

微小拡散過程の適応的推定

河合 哲弥

大阪大学大学院基礎工学研究科 博士前期課程2年

自己紹介

- ▶ 専門分野: 数理統計学, 計算機科学, 確率解析
- ▶ 学部時代は情報科学と数理科学を専攻.
- ▶ 3つの専門分野を融合したテーマとして, 確率微分方程式を用いて表現される統計モデルを対象に, 計算効率の良い統計的推測法の理論と数値シミュレーションについて研究を行っている.

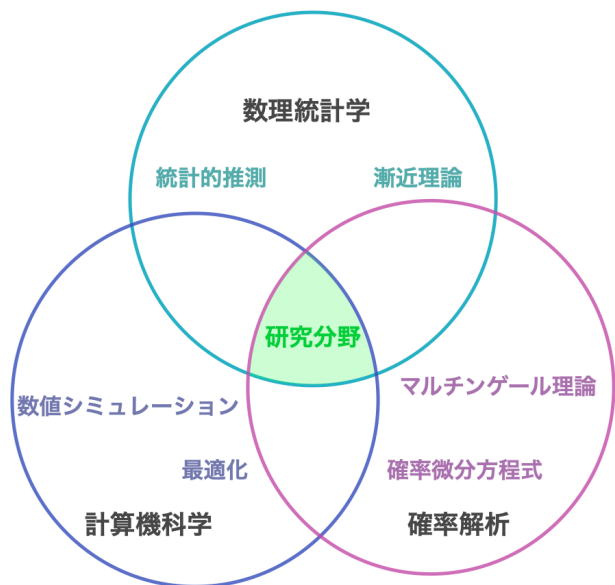


図 1: 研究分野の説明図

研究動機

- ▶ 確率微分方程式によって記述される拡散過程モデルは, 金融, 物理学, 生物学, 気象学など幅広い分野で利用されている.
- ▶ このような応用範囲の広い拡散過程モデルに対して, 専門分野の1つである数理統計学の立場からアプローチを行いたい.
- ▶ 先行研究では, 統計解析の手法が数多く提案されているが, その中には, 数学的理論からの視点のみを重視しており, 実際の解析における安定性や, 計算コストの面については課題があるような, 実用性の乏しい手法も多いことが現状.
- ▶ そこで, 計算機科学の視点も導入し, 実際の数値シミュレーションに対しても安定して効率的に計算できる, 実用的な手法を提案したい.
- ▶ 単に手法を提案することで終わるのではなく, 理論によって提案手法の数学的正当化も行いたい.

研究の位置付け

拡散過程に対する統計的推測の研究における,
計算効率を考慮した統計的手法の開発と数学的正当化

微小拡散過程とは

次の確率微分方程式で表されるモデルが d 次元微小拡散過程:

$$\begin{cases} dX_t = b(X_t, \alpha)dt + \varepsilon \sigma(X_t, \beta)dW_t, & t \in [0, T], \\ X_0 = x_0 \in \mathbb{R}^d, \end{cases}$$

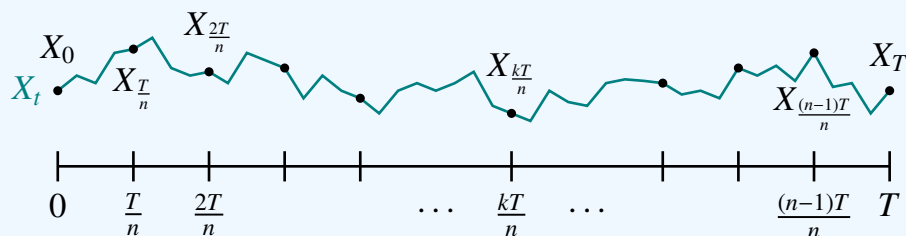
- ▶ W_t : r 次元標準 Wiener 過程
- ▶ $b: \mathbb{R}^d \times \Theta_\alpha \rightarrow \mathbb{R}^d, \sigma: \mathbb{R}^d \times \Theta_\beta \rightarrow \mathbb{R}^d \otimes \mathbb{R}^r$ (関数形は既知)
- ▶ $\varepsilon \in (0, 1)$: 微小攪乱パラメータ
- ▶ $\alpha \in \Theta_\alpha \subset \mathbb{R}^p, \beta \in \Theta_\beta \subset \mathbb{R}^q, \theta = (\alpha, \beta), \Theta := \Theta_\alpha \times \Theta_\beta$
(α : ドリフトパラメータ, β : 拡散パラメータ, それぞれ未知)

