

# A Tutorial on Hawkes Processes for Events in Social Media

## 1.1 引入

点过程是描述事件的时间和特性的统计语言。在金融领域，事件可以表示股票市场中的买卖交易，这些交易会影影响未来的价格和交易量；在地球物理领域，事件可以是一次地震，它可能暗示着附近区域在短期内发生另一场地震的可能性；在生态学中，事件数据通常是某个物种在不同地理位置的观测记录；在在线社交媒体分析中，事件可以是用户随时间发生的行为，每个事件可能具有用户影响力、兴趣主题以及社交网络连接性等属性。

## 1.2 Poisson process

### 1.2.1 定义点过程

在非负实数轴上的点过程（其中非负实数轴被用来表示时间）是一个随机过程，其实现由事件发生的时间  $T_1, T_2, \dots$  组成，这些事件沿时间轴分布。

计数过程（counting process） $N_t$  是一个定义在  $t \geq 0$  上的随机函数，其取值为正整数  $1, 2, \dots$ 。 $N_t$  的值表示在时间  $t$  前发生的点过程中的事件数量。因此，它由一系列非负随机变量  $T_i$  唯一确定，并满足  $T_i < T_{i+1}$ （若  $T_i < \infty$ ）。即  $N_t$  计算的是在时间  $t$  之前发生的事件总数，即

$$N_t := \sum_{i \geq 1} \mathbb{1}_{\{t \geq T_i\}} \quad (1)$$

其中  $\mathbb{1}_{\{\cdot\}}$  是指示函数。因此事件时间序列  $\{T_1, T_2, \dots\}$  与相应的计数过程  $N_t$  是点过程的等效表示。

### 1.2.2 Poisson process

最简单的一类点过程是泊松过程。

DEFINITION 1. (Poisson process): Let  $(\tau_i)_{i \geq 1}$  be a sequence of i.i.d. exponential random variables with parameter  $\lambda$  and event times  $T_n = \sum_{i=1}^n \tau_i$ . The process  $(N_t, t \geq 0)$  defined by  $N_t := \sum_{i \geq 1} \mathbb{1}_{\{t \geq T_i\}}$  is called a Poisson process with intensity  $\lambda$ .

$\tau_j$  被称作到达间隔时间（inter-arrival times），即第一个事件发生在时间  $\tau_1$ ，第二个事件与第一个事件间隔  $\tau_2$ 。 $\tau$  的概率密度函数为（ $\lambda > 0$ ）

$$f_\tau(t) = \begin{cases} \lambda \exp(-\lambda t), & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (2)$$

其期望易得

$$\mathbb{E}_\tau[\tau] = \frac{1}{\lambda} \quad (3)$$

直观上事件的到达速率平均为每单位时间  $\lambda$  次，因为时间间隔的期望值为  $\lambda^{-1}$ 。因此我们可以非正式地说泊松过程的事件强度为  $\lambda$ 。一般而言事件强度不必是常数，而可以是时间的函数，记作  $\lambda(t)$ ，这种更一般的情况称为非齐次泊松过程（non-homogeneous Poisson process）。

到达时间或者事件时间可以表示为：

$$T_n = \sum_{j=1}^n \tau_j \quad (4)$$

$$N_t = \begin{cases} 0, & \text{if } 0 \leq t < T_1 \\ 1, & \text{if } T_1 \leq t < T_2 \\ \dots \end{cases} \quad (5)$$

可见  $N_t$  为右连续而左极限。

### 1.2.3 泊松过程的无记忆性

无记忆性（memorylessness）意味着未来的到达间隔时间的分布仅取决于当前时间的相关信息，而不依赖于更早的历史信息。 $F_\tau(t)$  是随机变量  $\tau$  的累积分布函数，其定义为  $F_\tau(t) := \mathbb{P}\{\tau < t\}$ ，有

$$F_\tau(t) := \mathbb{P}(\tau < t) = \int_0^t \lambda \exp(-\lambda x) dx = [-\exp(-\lambda x)]_{x=0}^{x=t} = 1 - \exp(-\lambda t), \quad t \geq 0 \quad (6)$$

因此在  $\tau > t$  观测到事件的概率是

$$\mathbb{P}(\tau > t) = \exp(-\lambda t), \quad t \geq 0 \quad (7)$$

假设  $m$  个单位的时间已经过去，即在  $[0, m]$  时间段内没有事件发生，我们还需要再等  $t$  单位时间的概率是

$$\begin{aligned} \mathbb{P}(\tau > t + m \mid \tau > m) &= \frac{\mathbb{P}(\tau > t + m, \tau > m)}{\mathbb{P}(\tau > m)} \\ &= \frac{\mathbb{P}(\tau > t + m)}{\mathbb{P}(\tau > m)} = \frac{\exp(-\lambda(t + m))}{\exp(-\lambda m)} = \exp(-\lambda t) = \mathbb{P}(\tau > t) \end{aligned} \quad (8)$$

