

## 研究ノート

# Hawkes 過程の理論と実証 —東京金先物市場への応用—

砂田 洋志

(山形大学人文社会科学部)

### 1. はじめに

本稿の目的は、A. G. Hawkes が 1971 年に提案した Hawkes 過程 (Hawkes process) を理論と実証の両面から紹介すること、日本の金先物市場へ応用して、その結果を紹介することである。同過程を紹介するにあたっては、理論的な説明だけでは不十分であると考えて、東京金先物市場のティックデータの実証分析を応用例として加えてある。

Hawkes 過程とは、不規則な時間間隔で発生する事象の発生間隔をモデル化した確率過程である。しかも、強度が時変することによって間隔が短い区間と長い区間が混在できるようにモデル化されている。地震の発生間隔もこうした性質を備えているため、余震の発生間隔の分析に数多く利用されてきた。株式市場において取引が成立する時点の間隔が不規則なことから、近年では株式市場の分析にも応用され始めた。初期の研究として Bowsher(2007) を挙げることができる。株式市場への応用については Bacry 他 (2015) のサーベイ論文を参照されたい。

Hawkes 過程やポアソン過程といった確率過程は点過程と呼ばれ、所謂時系列モデルと区別されている。というのも、時系列モデルでは等しい時間間隔で事象が発生し、異時点間で発生する事象の統計的な関係を分析対象とする。それに対して、点過程では不規則な間隔で事象が発生する点が大きく異なり、事象の不規則な発生間隔を分析の主たる対象としているからである。

Hawkes 過程の推定方法としては最尤法を利用することが可能である。しかし、多変量になると推定するパラメータ数が急増するため、尤度関数が複雑となり、最適化が難しくなる。そのためノンパラメトリックな推定方法やベイズ統計学を用いた推定方法などが考案されている。

Hawkes 過程に関する研究であるが、推定方法として頻繁に利用される最尤推定法については Ogata(1978) において最尤推定量の漸近的正規性が証明された。Ozaki(1979) において、指数型の反応関数を仮定した一変量 Hawkes 過程のパラメータの最尤推定法とデータのシミュレーション方法が提案された。間引き (thinning) 法によるデータのシミュレーションについては、

Ogata(1981) で Ogata の修正間引き法が提案されている。

多変量の Hawkes 過程については Hawkes(1971) で既に提案されている他、Embrechts 他 (2011) でも提案されている。佐藤・今井 (2020) では、Embrechts 他 (2011) の多変量の Hawkes 過程を用いて日本の株式市場における注文板データをモデル化して分析している。

本稿の構成は以下のとおりである。<sup>1</sup> 第2節では、点過程の説明を行った後、基本的な点過程であるポアソン過程を説明する。第3節では、一変量 Hawkes 過程を説明する。第4節では、マーク付き点過程を説明した後、マーク付き Hawkes 過程の代表例である ETAS モデルを説明する。第5節では、一変量の Hawkes 過程を金先物市場のティックデータに応用する他、マーク付き Hawkes 過程として ETAS モデルも同データに応用する。その後でパラメータを推定した結果を検討する。最後の第6節で結論を述べる。

## 2. 点過程

### 2.1 点過程

点過程とは、空間上にランダムに分布する“点”の集合に関する確率過程である。<sup>2</sup> 分析者が興味を持っている事象の発生したことを“点”は表す。事象の例としては、地震の発生、ガソリンスタンドに給油に来る自動車、機器の故障の発生などである。ある観察期間  $[0, T]$  に  $n$  個の事象が発生し、その発生時点を時間順に  $t_1, t_2, t_3, \dots, t_n$  と記述するならば、点過程は時点の集合  $\mathbf{t}_n = \{t_1, t_2, t_3, \dots, t_n\}$  によって特徴づけられる。点過程では、ランダムに発生した時点の集合の確率的な構造を分析するが、その性質は強度関数と呼ばれる事象の発生し易さを表す関数によって決まる。

点過程を基にして、時点 0 から時点  $t$  までの間に発生した事象の回数を考えることができる。たとえば、 $t_1 < t < t_2$  であれば  $[0, t]$  の間に事象は 1 回発生している。もし、 $t$  が大きな値で  $t_{n-1} < t < t_n$  であれば  $[0, t]$  の間に事象は  $n-1$  回発生している。つまり、 $[0, t]$  の間に発生する事象の回数は  $t$  の関数である。 $[0, t]$  の間の事象の発生回数を  $N(0, t)$  と記述すれば、 $N(0, t)$  は 0 以上の整数を実現値とする確率変数であり、計数過程 (counting process) と呼ばれる。そして、時点の集合  $\mathbf{t}_n$  と  $N(0, t)$  は一対一に対応する。

点の発生間隔  $\tau_i = t_i - t_{i-1}$  は正值の値しかとらない時系列である。発生間隔  $\{\tau_i\}$  が独立かつ同一な確率分布に従う点過程は更新過程 (renewal process) と呼ばれる。さらに、発生間隔  $\{\tau_i\}$  が独立で指数分布に従うのであれば、ポアソン過程 (Poisson process) と呼ばれる。発生間隔  $\{\tau_i\}$  が (多重) マルコフ性を持つのであれば、Wold 過程と呼ばれる。

