Estimation, Simulation, and Prediction of Critical Exponential Sum Self-Exciting Processes

[1]

December 18, 2017 Stephen Crowley

Table of contents

1. Self-Exciting Processes

1.1. The Self-Exciting Critical Exponential Sum Process of Order P.

Let

$$\theta(t) = \begin{cases} 0 & t \leqslant 0 \\ 1 & t > 0 \end{cases} \tag{1}$$

be the Heaviside step function and $\{t_i : t_i < t_{i+1}\} \in \mathbb{R}$ be the time of occurance of the i-th event of a simple point process whose counting function is

$$N_t = \sum_{t_i < t} \theta(t - t_i) \tag{2}$$

and whose conditional intensity (event rate), also known in some contexts as the hazard function, is given by

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \to 0} \frac{P(N_{t+\Delta t} > N_t | \mathcal{F}_n)}{\Delta t} \\ &= \lambda_0(t) + \int_{-\infty}^{t} f(t-s) \mathrm{d}N_s \\ &= \lambda_0(t) + \sum_{T_k < t} f(t - T_k) \end{aligned} \tag{3}$$

where $\mathcal{F}_n = \{t_0, \ldots,_{t_n}\}$ is the filtration which is an increasing sequence of σ-algebras represented

1. stephencrowley214@gmail.com

by the ordered sequence of the unique occurance times of events of the process and $\lambda_0(t)$ is a deterministic function which will be regarded as a constant $\lambda_0(t) = \lambda_0 = E[\lambda_0(t)]$, [10][5][9][4] and $f: \mathbb{R}_+ \to \mathbb{R}_+$ is a kernel impulse response function which expresses the positive influence of past events T_i on the current value of the intensity process. The critical exponential sum self-exciting process of order P is then defined by the kernel function

$$f(t) = \frac{1}{Z} \sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \tag{4}$$

where Z is a normalization factor defined by

$$Z = \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}$$

so that the branching rate is equal to 1

$$\rho = \int_0^\infty f(t) \mathrm{d}t = \frac{\sum_{j=1}^{P} \alpha_j e^{-\beta_j t}}{\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}} = 1 \tag{5}$$

which puts the process in a state of criticality; precisely poised on the boundary between stationarity and non-stationarity. [3] The conditional intensity can then be expressed as

$$\begin{aligned} \lambda(t) &= \lambda_0(t) + \int_0^t \frac{1}{Z} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-s)} \mathrm{d}N_s \\ &= \lambda_0(t) + \sum_{j=1}^{P} \sum_{k=0}^{N_t} \frac{1}{Z} \alpha_j e^{-\beta_j (t-t_k)} \\ &= \lambda_0(t) + \sum_{j=1}^{P} \frac{\alpha_j}{Z} \sum_{k=0}^{N_t} e^{-\beta_j (t-t_k)} \\ &= \lambda_0(t) + \sum_{j=1}^{P} \frac{\alpha_j}{Z} B_j\big(\breve{N}_t\big) \end{aligned} \tag{6}$$

where $B_j(i)$ is given recursively by

$$\begin{aligned} B_j(i) &= \sum_{k=0}^{i-1} e^{-\beta_j (t_i - t_k)} \\ &= e^{-\beta_j (t_i - t_{i-1})} \sum_{k=1}^{i-1} e^{-\beta_j (t_{i-1} - t_k)} \\ &= e^{-\beta_j (t_i - t_{i-1})} \left(1 + \sum_{k=0}^{i-2} e^{-\beta_j (t_{i-1} - t_k)} \right) \\ &= e^{-\beta_j (t_i - t_{i-1})} (1 + B_j(i-1) \\ &= A_j(i) - 1 \end{aligned} \tag{7}$$

since $e^{-\beta_j (t_{i-1} - t_{i-1})} = e^{-\beta_j 0} = e^{-0} = 1$ and $A_j(i)$ is defined in Equation (9) . For consecutive events, the dual-predictable projection, also known as the compensator, is expressed by