### 1.2.4 非齐次泊松过程（注意这两个公式都是定义而不是推导）

DEFINITION 2. A point process  $\{N_t\}_{t>0}$  can be completely characterized by conditional intensity function, defined as

$$\lambda(t | \mathcal{H}_t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(N_{t+h} - N_t | \mathcal{H}_t)}{h} \quad (9)$$

where  $\mathcal{H}_t$  is the history of the process up to time  $t$ , containing the list of event times  $\{T_1, T_2, \dots, T_{N_t}\}$ .

在后续内容中，我们使用简化记号  $\lambda(t) = \lambda(t | \mathcal{H}_t)$ ，并始终假定在时间  $t$  之前存在隐式历史信息。上述定义提供了强度视角（intensity view）来描述点过程，这与之前基于事件时间或计数过程的视角是等价的。换句话说，事件强度  $\lambda(t)$  确定了事件发生的时间的分布，从而决定了相应的计数过程  $N_t$ 。

严格来说， $\lambda(t)$  与  $N_t$  通过小时间间隔  $h$  内事件发生的概率相关联，其中  $o(h)$  是满足  $\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$  的函数。换言之，在  $h \rightarrow 0$  时，时间区间  $[t, t+h]$  内观察到一个事件的概率为  $\lambda(t)h$ ，而观察到多个事件的概率可以忽略不计。

正式地， $\lambda(t)$  和  $N_t$  之间的关系由以下概率公式给出

$$\mathbb{P}(N_{t+h} = n + m | N_t = n) = \begin{cases} \lambda(t)h + o(h), & \text{if } m = 1 \\ o(h), & \text{if } m > 1 \\ 1 - \lambda(t)h + o(h), & \text{if } m = 0 \end{cases} \quad (10)$$

### 1.3 Hawkes processes

在前面描述的模型中，事件的到达是独立的，要么以恒定速率到达（泊松过程），要么由强度函数控制（非齐次泊松过程）。然而，对于某些场景，已知事件的发生会增加未来某段时间内观察到事件的可能性。例如，在地震模型中，余震的到达会增加未来发生地震的概率；在社交网络中，用户的交互可能会导致更多的交互行为。在本节中，我们引入了一类事件到达速率显式依赖于过去事件的过程——即自激励过程，并进一步详细介绍最著名的自激励过程——霍克斯过程。

#### 1.3.1 Self-exciting processes（自激励过程）

自激励过程是一种点过程，其中事件的到达会导致条件强度函数的增加。一个著名的自激励过程是霍克斯过程（Hawkes process），该过程由 Hawkes（1971）提出，基于一个计数过程，其中强度函数显式地依赖于所有先前发生的事件。霍克斯过程定义如下：

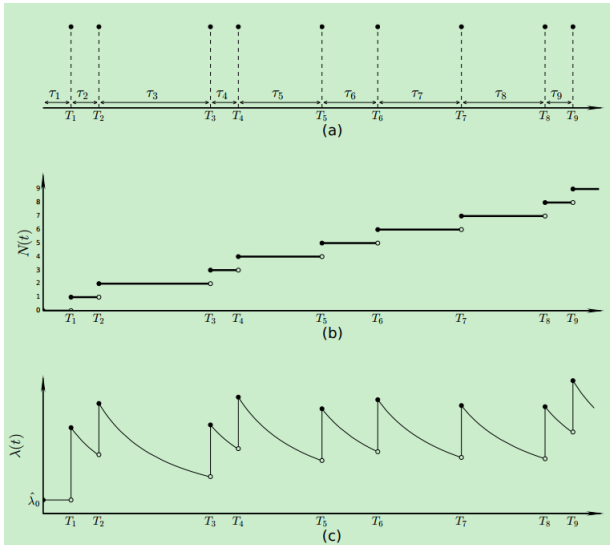
DEFINITION 3. (Hawkes process) Let  $\{N_t\}_{t>0}$  be a counting process with associated history  $\mathcal{H}_t, t > 0$ . The point process is defined by the event intensity function  $\lambda(t)$  with respects Eq 9 (the intensity view of a non-homogeneous Poisson process). The point process is said to be a Hawkes process if the conditional intensity function  $\lambda(t | \mathcal{H}_t)$  takes the form:

$$\lambda(t | \mathcal{H}_t) = \lambda_0(t) + \sum_{i: t > T_i} \phi(t - T_i) \quad (11)$$

