, is that Equation (61) actually has a single real solution and N number of complex solutions, where N is the number of points that have occurred in the process up until the time of prediction. This set of complex expected future event arrival times is deemed the *constellation* of the process, which becomes more detailed with the occurance of each event(the increasing σ-algebra filtration). We shall ignore the constellation for now, and single out the sole real valued element as the expected real time until the next event. After all, does it even make sense to say "something will probably happen around $9.8 + i7.2$ seconds from now?" where i is the imaginary unit, $i = \sqrt{-1}$. The recursive equations for the resemble the heta functions of number theory if you one extends from real valued $\beta \in \mathbb{R}$ to a complex $\beta = i$.

1.4. Calculation of the Expected Number of Events Any Given Time From Now.

The expected number of events given any time from now whatsoever can be calculated by integrating out ε since the process which is adapted to the compensator will be closer to being a unit rate Poisson process the closer the parameters are to being correct and the model actually being a good model of the phemenona it is being applied to. Let F_t be all points up until now, let

$$E(t_{n+1}) = \int_0^\infty \Lambda^{-1}(\varepsilon; \alpha, \beta, F_{t_n}) e^{-\varepsilon} \mathrm{d}\varepsilon$$

then iterate the process, by proceeding as if the next point of the process will occur at the predicted time, simply append the expectation to the current state vector, and project the next point, repeating the process as fast ast the computer will go until some sufficient stopping criteria is met. This equation seems very similiar to the infinite horizon discounted regulator of optimal control; see [1, 1.1].

1.4.1. Prediction.

The next event arrival time of the m-th dimension of a multivariate Hawkes process having the usual exponential kernel can be predicted in the same way as the uni-variate process in Section (1.3), by solving for the unknown t_{n+1} in the equation

$$\left\{ t_{n+1}^m: \varepsilon = \Lambda^m(t_n^m, t_{n+1}^m) = \int_{t_n^m}^{t_{n+1}^m} \lambda^m(s; \mathfrak{F}_s) \mathrm{d}s \right\} \tag{63}$$

where $\Lambda^m(t_n^m, t_{n+1}^m)$ is the compensator from Equation (86) and $\mathfrak{F}_s$ is the filtration up to time s and the parameters of λ^m are fixed. As is the case for the uni-variate Hawkes process, the idea is to average over all possible realizations of ε (of which there are an uncountable infinity) weighted according to an exponential unit distribution. Another idea for more accurate prediction is to model the deviation of the generalized residuals from a true exponential distribution and then include the predicted error when calculating this expectation.

Let the most recent arrival time of the pooled and m-th processes respectively be given by

$$T = \max(T_m : m = 1 \ldots M) \tag{64}$$

$$T_m = \max(t_n^m : n = 0 \ldots N^m - 1) = t_{N^m - 1}^m \tag{65}$$

and

$$\breve{N}_{T_m}^n = \sum_{k=0}^{\breve{N}^n} \begin{cases} 1 & t_k^n < T_m \\ 0 & \end{cases} \tag{66}$$

count the number of points occurring in the n-th dimension before the most recent point of the m-th dimension and

$$\breve{N}(t_j^m < t_k^n) \tag{67}$$

then the next arrival time for a given value of the exponential random variable ε of the m-th dimension of a multivariate Hawkes process having the standard exponential kernel is found by solving for the real root of

$$\varphi_m(x(\varepsilon); \mathcal{F}_T) = \tau_m(x, \varepsilon) + \sum_{l=1}^{P} \sum_{i=1}^{M} \phi_{m,i,l} \sum_{k=0}^{\breve{N}_{T_m}^i} (\sigma_{m,i,l,k}(x, x) - \sigma_{m,i,l,k}(x, T_m)) \tag{68}$$

