

# 線形確率微分方程式で表される状態空間モデルにおける推定

栗崎 正博

東京大学数理科学研究科

## 状態空間モデルとは？

観測できない潜在的な状態量を含むモデルのこと。例えば、以下のようなものがある。

- GPSの計測データは、観測できない真の位置に様々なノイズが加わったものになっている。
- 発表される新型コロナウイルス陽性者数は、観測できない真の感染者数にノイズが加わったものになっている。
- ブラックショールズモデルによると、株価 $S_t$ は

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

という方程式に従う。ここで、 $W_t$ はブラウン運動であり、 $\mu, \sigma$ は未知定数である。

しかし、ボラティリティ $\sigma$ は定数ではなく、時間変化する確率過程であることが知られている。これを考慮するとブラックショーするモデルは

$$dS_t = \mu S_t dt + \sigma_t S_t dW_t$$

と書き換えられる。この時 $\sigma_t$ は観測できない状態量である。このようなモデルを確率的ボラティリティモデルという。

## 問題設定

本研究においては、観測できない状態量 $X_t \in \mathbb{R}^{d_1}$ および観測量 $Y_t \in \mathbb{R}^{d_2}$ が次のような方程式に従う状態空間モデルについて考えた：

$$\begin{cases} dX_t = -a(\theta_2)X_t dt + b(\theta_2)dW_t^1 \\ dY_t = c(\theta_2)X_t dt + \sigma(\theta_1)dW_t^2 \end{cases} \quad (1)$$

ここで、 $W_t^1, W_t^2$ は独立な(多次元)ブラウン運動で、 $\theta_1, \theta_2$ は未知パラメータである。また、 $a, b, c, \sigma$ は既知の(行列値) $C^4$ -級関数で、

$$\begin{aligned} \inf_{\theta_2} \lambda_{\min}(a(\theta_2)) &> 0 \\ \inf_{\theta_2} \lambda_{\min}(\sigma(\theta_1)\sigma(\theta_1)^T) &> 0 \end{aligned}$$

を満たすとする。ただし、 $\lambda_{\min}$ は固有値の最小実部を表す。

このようなモデルにおいて、 $Y_t$ の観測値 $Y_{t_0}, Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ に基づいて、未知パラメータ $\theta_1, \theta_2$ を推定するという問題を考える。ただし、 $t_j = jh_n$  ( $j = 0, 1, 2, \dots, n$ )であり、観測幅 $h_n$ は非常に小さく、観測時間 $t_n$ は非常に大きいとする(つまり、高頻度データであるとする)。

## 今回のモデルについて

- $X_t$ はOrnstein-Uhlenbeck過程であり、確率微分方程式で表される確率過程の中でも最も基本的なものである。 $X_t$ は

$$X_t = \exp(-a(\theta)t)X_0 + \int_0^t \exp(-a(\theta)(t-s))dW_s^1$$

という式で書ける。

- $Y_t$ はOrnstein-Uhlenbeck過程に観測ノイズが加わったものと考えられる。
- 次のようなAR(1)モデルに観測ノイズが加わった状態空間モデルの連続化とも考えられる：

$$\begin{cases} x_{k+1} = Ax_k + v_k \\ y_k = Cx_k + w_k \end{cases} \quad (k = 1, 2, \dots).$$

ここで、 $x_k \in \mathbb{R}^{d_1}$ は観測できない状態量、 $y_k \in \mathbb{R}^{d_2}$ は観測量であり、 $A, C$ は定行列、 $v_1, v_2, \dots, w_1, w_2, \dots$ は独立な正規分布に従う確率変数である。

## 応用例

1次元運動を行う物体に弾性力と空気抵抗、及びランダムな力 $F(t)$ が働いているとすると、物体の運動 $Q(t)$ は次のような方程式に従う：

$$M \frac{d^2 Q(t)}{dt^2} + \gamma \frac{dQ(t)}{dt} + \omega^2 Q(t) = F(t) \quad (M, \gamma, \omega > 0).$$