其中  $T_i < t$  是在当前时间  $t$  之前发生的所有事件的事件时间，这些事件对时间  $t$  的事件强度做出贡献。 $\lambda_0(t) : \mathbb{R} \rightarrow \mathbb{R}_+$  是一个确定性的基础强度函数， $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$  被称为记忆核（memory kernel）。

#### 1.3.2 强度函数

量  $\lambda_0(t)$  是基础（或背景）强度，描述由外部来源触发的事件的到达。这些事件也称为外生（exogenous）或移民（immigrant）事件，它们的到达独立于过程中的先前事件。Hawkes过程的自激励特性通过公式（10）中的求和项体现，通常核函数  $\phi(\cdot)$  被认为是单调递减的，最近的事件对当前事件强度的影响较大，而距离较远的事件影响较小。下图展示了Hawkes过程的一个实例。显而易见，强度函数的值在事件  $T_i$  发生时突然增大，并且随着时间的推移，给定事件  $T_i$  的影响逐渐衰减。



核函数  $\phi(\cdot)$  并不一定是递减的，然而在本笔记中，我们将讨论限制在递减函数族内，因为我们自然认为事件的影响会随着时间的推移而衰减。一种常见的衰减函数是指数函数（Hawkes, 1971），其形式为

$$\phi(x) = \alpha \exp(-\delta x), \quad \alpha \geq 0, \delta > 0, \alpha < \delta \quad (12)$$

另外一种常用的核函数为幂律核

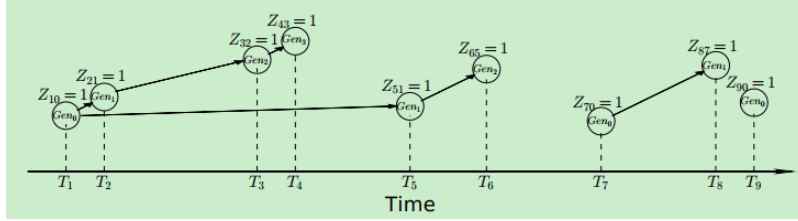
$$\phi(x) = \frac{\alpha}{(x + \delta)^{\eta+1}}, \quad \alpha \geq 0, \delta > 0, \eta > 0, \alpha < \eta\delta^\eta \quad (13)$$

这种核通常在地震学文献 (Ozaki 1979) 和社交媒体文献 (Rizoiu et al. 2017) 中使用。

已经提出了其他自激励点过程，这些过程遵循公式 (10) 给出的经典规范，并扩展了 Hawkes (1971) 提出的初始自激励过程。尽管我们在本笔记中不涉及这些过程，但建议读者参考霍克斯过程的扩展，例如非线性霍克斯过程 (Bremaud and Massoulié 1996, Daley and Vere-Jones 2003)、一般时空自激励点过程 (Ogata 1988, Veen and Schoenberg 2008)、具有指数基础事件强度的过程 (Dassios and Zhao 2011)，或者自抑制 (self-inhibiting) 过程 (Yang et al. 2015)。

### 1.3.3 分支结构

霍克斯过程的另一种等价视角是泊松簇过程解释 (Hawkes 和 Oakes 1974)，它将霍克斯过程中的事件分为两类：移民事件和后代事件。后代事件由过程中的现有（先前）事件触发，而移民事件独立到达，因此没有现有的父事件。后代事件被认为是结构化簇，与每个移民事件相关联，这种结构称为分支结构。在本节的其余部分，我们进一步详细讨论分支结构，并计算两个量：分支因子 (branching factor) ——霍克斯过程中特定事件直接触发的预期事件数，以及估计的后代簇中的事件总数。



我们考虑移民事件遵循具有基础强度  $\lambda_0(t)$  的齐次泊松过程的情况，而后代则通过自激励生成，受公式 (10) 中求和项的控制。下图说明了之前讨论的霍克斯过程的九个事件时刻的分支结构。事件时刻  $T_i$  由圆圈表示，事件之间的“父-子”关系通过箭头显示，我们引入随机变量  $Z_{ij}$ ，其中  $Z_{i0} = 1$  如果事件  $i$  是移民事件，而  $Z_{ij} = 1$  如果事件  $i$  是事件  $j$  的后代。每个圆圈中的文本表示事件所属的代数。例如  $T_3$  和  $T_6$  是移民事件  $T_2$  的直接后代，即数学上表示为  $Z_{32} = 1, Z_{62} = 1, Z_{20} = 1$ 。

