

多変量点過程の統計的推測とその多変量 Neyman-Scott 点過程への応用

塩谷天章

東京大学大学院 数理科学研究科

研究概要

- 点過程は、ランダムなイベントの発生時刻の数理モデル。
 - ▶ 例: 地震の発生、金融商品の取引、SNS への投稿など
- モデルの分布 (=データ発生構造) が有限次元のパラメータ $\theta \in \mathbb{R}^p$ で特定されるとして、観測データから真のパラメータを推定したい。
- 一般的な推定方法として最尤法がある。しかし、例えば多変量 Neyman-Scott 点過程 (MNSP) モデルでは、最尤法の計算コストが大きく、大規模データへの適用が大変。
- 1変量モデルに対しては、最尤法の代替手法として、擬似尤度最大化推定量 (QMLE) が提案され、理論的性質が調べられている。
- 本研究では、多変量モデルに対して QMLE を拡張し、その理論的性質を調べた。特にこの推定法は MNSP モデルに対して適用することができる。
- 結果として、MNSP モデルを、理論保証つきでビッグデータに対して手軽に適用できるようになった。また理論を一般的な設定で整備したため、MNSP 以外のモデルに対しても本研究の手法は適用可能である。

多変量 Neyman-Scott 点過程モデル

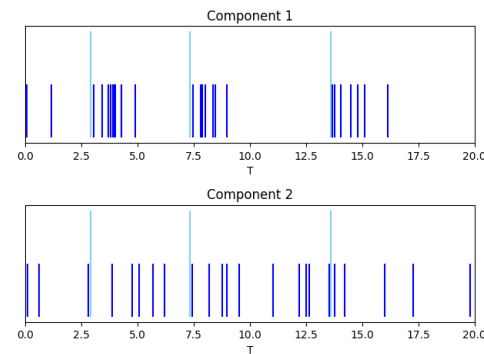


Figure: 2変量 MNSP のサンプルパス。
青線がイベントの起きた時刻を表す。

- 多変量点過程モデルの1種。
- 観測されない共通の「親」を起点に、観測可能な「子」が生成される。
- 子の個数や間隔の分布は成分ごとに異なる。

多変量 Neyman-Scott 点過程モデルは、以下のような利点をもつ。

- データに直接現れない量 (=親) からの影響を直接モデリングできる。
 - ▶ 例: 株取引が行われた時刻 (子)-企業に関するニュース等 (親)
- 子と親の間隔の分布を自由に決めることができ、柔軟に現象をモデリングできる。
 - ▶ 例: 指数分布、ガンマ分布、...

実際に、地震データや SNS データのモデリングに活用されている。(Hong and Shelton (2022))

MNSP モデルの正確な定義

MNSP モデルの第 j 成分は以下の式で与えられる。

$$N_j(I) = \sum_{c \in \mathcal{C}} \sum_{m=1}^{M_j(c)} \delta_{c+d_j(c,m)}(I), \quad I \in \mathcal{B}(\mathbb{R})$$

親 \mathcal{C} 強度 λ の齊次ポアソン過程

子の個数 $M_j(c)$ 平均 σ_j のポアソン分布からの i.i.d.

子の間隔 $d_j(c, m)$ 確率密度関数 $f_j(\cdot; \tau_j)$ をもつ分布からの i.i.d.

モデルの分布は、パラメータ $\theta = (\lambda, \{\sigma_j\}_{j=1}^J, \{\tau_j\}_{j=1}^J)$ によって決定される。

問題設定

- $\{N_j\}_{j=1}^J$: \mathbb{R} 上の定常多変量点過程
 - ▶ 定常性: 結合分布が並行移動不変
 - ▶ 動機は MNSP モデルにあるが、MNSP モデルに限らない一般的な設定で推定法の理論を整備した。
- 各成分の 1 次・2 次モーメント密度が $\theta \in \mathbb{R}^p$ でパラメトライズされているとする。

$$\lambda_1^j(\theta) = \lim_{|dx| \rightarrow 0} \frac{E[N_j(dx)]}{|dx|} \quad (const)$$

$$\lambda_2^j(y-x; \theta) = \lim_{|dx|, |dy| \rightarrow 0} \frac{E[N_j(dx)N_j(dy)]}{|dx||dy|}, \quad (x, y) \in \mathbb{R}^2, x \neq y$$

推定量の構成

$[0, T]$ 上の観測データとパラメータから計算される目的関数 (擬似尤度) $\mathbb{H}_T(\theta)$ を最大化することで推定量 (quasi-maximum likelihood estimator, QMLE) を得る。

$$\hat{\theta}_T \in \operatorname{argmax}_{\theta} \mathbb{H}_T(\theta)$$

本研究では 3 種類の擬似尤度を取り扱う。

- composite likelihood \mathbb{H}^C : 近い観測点のペア (x, y) を、正規化した 2 次のモーメント密度関数からの独立なサンプルと考えて尤度を書き下す。

$$\mathbb{H}_T^C(\theta) = \sum_{j=1}^J \sum_{(x,y), x \neq y} \log \left(\frac{\lambda_2^j(y-x; \theta)}{\iint \lambda_2^j(v-u; \theta) 1_{|v-u| \leq r} dudv} \right) 1_{|y-x| \leq r}$$

