

数理科学実践研究レター 2019–20 December 04, 2019

確率微分方程式モデルによるセンサーデータ解析

by

鈴木 拓海



UNIVERSITY OF TOKYO
GRADUATE SCHOOL OF MATHEMATICAL SCIENCES
KOMABA, TOKYO, JAPAN

確率微分方程式モデルによるセンサーデータ解析

鈴木拓海¹ (東京大学大学院数理科学研究科)

Takumi Suzuki (Graduate School of Mathematical Sciences, The University of Tokyo)

概要

センサーデータの解析では、多くの場合、離散時間確率変数列を用いたモデルが考えられている。しかし、各センサーを等間隔で同時刻観測するという仮定は強く、そのことを暗に仮定している従来のモデルでは、現実の問題に正しく適用できるか不明である。そこで、本論文では、センサーデータを連続時間確率過程とみなし、従来の離散時間確率変数列を用いたモデルの拡張になっていることや、必ずしも等間隔・同時観測では無い場合にも適用できる結果をまとめた。また、スパース推定の結果を用いることにより、高次元の解析にも応用できることを紹介する。

1 はじめに

近年、IT やセンシング技術の発展に伴い、大量のデータの収集が容易になった。そのことにより、様々な分野でデータの活用ということが重要になってきている ([1])。そこで、様々な解析手法が提案されているが、その多くはセンサーデータを離散時間確率変数列とみなしたモデルを考えている。離散時間確率変数列は理論的な性質が多く知られており扱いやすい面がある一方、暗に各センサーを等間隔・同時観測していることを仮定してしまっている。実用上、センサーデータは超高次元にもなり得る中でこの仮定は非常に強く、もし、実際には等間隔・同時観測していないデータをこの理論に当てはめてしまうと、当然ながら誤った結果を導いてしまう。そこで、我々は、データの種類を連続時間確率過程に拡張し、その下でデータの構造について調べた。この手法はこれまであまり議論されてこなかったことではあるが、従来の手法に対する一つの拡張として考えられるだけでなく、今後、センサーデータを高頻度観測できるようになった時に大きな意味を持つ。

センサーデータ解析においては、異常検知が最も重要なテーマの一つであるが、異常を検出する手がかりとなるのが、センサーデータ間の相関関係である。グラフィカル・ガウシアン・モデルを用いた、センサーデータ間の相関関係を調べる研究は、[2] などにもあるが、本論文では特に、確率微分方程式 (SDE : Stochastic differential equation) で表されるモデルについて調べることで、より一般的な状況を考えて。

また、データ間の相関関係を調べるときには、多くの場合、分散共分散行列の逆行列である精度行列の推定を行う。精度行列を用いる主な理由は、ある 2 変数の相関に注目した時に、それ以外の変数の影響を無視するためである。しかし、変数の数を n とした時に、精度行列を求めるためには、パラメータが $n(n+1)/2$ 必要になる。そこで、スパース推定の技術を応用する。このスパース推定について、感覚的に言うと、相関係数が 0 に近いものは真に 0 であるとみなすことで、超高次元の推定を限られたデータ数で可能にするものである。本来は、線形回帰モデルにおいて考えられていた手法だが、今回考えるモデルを含めた、広いモデルに対して適用できることも併せて紹介する。

2 連続時間確率過程モデル

本節では連続時間確率過程モデルとして、確率微分方程式を用いたモデルを考える。また、それに対応した推定手法としてよく知られている擬似尤度解析を用いて推定量を構成する。擬似尤度解析について、詳しくは [3] 参照。

センサーの数を M 個とする。このとき、センサーデータを M 次元連続時間確率過程 $X = (X^i)_{i=1, \dots, M}$ と考え、以下の SDE に従うものとする。

$$dX_t = a(X_t, \theta^2)dt + b(X_t, \theta^1)dW_t, \quad X_0 = x_0 \in \mathbb{R}^d. \quad (1)$$