$$\begin{aligned}\Lambda_i &= \Lambda(t_{i-1}, t_i)\\ &= \int_{t_{i-1}}^{t_i} \lambda(t)\mathrm{d}t\\ &= \int_{t_{i-1}}^{t_i} \lambda_0(t) + \sum_{j=1}^{P} \frac{\alpha_j}{Z} B_j(N_t)\mathrm{d}t\\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \int_{t_{i-1}}^{t_i} \frac{1}{Z}\sum_{j=1}^{P} \alpha_j \sum_{k=0}^{i-1} e^{-\beta_j(t-t_k)}\mathrm{d}t\\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \frac{1}{Z}\sum_{j=1}^{P} \alpha_j \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} e^{-\beta_j(t-t_k)}\mathrm{d}t\\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} f(t-t_k)\mathrm{d}t\\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{k=0}^{i-1} \frac{1}{Z}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(e^{-\beta_j(t_{i-1}-t_k)} - e^{-\beta_j(t_i-t_k)})\\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \frac{1}{Z}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(1 - e^{-\beta_j(t_i-t_{i-1})})A_j(i)\end{aligned} \tag{8}$$

where

$$\begin{aligned}A_j(i) &= \sum_{t_k \leqslant t_i} e^{-\beta_j(t_i-t_k)}\\ &= 1 + \sum_{t_k < t_i} e^{-\beta_j(t_i-t_k)}\\ &= 1 + \sum_{k=0}^{i-1} e^{-\beta_j(t_i-t_k)}\\ &= 1 + e^{-\beta_j(t_i-t_{i-1})}A_j(i-1)\\ &= 1 + B_j(i)\end{aligned} \tag{9}$$

with $A_j(-1) = 0$ and $A_j(0) = 1$ since the integral of $f(t)$ over an interval spanning consecutive points is

$$\begin{aligned}\int_{t_{i-1}}^{t_i} f(t)\mathrm{d}t &= \int_{t_{i-1}}^{t_i} \frac{1}{Z}\sum_{j=1}^{P} \alpha_j e^{-\beta_j(t-t_k)}\mathrm{d}t\\ &= \frac{1}{Z}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(e^{-\beta_j t_i} - e^{-\beta_j t_{i-1}})\end{aligned} \tag{10}$$

If $\lambda_0(t)$ is a constant function, $\lambda_0(t) = \lambda_0$ then (8) simplifies to

$$\begin{aligned}\Lambda(t_{i-1}, t_i) &= \lambda_0(t_i - t_{i-1}) + \sum_{k=0}^{i-1} \frac{1}{Z}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(e^{-\beta_j(t_{i-1}-t_k)} - e^{-\beta_j(t_i-t_k)})\\ &= \lambda_0(t_i - t_{i-1}) + \sum_{k=0}^{i-1} \frac{1}{Z}\int_{t_{i-1}-t_k}^{t_i-t_k} f(t)\mathrm{d}t\\ &= \lambda_0(t_i - t_{i-1}) + \frac{1}{Z}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(1 - e^{-\beta_j(t_i-t_{i-1})})A_j(i)\end{aligned} \tag{11}$$

1.1.1. Maximum Likelihood Estimation.

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

$$\begin{aligned}\ln\mathcal{L}(N(t)_{t\in[0,T]}) &= \int_0^T (1 - \lambda(s))\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s\\ &= T - \int_0^T \lambda(s)\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s\end{aligned} \tag{12}$$

which in the case of the sums-of-exponentials self-exciting process of order P can be explicitly written [7] as

$$\begin{aligned}
\ln \mathcal{L}(\{t_i\}_{i=1\ldots n}) &= T - \int_0^T \lambda(t)\mathrm{d}t + \sum_{i=1}^{n} \ln\lambda(t_i) \\
&= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\lambda(t_i) \\
&= T + \sum_{i=1}^{n} (\ln\lambda(t_i) - \Lambda_i) \\
&= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\lambda(t_i) \\
&= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\left(\lambda_0(t_i) + \frac{1}{Z}\sum_{j=1}^{P}\sum_{k=0}^{i-1} \alpha_j e^{-\beta_j(t_i - t_k)}\right) \\
&= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\left(\lambda_0(t_i) + \frac{1}{Z}\sum_{j=1}^{P} \alpha_j B_j(i)\right) \\
&= T - \int_0^T \lambda_0(s)\mathrm{d}s - \sum_{i=0}^{n} \frac{1}{Z}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(1 - e^{-\beta_j(t_n - t_i)}) \\
&+ \sum_{i=1}^{n} \ln\left(\lambda_0(t_i) + \frac{1}{Z}\sum_{j=1}^{P} \alpha_j B_j(i)\right)
\end{aligned} \tag{13}$$

