Hawkes型計数時系列モデル

小山　慎介　　モデリング研究系　准教授

要旨

本研究では，離散時間の計数時系列に対してHawkes過程に類似する時系列モデルを提案する．特にHawkes過程と分岐過程の対応に着目し，同様な性質を計数時系列モデルで構築する．このときに，計数時系列モデルの構成をHawkes過程とパラレルに展開し，両者を対比する．

Hawkes過程

時刻$t \in \mathbb{R}^+$までに生起したイベント数を$N(t)$とし，対応するイベント発生時刻をt_i $(i=1,2,\ldots)$とする．時刻tまでのイベント履歴$H_t = \{t_i : t_i < t\}$が与えられたもとで次の瞬間にイベントが発生する確率が，条件付き強度関数$\lambda^*(t)$を用いて

$$P\{N(t+\Delta t) - N(t) = 1 | \mathcal{H}_t\} = \lambda^*(t)\Delta t + o(\Delta t)$$
$$P\{N(t+\Delta t) - N(t) > 1 | \mathcal{H}_t\} = o(\Delta t)$$

で与えられるとする．Hawkes過程の条件付き強度関数は以下で与えられる．

$$\lambda^*(t) = \mu + \int_0^t \phi(t-u)dN(u)$$

期間$(0,T]$にnのイベントが時刻$\{t_1,\ldots,t_n\}$に発生する同時確率密度関数は

$$p(t_1,\ldots,t_n) = \left\{\prod_{i=1}^n \lambda^*(t_i)\right\} \exp\left\{-\int_0^T \lambda^*(t)dt\right\}$$

分枝過程

i番目のイベントを引き起こした親イベントの番号を$z_i \in \{0,1,\ldots,i-1\}$とする．$z_i = 0$の場合は親イベントを持たないとする．イベント発生時刻と親イベントの番号を合わせた時系列$\{(t_i,z_i)\}_{i\geq 1}$は以下のルールに従って生成されるとする．

(a) 親を持たないイベント$\{(t_i,z_i) : z_i = 0\}$は強度μの一様Poisson過程に従って発生する．

(b) $j(\geq 1)$番目のイベントを親に持つイベント$\{(t_i,z_i) : z_i = j\}$は強度$\phi(t_i - t_j)$の非一様Poisson過程に従って発生する．

イベント時系列全体は(a)と(b)の重ね合わせで与えられるとすると，期間$(0,T]$における$\{(t_i,z_i)\}_{i=1}^n$の同時確率密度関数は以下で与えられる．

$$p(t_1,\ldots,t_n,z_1,\ldots,z_n) = \left\{\prod_{i=1}^n \psi_{iz_i}\right\} \exp\left\{-\mu T - \sum_{j=1}^n \int_{t_j}^T \phi(t-t_j)dt\right\}$$

ここで

$$\psi_{ij} = \begin{cases} \mu & (j=0) \\ \phi(t_i - t_j) & (j=1,\ldots,i-1). \end{cases}$$

とおいた．これを親イベント$\{z_1,\ldots,z_n\}$について周辺化すると

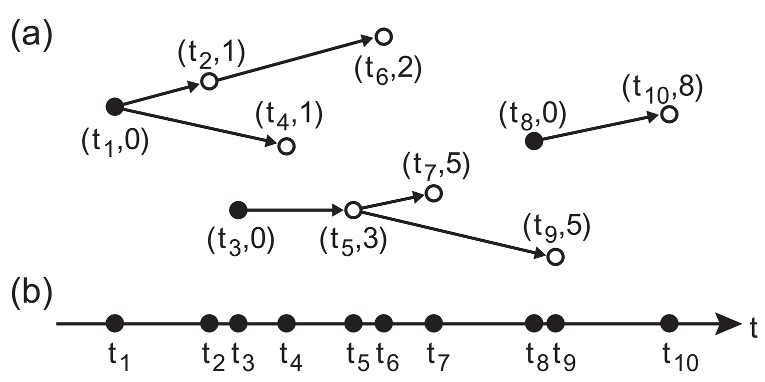
$$p(t_1,\ldots,t_n) = \sum_{z_1=0}^{0}\sum_{z_2=0}^{1}\cdots\sum_{z_n=0}^{n-1} p(t_1,\ldots,t_n,z_1,\ldots,z_n)$$
$$= \left\{\prod_{i=1}^n \lambda^*(t_i)\right\} \exp\left\{-\int_0^T \lambda^*(t)dt\right\}$$