さらに運動量 $P(t) = M \frac{dQ(t)}{dt}$ を用いると、上の式は

$$\frac{d}{dt} \begin{pmatrix} Q(t) \\ P(t) \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{M} \\ -\gamma & -\omega^2 \end{pmatrix} \begin{pmatrix} Q(t) \\ P(t) \end{pmatrix} + \begin{pmatrix} 0 \\ F(t) \end{pmatrix}$$

と書ける。この式を、数学的に厳密に表現すると次のようになる。

$$\begin{pmatrix} dQ(t) \\ dP(t) \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{M} \\ -\gamma & -\omega^2 \end{pmatrix} \begin{pmatrix} Q(t) \\ P(t) \end{pmatrix} dt + \begin{pmatrix} 0 \\ dW_t^1 \end{pmatrix}.$$

さらに、カメラや計測誤差による影響を想定すると、次のような観測方程式が得られる。

$$dZ(t) = \begin{pmatrix} Q(t) \\ P(t) \end{pmatrix} dt + \sigma dW_t^2.$$

# 線形確率微分方程式で表される状態空間モデルにおける推定

栗崎 正博

東京大学数理科学研究科

## Kalman-Bucy フィルター

$Y_t$  の(連続)観測に基づく  $X_t$  の推定量は、次の定理により計算できる。

### 定理

$m_t = E[X_t | \{Y_s\}_{0 \leq s \leq t}]$  および  $\gamma_t = E[(X_t - m_t)(X_t - m_t)^T]$  は次の式に従う:

$$dm_t = -a(\theta_2)m_t dt + \gamma_t c(\theta_2)^T \{\sigma(\theta_1)\sigma(\theta_1)^T\}^{-1} \{dY_t - c(\theta_2)m_t dt\} \quad (2)$$

$$\frac{d\gamma_t}{dt} = -a(\theta_2)\gamma_t - \gamma_t a(\theta_2)^T - \gamma_t c(\theta_2)^T \{\sigma(\theta_1)\sigma(\theta_1)^T\}^{-1} c(\theta_2)\gamma_t + b(\theta_2)b(\theta_2)^T. \quad (3)$$

- これはKalman-Bucy フィルターと呼ばれ、離散Kalman フィルターの連続版とみることができる。
- 離散Kalman フィルターと少し異なる形をしているようにも見えるが、離散Kalman フィルターの連続化極限を取ることで、連続時間Kalman-Bucyの式を導出することが可能である。

## 先行研究

- 確率微分方程式で表される状態空間モデルにおける推定問題に関する研究はまだ少ない。
- 上述のように今回のモデルは非常に基本的なものであるが、これについても先行研究は1次元 ( $d_1 = d_2 = 1$ ) で  $Y_t$  の連続観測が可能な場合に限られる。
  - 1次元かつ連続観測の場合には、 $\sigma(\theta_1)$  の値が直接計算できるため、 $\theta_1$  の推定問題は発生しない。
  - 離散観測の場合、方程式(1)の解は計算できなくなるため、近似を行う必要がある。
  - 多次元の場合には方程式(2)が行列型のRiccati方程式となるため、1次元の場合に比べて扱いが難しくなる。
- 参考文献
  - Kutoyants, Y.A.: Identification of Dynamical Systems with Small Noise. Springer (1994)
  - Kutoyants, Y.A.: Statistical Inference for Ergodic Diffusion Processes. Springer (2004)
  - Kutoyants, Y.A.: On parameter estimation of the hidden ornstein-uhlenbeck process. Journal of Multivariate Analysis 169, 248–263 (2019)

## 本研究の成果

- $X_t, Y_t$  が多次元の場合において、 $Y_t$  の離散観測に基づく  $\theta_1, \theta_2$  の推定量を構成し、その推定量が漸近正規性を満たすことを示した。さらに、その漸近分散を求めた。
- 行列型Riccati方程式(2)が指数関数オーダーである定常解に収束することを示し、さらにその定常解の微分可能性を示した。
- 推定量を構成する際の初期値によって推定の精度が悪化し得ることをシミュレーションにより確認し、それを解消する方法を提案した。