<sup>1</sup> 点過程、ポアソン過程、(マーク付き)Hawkes 過程の説明は近江・野村 (2019) に大きく負っている。

<sup>2</sup> Daley=Vere-Jones(2003) に点過程に関する包括的な説明が記載されている。

## 2.2 ポアソン過程

点過程の代表的なものがポアソン過程である。ポアソン過程には定常なものと同定常的なものがある。前者は強度が時間に関係なく一定であるが、後者は時間とともに変化する。ここでは定常なポアソン過程を紹介する。

### 強度関数 (intensity function)

前述したとおり、点過程は事象の発生した時点の集合である。事象は発生確率が高いほど発生しやすいので、発生確率が高いと発生間隔は短くなりやすい。逆に、発生確率が低いと発生間隔は長くなりやすい。瞬間的な発生確率は強度あるいは強度関数と呼ばれ、点過程の性質を決める。<sup>3</sup> 定常なポアソン過程では、以下に示すとおり、強度関数 $\lambda(t)$ は過去の影響や時点の影響を受けずに一定値 $\lambda$ である。

$$\lambda(t)=\lambda \quad (1)$$

定常なポアソン過程において $[t, t+\Delta]$ の間に事象の発生する確率と発生しない確率を記述すると、以下のとおりである。

$$\Pr(N(t, t+\Delta)=1 | N(s) (s \leq t)) = \lambda \Delta + o(\Delta) \quad (2)$$

$$\Pr(N(t, t+\Delta) > 1 | N(s) (s \leq t)) = o(\Delta)$$

定常なポアソン過程の場合、 $\Delta$ を固定すると時点に関係なく発生間隔は平均 $\lambda \Delta$ のポアソン分布に従う。

### 尤度関数

一般的な点過程の確率密度関数は以下のとおりである。

$$p_{[0, T]}(\{t_i\}_{i=1}^n) = \prod_{i=1}^n \lambda(t_i | H_{t_i}) \times \exp \left[ - \int_0^T \lambda(s | H_s) ds \right] \quad (3)$$

ただし、 $\lambda(t_i | H_{t_i})$ は時点 $t_i$ の直前までに発生した事象の発生履歴 $H_{t_i}$ が与えられた時の時点 $t_i$ における発生強度である。(3)に定常なポアソン過程の強度 $\lambda$ を代入することによって、尤度関数(4)が得られる。なお、 $\theta$ は未知パラメータを表し、この場合は $\lambda$ を指す。

$$\begin{aligned} L(\theta | \{t_i\}_{i=1}^n) &= \prod_{i=1}^n \lambda \times \exp \left[ - \int_0^T \lambda ds \right] = \lambda^n \times \exp \{ - \lambda T \} \\ &= \lambda^n \times \exp \{ - \lambda T \} \end{aligned} \quad (4)$$

したがって、対数尤度関数は以下のとおりである。

$$\ln L(\theta | \{t_i\}_{i=1}^n) = \ln(n\lambda) - \lambda T \quad (5)$$

<sup>3</sup> 代表的な点過程である定常ポアソン過程では強度関数が一定値である。時点 $t$ の関数ならば非同定常ポアソン過程である。一方、Hawkes 過程では過去の値に依存する形になっている。尾形(1998)を参照。

### 3. 一変量 Hawkes 過程

強度が一定の点過程から得られるデータの場合、事象の発生間隔は安定していると考えられる。したがって、このような点過程において、発生間隔が大きく変化する場合は外生的なショックが原因の場合が多いと考えられる。一方、自己励起性がある点過程ならば、内部に大きな変動を生み出す仕組みが内包されている。したがって、自己励起性があれば、外生的なショックが無くても比較的短い時間内に複数の事象が集中して発生する現象が起きうようになる。自己励起性を有するタイプの点過程として Hawkes 過程を紹介する。

ポアソン過程は強度が一定なので、発生間隔が大きく変化する可能性は低い。したがって、発生間隔が大きく変化するような事象をモデル化するには、ポアソン過程よりも Hawkes 過程の方が適切であると考えられる。Hawkes 過程には一変量過程だけでなく、多変量過程もあるが、本稿では基本的な一変量過程を扱う。

#### 強度関数

Hawkes 過程では強度関数が過去の値に依存する。つまり、時点  $t$  における強度関数  $\lambda(t)$  は時点  $t$  までの情報  $H_t$  を条件とした条件付き強度関数である。カーネル関数  $g(t)$  を用いて以下のように記述できる。

$$\lambda(t) = \lambda(t|H_t) = \mu + \int_0^t g(t-u)dN(u) \quad (6)$$

Ogata(1978)によれば、(6)は以下のとおりに式変形できる。<sup>4</sup>

$$\lambda(t|H_t) = \mu + \sum_{t_i < t} g(t - t_i) \quad (6')$$

Hawkes 過程において  $[t, t+\Delta]$  の間に事象の発生する確率と発生しない確率を記述すると、以下のとおりである。

$$\begin{aligned} \Pr(N(t, t+\Delta)=1 | N(s) (s \leq t)) &= \lambda(t)\Delta + o(\Delta) \\ \Pr(N(t, t+\Delta)>1 | N(s) (s \leq t)) &= o(\Delta) \end{aligned} \quad (2')$$

#### 尤度関数

一般的な点過程の確率密度(3)に Hawkes 過程の条件付き強度関数(6')を代入することによって、Hawkes 過程の尤度関数が得られる。

$$\begin{aligned} L(\theta | \{t_i\}_{i=1}^n) &= \prod_{i=1}^n [\mu + \sum_{j < i} g(t_i - t_j)] \times \exp \left[ - \int_0^T \{ \mu + \sum_{t_i < t} g(t - t_i) \} dt \right] \\ &= \prod_{i=1}^n [\mu + \sum_{j < i} g(t_i - t_j)] \times \exp \left[ - \mu T - \sum_{i=1}^n \int_{t_i}^T g(s - t_i) ds \right] \end{aligned}$$