¹takumi@ms.u-tokyo.ac.jp

ここで、 $\theta = (\theta^1, \theta^2)$ は推定したいパラメータ、 $W = (W_t)_{t \in [0, \infty)}$ は M 次元標準ブラウン運動である。このセンサーデータ X から得られるデータセットを $\mathcal{D} = \{(X_{t_n^i}^i)_{n=1, \dots, N_i}; i = 1, \dots, M\}$ あるいは $\mathcal{D}' = \{(X_{t_n^i}^i - X_{t_{n-1}^i}^i)_{n=2, \dots, N_i}; i = 1, \dots, M\}$ とおく。ここで、 $X^i = (X_t^i)_{t \in [0, \infty)}$ は i 番目のセンサーデータ、 N_i を i 番目のセンサーのデータサイズ、 $0 \leq t_1^i < t_2^i \cdots < t_{N_i}^i$ を i 番目のセンサーの観測時刻である。データセット \mathcal{D}' はデータセット \mathcal{D} の差分をとっただけのものであるが、データが等間隔・同時刻観測であって独立（同分布）確率変数列である場合はこのような形をしている。推定手法に影響は無いため、以下では主に \mathcal{D} を用いる。

注意 1 $a \equiv 0, b(x, \theta^1) \equiv \Lambda^{-\frac{1}{2}} \in \mathbb{R}^{M \times M}$ であり、等間隔・同時刻観測、つまり、データサイズ N_i や観測時刻 $(t_n^i)_{n=1, \dots, N_i}$ が共通の時、データセット \mathcal{D}' は、多変量正規分布 $N_M(0, \Lambda^{-1})$ に従う確率変数列を表す。

今回はセンサー間の相関に興味があるので、簡単な場合として、以下の状況を考える：

$$a \equiv 0, \quad b(x, \theta^1) \equiv \Lambda^{-\frac{1}{2}} \in \mathbb{R}^{M \times M}. \quad (2)$$

2.1 同期観測

ここでは、同期観測されている場合を考える。この場合、 $(t_n^i)_n$ および N_i は共通に取れるため、それぞれ $(t_n)_n$ および N とおく。このとき、擬似対数尤度関数 $\mathbb{H}_N^{(1)}(\Lambda)$ を次のように定義する ([3])。

$$\mathbb{H}_N^{(1)}(\Lambda) = -\frac{1}{2} \sum_{k=2}^N \left\{ \log \det \Lambda^{-1} + N(\Delta_k X)' \Lambda (\Delta_k X) \right\}$$

また、 $\mathbb{H}_N^{(1)}(\Lambda)$ を最大にするものとして、擬似最尤推定量 $\tilde{\Lambda}^{(1)}$ が得られる：

$$\mathbb{H}_N^{(1)}(\tilde{\Lambda}^{(1)}) = \sup_{\Lambda} \mathbb{H}_N^{(1)}(\Lambda).$$

ここで、 $\Delta_k X$ は X の差分を表しており、 $\Delta_k X = X_{t_k} - X_{t_{k-1}}$ で与えられる。

2.2 非等間隔・非同期観測

ここでは、非同期・非等間隔な観測について考える。簡単のため、今回はセンサーデータが2個の場合についてのみ議論する。このとき、推定すべき精度行列 Λ は 2×2 行列となるので、

$$\Lambda = \begin{bmatrix} \lambda_1 & \lambda_0 \\ \lambda_0 & \lambda_2 \end{bmatrix}$$

とおく。つまり、推定したいパラメータは $[\lambda_1, \lambda_2, \lambda_0]' \in \mathbb{R}^3$ となる。また、時間を表す変数 T を考え、各 $i = 1, 2$ に対して $N_i = N_i(T)$ とし、 $T \rightarrow \infty$ の時、 $N_i(T) \rightarrow \infty$ とする。このとき、次の擬似対数尤度関数 $\mathbb{H}_T^{(2)}(\Lambda)$ を考える。

$$\mathbb{H}_T^{(2)}(\Lambda) = -\frac{1}{2} \begin{bmatrix} \Delta^1 X \\ \Delta^2 X \end{bmatrix}' S^{-1} \begin{bmatrix} \Delta^1 X \\ \Delta^2 X \end{bmatrix} - \frac{1}{2} \log \det S$$