where $T = t_n$ and $B_j(i)$ [6] is defined by (7). If the baseline intensity is constant $\lambda_0(t) = \lambda_0$ then the log-likelihood can be written

$$\begin{aligned}
\ln \mathcal{L}(\{t_0, \ldots, t_n\}) &= (1-\lambda_0)t_n + \sum_{i=1}^{n} \ln\left(\lambda_0 + \frac{1}{Z}\sum_{j=1}^{P} \alpha_j B_j(i)\right) - \sum_{i=1}^{n} \frac{1}{Z}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(1 - e^{-\beta_j(t_n - t_i)}) \\
&= (1-\lambda_0)t_n + \sum_{i=1}^{n} \left(\ln\left(\lambda_0 + \frac{1}{Z}\sum_{j=1}^{P} \alpha_j B_j(i)\right) - \frac{1}{Z}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(1 - e^{-\beta_j(t_n - t_i)})\right) \\
&= (1-\lambda_0)t_n + \sum_{i=1}^{n} \left(\ln\left(\lambda_0 + \frac{1}{Z}\sum_{j=1}^{P} \alpha_j B_j(i)\right) - \frac{1}{Z}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}(1 - e^{-\beta_j(t_n - t_i)})\right)
\end{aligned} \tag{14}$$

1.2. The Inverse Compensator.

Define the inverse compensator

$$\begin{aligned}
\Lambda^{-1}_{t_{n+1}}(t_n, y|\mathcal{F}_{T_n}) &= \{t_{n+1} \colon \Lambda(t_n, t_{n+1}|\mathcal{F}_{t_n}) = y\} \\
&= t_n + \{t \colon \varphi(t, y|\mathcal{F}_{t_n}) = 0\}
\end{aligned} \tag{15}$$

where

$$\varphi(t, y) = \sum_{j=1}^{P} (e^{-\beta_j t} - 1)\gamma(j) A_j(i) + \mathrm{yZy}\prod_{j=1}^{P} \beta_j \tag{16}$$

and

$$\gamma(k) = \prod_{j=1}^{P} \begin{cases} \alpha_j & j = k \\ \beta_j & j \neq k \end{cases} \tag{17}$$

with the time-derivative

$$\frac{\partial}{\partial t}\varphi(t, y) = -\sum_{j=1}^{P} \beta_j e^{-\beta_j t}\gamma(j) A_j(i) \tag{18}$$

then the root of $\varphi(t, y|\mathcal{F}_{t_n})$; for a specific value of the exponentially distributed random variable y; is such that

$$t_n + \{t \colon \varphi(t, y) = 0\} \tag{19}$$

is the inverse of $\Lambda(t_n, t_{n+1}|\mathcal{F}_{t_n}) = y$ in t_{n+1} so that

$$\Lambda\big(t_n, \Lambda^{-1}_{t_{n+1}}(t_n, y|\mathcal{F}_{t_n})\big|\mathcal{F}_{t_n}\big) = y \tag{20}$$

Therefore this slightly modified Newton iteration function is an automorphism of the real line expressed by

$$\begin{aligned} N_\varphi(t_{n+1}, y) &= t_n - \frac{\varphi(t_{n+1} - t_n, y)}{\frac{\partial}{\partial t}\varphi(t_{n+1} - t_n, y)} \\ &= t_n - \frac{\sum_{j=1}^P (e^{-\beta_j\,(t_{n+1} - t_n)} - 1)\gamma(j) A_j(i) + Zy\prod_{j=1}^P \beta j}{-\sum_{j=1}^P \beta_j\, e^{-\beta_j\,(t_{n+1} - t_n)}\gamma(j) A_j(i)} \end{aligned} \tag{21}$$

which has a unique real-valued fixed-point at the point $t_{n+1} = N_\varphi(t_{n+1}, y)$ where $\varphi(t_{n+1} - t_n, y) = 0$ which is the exact time of the next point of the process t_{n+1} such that $\Lambda_{n+1} = \Lambda(t_n, t_{n+1}) = y$. The expected duration until the $(n+1)$-th point of the process occurs, conditional upon the points $\{t_n\} = \{t_i : i = 0, 1, 2, ..., n\}$ and parameters $\theta = \{\alpha_{1\ldots P}, \beta_{1\ldots P}\}$ in the filtration