したがって、対数尤度関数は以下のとおりである。

<sup>4</sup> Ogata(1978)のp248の(2.4)を参照されたい。

$$\ln L(\theta|\{t_i\}_{i=1}^n) = \sum_{i=1}^n \ln[\mu + \sum_{j<i} g(t_i - t_j)] - \mu T - \sum_{i=1}^n \int_{t_i}^T g(s - t_i) ds \quad (7)$$

実際の推定ではさらに、カーネル関数  $g(t)$  の形状も具体的に仮定する必要がある。効率的な計算が可能になるため、カーネル関数  $g(t)$  として、

- ・ 指数関数… $g(t) = a \exp(-\beta t)$
- ・ ラゲール型多項式… $\sum_{j=0}^J \alpha_j t^j \exp(-\beta t)$
- ・ 冪関数… $g(t) = \frac{K}{(t+c)^p}$

が仮定されることが多い。ラゲール型多項式を仮定すると条件付き強度関数がマルコフ性を持ち、尤度の計算速度を速めることが可能なため、カーネル関数として仮定される場合は多い。

本稿ではカーネル関数  $g(t)$  として指数関数  $g(t) = a \exp(-\beta t)$  を仮定する。 $a \exp(-\beta t)$  は逐次的な関係を持つラゲール型多項式  $\sum_{j=0}^J \alpha_j t^j \exp(-\beta t)$  の中で最も単純な形をした関数である。点過程では事象の発生によって強度は大きく跳ね上がるが、そのジャンプの大きさを  $\alpha$  が表す。また、一旦大きくなった強度は時間とともに減衰するが、減衰の程度を  $\beta$  が表す。 $g(t) = a \exp(-\beta t)$  と仮定した場合の尤度関数は以下のとおりである。

$$\begin{aligned} L(\theta|\{t_i\}_{i=1}^n) &= \prod_{i=1}^n [\mu + \sum_{j<i} \alpha \times \exp\{-\beta(t_i - t_j)\}] \\ &\quad \times \exp\left[-\mu T - \sum_{i=1}^n \int_{t_i}^T \alpha \times \exp\{-\beta(s - t_i)\} ds\right] \\ &= \prod_{i=1}^n [\mu + \alpha \sum_{j<i} \exp\{-\beta(t_i - t_j)\}] \\ &\quad \times \exp\left[-\mu T - \alpha \sum_{i=1}^n \int_{t_i}^T \exp\{-\beta(s - t_i)\} ds\right] \end{aligned}$$

ただし、 $\theta = \{\mu, \alpha, \beta\}$  である。指数部分の積分を計算すると以下のように式変形できる。<sup>5</sup>

$$\begin{aligned} &= \prod_{i=1}^n [\mu + \alpha \sum_{j<i} \exp\{-\beta(t_i - t_j)\}] \\ &\quad \times \exp\left[-\mu T - \frac{\alpha}{\beta} \sum_{i=1}^n [1 - \exp\{-\beta(T - t_i)\}]\right] \end{aligned}$$

以上から対数尤度関数は以下のとおりである。

$$\begin{aligned} \ln L(\theta|\{t_i\}_{i=1}^n) &= \sum_{i=1}^n \ln[\mu + \alpha \sum_{j<i} \exp\{-\beta(t_i - t_j)\}] \\ &\quad - \left[ \mu T + \frac{\alpha}{\beta} \sum_{i=1}^n [1 - \exp\{-\beta(T - t_i)\}] \right] \end{aligned} \quad (8)$$

さらに、計算を簡単にするためにこの過程の逐次的な構造を反映させた対数尤度関数は以下のとおりである。<sup>6</sup>

$$\ln L(\theta|\{t_i\}_{i=1}^n) = \sum_{i=1}^n \ln[\mu + \alpha A(i)] - \left[ \mu T + \frac{\alpha}{\beta} \sum_{i=1}^n [1 - \exp\{-\beta(T - t_i)\}] \right] \quad (9)$$

ただし、 $A(i) = \sum_{j<i} \exp\{-\beta(t_i - t_j)\}$ ,  $i \geq 2$ ,  $A(1) = 0$  である。<sup>7</sup>

さらに、カーネル関数  $g(t)$  が複数の指数関数を組み合わせたラゲール型多項式の場合、

<sup>5</sup>  $\sum_{i=1}^n \int_{t_i}^T \exp\{-\beta(s - t_i)\} ds = \sum_{i=1}^n \left[ \frac{1}{-\beta} \exp\{-\beta(s - t_i)\} \right]_{t_i}^T = \sum_{i=1}^n \frac{-1}{\beta} \exp\{-\beta(T - t_i)\} + \frac{1}{\beta} \exp\{-\beta(t_i - t_i)\}$

$= \frac{-1}{\beta} \sum_{i=1}^n \exp\{-\beta(T - t_i)\} + \frac{1}{\beta} = \frac{1}{\beta} \sum_{i=1}^n [1 - \exp\{-\beta(T - t_i)\}]$  と式変形できる。