ここで、各 $i = 1, 2$ に対し、

$$\Delta^i X = \begin{bmatrix} \Delta_1^i X \\ \vdots \\ \Delta_{N_i}^i X \end{bmatrix} \in \mathbb{R}^{N_i}, \quad \Delta_n^i X = (t_n^i - t_{n-1}^i)^{-1/2} (X_{t_n^i}^i - X_{t_{n-1}^i}^i), \quad n = 1, \dots, N_i$$

であり, $S = [S_{ij}]_{i,j=1,2} \in \mathbb{R}^{(N_1+N_2) \times (N_1+N_2)}$ とおくと, $i = 1, 2$ に対して,

$$S_{ii} = \text{diag} \left(\left[(\lambda_i^2 + \lambda_0^2)(t_n^i - t_{n-1}^i) \right]_{n=1, \dots, N_i} \right) \in \mathbb{R}^{N_i \times N_i},$$

$$S_{12} = S'_{21} = 2\lambda_0^2 \left[\frac{0 \vee (t_n^1 \wedge t_m^2 - t_{n-1}^1 \vee t_{m-1}^2)}{\sqrt{(t_n^1 - t_{n-1}^1)(t_m^2 - t_{m-1}^2)}} \right]_{n=1, \dots, N_1, m=1, \dots, N_2} \in \mathbb{R}^{N_1 \times N_2}$$

を表すものとする. このとき, 擬似最尤推定量 $\tilde{\Lambda}^{(2)}$ を擬似対数尤度関数 $\mathbb{H}_T^{(2)}(\Lambda)$ を最大化するものとして定義する:

$$\mathbb{H}_T^{(2)}(\tilde{\Lambda}^{(2)}) = \sup_{\Lambda} \mathbb{H}_T^{(2)}(\Lambda).$$

注意 2 例えば, T は観測を始めてからの時間を表す. この時, $N_i = N_i(T)$ は, i 番目のセンサーの時刻 T までの観測点の個数に相当する.

3 結果

推定したいパラメータ Λ の真の値を Λ^* とおく. 次に, 正の定数 c および C に対して, $\Theta_{c,C}$ を次で定義する:

$$\Theta_{c,C} = \{ \Lambda \in \mathbb{R}^{M \times M}; \Lambda \text{ は対称行列で任意の } u \in \mathbb{R}^M \text{ に対して, } c|u|^2 < u' \Lambda u < C|u|^2 \}$$

この時, 真値 Λ^* が $\Theta_{c,C}$ の内点になるように c および C をとり固定する. これ以降はパラメータ空間として $\Theta_{c,C}$ を考える. このとき, 適当な条件のもとで, 次が得られる.

定理 1 ([3] の Theorem 13 および [4] の Theorem 1, 2)

1) 一致性:

(a) $\tilde{\Lambda}^{(1)} \rightarrow^p \Lambda^*$.

(b) $\tilde{\Lambda}^{(2)} \rightarrow^p \Lambda^*$.

2) 漸近正規性:

(a) $\sqrt{N}(\tilde{\Lambda}^{(1)} - \Lambda^*) \rightarrow^d \text{normal}$.

(b) $b_T \rightarrow \infty (T \rightarrow \infty)$ となるような適当な数列 $(b_T)_T$ に対して, $b_T(\tilde{\Lambda}^{(2)} - \Lambda^*) \rightarrow^d \text{normal}$.

3) L^∞ -有界性:

(a) $(\sqrt{N}(\tilde{\Lambda}^{(1)} - \Lambda^*))_N$ は L^∞ -有界である.

(b) $(b_T(\tilde{\Lambda}^{(2)} - \Lambda^*))_T$ は L^∞ -有界である.

ここで, \rightarrow^p および \rightarrow^d は, それぞれ確率収束および分布収束を意味する. また, 確率変数列 $(X_n)_n$ が L^∞ -有界であるとは, 任意の $p > 1$ に対して, $\sup_n E[|X_n|^p] < \infty$ である時にいう. 証明は各論文の定理の条件を調べれば良い. 本論文では, SDE モデル (1) において, (2) を仮定しているので, 容易に示すことができる.