## $\theta_1$ の推定

$\theta_1$  の推定量  $\hat{\theta}_1^n$  は、目的関数

$$\mathbb{H}_n^1(\theta_1) = -\frac{1}{2} \sum_{j=1}^n \left\{ \frac{1}{h} \Delta_j Y^T \Sigma^{-1}(\theta_1) \Delta_j Y + \log \det \Sigma(\theta_1) \right\}$$

を最大化する  $\theta_1$  として定義する。ここで、 $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$  である。

## $\theta_2$ の推定

まず、方程式(2)の定常解を考えるため、(2)に対応する代数的Riccati方程式

$$0 = -a(\theta_2)\gamma - \gamma a(\theta_2)^T - \gamma c(\theta_2)^T \{\sigma(\theta_1)\sigma(\theta_1)^T\}^{-1} c(\theta_2)\gamma + b(\theta_2)b(\theta_2)^T \quad (4)$$

を考える。適切な仮定の下で、この方程式は「極大解」を持つことが知られているので、それを  $\gamma(\theta_1, \theta_2)$  とする。

さらに、(2)で定義される  $m_{t_i}$  を、 $Y_t$  の離散観測  $Y_{t_0}, Y_{t_1}, \dots, Y_{t_n}$  に基づいて次の値で近似する。

$$\hat{m}_i^n(\theta_2; m_0) = \exp\left(-\alpha(\hat{\theta}_1^n, \theta_2)t_i\right) m_0 + \sum_{j=1}^i \exp\left(-\alpha(\hat{\theta}_1^n, \theta_2)(t_i - t_{j-1})\right) \gamma(\hat{\theta}_1^n, \theta_2) c(\theta_2)^T \Sigma(\hat{\theta}_1^n)^{-1} \Delta_j Y. \quad (5)$$

ここで、

- $\alpha(\theta_1, \theta_2) = a(\theta_2) + c(\theta_2)^T \gamma(\theta_1, \theta_2) \Sigma(\theta_1) c(\theta_2)$
- $\Sigma(\theta_1) = \sigma(\theta_1)\sigma(\theta_1)^T$
- $m_0$  は任意の初期値で、 $X_0$  の推定量に相当する

である。

これを用いて、 $\theta_2$  を推定するための目的関数を

$$\mathbb{H}_n^2(\theta_2; m_0) = -\frac{1}{2} \sum_{i=1}^m \{c(\theta_2)\hat{m}_{j-1}^n(\theta_2) - \Delta_j Y\}^T \Sigma(\hat{\theta}_1^n)^{-1} \{c(\theta_2)\hat{m}_{j-1}^n(\theta_2) - \Delta_j Y\}$$

によって定義し、 $\mathbb{H}_n^2(\theta_2; m_0)$  を最大化する  $\hat{\theta}_2^n$  によって、 $\theta_2$  を推定する。

## $\mathbb{H}_n^1$ や $\mathbb{H}_n^2$ の意味について

- $\mathbb{H}_n^1$  や  $\mathbb{H}_n^2$  は、方程式(1)を離散近似したモデルから得られる対数尤度関数に相当する。
- $\mathbb{H}_n^1$  は通常の拡散過程における推定問題に用いられるものと同様である。
- $\hat{m}_i^n$  は  $Y_t$  の離散観測  $Y_{t_0}, Y_{t_1}, \dots, Y_{t_i}$  に基づく  $m_{t_i}$  の近似値である。
- 方程式(1)で誤差を無視すると  $\Delta_j Y \approx c(\theta_2)X_{t_j} \approx c(\theta_2)\hat{m}_j^n$  と近似できるので、 $\mathbb{H}_n^2$  を最大化することは(重み付き)最小二乗法と解釈することもできる。