<sup>6</sup> Ozaki(1979) を参照されたい。

<sup>7</sup>  $A(i+1) = \{A(i) + 1\} \exp\{-\beta(t_{i+1} - t_i)\}$  という逐次的な構造を有している。

$$g(t) = \sum_{j=0}^p \alpha_j t^j \times \exp\{-\beta t\}$$

であり、尤度関数と対数尤度関数は以下のとおりである。<sup>8</sup>

$$L(\theta|\{t_i\}_{i=1}^n) = \prod_{i=1}^n \left[ \mu + \sum_{j < i} \sum_{k=0}^p \alpha_k (t_i - t_j)^k \times \exp\{-\beta(t_i - t_j)\} \right] \\ \times \exp \left[ -\mu T - \sum_{i=1}^n \int_{t_i}^T \sum_{k=0}^p \alpha_k (s - t_i)^k \times \exp\{-\beta(s - t_i)\} ds \right]$$

$$\ln L(\theta|\{t_i\}_{i=1}^n) = \sum_{i=1}^n \ln \left[ \mu + \sum_{j=0}^p \alpha_j \times R_j(i) \right] \\ - \left[ \mu T + \sum_{i=1}^n \sum_{j=0}^p \alpha_j \times S_j(T - t_i) \right]$$

ただし、 $\theta = \{\mu, \{\alpha_j\}_{j=1}^p, \beta\}$ である。そして、 $R_j(i)$ と $S_{j+1}(t)$ は以下のとおりである。

$$R_j(i) = A_j(t_i - t_{i-1}) + \sum_{k=0}^j {}_j C_k A_{j-k}(t_i - t_{i-1}) R_k(i-1), R_0(1)=0, t_0=0, \\ S_{j+1}(t) = \frac{(j+1)S_j(t) - A_{j+1}(t)}{\beta}, S_0(t) = \frac{1 - \exp(\beta t)}{\beta}, A_j(t) = t^j \exp(-\beta t)$$

## 4. マーク付き点過程

### 4.1 マーク付き点過程

ある事象が発生する時点を記述する一変量の点過程がある場合、その発生時点で別の事象も発生し、相互に影響しあっている場合がある。たとえば、株式の取引であれば、株の売買が行われる時点が点過程で記述されるだけでなく、取引に伴い取引量も同じ時点で記録される。また、地震であれば、地震が発生する時点が点過程で記述されるだけでなく、地震の規模や震源の位置が同じ時点で記録される。事象の発生時点 $t_n = \{t_1, t_2, \dots, t_n\}$ があり、各時点で発生した事象に対応する情報や特性はマーク (mark) と呼ばれる。

マークはある空間 $\mathcal{X}$ 上の確率変数 $\mathbf{x}_n = \{x_1, x_2, \dots, x_n\}$ として記述される。株の場合は取引量、地震の場合は地震の規模と震源の位置がマークの例として挙げられる。それらの情報を含めた点過程はマーク付き点過程 $\{t_n, \mathbf{x}_n\}$ と呼ばれる。

### 強度関数

マーク付き点過程 $\{t_n, \mathbf{x}_n\}$ は事象の発生時点とそのマークを表す。地震であれば、 $t_i$ が*i*番目の地震発生時点、 $x_i$ が*i*番目の地震の規模などである。この場合の強度関数は $\lambda(t, \mathbf{x}|H_t)$ と記述される。ただし、 $H_t$ は時点*t*までの事象とマークの履歴 $H_t = \{(t_i, x_i) | t_i < t\}$ である。

<sup>8</sup> Ogata(1981)を参照されたい。

### 尤度関数

マーク付き点過程の確率密度関数は以下のとおりに記述される。

$$p_{[0,T]}[\{t_i\}_{i=1}^n, \{x_i\}_{i=1}^n] = \prod_{i=1}^n \lambda(t_i, x_i | H_{t_i}) \times \exp \left[ - \int_0^T \lambda(s | H_s) ds \right]$$

マークが事象の発生時点や過去の発生履歴に依存せず、独立かつ同一の確率分布に従うと仮定される場合も多い。この場合の強度関数  $\lambda(t_i, x_i | H_{t_i})$  はマークの確率 (密度) 関数を  $\psi(x)$  と記述すると、以下のように分解できる。

$$\lambda(t_i, x_i | H_{t_i}) = \lambda(t_i | H_{t_i}) \psi(x_i) \quad (10)$$

(10) を上述した確率密度関数へ代入すれば尤度関数は以下のとおりである。

$$L(\theta | \{t_i\}_{i=1}^n, \{x_i\}_{i=1}^n) = \prod_{i=1}^n \psi(x_i) \prod_{i=1}^n \lambda(t_i | H_{t_i}) \times \exp \left[ - \int_0^T \lambda(s | H_s) ds \right] \quad (11)$$

対数尤度関数は以下のとおりである。

$$\ln L(\theta | \{t_i\}_{i=1}^n, \{x_i\}_{i=1}^n) = \sum_{i=1}^n \ln \psi(x_i) + \sum_{i=1}^n \ln \lambda(t_i | H_{t_i}) - \int_0^T \lambda(s | H_s) ds \quad (12)$$

(10) ではマークの確率 (密度) 関数  $\psi(x)$  が他と分離されているため、 $\psi(x)$  の推論、そしてマークの値が与えられた下で事象が発生する推論を分けて考えることができる。