which is similar to the uni-variate case

$$\varphi_P(x(\varepsilon)) = \tau(x, \varepsilon) + \sum_{j=1}^{P} \phi_j \sum_{k=0}^{\breve{N}_T} (\sigma_{j,k}(x, x) - \sigma_{j,k}(x, T)) \tag{69}$$

where

$$\mathcal{F}_T = \{\kappa_{\ldots}, \alpha_{\ldots}, \beta_{\ldots}, t_0^1 \ldots t_{N^1}^1 \leqslant T, \ldots, t_0^m \ldots t_{N^m}^m \leqslant T, \ldots, t_0^M \ldots t_{N^M}^M \leqslant T\} \tag{70}$$

is the filtration up to time T, to be interpreted as the set of available information, here denoting fitted parameters and observed arrival times of all dimensions, and where

$$\tau_m(x, \varepsilon) = ((x - T_m)\kappa_m - \varepsilon)\upsilon_m \eta_m(x) \tag{71}$$

$$\eta_m(x) = e^{(x + T_m)\sum_{j=1}^{P} \sum_{n=1}^{M} \beta_{m,n,j}} \tag{72}$$

can be seen to be similar to the uni-variate equations $\tau(x, \varepsilon) = ((x - T)\kappa - \varepsilon)\upsilon\eta(x)$ and $\eta(x) = e^{(x+T)\sum_{k=1}^{P} \beta_k}$ and

$$\upsilon_m = \prod_{j=1}^{P} \prod_{n=1}^{M} \beta_{m,n,j} \tag{73}$$

$$\phi_{m,p,k} = \prod_{j=1}^{P} \prod_{n=1}^{M} \begin{cases} \alpha_{m,n,j} & n = p \,\text{and}\, j = k \\ \beta_{m,n,j} & n \neq p \,\text{or}\, j \neq k \end{cases} \tag{74}$$

$$\sigma_{m,i,l,k}(x, a) = e^{\sum_{j=1}^{P} \sum_{n=1}^{M} \beta_{m,n,j} \begin{cases} a + t_k^n & n = i \,\text{and}\, j = l \\ x + T_n & n \neq i \,\text{or}\, j \neq l \end{cases}} \tag{75}$$

For comparison, the uni-variate case is Equation (45) where

$$\sigma_{m,k}(x, a) = e^{(a + t_k)\beta_m + (x + T)\sum_{\substack{j=1 \\ j \neq m}}^{P} \beta_j} = e^{\sum_{j=1}^{P} \beta_j \begin{cases} a + t_k & j = m \\ x + T & j \neq m \end{cases}} \tag{76}$$

Bibliography

[1] Martino Bardi and Italo Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations (Systems & Control: Foundations & Applications)*. Birkauser Boston, 1 edition, 1997.

[2] A.T. Bharucha-Reid. *Random Integral Equations*, volume 96 of *Mathematics in Science and Engineering*. Academic Press, 1972.

[3] J. Boyd. *Solving Transcendental Equations*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[4] V. Chavez-Demoulin and JA McGill. High-frequency financial data modeling using hawkes processes. *Journal of Banking & Finance*, 2012.

[5] A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[6] N. Katzourakis. *An Introduction To Viscosity Solutions for Fully Nonlinear PDE with Applications to Calculus of Variations in L^∞*. SpringerBriefs in Mathematics. Springer International Publishing, 2014.

[7] Y. Ogata. On lewis' simulation method for point processes. *Information Theory, IEEE Transactions on*, 27(1):23–31, 1981.

[8] T. Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.

[9] M.M. Rao. *Measure Theory and Integration*, volume 265 of *Pure and Applied Mathematics*. Marcel Dekker, 2nd, Revised and Expanded edition, 2004.

[10] H. Shek. Modeling high frequency market order dynamics using self-excited point process. *Available at SSRN 1668160*, 2010.

[11] Ioane Muni Toke. An introduction to hawkes processes with applications to finance. *???*, `http://fiquant.mas.ecp.fr/ioane_files/HawkesCourseSlides.pdf`, 2012.