$$\mathcal{F}_n = \{t_n\} \cup \theta \tag{22}$$

is expressed by

$$\begin{aligned} E_y(t_{n+1} - t_n|\mathcal{F}_n) &= E_y\big(\Lambda^{-1}_{t_{n+1}}(t_n, y|\mathcal{F}_n)\big) \\ &= \int_0^\infty y\, e^{-y}\Lambda^{-1}_{t_{n+1}}(t_n, y|\mathcal{F}_n)\mathrm{d}y \end{aligned} \tag{23}$$

where the probability density f_y of y is $f_y = e^{-y}$. The expected inverse value of the marginal distribution of the duration $\Lambda^{-1}_{t_{n+1}}(t_n, y|\mathcal{F}_n)$ until the next point t_{n+1} of the process when y is not known ahead-of-time and is therefore marginalized, or integrated out, can then shown to be equal to

$$\begin{aligned} E_y^{-1}\big(\Lambda^{-1}_{t_{n+1}}(t_n, y|\mathcal{F}_n)\big) &= \int_0^\infty e^{-y}\Lambda^{-1}_{t_{n+1}}(t_n, y|\mathcal{F}_n)\mathrm{d}y \\ &= \varphi(t_{n+1} - t_n, 1) \\ &= \sum_{j=1}^P \gamma(j)(-1 - B_j(n))(1 - e^{-\beta_j(t_{n+1} - t_n)}) + Z\prod_{j=1}^P \beta_j \\ &= \sum_{j=1}^P \gamma(j)(-A_j(n))(1 - e^{-\beta_j(t_{n+1} - t_n)}) + Z\prod_{j=1}^P \beta_j \\ &= \sum_{j=1}^P (e^{-\beta_j\,(t_{n+1} - t_n)} - 1)\gamma(j) A_j(i) + Z\prod_{j=1}^P \beta_j \end{aligned} \tag{24}$$

where the subscript y indicates that the expectation is taken with respect to the marginal distribution of the inverse compensator with respect to the standard/unit exponentially distributed random variable y and $B_j(i) = \sum_{k=0}^{i-1} e^{-\beta_j(t_i - t_k)}$ is defined in Equation (7) so that the expected value when y is not known is given by the root of the inverse expectation functional

$$\begin{aligned} E_y\big(\Lambda^{-1}_{t_{n+1}}(t_n, y|\mathcal{F}_n)\big) &= \big\{t_{n+1} : E_y^{-1}\big(\Lambda^{-1}_{t_{n+1}}(t_n, y|\mathcal{F}_n)\big) = 0\big\} \\ &= \{t_{n+1} : \varphi(t_{n+1} - t_n, 1) = 0\} \end{aligned} \tag{25}$$

1.2.1. The case when P=1.

1.3. An Expression for the Density of the Duration Until the Next Event.

The simplest case occurs when the deterministic intensity $\lambda_0(t) = \lambda_0$ is constant and $P = 1$ where we have

$$\lambda(\{t_i\}) = \lambda_0 + \sum_{t_i < t}\sum_{j=1}^1 \alpha_j\, e^{-\beta_j\,(t - t_i)} = \lambda_0 + \sum_{t_i < t} \alpha\, e^{-\beta\,(t - t_i)} \tag{26}$$

and

$$\lambda = E[\lambda(t)] = \frac{\kappa}{1 - \frac{\alpha}{\beta}} \tag{27}$$

is the expected value of the unconditional mean intensity.