となり，Hawkes過程が導かれる．すなわちイベント間の因果関係は観測されず，イベント発生時刻だけ観測されたものがHawkes過程であると解釈することができる．

また，イベント時刻列$\{t_1,\ldots,t_i\}$が与えられたもとでのz_iの条件付き確率は

$$p(z_i \mid t_1,\ldots,t_i) = \frac{\psi_{iz_i}}{\lambda^*(t_i)}$$

と求められる．



Hawkes型計数時系列モデル

時点$i(=1,2,\ldots,)$におけるイベント数をn_iとする．期待値$\lambda = E(n)$をパラメータに持つ**加法的な確率分布**$p(n;\lambda)$を用いて，過去のイベント数$H_{i-1} = \{n_1,\ldots,n_{i-1}\}$が与えられたもとで$n_i$の条件付き確率が

$$P(n_i \mid H_{i-1}) = p(n_i;\lambda_i)$$

で与えられるとする．ここで期待値パラメータは以下で与えられる．

$$\lambda_i = \mu + \sum_{j=1}^{i-1} \phi_{i-j} n_j$$

T時点までの時系列$\{n_1,\ldots,n_T\}$の同時確率分布は

$$P(n_1,\ldots,n_T) = \prod_{i=1}^T P(n_i \mid H_{i-1}) = \prod_{i=1}^T p(n_i;\lambda_i)$$

親イベントの割り当て

時点iのイベント数n_iのうち，過去の時点$j(<i)$のイベントに引き起こされたイベント数をy_{ij}とし，親を持たないイベント数をy_{i0}とする．$Y_i \equiv \{y_{i0},\ldots,y_{i,i-1}\}$の確率分布について以下を仮定する．

(a) 親を持たないイベント数y_{i0}は期待値がμの確率分布$p(y_{i0};\mu)$に従う．

(b) 過去の時点$j(<i)$のイベントに引き起こされたイベント数y_{ij}は期待値が$\phi_{i-j}n_j$の確率分布$p(y_{ij};\phi_{i-j}n_j)$に従う．

過去のイベントH_{i-1}が与えられたもとで$y_{i0},\ldots,y_{i,i-1}$は互いに条件付き独立であるとすると，時点Tまでの時系列の同時確率分布は以下で与えられる．

$$P(Y_1,\ldots,Y_T) = \prod_{i=1}^T \prod_{j=0}^{i-1} p(y_{ij};\psi_{ij})$$

ここで

$$\psi_{ij} = \begin{cases} \mu & (j=i) \\ \phi_{i-j}n_j & (j=1,\ldots,i-1) \end{cases}$$

とおいた．これを$\{Y_1,\ldots,Y_T\}$について周辺化すると，確率分布の**加法性**により

$$P(n_1,\ldots,n_T) = \sum_{Y_1\in\mathcal{Y}_1}\cdots\sum_{Y_T\in\mathcal{Y}_T} P(Y_1,\ldots,Y_T) = \prod_{i=1}^T p(n_i;\lambda_i)$$

となり，もとの計数時系列モデルが導かれる．

また，$H_i = \{n_1,\ldots,n_i\}$が与えられたもとでのY_iの条件付き確率は

$$P(Y_i \mid H_i) = \frac{\prod_{j=1}^i p(y_{ij};\psi_{ij})}{p(n_i;\lambda_i)}$$

と求められる．

例1: Poisson分布

$$p(n;\lambda) = \frac{\lambda^n}{n!}e^{-\lambda} \quad (n=0,1,\ldots)$$

Y_iの条件付き確率は多項分布:

$$P(Y_i \mid H_i) = \frac{n_i!}{\prod_{j=1}^i y_{ij}!}\prod_{j=1}^i \left(\frac{\psi_{ij}}{\lambda_i}\right)^{y_{ij}}$$

例2: 負の二項分布

$$p(n;\lambda,\rho) = \frac{\Gamma(n+\frac{\lambda}{\rho})}{\Gamma(n+1)\Gamma(\frac{\lambda}{\rho})}\left(\frac{\rho}{1+\rho}\right)^n\left(\frac{1}{1+\rho}\right)^{\frac{\lambda}{\rho}} \quad (n=0,1,\ldots)$$

Y_iの条件付き確率はディリクレ多項分布:

$$P(Y_i \mid H_i,\rho) = \frac{\Gamma(n_i+1)\Gamma(\frac{\lambda_i}{\rho})}{\Gamma(n_i+\frac{\lambda_i}{\rho})}\prod_{j=1}^i \frac{\Gamma(y_{ij}+\frac{\psi_{ij}}{\rho})}{\Gamma(y_{ij}+1)\Gamma(\frac{\psi_{ij}}{\rho})}$$