目的: 観測される X_t のデータから, パラメータ θ を統計的に推定する

観測データに関する設定

観測されるデータは初期値と n 個の等間隔な離散時間観測

- ▶ データ: $(X_{t_k^n})_{0 \leq k \leq n}, t_k^n = kh_n, k = 0, 1, \dots, n$ ($h_n = T/n$: 観測幅)



- ▶ $n \rightarrow \infty$ のとき, $\varepsilon = \varepsilon_n \rightarrow 0$ が成り立ち, さらに, ある定数 $\rho \geq \frac{1}{2}$ が存在し, $\lim_{n \rightarrow \infty} (\varepsilon n^\rho)^{-1} < +\infty$ となることを仮定

例. 感染症モデル (SIR モデルの微小拡散化)

感染症伝播の数理モデルの一つである SIR モデルは次の微分方程式で表現される: $X_t = (S_t, I_t), (S_0, I_0) = (s_0, i_0) \in (0, 1)^2$,

$$d \begin{pmatrix} S_t \\ I_t \end{pmatrix} = \begin{pmatrix} -\alpha_1 S_t I_t \\ \alpha_1 S_t I_t - \alpha_2 I_t \end{pmatrix} dt, \quad t \in [0, T].$$

(S_t は非感染者, I_t は感染者, $R_t = 1 - S_t - I_t$ は回復 or 死亡者の割合)

SIR モデルは確定的な挙動をするが, このモデルに微小拡散の項を追加することで, 小さなランダム挙動を表現することができ, さらに観測データからパラメータの値を推定することが可能.

微小拡散 SIR モデルの構成例 ([2] の提案モデル)

$$d \begin{pmatrix} S_t \\ I_t \end{pmatrix} = \begin{pmatrix} -\alpha_1 S_t I_t \\ \alpha_1 S_t I_t - \alpha_2 I_t \end{pmatrix} dt + \varepsilon \begin{pmatrix} \sqrt{\alpha_1 S_t I_t} & 0 \\ -\sqrt{\alpha_1 S_t I_t} & \sqrt{\alpha_2 I_t} \end{pmatrix} d \begin{pmatrix} W_{t,1} \\ W_{t,2} \end{pmatrix}, \quad t \in [0, T].$$

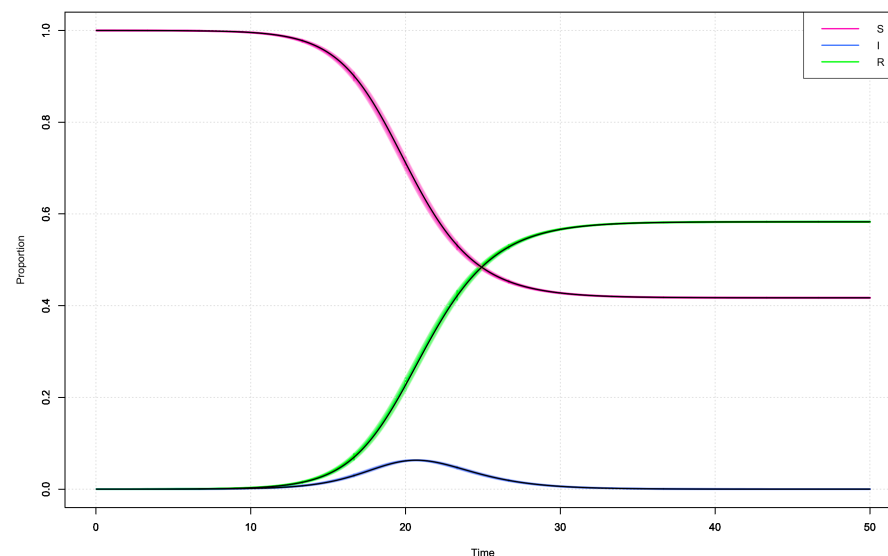


図 2: 微小拡散 SIR モデルのサンプルパスの重ね書き (30 本ランダムに生成). 黒線は通常の SIR モデルのパス. 微小拡散項を追加したサンプルパスは, 黒線付近を小さく振動するような値をとる. ($\alpha_1 = 1.5, \alpha_2 = 1.0, \varepsilon = 10^{-4}, h_n = 1/30, T = 50$)

従来のパラメータ推定法の課題

先行研究 [1] で提案されているパラメータの推定法は、2 種類のパラメータ (α, β) を同時に最適化することによって推定を行っている。この手法では、パラメータの次元が大きくなった場合に、数値計算の面で以下の課題が発生:

- ▶ 最適化の計算時間の爆発的増加
- ▶ 複雑な最適化に失敗する可能性の増加

この課題を改善する手法… 適応的推定法

適応的推定法

パラメータの最適化を α, β それぞれ別々に分けて行い、複雑な最適化を段階的に行うことによって、効率的にパラメータを推定する手法。

記号の定義

- ▶ $S(x, \beta) := \sigma(x, \beta)\sigma(x, \beta)^\top$ (\top は転置)
- ▶ $\nu := \lceil \rho + \frac{1}{2} \rceil$, ここで, $\lceil y \rceil := \min\{z \in \mathbb{Z} \mid y \leq z\}$
- ▶ $\mathcal{L}_\alpha^0(f)(x) := \sum_{i=1}^d b^{(i)}(x, \alpha) \frac{\partial}{\partial x_i} f(x)$, ($f \in C^1$)
- ▶ $P_{1,k}(\alpha) := X_{t_k^n} - X_{t_{k-1}^n} - h_n b(X_{t_{k-1}^n}, \alpha)$
- ▶ $P_{l,k}(\alpha) := P_{1,k}(\alpha) - Q_{l,k}(\alpha)$, ($2 \leq l \leq \nu$)
- ▶ $Q_{l,k}(\alpha) := \sum_{j=1}^{l-1} \frac{h_n^{j+1}}{(j+1)!} (\mathcal{L}_\alpha^0)^j b(X_{t_{k-1}^n}, \alpha)$

適応的推定法のアルゴリズム [3]

▶ Step 1.

$$V_{\varepsilon, n, \nu}^{(1)}(\alpha) := \varepsilon^{-2} h_n^{-1} \sum_{k=1}^n P_{1,k}(\alpha)^\top P_{1,k}(\alpha),$$

$$\hat{\alpha}_{\varepsilon, n}^{(1)} \leftarrow \operatorname{argmin}_{\alpha \in \Theta_\alpha} V_{\varepsilon, n, \nu}^{(1)}(\alpha).$$

▶ Step l ($2 \leq l \leq \nu$).

$$V_{\varepsilon, n, \nu}^{(l)}(\alpha | \bar{\alpha}) := \varepsilon^{-2} h_n^{-1} \sum_{k=1}^n (P_{1,k}(\alpha) - Q_{l,k}(\bar{\alpha}))^\top (P_{1,k}(\alpha) - Q_{l,k}(\bar{\alpha})),$$

$$\hat{\alpha}_{\varepsilon, n}^{(l)} \leftarrow \operatorname{argmin}_{\alpha \in \Theta_\alpha} V_{\varepsilon, n, \nu}^{(l)}(\alpha | \hat{\alpha}_{\varepsilon, n}^{(l-1)}).$$

▶ Step $\nu + 1$.

$$V_{\varepsilon, n, \nu}^{(\nu+1)}(\beta | \bar{\alpha}) := \sum_{k=1}^n \left\{ \log \det S(X_{t_{k-1}^n}, \beta) + \varepsilon^{-2} h_n^{-1} P_{\nu, k}(\bar{\alpha})^\top S^{-1}(X_{t_{k-1}^n}, \beta) P_{\nu, k}(\bar{\alpha}) \right\},$$

$$\hat{\beta}_{\varepsilon, n} \leftarrow \operatorname{argmin}_{\beta \in \Theta_\beta} V_{\varepsilon, n, \nu}^{(\nu+1)}(\beta | \hat{\alpha}_{\varepsilon, n}^{(\nu)}).$$

▶ Step $\nu + 2$.

$$V_{\varepsilon, n, \nu}^{(\nu+2)}(\alpha | \bar{\alpha}, \bar{\beta}) := \varepsilon^{-2} h_n^{-1} \sum_{k=1}^n (P_{1,k}(\alpha) - Q_{\nu, k}(\bar{\alpha}))^\top S^{-1}(X_{t_{k-1}^n}, \bar{\beta}) (P_{1,k}(\alpha) - Q_{\nu, k}(\bar{\alpha})),$$

$$\hat{\alpha}_{\varepsilon, n} \leftarrow \operatorname{argmin}_{\alpha \in \Theta_\alpha} V_{\varepsilon, n, \nu}^{(\nu+2)}(\alpha | \hat{\alpha}_{\varepsilon, n}^{(\nu)}, \hat{\beta}_{\varepsilon, n}).$$

▶ Output $(\hat{\alpha}_{\varepsilon, n}, \hat{\beta}_{\varepsilon, n})$.