# 線形確率微分方程式で表される状態空間モデルにおける推定

栗崎 正博

東京大学数理科学研究科

## 主結果

$\theta_1^*, \theta_2^*$  を  $\theta_1, \theta_2$  の初期値とし, 確率変数  $\hat{\theta}_1^n, \hat{\theta}_2^n$  が

$$\mathbb{H}_n^1(\hat{\theta}_1^n) = \sup_{\theta_1} \mathbb{H}_n^1(\theta_1)$$

$$\mathbb{H}_n^2(\hat{\theta}_2^n) = \sup_{\theta_2} \mathbb{H}_n^2(\theta_2)$$

を満たすとする. この時, 適切な仮定の下で, 任意の初期値  $m_0$  に対して

$$\hat{\theta}_1^n \xrightarrow{P} \theta_1^*, \hat{\theta}_2^n \xrightarrow{P} \theta_2^*$$

および

$$\sqrt{n}(\hat{\theta}_1^n - \theta_1^*) \xrightarrow{d} N(0, (\Gamma^1)^{-1})$$

$$\sqrt{n}(\hat{\theta}_2^n - \theta_2^*) \xrightarrow{d} N(0, (\Gamma^2)^{-1})$$

が成り立つ ( $n \rightarrow \infty$ ).

ただし, フィッシャー情報量  $\Gamma^1$  および  $\Gamma^2$  は,  $(i, j)$  成分がそれぞれ

$$\Gamma_{ij}^1 = \frac{1}{2} \text{Tr} \left\{ \Sigma^{*-1} \frac{\partial}{\partial \theta_1^i} \Sigma(\theta_1^*) \right\} \text{Tr} \left\{ \Sigma^{*-1} \frac{\partial}{\partial \theta_1^j} \Sigma(\theta_1^*) \right\}$$

および

$$\Gamma_{ij}^2 = \text{Tr} \int_0^\infty \frac{\partial}{\partial \theta_2^i} f(s; \theta_2^*) \Sigma^{*-1} \frac{\partial}{\partial \theta_2^j} f(s; \theta_2^*) ds$$

で与えられる. ただし,  $\theta_1, \theta_2$  の第  $j$  成分が  $\theta_1^j, \theta_2^j$  であり,

$$\begin{aligned} f(s; \theta_2) &= \int_0^s c(\theta_2) \exp(-\alpha(\theta_2)u) \gamma_+(\theta_2) c(\theta_2)^T \Sigma^{*-1} c^* \\ &\quad \times \exp(-a^*(s-u)) \gamma_+(\theta^*) c^{*T} du \\ &\quad + c(\theta_2) \exp(-\alpha(\theta_2)s) \gamma_+(\theta_2) c(\theta_2)^T \end{aligned}$$

である.

## Riccati方程式に関する結果

適切な仮定の下で, 以下のことを示した.

- 代数的Riccati方程式(3)の解  $\gamma(\theta_1, \theta_2)$  が  $\theta_1, \theta_2$  に関して  $C^4$ -級となることを示した.
- 任意の  $(\theta_1, \theta_2)$  に対し, ある定数  $C_1, C_2 > 0$  が存在して, Riccati型微分方程式(2)の解が

$$\|\gamma_t(\theta_1, \theta_2) - \gamma(\theta_1, \theta_2)\| \leq C_1 e^{-C_2 t}$$

となることを示した.

## $\gamma(\theta_1, \theta_2)$ の計算

$\gamma(\theta_1, \theta_2)$  は以下の手順により計算できる.

- 行列  $H(\theta_1, \theta_2) = \begin{pmatrix} a(\theta_2)' \Sigma(\theta_1)^{-1} [c(\theta_2)^{\otimes 2}] \\ b(\theta_2)^{\otimes 2} & -a(\theta_2) \end{pmatrix}$  の(一般化)固有ベクトルで, 実部が正である固有値に対応するものを  $v_1, v_2, \dots, v_{d_1}$  とする.  $H$  はハミルトニアン行列なので, このようなものはちょうど  $d_1$  個であることがわかる.
- 行列  $X_1, X_2$  を

$$(v_1 \cdots v_{d_1}) = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

により定める. この両辺は  $2d_1 \times d_1$  行列である.

- この時  $\gamma(\theta_1, \theta_2) = X_2 X_1^{-1}$  である.