4 スパース推定

本章では、上で述べた擬似最尤推定にも適用できるスパース推定について述べる。なお、前章の仮定は満たしているものとし、変数 S を、2.1 節の場合を考えると $S = N$ 、2.2 節の場合を考えると $S = T$ とおく。また、真値 Λ^* のいくつかの成分が 0 であることを仮定する。

まず、上記の擬似最尤推定量 $\tilde{\Lambda}$ を用いて、新しい目的関数 Q_S を構成する。簡単のため、パラメータ Λ はベクトルとみなす： $\Lambda = [\lambda_j]_j \in \mathbb{R}^{M(M+1)/2}$ 。例えば、 $\Lambda = [\lambda_{ij}]_{i,j} \in \Theta_{c,C}$ に対して、 $\lambda_{ij} (i \geq j)$ を添字 i, j について辞書式順序により並べ替えたものを考えれば良い。この時、目的関数 Q_S を次で定義する：

$$Q_S(\Lambda) = (\Lambda - \tilde{\Lambda})' \hat{G}(\Lambda - \tilde{\Lambda}) + \sum_j \hat{\kappa}_S^j |\lambda_j|^q,$$

ここで、 \hat{G} はある（ランダムな）正定値行列、 $\hat{\kappa} = [\hat{\kappa}^j]_j$ はある（ランダムな）確率変数列、 q は $0 < q \leq 1$ を満たす定数とする。さらに、推定量 $\hat{\Lambda}$ を目的関数 Q_S を最小化するものとして定義する：

$$Q_S(\hat{\Lambda}) = \inf_{\Lambda} Q_S(\Lambda)$$

このとき、 $\hat{\Lambda} = [\hat{\lambda}_j]_j, \Lambda^* = [\lambda_j^*]_j$ とおくと、[5] より、適当な条件のもとで、 $\lambda_j^* = 0$ となるような j に対して、 $\hat{\lambda}_j = 0$ となる確率が、 $S \rightarrow \infty$ の時に、1 に収束する。

5 終わりに

本論文では、従来、離散時間確率過程モデルで考えられることが多かった問題を連続時間確率過程モデルの問題に置き換え、簡単な SDE モデルに関して漸近正規性を持つ推定量の構成や、スパース推定への応用について考察した。本論文で議論したことは、より複雑なモデルにも対応できるが、従来の手法を連続時間確率過程モデルに拡張することが目的であったため、簡単なモデルについてのみ考えた。

今後の課題は、理論面では観測時刻の一般化、実用面ではシミュレーションや実際のセンサーデータを用いた解析を行うことである。観測時刻に関しては、本論文ではかなり限定的な状況であったため、非同期高次元伊藤過程の分散行列の推定問題を扱った [6] などのような、より一般的な状況における理論的考察が望まれる。観測時刻のランダム化については、本論文では触れなかったが、この分野の重大な課題である欠損データの扱いに対応する一つの方法である。また、センサーデータの解析は、主に産業分野での問題であるため、実データに応用できてこそ価値のあるものとなる。

参考文献

- [1] 井手剛. Ibm プロフェッショナル論文スパース構造学習によるセンサー・データの変化点検出と異常解析. *Provision*, No. 65, pp. 71–76, 2010.
- [2] 井手剛ほか. 疎な相関グラフの学習による相関異常の検出. データマイニングと統計数理研究会 (第 9 回).
- [3] Nakahiro Yoshida. Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Annals of the Institute of Statistical Mathematics*, Vol. 63, No. 3, pp. 431–479, 2011.
- [4] Teppei Ogihara and Nakahiro Yoshida. Quasi-likelihood analysis for nonsynchronously observed diffusion processes. *Stochastic Processes and their Applications*, Vol. 124, No. 9, pp. 2954–3008, 2014.

- [5] Takumi Suzuki and Nakahiro Yoshida. Penalized least squares approximation methods and their applications to stochastic processes. *arXiv preprint arXiv:1811.09016*, 2018.
- [6] Jianqing Fan and Donggyu Kim. Structured volatility matrix estimation for non-synchronized high-frequency financial data. *Journal of Econometrics*, Vol. 209, No. 1, pp. 61–78, 2019.