従来の同時推定法のイメージ

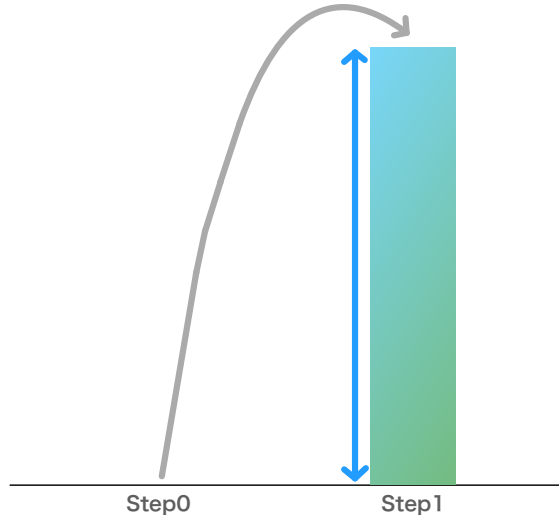


図 3: 同時推定法のイメージ図

- ステップ数は1つだが、最適化が難しい(失敗する可能性が高い).

適応的推定法のイメージ

($v = 3$ の場合)

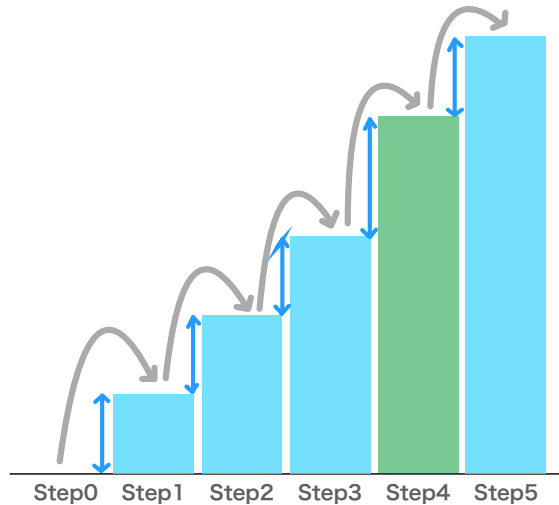


図 4: 適応的推定法のイメージ図

- ステップ数は増加するが、各ステップの最適化は比較的容易。
数値計算ではこちらの手法の方が速く、安定する。

理論結果

パラメータの真値を $\theta_0 = (\alpha_0, \beta_0)$ とし, $(X_t^0)_{t \in [0, T]}$ は次の常微分方程式の解とする:

$$\begin{cases} dX_t^0 = b(X_t^0, \alpha_0)dt, & t \in [0, T], \\ X_0^0 = x_0. \end{cases}$$

さらに, 以下を定義する.

$$I(\theta_0) := \begin{pmatrix} (I_b^{i,j}(\theta_0))_{1 \leq i, j \leq p} & 0 \\ 0 & (I_\sigma^{i,j}(\beta_0))_{1 \leq i, j \leq q} \end{pmatrix},$$

$$I_b^{i,j}(\theta_0) = \int_0^T \left(\frac{\partial}{\partial \alpha_i} b(X_s^0, \alpha_0) \right)^\top S^{-1}(X_s^0, \beta_0) \left(\frac{\partial}{\partial \alpha_j} b(X_s^0, \alpha_0) \right) ds,$$

$$I_\sigma^{i,j}(\beta_0) = \frac{1}{2T} \int_0^T \text{tr} \left[\left\{ \left(\frac{\partial}{\partial \beta_i} S \right) S^{-1} \left(\frac{\partial}{\partial \beta_j} S \right) S^{-1} \right\} (X_s^0, \beta_0) \right] ds.$$

このとき, 適応的推定法によって得られるパラメータの推定量 $\hat{\theta}_{\varepsilon, n} = (\hat{\alpha}_{\varepsilon, n}, \hat{\beta}_{\varepsilon, n})$ に対して, 次の漸近的結果が成り立つ. [3]
(\xrightarrow{P} , \xrightarrow{d} はそれぞれ確率収束, 分布収束を表すとする)

定理

ある正則条件の下で, $\varepsilon \rightarrow 0, n \rightarrow \infty$ としたとき,

$$\hat{\theta}_{\varepsilon, n} \xrightarrow{P} \theta_0,$$

すなわち, $\hat{\theta}_{\varepsilon, n}$ は θ の一致推定量であり, さらに,

$$\begin{pmatrix} \varepsilon^{-1}(\hat{\alpha}_{\varepsilon, n} - \alpha_0) \\ \sqrt{n}(\hat{\beta}_{\varepsilon, n} - \beta_0) \end{pmatrix} \xrightarrow{d} N_{p+q}(0, I(\theta_0)^{-1})$$

の漸近正規性が成立する.