- Palm likelihood \mathbb{H}^P : 各観測点 x に対し、近い点 y との差 $D_x = \{y-x; |y-x| \leq r, x \neq y\}$ を、強度 $\lambda_o(u) = \lambda^2(u)/\lambda^1$ の非齊次ポアソン過程と思って、尤度を書き下す。 $\lambda_o(u-x)$ は、1 点 x を与えたときの $x+u$ における点の発生しやすさと解釈できる。

$$\mathbb{H}_T^P(\theta) = \sum_{j=1}^J \sum_x \left(\sum_{y, y \neq x} \log \lambda_o^j(y-x; \theta) 1_{|y-x| \leq r} - \int_{|u| \leq r} \lambda_o^j(u; \theta) du \right)$$

- \mathbb{H}^W : 2 次のモーメント密度関数で Poisson 過程の尤度を真似る。

$$\mathbb{H}_T^W(\theta) = \sum_{j=1}^J \sum_{(x,y), x \neq y} \log \left(\lambda_2^j(y-x; \theta) \right) 1_{|y-x| \leq r} - \int_r^{T-r} dt \int_{|u| \leq r} \lambda_2^j(u; \theta) du$$

これらの擬似尤度は、1 変量モデルの場合の擬似尤度の成分ごとの和をとって定義した。特に MNSP モデルではモーメント密度関数が明示的に書けるので、擬似尤度は簡単に計算できる。各擬似尤度に基づく推定量を、 $\hat{\theta}_T^C, \hat{\theta}_T^P, \hat{\theta}_T^W$ で書く。

主結果

Theorem

定常な多変量点過程モデルが、正則性や識別可能性等の適切な条件に加え以下の仮定を満たすとす。

- α -mixing 係数が十分速く減少する

このとき、各 QMLE $\hat{\theta}_T^A$ は以下の性質を満たす。

- 強一貫性

$$\hat{\theta}_T^A \xrightarrow{a.s.} \theta_0 \quad (T \rightarrow \infty)$$

- 漸近正規性

$$\sqrt{T}(\hat{\theta}_T^A - \theta_0) \rightarrow^d N(0, \Sigma) \quad (T \rightarrow \infty)$$

- モーメント収束

$$\forall L > 1: \sup_T E[|\sqrt{T}(\hat{\theta}_T^A - \theta_0)|^L] < \infty$$

- α -mixing 係数: 「時間的に離れたサンプルの独立度合い」
 - ▶ 減少の速さに応じて、エルゴード定理・中心極限定理などが出る。
- MNSP モデルでは、 α -mixing 係数を子と親の間隔の分布 f_j の裾確率で上から評価でき、主結果を適用可能。

先行研究との比較

- 先行研究では、1 変量の場合に強一貫性と漸近正規性が示されているのみである。(Guan (2006), Prokešová & Jensen (2013))
- モーメント収束は、Yoshida (2011) の PLD 不等式によって証明される。

MNSP モデルへの適用

QMLE $\hat{\theta}$ を求めるには高次元空間上の数値最適化が必要。

- J が大きいときに数値計算が大変。

$$\hat{\theta} \in \operatorname{argmax}_{\theta=(\lambda, \{\sigma_j\}_j, \{\tau_j\}_j)} \sum_{j=1}^J \mathbb{H}^j(\lambda, \sigma_j, \tau_j)$$

そこで、MNSP モデル限定の方法ではあるが、高次元空間上の最適化を回避できる**適応的推定量**と、それをさらに発展させた**適応的ワンステップ推定量**を提案し、主結果を応用して、これらの推定量の強一貫性・漸近正規性・モーメント収束も証明した。あらかじめ平均レートを推定して擬似尤度に代入することで、成分ごとに最適化を行えるというのが基本的なアイデアである。

アルゴリズム

- 1 成分ごとに平均レート $\lambda\sigma_j$ を推定

$$\widehat{\lambda\sigma_j} = T^{-1}N_j([0, T])$$

- 2 成分ごとの擬似尤度 $\mathbb{H}^j(\lambda, \sigma_j, \tau_j)$ に $\widehat{\lambda\sigma_j}$ を挿入して推定

$$(\{\hat{\sigma}_j^0\}_{j=1}^J, \{\hat{\tau}_j^0\}_{j=1}^J) \underset{\text{j-wise maximization}}{\operatorname{maximize}} \sum_{j=1}^J \mathbb{H}^j\left(\frac{\widehat{\lambda\sigma_j}}{\sigma_j}, \sigma_j, \tau_j\right)$$