$$a_n = \sum_{k=0}^{n} e^{\beta t_k} \tag{28}$$

$$b_n = \sum_{k=0}^{n} e^{\beta(t_k - t_n)} \tag{29}$$

$$c_n = \sum_{k=0}^{n} \sum_{l=0}^{n} e^{\beta(t_k + t_l - t_n)} \tag{30}$$

The expected time until the next point can be obtained by integrating over the unit exponentially distributed parameter ε appearing in the inverse of the compensator

$$\Lambda^{-1}(\varepsilon, \alpha, \beta) = e^{-\beta T}\left(\frac{T a_n + \frac{a_n}{\beta} W\left(\frac{\alpha}{\lambda_0} A_1(n) \cdot e^{\frac{\alpha b n - \beta \varepsilon}{\lambda_0}}\right) + \frac{e^{-\beta T}}{\lambda_0}\left(a_n \varepsilon - \frac{\alpha}{\beta} c_n\right)}{A_1(n)}\right) \tag{31}$$

where $A_j(i) = \sum_{k=0}^{i-1} e^{-\beta_j(t_i - t_k)}$ is defined recursively in Equation (9) so that

$$\mathrm{E}_{\varepsilon}[\Lambda^{-1} | \mathcal{F}_t, \alpha, \beta] = \int_0^{\infty} e^{-\varepsilon} \Lambda^{-1}(\varepsilon, \alpha, \beta) \mathrm{d}\varepsilon \tag{32}$$

$$\begin{aligned} a_n &= a_{n-1} e^{-\beta \Delta t_n} + 1 \\ b_n &= b_{n-1} e^{-\beta \Delta t_n} + 1 \\ c_n &= c_{n-1} e^{-\beta \Delta t_n} + e^{\beta t_n} + 2 a_{n-1} \end{aligned} \tag{33}$$

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

Let $S = t_n$ where n is the number of points that have occured so far and

$$\begin{aligned} \varphi(P, \varepsilon, x, \alpha, \beta, S) = & \left(\prod_{k=1}^{P} \beta_k\right)(\lambda_0 x - (\varepsilon + \lambda_0 S)) e^{\sum_{k=1}^{P} \beta_k (x + S)} + \\ & \sum_{m=1}^{P} \gamma(m) \sum_{k=0}^{n} e^{\sum_{j=1}^{P} \beta_j \left(x + \left\{\begin{smallmatrix} S & j \neq m \\ t_k & j = m \end{smallmatrix}\right.\right)} - e^{\sum_{j=1}^{P} \beta_j \left(S + \left\{\begin{smallmatrix} x & j \neq m \\ t_k & j = m \end{smallmatrix}\right.\right)} \\ = & \tau(x, \varepsilon) + \sum_{j=1}^{P} \phi_j \sum_{k=0}^{N_T} (\sigma_{j,k}(x, x) - \sigma_{j,k}(x, S)) \end{aligned} \tag{34}$$

then the time of occurance of the next point of the process is equal to the value of x which solves $\varphi(P, \varepsilon, x, \alpha, \beta, S) = 0$ for a given value of ε which is not known ahead-of-time. In simulation, ε is a randomly chosen value drawn from a unit exponential distribution. In operation, ε is the integral of the conditional intensity accumulated over the interval spanning the last point and the arrival of the next point. Let $x(\varepsilon)$ be the implicity defined function

$$x(\varepsilon) = \{x \colon \varphi(P, \varepsilon, x, \alpha, \beta, S) = 0\} \tag{35}$$

The expected value $\bar{x} = E_{\varepsilon}[x(\varepsilon)]$ of the solution over all possible values of ε is equal to

$$\bar{x} = \int_0^{\infty} \varepsilon e^{-\varepsilon} x(\varepsilon) \mathrm{d}\varepsilon \tag{36}$$

where $x(\varepsilon) = x(\varepsilon; P, x, \alpha, \beta, S)$ is shortened notation to indicate that x is a univariate function of ε only and P, x, α, β, S are constants. Note that $\gamma(k)$ can be written as

$$\begin{aligned} \gamma(m) &= \alpha_m \left(\prod_{k=1}^{m-1} \beta_k\right)\left(\prod_{k=m+1}^{P} \beta_k\right) \\ &= \alpha_m \prod_{\substack{k=1 \\ k \neq m}}^{P} \beta_k \end{aligned} \tag{37}$$

where

$$\begin{aligned}\sigma_{m,k}(x,x) &= \sum_{j=1}^{P} \beta_j\left(x + \begin{cases} T & j \neq m \\ t_k & j = m \end{cases} \right) \\ &= \beta_m(x+t_k) + \sum_{j=1}^{m-1} \beta_j(x+T) + \sum_{j=m+1}^{P} \beta_j(x+T) \\ &= \beta_m(x+t_k) + \sum_{\substack{j=1 \\ j\neq m}}^{P} \beta_j(x+T)\end{aligned} \tag{38}$$