分支因子 (或分支比率)  $n^*$  是描述霍克斯过程的一个关键量，定义为 **由单个事件直接生成的预期后代数**。分支因子直观地描述了过程中新出现的事件数量，或者在社交媒体背景下，非正式地裂解为病毒传播性。此外，分支因子还可以指示与移民事件相关的后代簇是否为无限集。当  $n^* < 1$  时，过程处于亚临界状态：任何簇中的事件总数都是有限的。移民事件根据基础强度  $\lambda_0(t)$  到达，但每个移民事件都有一个有限的后代簇，其数量和时间都是有限的。当  $n^* > 1$  时，过程处于所谓的超临界状态，强度函数  $\lambda(t)$  增加，每个簇中的事件总数是无界的。我们通过对  $\phi(t)$  进行积分来计算分支因子：

$$n^* = \int_0^\infty \phi(\tau) d\tau \quad (14)$$

分支因子  $n^*$  表示每个移民事件的后代数量是否有限 ( $n^* < 1$ ) 或无限 ( $n^* > 1$ )。当  $n^* < 1$  时，可以获得每个簇大小的更准确估计。设  $A_i$  为第  $i$  代中事件的预期数量，且  $A_0 = 1$  (因为每个簇只有一个移民事件)。簇中事件的总数  $N_\infty$  被定义为：

$$N_\infty = \sum_{i=0}^\infty A_i \quad (15)$$

为了计算  $A_i, i > 1$ ，我们注意到，上一代中的每个  $A_{i-1}$  事件平均生成  $n^*$  个子事件。这导致了递推关系  $A_i = A_{i-1}n^*$ 。已知  $A_0 = 1$ ，我们可以推导出：

$$A_i = A_{i-1}n^* = A_{i-2}(n^*)^2 = \dots = A_0(n^*)^i = (n^*)^i, \quad i \geq 1 \quad (16)$$

我们得到每个移民簇的大小估计  $N_\infty$ ，他是一个收敛几何级数的和 (假设  $n^* < 1$ )：

$$N_\infty = \sum_{i=0}^\infty A_i = \frac{1}{1 - n^*}, \quad n^* < 1 \quad (17)$$

## 1.4 模拟 Hawkes processes 中的事件

在本节中，我们关注如何根据给定霍克斯过程的设置模拟一系列随机事件。我们介绍两种霍克斯过程的模拟技术。第一种技术是截取 (thinning) 算法 (Ogata 1981)，适用于所有非齐次泊松过程，并可用于任意核函数  $\phi(\cdot)$  的霍克斯过程。第二种技术由 Dassios 和 Zhao (2013) 提出，计算效率更高，因为它针对指数衰减核的霍克斯过程设计了一种变量分解技术。

### 1.4.1 截取算法

采样算法的基本目标是根据给定的强度函数  $\lambda(t)$  模拟事件的相邻到达时间 (inter-arrival times)  $t_i, i = 1, 2, \dots$ 。我们首先回顾其次泊松过程的采样方法，然后介绍泊松过程的截取 (或可加 additive) 特性，并利用该特性推导霍克斯过程的采样算法。

在齐次泊松过程中，事件的相邻到达时间服从指数分布，其概率密度函数为  $f_\tau(t) = \lambda \exp(-\lambda t), t > 0$ ，其累积分布函数为  $F_\tau(t) = 1 - \exp(-\lambda t)$ 。由于  $F_\tau(t)$  及其逆函数  $F_\tau^{-1}(t)$  具有解析解，因此可以使用逆变换采样法 (inverse transform sampling) 来生成等待时间。因为如果随机变量  $X$  具有累积分布函数  $F_X$ ，且  $Y = F_X(X)$  服从均匀分布  $U(0, 1)$ ，那么  $X^* = F_X^{-1}(Y)$  服从与  $X$  相同的分布，换言之，采样  $X^* = F_X^{-1}(Y)$ ， $Y \sim U(0, 1)$  与直接从  $X$  采样是等价的。对于泊松过程的指数分布等待时间，逆累积分布函数为  $F_\tau^{-1}(u) = \frac{-\ln u}{\lambda}$ 。因

此，在泊松过程中采样等待时间  $\tau$  只需：

$$\text{Sample } u \sim U(0, 1), \text{ then compute } \tau = \frac{-\ln u}{\lambda} \quad (18)$$

泊松过程的截取特性表明，具有强度  $\lambda$  的泊松过程可以拆分为两个独立的泊松过程，其强度分别为  $\lambda_1$  和  $\lambda_2$ ，满足  $\lambda = \lambda_1 + \lambda_2$ 。换句话说，原始过程的每个事件可以独立地分配给两个新过程中的一个。利用这一性质，我们可以通过截取一个齐次泊松过程（强度  $\lambda^*$ ）来模拟具有强度函数  $\lambda(t)$  的非齐次泊松过程，只要满足  $\lambda^* \geq \lambda(t), \forall t$ 。