- 3 全ての成分での λ の推定量の平均をとり、まとめて適応的推定量 $\hat{\theta}^0$ を得る。

$$\hat{\theta}^0 = \left(\frac{1}{J} \sum_{j=1}^J \frac{\widehat{\lambda\sigma_j}}{\hat{\sigma}_j^0}, \{\hat{\sigma}_j^0\}_{j=1}^J, \{\hat{\tau}_j^0\}_{j=1}^J\right)$$

- 4 さらに全体の擬似尤度 $\mathbb{H}(\theta)$ を用いて適応的ワンステップ推定量 $\hat{\theta}^1$ を得る。

$$\hat{\theta}^1 = \hat{\theta}^0 - (\partial_{\theta}^2 \mathbb{H}(\hat{\theta}^0))^{-1} \partial_{\theta} \mathbb{H}(\hat{\theta}^0)$$

数値実験

MNSP モデルに対しては9種類の推定量

- 同時推定量 (通常の QMLE) $\hat{\theta}^C, \hat{\theta}^P, \hat{\theta}^W$
- 適応的推定量 $\hat{\theta}^{0C}, \hat{\theta}^{0P}, \hat{\theta}^{0W}$
- 適応的ワンステップ推定量 $\hat{\theta}^{1C}, \hat{\theta}^{1P}, \hat{\theta}^{1W}$

を与えたことになる。(実は、うち2つ $\hat{\theta}^C$ と $\hat{\theta}^{1C}$ は識別可能条件を満たさないため使用不能であり、実質7種類。) これらの推定量のパフォーマンスを数値実験により比較した。

実験の設定

- $J = 2, T = 1000, r = 1.0, \#MC = 1000$
- 指数分布 $f_j(x; \tau_j) = 1_{x \in (0, \infty)} \tau_j e^{-\tau_j x}$ (この場合は定理の仮定を全て満たす)
- 推定すべきパラメータ $(\lambda, \sigma_1, \sigma_2, \tau_1, \tau_2)$
- 真値 $(\lambda_0, \sigma_{01}, \sigma_{02}, \tau_{01}, \tau_{02}) = (0.25, 10, 20, 25, 50)$

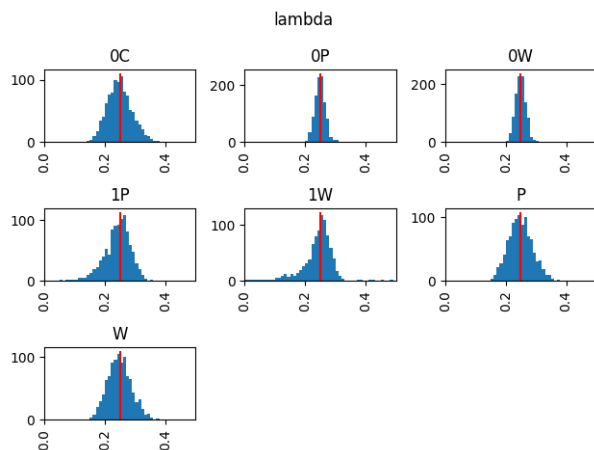


Figure: λ の推定値の分布

- 例えば λ の推定値の分布は左図のようになった。
- どの推定量も真値 $\lambda_0 = 0.25$ の周りに分布しており、漸近正規性も読み取れる。

数値実験

各パラメータに対する各推定量の平均二乗誤差を比較したところ、Palm 型擬似尤度 \mathbb{H}^P に基づく適応的推定量 $\hat{\theta}^{0P}$ が最もパフォーマンスが良かった。

	λ	σ_1	σ_2	τ_1	τ_2
$\hat{\theta}^P$	0.0400	0.371	0.565	1.39	2.25
$\hat{\theta}^W$	0.0399	0.849	1.67	1.38	2.26
$\hat{\theta}^{0C}$	0.0410	1.88	3.16	1.55	2.25
$\hat{\theta}^{0P}$	0.0172	0.386	0.527	1.45	2.17
$\hat{\theta}^{0W}$	0.0172	0.387	0.529	1.44	2.17
$\hat{\theta}^{1P}$	0.0490	0.483	0.849	1.44	2.35
$\hat{\theta}^{1W}$	0.481	13.1	26.7	1.38	11.5

Table: 平均二乗誤差。各列で最も誤差が小さいものが太字になっている。

今後の課題

- より多様な設定での数値実験。
- 各推定法の誤差の理論的比較。
- 大規模データ (e.g. 高頻度金融取引データ) への応用。

参考文献

- Hong, C., & Shelton, C. (2022). Deep Neyman-Scott Processes. In International Conference on Artificial Intelligence and Statistics (pp. 3627-3646). PMLR.
- Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. Annals of the Institute of Statistical Mathematics, 63(3), 431-479.
- Prokešová, M., & Jensen, E. B. V. (2013). Asymptotic Palm likelihood theory for stationary point processes. Annals of the Institute of Statistical Mathematics, 65(2), 387-412.
- Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. Journal of the American Statistical Association, 101(476), 1502-1512.