### 4.2 ETAS モデル

前小節では点過程の詳細を特定していなかったが、この小節では Hawkes 過程に従うと仮定する。マーク付き Hawkes 過程の重要なモデルとして Ogata(1988) で提案された ETAS (Epidemic-Type Aftershock Sequence) モデルがある。この小節ではマーク付き Hawkes 過程として ETAS モデルを紹介する。ETAS モデルでは、マーク  $x$  が強度に与える影響を表す衝撃関数を  $\phi(x)$  と記述すれば、 $\lambda(t | H_t)$  は以下のように  $\phi(x)$  を通じて  $x$  の関数となり、マークの影響を受ける。

$$\lambda(t | H_t) = \mu + \sum_{t_i < t} g(t - t_i) \phi(x_i) \quad (13)$$

ETAS モデルはもともと地震活動を記述するために考案されたモデルであり、 $\{t_i\}$  は地震の発生時点、 $\{x_i\}$  は地震の震度である。 $\phi(x_i)$  は  $\exp\{\gamma(x_i - X)\}$ 、ただし、 $x_i \geq X$  と定式化される。この定式化によって震度  $X$  以上の地震を対象に震度が大きいほどその後多くの地震を引き起こすという性質が強度関数に組み込まれる。

なお、ETAS モデルでも強度関数  $\lambda(t, x | H_t)$  が (10) のようにマークの確率 (密度) 関数  $\psi(x)$  と  $\lambda(t | H_t)$  に分解される。本稿ではマークの確率密度関数  $\psi(x)$  として平均が  $\kappa$  の指数関数を仮定する。

(11) に条件付き強度関数 (13) とマークの確率 (密度) 関数  $\psi(x_i)$  を代入することにより、マーク付き Hawkes 過程の尤度関数が以下のとおりに得られる。

$$L(\theta | \{t_i\}_{i=1}^n, \{x_i\}_{i=1}^n) = \prod_{i=1}^n \left[ \mu + \sum_{t_j < t_i} \phi(x_j) g(t_i - t_j) \right] \psi(x_i) \times \exp \left[ - \mu T - \sum_{i=1}^n \phi(x_i) \int_{t_i}^T g(s - t_i) ds \right]$$

さらに、カーネル関数  $g(t)$  として指数的に減衰する  $g(t) = a \exp(-\beta t)$ 、衝撃関数として下限  $X$  を

設けた $\phi(x_i)=\exp\{\gamma(x_i-X)\}$ 、マークの確率密度関数として $\psi(x_i)=\frac{1}{\kappa}\exp\left\{-\frac{x_i}{\kappa}\right\}$ を仮定し、上式へ代入した後で対数変換することにより、以下の対数尤度関数が得られる。

$$\begin{aligned} \ln L(\theta|\{t_i\}_{i=1}^n, \{x_i\}_{i=1}^n) &= \sum_{i=1}^n \log \left[ \mu + \alpha \sum_{j<i} \exp\{-\beta(t_i - t_j)\} \exp\{\gamma(x_j - X)\} \right] \\ &\quad - \sum_{i=1}^n \left( \log \kappa + \frac{x_i}{\kappa} \right) \\ &\quad - \left[ \mu T + \frac{\alpha}{\beta} \sum_{i=1}^n \exp\{\gamma(x_i - X)\} [1 - \exp\{-\beta(T - t_i)\}] \right] \end{aligned} \quad (14)$$

ただし、 $\theta=\{\mu, \alpha, \beta, \gamma, \kappa\}$ である。推定を行う際の計算を簡単にするため、上記の過程の逐次的な構造を反映させた対数尤度関数は以下のとおりである。

$$\begin{aligned} \ln L(\theta|\{t_i\}_{i=1}^n, \{x_i\}_{i=1}^n) &= \sum_{i=1}^n \log[\mu + \alpha A(i)] - \sum_{i=1}^n \left( \log \kappa + \frac{x_i}{\kappa} \right) \\ &\quad - \left[ \mu T + \frac{\alpha}{\beta} \sum_{i=1}^n \exp\{\gamma(x_i - X)\} [1 - \exp\{-\beta(T - t_i)\}] \right] \end{aligned} \quad (15)$$

ただし、 $A(i) = \sum_{j<i} \exp\{-\beta(t_i - t_j)\} \exp\{\gamma(x_j - X)\}$ ,  $i \geq 2$ ,  $A(1)=0$ である。<sup>9</sup>

## 5. Hawkes 過程の推定

### 5.1 分析する過程

本稿では最初に一変量の Hawkes 過程を取引時点のみの金先物市場のティックデータに応用して、同過程のパラメータを推定し、推定結果を検討する。次に出来高を強度関数に組み入れたマーク付き Hawkes 過程を、取引時点と出来高の金先物データに応用して、同過程のパラメータを推定し、推定結果を検討する。

今回の分析では、ティックデータなので時点数は多いものの、扱うデータの種類は1種類か2種類と少ないので、推定に当たっては最尤法を利用する。

一変量の Hawkes 過程の強度関数は、カーネル関数 $g(t)$ として指数関数 $g(t)=\alpha \exp(-\beta t)$ を用いる。したがって、強度関数は以下のように定式化される。

$$\lambda(t_i | H_{t_i}) = \mu + \sum_{j<i} g(t_i - t_j) = \mu + \sum_{j<i} \alpha \exp\{-\beta(t_i - t_j)\}$$