霍克斯过程的截取采样算法如算法1所示。对于任何有界的  $\lambda(t)$ ，可以找到一个常数  $\lambda^*$  使得在给定时间区间内  $\lambda(t) \leq \lambda^*$ 。特别地，对于具有单调递减核函数  $\phi(t)$  的霍克斯过程，在两个连续事件时间  $[T_i, T_{i+1})$  之间， $\lambda(T_i)$  是事件强度的上界。因此，我们示例如何在已经采样出事件时间  $T_1, T_2, \dots, T_i$  之后生成下一个事件时间  $T_{i+1}$ ：

1. 令当前时间计数器  $T = T_i$
2. 以公式 (17) 采样一个相邻到达时间  $\tau$ ，使用  $\lambda^* = \lambda(T)$
3. 更新时间计数器  $T = T + \tau$
4. 计算实际事件率  $\lambda(T)$  和截取率  $\lambda^*$  的比值，并以此决定是否接受该事件：
  - 若接受，则记录事件时间  $T_{i+1} = T$
  - 若拒绝，则重新采样，直到接受一个新的事件时间

#### ALGORITHM 1: Simulation by Thinning

1. Given Hawkes process as in Eq (10).
2. Set current time  $T = 0$  and event counter  $i = 1$ .
3. While  $i \leq N$ :
  - (a) Set the upper bound of Poisson intensity  $\lambda^* = \lambda(T)$  (using Eq (10)).
  - (b) Sample inter-arrival time: draw  $u \sim U(0, 1)$  and let  $\tau = -\frac{\ln(u)}{\lambda^*}$  (as described in Eq (18)).
  - (c) Update current time:  $T = T + \tau$ .
  - (d) Draw  $s \sim U(0, 1)$ .
  - (e) If  $s \leq \frac{\lambda(T)}{\lambda^*}$ , accept the current sample: let  $T_i = T$  and  $i = i + 1$ . Otherwise, reject the sample and return to step (a).

需要注意的是，即使某个相邻到达时间  $\tau$  被拒绝，时间计数器  $T$  仍然会更新。这正是截取一个高强度值的齐次泊松过程的基本原理。此外，由于  $\lambda(t)$  在事件时间之间是严格单调的，因此即使在拒绝某次采样后，也可以动态更新上界  $\lambda^*$  以提高计算效率。

对于采样  $N$  个事件，该算法的时间复杂度为  $O(N^2)$ ，因为直接计算事件强度（公式 (10)）需要  $O(N)$  的计算量。此外，如果事件率衰减较快，在接受一个新事件之前可能会经历多次拒绝采样，进一步影响计算效率。

### 1.4.2 高效分解采样

我们介绍一种更高效的霍克斯过程采样算法，该算法采用指数核函数并避免了拒绝采样。该方法由Dassios 和 Zhao (2013) 提出，能够线性扩展到所生成的事件数量。

首先，该算法适用于具有指数型移民率（exponential immigrant rates）和指数记忆核（exponential memory kernel）的霍克斯过程。这比我们在 1.3.2 节中定义的形式更为一般化。其中，移民率由一个非齐次泊松过程描述，并服从指数函数  $a + (\lambda_0 - a) \exp(-\delta t)$ 。对于每个新事件，其引入的事件强度增量由一个常数  $\gamma$  给出，因此霍克斯过程的强度函数可表示为

$$\lambda(t) = a + (\lambda_0 - a) \exp(-\delta t) + \sum_{T_i < t} \gamma \exp(-\delta(t - T_i)), \quad t \geq 0 \quad (19)$$

我们可以进一步推广该模型，通过引入一个对  $\gamma$  进行建模的分布，但这超出了本教程的讨论范围。

需要注意的是，如果一个过程具有这样的性质：在给定当前状态条件下，其未来状态独立于过去状态，则该过程是一个马尔可夫过程。Ogata (1981) 证明，当核函数  $\phi$  采用指数形式时，Hawkes 过程的强度函数是马尔可夫过程。直观上，这可以通过公式 (18) 理解，因为  $\lambda(t_2) = \exp(-\delta(t_2 - t_1))\lambda(t_1), \forall t_2 > t_1$ （没有移民率的情况）。也就是说，在给定当前事件强度  $\lambda(t_1)$  的情况下，未来的强度仅依赖于从  $t_1$  以来经过的时间。

我们利用这一马尔可夫性质，将事件间隔时间分解为两个相互独立的随机变量，

- 第一个随机变量  $s_0$  表示下一个事件的发生时间（若该事件来自于常数背景率  $a$ ）。这一间隔时间的采样遵循公式 (17)。
- 第二个随机变量  $s_1$  表示下一个事件的发生时间（若该事件来自于指数型移民核  $(\lambda_0 - a) \exp(-\delta t)$  或霍克斯自激发核  $\sum_{T_i < t} \exp(\delta(t - T_i))$ ）。由于强度函数的马尔可夫性质，该随机变量的累积分布函数可以显式求逆，完整推导可参考Dassios 和 Zhao (2013)。直观上，最终采样的事件间隔时间为这两个变量中的最小值。值得注意的是，第二个事件间隔可能是无穷大的，这是合理的，因为指数核函数衰减较快，在这种情况下，下一个事件必然来自于常数背景率。该算法的详细步骤见算法2。