この定理によって推定量の誤差分布を近似的に求めることができ, パラメータの区間推定や統計的仮説検定を行うことも可能.

設定

$$\text{モデル: } dX_t = \begin{pmatrix} -\alpha_1 X_{t,1} + 2 \cos(1 + \alpha_2 X_{t,2}) \\ 2 \sin(1 + \alpha_3 X_{t,1}) - \alpha_4 X_{t,2} \end{pmatrix} dt + \varepsilon \begin{pmatrix} \beta_1(1 + X_{t,1}^2)^{-1} & 0.1 \\ 0.1 & \beta_2(1 + X_{t,2}^2)^{-1} \end{pmatrix} dW_t, \quad (t \in [0, 1]), \quad X_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

パラメータ: $\theta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2)$, 真値: $\theta_0 = (3, 6, 5, 4, 1, 0.5)$, パラメータ空間: $\Theta = [0.01, 50]^6$, 適応的推定のステップ: $v=2$.

最適化におけるパラメータの初期値: $\theta_{\text{init}} = (6, 4, 6, 8, 2, 1)$; 真値から少し離れた初期値を設定.

$(\varepsilon, n) = (0.01, 100)$ or $(0.01, 1000)$ の2通りの設定で行い, それぞれ 10000 回ずつデータを発生させ, 同時推定と適応的推定の結果を比較.

表 1: 計算された推定値の平均 (標準偏差) と総計算時間

(ε, n)	手法	α_1	α_2	α_3	α_4	β_1	β_2	時間 (m)
(0.01,100)	同時	2.6508 (0.0919)	46.418 (1.9025)	42.110 (2.6332)	4.3134 (0.4090)	22.673 (0.5207)	17.687 (0.3332)	22
	適応的	2.9997 (0.0168)	6.0004 (0.0158)	5.0000 (0.0068)	4.0002 (0.0134)	0.9717 (0.0667)	0.4925 (0.0346)	13
(0.01,1000)	同時	2.5863 (0.0839)	30.540 (1.5777)	4.9727 (0.0340)	3.9766 (0.0488)	6.8759 (0.7533)	0.5218 (0.0422)	119
	適応的	3.0001 (0.0167)	6.0002 (0.0155)	5.0001 (0.0068)	4.0000 (0.0133)	0.9971 (0.0215)	0.4991 (0.0109)	45

推定精度, 計算時間ともに適応的推定法の方が良い結果となっている.

詳細や, その他のモデルに対するシミュレーション結果については, [3] を参照.

参考文献

- [1] Gloter, A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Processes and their Applications*. 150:84-110.
- [2] Guy, R., Larédo, C., and Vergu, E. (2014). Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient. *Stochastic Processes and their Applications*. 124(1):51-80
- [3] Kawai, T. and Uchida, M. (2021). Adaptive estimation for small diffusion processes based on sampled data. arXiv:2108.01397.

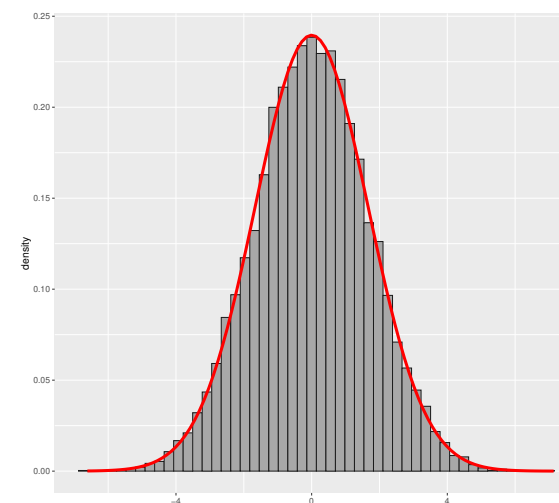


図 5: α_1 の推定値のヒストグラム. 赤線は理論分布の密度関数. (適応的推定, $(\varepsilon, n) = (0.01, 1000)$ の場合における結果)