そして、対数尤度関数は(8)である。

次に、出来高 $VOL_i$ を強度関数に組み入れたマーク付き Hawkes 過程の強度関数は、出来高を対数変換した $vol_i = \ln(VOL_i)$ を指数関数の形で組み込み、以下のように定式化される。

$$\lambda(t_i | H_{t_i}) = \mu + \sum_{j<i} \alpha \exp\{-\beta(t_i - t_j)\} \exp\{\gamma vol_j\}$$

出来高と取引時間間隔では大きさに大きな違いがあるので、出来高を対数変換した $vol_i$ を推定に利用した。 $\phi(\ln(VOL_i)) = \exp\{\gamma[\ln(VOL_i) - 0]\} = \exp(\gamma vol_i)$ であるから、ETAS モデルと比較す

<sup>9</sup>  $A(i+1) = [A(i) + \exp\{\gamma(x_i - X)\}] \exp\{-\beta(t_{i+1} - t_i)\}$ という逐次的な構造を有している。

ると、 $\ln(VOL_i)$ の下限は0である。よって、出来高の下限は1である。したがって、推定するモデルはETASモデルであり、その対数尤度は以下のとおりである。<sup>10</sup>

$$\ln L(\theta|\{t_i\}_{i=1}^n, \{vol_i\}_{i=1}^n) = \sum_{i=1}^n \log[\mu + \alpha A(i)] - \sum_{i=1}^n \left( \log \kappa + \frac{vol_i}{\kappa} \right) - \left[ \mu T + \frac{\alpha}{\beta} \sum_{i=1}^n \exp\{\gamma vol_i\} [1 - \exp\{-\beta(T - t_i)\}] \right] \quad (16)$$

ただし、 $A(i) = \sum_{j < i} \exp\{-\beta(t_i - t_j)\} \exp\{\gamma vol_j\}$ ,  $i \geq 2$ ,  $A(1)=0$ ,  $\theta = \{\mu, \alpha, \beta, \gamma, \kappa\}$ である。

## 5.2 分析するデータ

本稿では、東京商品取引所で取引されている金先物市場（現物受渡用）のティックデータを加工して取引間隔のデータを作成し、そのデータを用いて上記のモデルを推定した。分析する限月については、金先物も含めて日本の商品先物市場では一般に期先物の出来高が最も多い。そこで、期先物を分析対象にすることにした。分析する期間については以下のように決めた。2018年7月から2020年6月までの24ヶ月の期先物の日次データを調べたところ、2019年2月は少ない方から12位、2018年9月は13位であった。日次の終値データを用いて計算した価格変化率の月間の標準偏差は2019年2月が小さい方から11位、2018年9月は9位であった。この2つの月は出来高でもボラティリティーでも中間的な月であると考えられる。今回はより中間的な月である2019年2月を分析対象とした。さらに、出来高が2月の22営業日で少ない方から12番目の2月1日のティックデータを分析対象とした。なお、この日の期先物は2019年12月限である。この日の日中の取引時間内に取引された期先物のデータは2,523個であった。

東京商品取引所の金先物市場における日中の取引時間は午前8時45分～15時15分である。午前8時45分に板合せ取引が行なわれて8時45分に始値が付く。その後は休みなくザラバ取引が15時10分まで行なわれ、5分間の取引停止時間を経て、15時15分に板合せ取引が行なわれる。この後、夜間取引も行われているが、本稿の分析では日中のザラバ取引のデータだけを利用してパラメータの推定を行った。

寄り付きと大引けには取引が集中するので、寄り付きの5分間と大引けの10分間（ザラ場に限定すると5分間）を除いた8時50分から15時5分までのデータを用いて実証分析した。最終的な分析に用いたデータは2,428個であった。

本稿で分析するティックデータとは、1日の中で取引される一つ一つの取引に関するデータを集めたものである。ティックデータには、取引時刻（時分秒）、取引価格（円）、取引数量（枚）が含まれている。東京商品取引所から得たティックデータには、日付（年月日）、取引種別（先物かオプションか）、商品番号、限月（年月）、約定時刻（時分秒）、約定価格（円）、出来高（枚）が入力されている。

<sup>10</sup>  $\phi(vol_i) = \exp\{\gamma vol_i\}$ は $(VOL_i - 0)^{\gamma}$ と記述することもできる。

ティックデータでは1秒間に複数の取引が執行される場合もある。この場合の例が表1の左側に示されている。ここに示すとおり、同じ時点に複数の取引情報が表示される。取引間隔を計算する場合には同じ時点(秒)で執行された取引が1秒間に等間隔で執行されたとみなす。つまり、表1の右側のように処理する。

表1 取引データの変更例

取引データ変更前			取引データ変更後		
取引時刻	取引価格	出来高	取引時刻	取引価格	出来高
9:05:10	1000	1	9:05:10:00	1000	1
9:05:10	1000	3	9:05:10:30	1000	3
9:05:11	1002	2	9:05:11:00	1002	2
9:05:11	1002	4	9:05:11:33	1002	4
9:05:11	1002	3	9:05:11:66	1002	3
9:05:13	1003	2	9:05:13	1003	2

### 5.3 推定結果

#### (1) パラメータの推定結果

最尤法を用いて一変量 Hawkes 過程のパラメータを推定した結果は表2に示したとおりである。推定に当たっては、勾配やヘッセ行列を事前に計算して利用した。 $\mu$  の最尤推定値は0.0528であり、有意に0と異なると判断される。取引発生からの経過時間に関係なく、強度は少なくとも $\mu$  (0.0528) だけはあることが分かった。そして、 $\alpha$  と  $\beta$  の最尤推定値は1.1383と2.2282であり、有意に0と異なると判断される。 $\beta$  が正值であることから、取引から時間が経つにつれて発生確率が減少することが分かった。また、取引発生によって強度が $\alpha$  (1.1383) だけ上昇することも分かった。