#### ALGORITHM 2: Simulation of Hawkes with Exponential Kernel

1. Set  $T_0 = 0$ , initial event rate  $\lambda(T_0) = \lambda_0$ .
2. For  $i = 1, 2, \dots, N$ :
  - (a) Draw  $u_0 \sim U(0, 1)$  and set  $s_0 = -\frac{1}{a} \ln(u_0)$ .
  - (b) Draw  $w_1 \sim U(0, 1)$ . Set  $d = 1 + \frac{\delta \ln u_1}{\lambda(T_{i-1}) - a}$ .

(c) If  $d > 0$ , set  $s_1 = -\frac{1}{\delta} \ln(d)$ ,  $\tau_i = \min\{s_0, s_1\}$ . Otherwise, set  $\tau_i = s_0$ .

(d) Record the  $i$ -th jump time  $T_i = T_{i-1} + \tau_i$ .

(e) Update event intensity at the left side of  $T_i$  with exponential decay:

$$\lambda(T_i^-) = (\lambda(T_{i-1}^+) - a) \exp(-\delta \tau_i) + a.$$

(f) Update event intensity at the right side of  $T_i$  with a jump from the  $i$ -th event:

$$\lambda(T_i^+) = \lambda(T_i^-) + \gamma.$$

该算法之所以高效，是因为在每个事件发生后，其强度函数的更新仅需常数时间，且该方法不依赖于拒绝采样。然而，这种分解方法难以直接应用于幂律核函数，因为幂律核函数不具备马尔可夫性质。

## 1.5 估计霍克斯过程的参数

在使用自激励点过程建模时，一个主要挑战是如何从观测数据中估计模型参数。在指数核霍克斯过程的情况下，通常需要确定基准强度函数  $\lambda_0(t)$ ，以及衰减核函数  $\phi(t)$  的参数  $\alpha$  和  $\delta$ 。一种常见的方法是通过最大化观测数据的对数似然函数来估计这些参数。

### 1.5.1 霍克斯过程的似然函数

设  $N(t)$  是定义在区间  $[0, T]$  上的点过程（其中  $T < \infty$ ），且  $\{T_1, T_2, \dots, T_n\}$  是  $N(t)$  在时间区间  $[0, T]$  内的观测事件时间集合。则参数集合  $\theta$  下的数据似然函数  $L$  可以表示为

$$L(\theta) = \prod_{i=1}^n \lambda(T_i) \exp\left(-\int_0^T \lambda(t) dt\right) \quad (20)$$

我们按照 (Daley and Vere-Jones 2003)、(Laub et al. 2015) 和 (Rasmussen 2013) 的方法大致推导该似然函数的表达式。假设当前时间为  $t$ ，回顾历史  $\mathcal{H}_t$ ，即所有事件发生的时间列表  $\{T_1, T_2, \dots, T_n\}$ （不包括当前时间  $t$ ）。定义  $f^*(t) := f(t | \mathcal{H}_t)$  为下一个事件  $T_{n+1}$  在给定历史  $\mathcal{H}_t$  情况下的条件概率密度函数。根据概率定义有  $\mathbb{P}(T_{n+1} \in (t, t+dt)) = f_{T_{n+1}}(t)dt$ 。我们有

$$f(T_1, T_2, \dots, T_n) = \prod_{i=1}^n f(T_i | T_1, T_2, \dots, T_{i-1}) = \prod_{i=1}^n f^*(T_i) \quad (21)$$

进一步地，事件强度函数  $\lambda(t)$  可以通过条件密度函数  $f^*(t)$  及其对应的累积分布函数  $F^*(t)$  表达为 (Rasmussen 2013)：

$$\lambda(t) = \frac{f^*(t)}{1 - F^*(t)} \quad (22)$$

**上述公式虽未给出严格证明，但从直观上理解<sup>[1]</sup>**：在一个无穷小区间  $dt$  内， $f^*(t)dt$  表示在  $dt$  内发生事件的概率，而  $1 - F^*(t)$  表示在时间  $t$  之前没有发生新事件的概率。利用贝叶斯公式进行推导 (Rasmussen 2013)，可以证明这个比率等价于计数过程  $N_{t+dt} - N_t$  的期望增量，而根据公式 (9)，这本质上等同于  $\lambda(t)dt$ 。