## 1次元の場合

$d_1 = d_2 = 1$  の場合は, 結果はより簡明になる. この時は  $\sigma$  はスカラー量となるので,  $\sigma(\theta_1) = \theta_1$  とすれば, 推定に関する量はより具体的な形で書ける.

—  $\theta_1$  の推定 —

$\theta_1$  の推定量は

$$\hat{\theta}_1^n = \sqrt{\frac{1}{t_n} \sum_{j=1}^n (\Delta_j)^2}$$

となり, フィッシャー情報量は

$$\Gamma^1 = \frac{2}{\theta_1^{*2}}$$

と書ける.

—  $\theta_2$  の推定 —

$\mathbb{H}_n^2$  の計算に必要な  $\gamma$  や  $\alpha$  は

$$\gamma(\theta_1, \theta_2) = \frac{\theta_1^2 a(\theta_2)}{c(\theta_2)^2} \left( \sqrt{1 + \frac{b(\theta_2)^2 c(\theta_2)^2}{\theta_1^2 a(\theta_2)^2}} - 1 \right),$$

$$\alpha(\theta_1, \theta_2) = \sqrt{a(\theta_2)^2 + \frac{b(\theta_2)^2 c(\theta_2)^2}{\theta_1^2}}$$

と書くことができ, フィッシャー情報量は

$$\begin{aligned} \Gamma^2 &= \frac{\{\partial_{\theta_2} \alpha(\theta^*)\}^{\otimes 2}}{2\alpha(\theta_2^*)} + \frac{\{\partial_{\theta_2} a(\theta^*)\}^{\otimes 2}}{2a(\theta_2^*)} \\ &\quad - \frac{\partial_{\theta_2} \alpha(\theta^*) \otimes \partial_{\theta_2} a(\theta^*) + \partial_{\theta_2} a(\theta^*) \otimes \partial_{\theta_2} \alpha(\theta^*)}{\alpha(\theta_2^*) + a(\theta_2^*)} \end{aligned}$$

となる.

# 線形確率微分方程式で表される状態空間モデルにおける推定

栗崎 正博

東京大学数理科学研究科

## シミュレーション

次のような1次元のモデルについてシミュレーションを行った。

$$\begin{cases} dX_t = aX_t dt + b dW_t^1 \\ dY_t = X_t dt + \sigma dW_t^2 \end{cases}, \begin{cases} X_0 = 0 \\ Y_0 = 0 \end{cases}$$

- 推定すべきパラメータは  $\theta_1 = \sigma, \theta_2 = (a, b)$  である。
- データ数  $n = 10^6$ , 分割幅  $h = 0.0001$  とし, 10000回のシミュレーションを行った。
- $\sigma, a, b$ の真値はそれぞれ0.02, 1.5, 0.3とした。

## シミュレーション結果

まず,  $X$ の推定初期値を  $m_0 = 0$ とした場合の結果を示す。

Table:  $m_0 = 0$ の場合の推定結果

	$\sigma$	$a$	$b$
真値	0.02	1.5	0.3
(推定標準誤差)	$(1.414 \times 10^{-5})$	(0.2115)	(0.01324)
推定値	0.02007	1.495	0.3011
(標準偏差)	$(1.640 \times 10^{-5})$	(0.2123)	(0.01338)