表2 一変量 Hawkes 過程の推定結果

パラメータ	推定値	標準誤差	推定値／標準誤差
$\mu$	0.0528	0.0016	32.479
$\alpha$	1.1383	0.0564	20.170
$\beta$	2.2282	0.0995	22.388
対数尤度	-5,252.9525	AIC	10,511.905

最尤法を用いて、マーク付き一変量 Hawkes 過程として ETAS モデルを推定した結果は表 3 に示したとおりである。推定に当たっては、勾配やヘッセ行列を事前に計算して利用した。また、マーク無しの Hawkes 過程を推定した結果を初期値として利用している。 $\mu$  の最尤推定値は 0.0528 であり、有意に 0 と異なると判断される。そして、 $\alpha$  と  $\beta$  の最尤推定値は 1.0012 と 2.2345 であり、有意に 0 と異なると判断される。 $\mu, \alpha, \beta$  の推定値は出来高の無い Hawkes 過程のパラメータの推定結果とほぼ同じである。 $\gamma$  の最尤推定値は 0.1352 であり、有意に 0 と異なると言える。出来高の係数が正值なので、出来高が大きくなると取引の発生確率が上がることになる。 $\kappa$  の最尤推定値は 4.136 であり、有意に 0 と異なると判断される。

本稿で分析するマーク付き一変量 Hawkes 過程においてマークとしては出来高の対数値を採用している。そして、出来高の確率分布として指数分布を仮定している。

表 3 マーク付き一変量 Hawkes 過程の推定結果

パラメータ	推定値	標準誤差	推定値／標準誤差
$\mu$	0.0528	0.0016	32.495
$\alpha$	1.0012	0.0630	15.894
$\beta$	2.2345	0.0998	22.394
$\gamma$	0.1352	0.0382	3.538
$\kappa$	4.1360	0.0839	49.275
対数尤度、( )内は出来高の対数尤度	-8,924.3996 (-3,677.4338)	AIC、( )内は出来高の AIC	17,858.7992 (7,356.8676)

表 3 に記載されている対数尤度には、出来高に関する対数尤度 $-\sum_{i=1}^n \left( \log \kappa + \frac{x_i}{\kappa} \right)$ が含まれていて、時間間隔と出来高という 2 変数の対数尤度になっている。一方、表 2 の対数尤度は、マークのない一変量 Hawkes 過程における時間間隔という 1 変数だけの対数尤度である。両者を比較するためには、表 3 の対数尤度から出来高に関する対数尤度 $-\sum_{i=1}^n \left( \log \kappa + \frac{x_i}{\kappa} \right)$ を計算して差し引く必要がある。出来高の対数尤度は-3,677.4338 であったから、この分を考慮した対数尤度は-5,246.9658 である。<sup>11</sup> したがって、マーク無しの対数尤度-5,252.9525 よりも大きいことが分かる。この対数尤度から計算した AIC は  $10,493.9316 + 2 \times 4 = 10,501.9316$  であり、マーク無しの場合の 10,511.905 と比べて小さくなることが分かる。したがって、モデルとしてはマーク付きの方が適切であると考えられる。

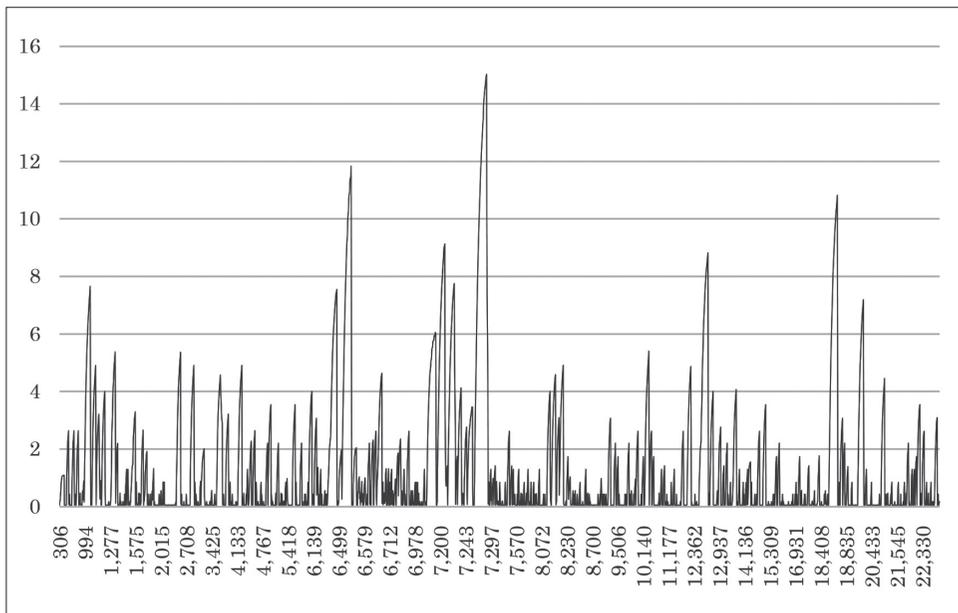
<sup>11</sup> パラメータの最尤推定に当たっては出来高に大きな差異があったため、対数変換した出来高を用いて推定した。したがって、 $y_i = \ln(VOL_i)$  の確率密度関数は以下のとおりである。

$$p(y_i) = \sum_{i=1}^n \frac{1}{\kappa} \exp \left( y_i - \frac{1}{\kappa} \exp(y_i) \right)$$