$$\begin{aligned}\sigma_{m,k}(x,T) = & \sum_{j=1}^{P} \beta_j\left(T + \begin{cases} x & j \neq m \\ t_k & j = m \end{cases} \right) \\ &= \beta_m(T+t_k) + \sum_{j=1}^{m-1} \beta_j(x+T) + \sum_{j=m+1}^{P} \beta_j(x+T) \\ &= \beta_m(T+t_k) + \sum_{\substack{j=1 \\ j\neq m}}^{P} \beta_j(x+T)\end{aligned} \tag{39}$$

$$\tau(x,\varepsilon) = ((x-T)\kappa - \varepsilon)\upsilon\eta(x) \tag{40}$$

$$\eta(x) = e^{(x+T)\sum_{k=1}^{P}\beta_k} \tag{41}$$

$$\upsilon = \prod_{k=1}^{P} \beta_k \tag{42}$$

$$\bar{\upsilon}_m = \sum_{\substack{k=1 \\ k\neq m}}^{P} \beta_k \tag{43}$$

so that (34) can be rewritten as

$$\varphi_P(x(\varepsilon)) = \tau(x,\varepsilon) + \sum_{j=1}^{P} \phi_j \sum_{k=0}^{N_T} (\sigma_{j,k}(x,x) - \sigma_{j,k}(x,T)) \tag{44}$$

The derivative

$$\varphi_P'(x(\varepsilon)) = \upsilon(\lambda_0\eta(x) + \tau(x,\varepsilon)) + \sum_{m=1}^{P} \phi_m \sum_{k=0}^{n} (\upsilon\sigma_{m,k}(x) - \bar{\upsilon}_m \sigma_{m,k}(T)) \tag{45}$$

is needed so that the Newton sequence can be expressed as

$$\begin{aligned}x_{i+1} &= x_i - \frac{\varphi_P(x_i)}{\varphi_P'(x_i)} \\ &= x_i - \frac{\tau(x_i,\varepsilon) + \sum_{m=1}^{P} \phi_m \sum_{k=0}^{n} (\sigma_{m,k}(x_i,x_i) - \sigma_{m,k}(x_i,T))}{\upsilon(\kappa\eta(x_i) + \tau(x_i,\varepsilon)) + \sum_{m=1}^{P} \phi_m \sum_{k=0}^{n} (\mu\,\sigma_{m,k}(x_i) - \mu_m\,\sigma_{m,k}(T))}\end{aligned} \tag{46}$$

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

$$\rho(x,d) = \sum_{m=1}^{P} \phi_m \sum_{k=0}^{n} \left(\sigma_{m,k}(x) \begin{cases} 1 & d=0 \\ \upsilon & d=1 \end{cases} - \sigma_{m,k}(T) \begin{cases} 1 & d=0 \\ \bar{\upsilon}_m & d=1 \end{cases} \right) \tag{47}$$

then

$$\begin{aligned}x_{i+1}(\varepsilon) &= x_i(\varepsilon) - \frac{\varphi_P(x_i(\varepsilon))}{\varphi_P'(x_i(\varepsilon))} \\ &= x_i - \frac{\tau(x_i(\varepsilon),\varepsilon) + \rho(x_i(\varepsilon),0)}{\upsilon(\kappa\eta(x_i(\varepsilon)) + \tau(x_i(\varepsilon),\varepsilon)) + \rho(x_i(\varepsilon),1)}\end{aligned} \tag{48}$$

so that

$$\Lambda_P^{-1}(\varepsilon; t_0 \ldots T) = \lim_{m\to\infty} x_m(\varepsilon) \tag{49}$$

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

$$\int_0^\infty \Lambda_P^{-1}(\varepsilon; t_0 \ldots T) e^{-\varepsilon} \mathrm{d}\varepsilon = \int_0^\infty \lim_{m\to\infty} x_m(\varepsilon) e^{-\varepsilon} \mathrm{d}\varepsilon \tag{50}$$

Fatou's lemma[8] can probably be invoked so that the order of the limit and the integral in Equation (50) 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

$$\begin{aligned} \lim_{x\to\infty} \frac{\varphi_P(x_i(\varepsilon))}{\varphi_P'(x_i(\varepsilon))} &= \lim_{x\to\infty} \frac{\tau(x_i(\varepsilon),\varepsilon) + \rho(x_i(\varepsilon),0)}{\upsilon(\kappa\eta(x_i(\varepsilon)) + \tau(x_i(\varepsilon),\varepsilon)) + \rho(x_i(\varepsilon),1)} \\ &= \frac{1}{\lambda} \end{aligned} \tag{51}$$

which converges to the inverse of the stationary rate.