我们可以用累积分布函数  $F^*$  来表示条件强度函数：

$$\lambda(t) = \frac{f^*(t)}{1 - F^*(t)} - \frac{\frac{\partial}{\partial t} F^*(t)}{1 - F^*(t)} = -\frac{\partial}{\partial t} \log(1 - F^*(t)) \quad (23)$$

将  $t$  之前最后一个已知事件的时间表示为  $T_n$ ，在两边对  $(T_n, t)$  进行积分，可得

$$\int_{T_n}^t \lambda(s) ds = -[\log(1 - F^*(s))]_{T_n}^t = -[\log(1 - F^*(t)) - \log(1 - F^*(T_n))] \quad (24)$$

注意到  $F^*(T_n) = 0$ ，因为  $T_{n+1} > T_n$ ，所以

$$\int_{T_n}^t \lambda(s) ds = -\log(1 - F^*(t)) \quad (25)$$

对上式进行变换，得到

$$F^*(t) = 1 - \exp\left(-\int_{T_n}^t \lambda(s) ds\right) \quad (26)$$

结合公式 (22) 可得

$$f^*(t) = \lambda(t)(1 - F^*(t)) = \lambda(t) \exp\left(-\int_{T_n}^t \lambda(s) ds\right) \quad (27)$$

将公式 (27) 代入似然函数，我们可以得到似然的表达式

$$L(\theta) = \prod_{i=1}^n f^*(T_i) = \prod_{i=1}^n \lambda(T_i) \exp\left(-\int_{T_{i-1}}^{T_i} \lambda(u) du\right) = \prod_{i=1}^n \lambda(T_i) \exp\left(-\int_0^{T_n} \lambda(u) du\right) \quad (28)$$

### 1.5.2 极大似然估计

设  $\theta$  为霍克斯过程的参数集合，其最大似然估计可通过在参数空间  $\Theta$  上最大化公式 (20) 计算得出。具体地，极大似然估计  $\hat{\theta}$  定义为  $\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta)$ 。在计算和数值复杂度方面，需要注意加法运算比乘法运算的计算成本更低。此外，直接计算似然函数可能导致数值下溢

(underflow)，因为似然值可能非常小，从而超出浮点数表示范围。因此，通常采用对数似然函数进行优化：

$$l(\theta) = \log L(\theta) = - \int_0^T \lambda(t) dt + \sum_{i=1}^{N(T)} \log \lambda(T_i) \quad (29)$$

在最大化对数似然时，可能会遇到多个局部最大值问题。负对数似然函数的形状可能较为复杂，甚至可能不是全局凸的。这意味着最大似然估计可能会陷入局部极大值，而非全局最优值。常见的解决方法是采用多个不同的初始值进行优化，尽管这不能完全避免局部极大值问题。此外，还可以结合不同的优化方法，如果不同优化方法的结果一致，则可以更确信得到的是全局最优解。

霍克斯过程中的事件通常呈时间簇性：包括初始（移民）事件及其后代事件。在实际应用中，该过程可能在我们开始观测之前的某个时间点就已开始，即  $t = 0$  之前，因此，可能存在发生在  $t = 0$  之前但未被观测到的事件，而这些事件可能在区间  $[0, T]$  内引发新的后代事件。这些未观测到的事件可能对观测期（即  $t > 0$  之后）产生影响，但由于我们无法得知它们的存在，其对事件强度（intensity）的贡献无法被记录。这种现象被称为边界效应（edge effect），相关讨论可见 Daley 和 Vere-Jones (2003) 以及 Rasmussen (2013)。应对这一问题的一种方法是假设初始强度等于基础强度，并忽略发生在观测期之前的事件所带来的边界效应（Daley & Vere-Jones, 2003）。在大多数霍克斯过程的应用中，这种假设是标准建模方式。如 Rasmussen (2013) 所指出的，如果数据集足够大，边界效应对模型估计的影响通常可以忽略不计。在本教程中，我们将基础强度设定为常数  $\lambda(0) = \lambda_0$ ，并忽略在观测期开始之前发生的事件带来的边界效应。

霍克斯过程极大似然估计（MLE）面临的主要问题之一是计算对数似然（log-likelihood）函数的计算成本，特别是事件强度函数的计算，如下所示。注意，在公式 (29) 中，对数似然的两个部分可以分别最大化（前提是它们没有共同项），相关讨论可见 Daley 和 Vere-Jones (2003)、Ogata (1988)、Zipkin 等 (2016)。计算复杂度的主要来源是双重求和运算，这一运算来自对数似然函数的第二部分：

$$\sum_{i=1}^{N_T} \log \lambda(T_i) = \sum_{i=1}^{N_T} \left( \log \left( a + (\lambda_0 - a) \exp(-\delta t) + \sum_{j: T_j < T_i} \alpha \exp(-\delta(T_i - T_j)) \right) \right) \quad (30)$$