この密度関数を用いて対数尤度を計算した。

(2) 強度関数の推定

マーク付き Hawkes 過程のパラメータを最尤推定した結果を利用して、2019年2月1日の強度関数を計算した。ただし、寄り付きと大引けを除外している。横軸に寄り付きからの経過時間（秒）を用いて、その結果を示したのが図1である。図1から、この営業日は取引開始から7,284秒後、つまり10時46分頃に強度が高まっていることが読み取れる。また、6,503秒後（10時33分頃）と18,755秒後（13時57分頃）にも高くなっていることが読み取れる。



※横軸は寄り付きからの経過時間（単位は秒）である。8時50分以降を示してある。

図1： 2019年2月1日の金先物（期先物）の強度関数の推移

6. 結論

本稿では、点過程、そして Hawkes 過程の説明を最初に行った。その後、東京金先物取引のティックデータに Hawkes 過程を応用して、パラメータを最尤推定した。推定結果は表2と表3に記したとおりである。

一変量 Hawkes 過程を推定した結果、 $\mu$  の最尤推定値は 0.0528 であり、取引発生からの経過時間に関係なく、強度は少なくとも  $\mu$  (0.0528) だけはあることが分かった。そして、 $\alpha$  と  $\beta$  の最尤推定値は 1.1383 と 2.2282 である。 $\beta$  が正值であることから、取引から時間が経つにつれて発生確率が減少することが分かった。また、取引発生によって強度が  $\alpha$  (1.1383) だけ上昇する

ことも分かった。

出来高をマークとする、マーク付き一変量 Hawkes 過程として ETAS モデルを推定した結果、 $\mu$ 、 $\alpha$  と  $\beta$  の最尤推定値はそれぞれ 0.0528、1.0012 と 2.2345 であり、マーク無しの場合とはほぼ変わらなかった。 $\gamma$  の最尤推定値は 0.1352 であった。出来高の係数が正值なので、出来高が大きくなると取引の発生確率が上がることになる。

マーク付き一変量 Hawkes 過程 (ETAS モデル) の対数尤度は、出来高に関する対数尤度を考慮すると -5,246.9658 であったから、マーク無しの対数尤度 -5,252.9525 よりも大きいことが分かる。この対数尤度から計算した AIC は 10,501.9316 であり、マーク無しの場合の 10,511.905 と比べて小さくなることが分かる。したがって、モデルとしてはマーク付きの方が適切であると考えられる。

本稿では、一変量 Hawkes 過程を扱った。今後は多変量 Hawkes 過程を用いて分析を行う予定である。その際には、板情報などを利用した分析を行いたいと考えている。

## 参考文献

- [1] Bacry E., Mastromatteo I., Muzy J.F., 2015, "Hawkes processes in finance," *Market Microstructure and Liquidity*, 1(1).
- [2] Bowsher, C.G., 2007, "Modelling security market events in continuous time: Intensity based, multivariate point process models," *Journal of Econometrics*, vol.141, pp.876-912.
- [3] Daley, D.J. and D. Vere-Jones, 2003, *An Introduction to the theory of point processes*, Springer.
- [4] Embrechts, P. Liniger, T., Lin, L., 2011, "Multivariate Hawkes process: An application to financial data," *Journal of Applied Probability*, Vol.48, pp.367-378.
- [5] Hawkes, A.G., 1971, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, Vol.58(1), pp.83-90.
- [6] Hawkes, A.G., 1971, "Point spectra of some mutually exciting point processes," *Journal of Royal statistical Society B* Vol.33, pp.438-443.
- [7] Ogata, Y., 1978, "Asymptotic behavior of maximum likelihood estimators for stationary point process," *Ann. Inst. Statist. Math.* vol.30(2)A, pp.243-261.
- [8] Ogata, Y., 1981, "On Lewis's simulation method for point process," *IEEE Transaction on Information Theory*, vol.IT-27(1), pp.23-31.
- [9] Ogata, Y., Akaike, H., Katsura, K., 1982, "The application of linear intensity models to the investigation causal relations between a point process and another stochastic process," *Ann. Inst. Statist. Math.*, vol.34, pp.373-387.

- [10] 尾形良彦,1998,「点過程モデル」,尾崎統・北川源四郎編,『時系列解析の方法』所収,朝倉書店,pp.168-179.
- [11] 尾形良彦,2008,「地震活動の統計科学」,小西貞則・国友直人編,『21世紀の統計科学Ⅱ』所収,東京大学出版会,pp.3-41.
- [12] 近江崇宏,2018,“点過程モデリングを用いた金融時系列データの解析” 『生産研究』,第70巻第3号,pp.151-155.
- [13] 近江崇宏・野村俊一,2019,『点過程の時系列解析』,共立出版.
- [14] Ozaki,T.,1979,“Maximum likelihood estimation of Hawkes' self-exciting point process,” *Ann.Inst.Statist.Math.*,vol.31,pp.145-155.
- [15] 佐藤正崇・今井潤一,2020,“マーク付き多次元 Hawkes 過程を用いた高頻度注文板データの分析,ジャフィー・ジャーナル,第18巻,pp.63-85.