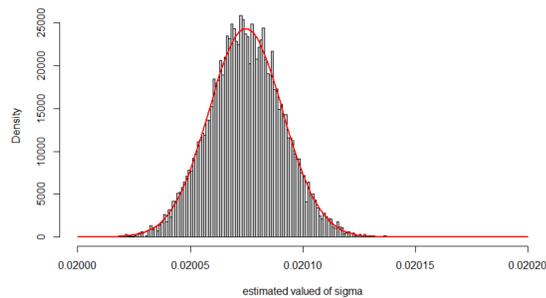


Figure:  $\sigma$ の推定結果

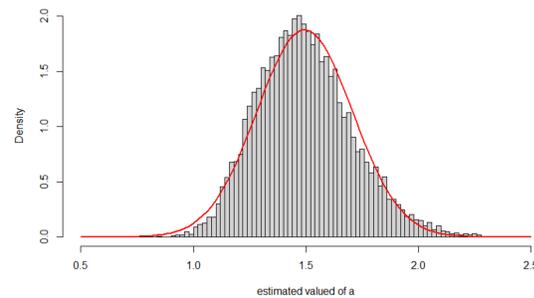


Figure:  $a$ の推定結果

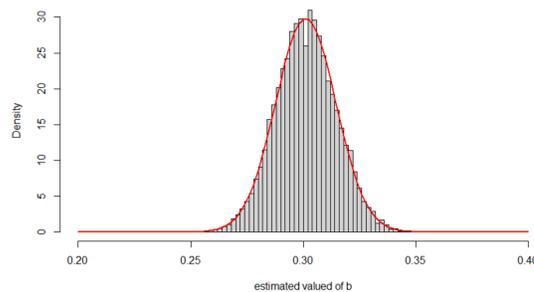


Figure:  $b$ の推定結果

- いずれもそれなりによく推定できていて, 標準誤差も実際の結果と一致している。

## 初期推定量が正しくない場合

$X_0$ の初期推定量を  $m_0 = 1$ として同様のシミュレーションを行った場合, 次のような結果となり,  $m_0 = 0$ とした場合に比べて有意に精度が悪化していることがわかる。

Table:  $m_0 = 1$ とした場合

	$\sigma$	$a$	$b$
推定値	0.02007	1.715	0.3249
(標準偏差)	$(1.640 \times 10^{-5})$	(0.2452)	(0.01338)

- 実際に  $X_t$ と推定値  $\hat{m}_t$ のsample pathを発生させてみると, (5)の中に現れる  $\exp(-\alpha(\hat{\theta}_1^n, \hat{\theta}_2^n)t_i) m_0$ という項の影響により, 最初の部分でずれが発生していることが見て取れる。
- この部分は指数関数的に減少するが,  $\mathbb{E}_n^2$ の計算でこの項を足しこむことになるため, 無視できない影響が出てしまっていると考えられる。

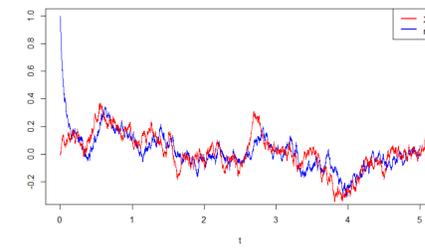


Figure:  $X_t$ と  $\hat{m}_t^n(\theta^*)$  with  $m_0 = 1$ のsample path.

そこで,  $\hat{m}_t^n$ の最初の1000個の値を取り除いて再度推定を行った結果, 次のようになり, 精度が回復した。

Table: 再推定の結果

	$\sigma$	$a$	$b$
推定値	0.02007	1.535	0.3059
(標準偏差)	$(1.640 \times 10^{-5})$	(0.2177)	(0.01304)

## 今後の課題

- モデルの拡張について
  - $X_t$ が回転や周期運動などの場合など,  $\inf_{\theta_2} \lambda_{\min}(a(\theta_2)) > 0$ という条件を  $\inf_{\theta_2} \lambda_{\min}(a(\theta_2)) \geq 0$ に緩める拡張が考えられる。
  - $a(\theta_2)$ が正則などの条件を追加すれば緩めることは可能かもしれないが, 数学的にはより難しくなる。
- 初期推定量の影響について
  - フィルターを用いた推定の場合, 一般的に時間が経過するにつれて状態量の推定精度が向上すると考えられる。
  - 従って, 本研究で用いた最初のいくつかを取り除くという方法は, フィルターを用いた推定問題において一般に応用可能であると考えられる。
  - 何個の項を取り除くべきかという問題が今後の研究対象となり得る。
  - 実験的な方法としては, 取り除く個数を徐々に大きくしてゆき, 推定量が一定値に収束してきたところでやめるという方法が考えられる。