1.4. 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

$$\Lambda_{\text{next}}(t_n,\delta) = \{t_{n+1} : \Lambda(t_n,t_{n+1}) = \delta\} \tag{52}$$

where

$$\Lambda(t_n,t_{n+1}) = \int_{t_n}^{t_{n+1}} \lambda(s;\mathfrak{F}_s)\mathrm{d}s \tag{53}$$

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 (1.5.1). The idea is to integrate over the solution of Equation (52) 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 (52) 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 $\beta = i$.

1.5. 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].

1.5.1. 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 (1.4), by solving for the unknown t_{n+1} in the equation

$$\left\{t_{n+1}^m : \varepsilon = \Lambda^m(t_n^m, t_{n+1}^m) = \int_{t_n^m}^{t_{n+1}^m} \lambda^m(s; \mathfrak{F}_s) \mathrm{d}s\right\} \tag{54}$$

where $\Lambda^m(t_n^m, t_{n+1}^m)$ is the compensator from Equation (?) 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 uncountably infinite) weighted by a standard exponential distribution to calculate the expected value of the next point of the process. 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

$$T = \max(T_m : m = 1 \ldots M) \tag{55}$$

$$T_m = \max(t_n^m : n = 0 \ldots N^m - 1) = t_{N^m - 1}^m \tag{56}$$

and

$$\breve{N}_{T_m}^n = \#(t_k^n < T_m) \tag{57}$$

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

$$\varphi_m(x(\varepsilon); \mathcal{F}_T) = \tau_m(x, \varepsilon) + \sum_{l=1}^{P} \sum_{i=1}^{M} \phi_{m,i,l} \sum_{k=0}^{\breve{N}_{T_m}^i} (\sigma_{m,i,l,k}(x, x) - \sigma_{m,i,l,k}(x, T_m)) \tag{58}$$

which is similar to the uni-variate case

$$\varphi_P(x(\varepsilon)) = \tau(x, \varepsilon) + \sum_{j=1}^{P} \phi_j \sum_{k=0}^{\breve{N}_T} (\sigma_{j,k}(x, x) - \sigma_{j,k}(x, T)) \tag{59}$$

where

$$\mathcal{F}_T = \{\kappa_{\ldots}, \alpha_{\ldots}, \beta_{\ldots}, t_0^1 \ldots t_{N^1}^1 \leqslant T, \ldots, t_0^m \ldots t_{N^m}^m \leqslant T, \ldots, t_0^M \ldots t_{N^M}^M \leqslant T\} \tag{60}$$

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

$$\tau_m(x, \varepsilon) = ((x - T_m)\kappa_m - \varepsilon)\upsilon_m \eta_m(x) \tag{61}$$

$$\eta_m(x) = e^{(x + T_m)\sum_{j=1}^{P} \sum_{n=1}^{M} \beta_{m,n,j}} \tag{62}$$

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

$$\upsilon_m = \prod_{j=1}^{P} \prod_{n=1}^{M} \beta_{m,n,j} \tag{63}$$

$$\phi_{m,p,k} = \prod_{j=1}^{P} \prod_{n=1}^{M} \begin{cases} \alpha_{m,n,j} & n = p \text{ and } j = k \\ \beta_{m,n,j} & n \neq p \text{ or } j \neq k \end{cases} \tag{64}$$

$$\sigma_{m,i,l,k}(x, a) = e^{\sum_{j=1}^{P} \sum_{n=1}^{M} \beta_{m,n,j} \begin{cases} a + t_k^n & n = i \text{ and } j = l \\ x + T_n & n \neq i \text{ or } j \neq l \end{cases}} \tag{65}$$

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] Pierre Brémaud and Laurent Massoulié. Hawkes branching point processes without ancestors. *J. Appl. Probab.*, 38(1):122–135, 03 2001.

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

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

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

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

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