对于大多数霍克斯过程，该计算的复杂度通常为  $O(N_T^2)$ ，其中  $N_T$  为事件数。因此，当  $N_T$  很大时，参数估计可能会变得相当缓慢，特别是在无法避免循环计算的情况下。

在指数核的情况下，利用递推公式可将计算复杂度从  $O(N_T^2)$  降至  $O(N_T)$ （见 Ogata, 1981）。然而，对于更复杂的霍克斯过程，例如涉及幂律衰减核的模型，该策略不再适用。

## 1.6 预测

### 1.6.1 生存函数

生存函数（Survival Function）在点过程和生存分析中是一个核心概念，它表示从某个起点（通常是最后一个事件发生的时间  $t_0$ ）到时间  $t$  之间没有事件发生的概率。

在点过程中，强度函数  $\lambda(t)$  表示在时间  $t$  发生事件的瞬时概率密度。具体来说，在一个很小的时间间隔  $[t, t + \Delta t)$  内，事件发生的概率近似为  $\lambda(t)\Delta t$ （当  $\Delta t$  趋于 0 时）。

将时间区间  $[t_0, t)$  划分为  $n$  个小的等长时间段，每个时间段长度为  $\Delta t = \frac{t-t_0}{n}$ 。记第  $k$  个时间点为  $t_k = t_0 + k\Delta t$ ，在每个小时间段  $[t_k, t_k + \Delta t)$  内，事件发生的概率近似为  $\lambda(t_k)\Delta t$ ，没有事件发生的概率近似为  $1 - \lambda(t_k)\Delta t$ 。

假设在不同时间段内事件发生的条件是独立的（在给定历史的情况下，例如点过程中条件于过去的强度函数）<sup>[2]</sup>，那么在整个区间  $[t_0, t)$  内没有事件发生的概率  $S(t)$  可以表示为所有小时间段内无事件概率的乘积：

$$S(t) = \prod_{k=0}^{n-1} (1 - \lambda(t_k)\Delta t) \quad (31)$$

为了将这个乘积转化为连续形式，我们对  $S(t)$  取自然对数

$$\log S(t) = \sum_{k=0}^{n-1} \log(1 - \lambda(t_k)\Delta t) \quad (32)$$

当  $\Delta t$  很小时， $\lambda(t_k)\Delta t$  是一个很小的值，根据泰勒展开， $\log(1 - x) \approx -x$ （当  $x \rightarrow 0$ ），所以  $\log(1 - \lambda(t_k)\Delta t) \approx -\lambda(t_k)\Delta t$ ，代入公式 (32)

$$\log S(t) \approx - \sum_{k=0}^{n-1} \lambda(t_k)\Delta t \quad (33)$$

当  $n \rightarrow \infty, \Delta t \rightarrow 0$  时，此离散和趋近于积分  $\log S(t) = \left( - \int_{t_0}^t \lambda(s) ds \right)$ ，即

$$S(t) = \exp \left( - \int_{t_0}^t \lambda(s) ds \right) \quad (34)$$

### 1.6.2 下一事件时间预测

下一事件在时间  $\tau$  之前发生的累积概率是  $F(\tau)$ ，它等于 1 减去生存概率

$$F(\tau) = P(\text{Next event time} \leq \tau) = 1 - S(\tau) = 1 - \exp\left(-\int_{\tau_0}^{\tau} \lambda(s) ds\right) \quad (35)$$

概率密度函数  $f(\tau)$  是累积分布函数  $F(\tau)$  对  $\tau$  的导数  $f(\tau) = \frac{d}{d\tau} F(\tau)$ , 可得

$$\begin{aligned} f(\tau) &= \frac{d}{d\tau} \left[ 1 - \exp\left(-\int_{\tau_0}^{\tau} \lambda(s) ds\right) \right] \\ &= \lambda(\tau) \exp\left(-\int_{\tau_0}^{\tau} \lambda(s) ds\right) \end{aligned} \quad (36)$$

进一步地, 可以计算下一事件时间的期望值

$$\hat{t} = \mathbb{E}[t \mid \mathcal{H}] = \int_{t_0}^{\infty} \tau \cdot f(\tau) d\tau = \int_{t_0}^{\infty} \tau \cdot \lambda(\tau) \exp\left(-\int_{\tau_0}^{\tau} \lambda(s) ds\right) d\tau \quad (37)$$

- 
1. 是否能严格证明。↩
  2. 这与非齐次泊松过程的假设是否相悖, 如果在前一个时间段内发生事件, 会影响接下来所有时间的事件发生。↩