

微小拡散モデルにおける推定量の 高次漸近理論と情報幾何

東京大学・情報理工学系研究科 清 智也 (Tomonari Sei)
Graduate School of Information Science and Technology,
University of Tokyo

東京大学・情報理工学系研究科 駒木 文保 (Fumiyasu Komaki)
Graduate School of Information Science and Technology,
University of Tokyo

Abstract

攪乱項が小さい拡散過程モデルにおいて、計量テンソルや接続係数などの情報幾何学的な量を求めた。また、バイアス補正推定量の漸近的性質を情報幾何学の観点から調べた。推定量の漸近正規性により、独立同一分布モデルの場合と類似した結果が得られる。独立同一分布モデルと異なる点は、幾何学量が攪乱の大きさに依存することである。また、攪乱の大きさがゼロに近づく際、ある種の特異性が現れるが、その問題は回避される。ただしこれらの結果は形式的な漸近展開に基づいており正当性は示していない。

キーワード：曲指数型分布族；情報幾何学；二次漸近有効性；微小拡散モデル

1 はじめに

次の1次元拡散過程を考える：

$$dX_t^\epsilon = \mu(X_t^\epsilon, u)dt + \epsilon\sigma(X_t^\epsilon)dW_t, \quad X_0^\epsilon = x_0, \quad t \in [0, T], \quad (1)$$

ここで $W = \{W_t\}$ は1次元標準ブラウン運動、 $u = (u^1, \dots, u^m) \in U \subset \mathbf{R}^m$ は未知パラメータ、 $\epsilon \in [0, 1]$ は攪乱の大きさを表す微小なパラメータ、 $\mu(\cdot, \cdot) : \mathbf{R} \times U \rightarrow \mathbf{R}$ と $\sigma(\cdot) : \mathbf{R} \rightarrow \mathbf{R}$ は滑らかな関数とする。初期値 x_0 は u によらず固定されているものとする。観測期間 T も固定する。 X_t^ϵ は $\epsilon = 0$ のとき決定論的になる。(1) 式で与えられるモデルを微小拡散モデルと呼ぶことにする。

連続観測値 $X = \{X_t | t \in [0, T]\}$ に基づく u の推定量を考える。なお、 ϵ は観測値 X から確率1で定まるので、既知としてよい。推定量の $\epsilon \rightarrow 0$ における漸近的性質、特

に2次漸近有効性に関する結果を示す。ここでいう2次漸近有効な推定量とは、1次漸近有効なバイアス補正推定量の中で平均二乗誤差を最小にする推定量のことをいう(Rao 1962)。独立同一標本のモデルでは、2次漸近有効性は情報幾何学の枠組みで自然に述べられる。この結果はTaniguchi & Watanabe (1994)により時系列モデルに拡張された。微小拡散モデルについて、他の定義の(高次)漸近有効性に関する理論はKutoyants (1994)やYoshida (1992a)に与えられている。

発表者らは、微小拡散モデルについても情報幾何の観点から二次漸近有効性の性質が明らかになることを示した(Sei & Komaki 2003)。その内容を報告する。

まず2節で微小拡散モデルに対する幾何学的な量を漸近展開の形で与える。次に3節で推定量の平均二乗誤差の漸近展開を与え、簡単な例について、最尤推定量と他の推定量を数値的に比較した結果を示す。数値計算の結果は漸近展開による結果と整合的であることが示された。

2 微小拡散モデルの幾何

$\mathcal{C}_T$ は $[0, T]$ 上の連続関数全体からなる空間、 P_u^ϵ は (1) 式の解 X^ϵ によって $\mathcal{C}_T$ に定義される測度とする。 $E_u^\epsilon[\cdot]$ または $E_u[\cdot]$ は P_u^ϵ に関する期待値を表す。尤度関数は

$$\frac{dP_u^\epsilon}{dP_{u_0}^\epsilon}(X) = \frac{\exp(\ell_\epsilon(X, u))}{\exp(\ell_\epsilon(X, u_0))}, \quad \text{for } X \in \mathcal{C}_T \text{ and } u \in U,$$

と表される。ただし、 $u_0 \in U$ は固定したパラメータ値とし、 ℓ_ϵ は

$$\ell_\epsilon = \ell_\epsilon(X, u) = \frac{1}{\epsilon^2} \int_0^T \frac{\mu(X_t, u)}{\sigma^2(X_t)} dX_t - \frac{1}{2\epsilon^2} \int_0^T \frac{\mu^2(X_t, u)}{\sigma^2(X_t)} dt. \quad (2)$$

と定義される。確率積分 $\int \{\cdot\} dX_t$ は Riemann-Stieltjes 式に計算できる (Kutoyants 1994)。以下では ℓ_ϵ を対数尤度関数として使う。統計的モデルを $\mathcal{P} = \mathcal{P}^\epsilon = \{\ell_\epsilon(\cdot, u) | u \in U\}$ とおく。

Amari (1985) に従い情報幾何学の記法を準備する。モデル $\mathcal{P}$ はパラメータ $u = (u^a) \in U$ を局所座標とする多様体と見ることができる。 $\mathcal{P}$ の $u \in U$ における接ベクトル空間は

$$\mathcal{T}_u = \text{span}\{\partial_a \ell_\epsilon(\cdot, u) | a = 1, \dots, m\}, \quad \partial_a = \partial / \partial u^a, \quad a = 1, \dots, m$$

である。 $\mathcal{T}_u$ 上の内積は $E_u[A(X)B(X)]$ for $A(X), B(X) \in \mathcal{T}_u$ とする。 $\mathcal{P}$ 上の Fisher 計量 g_{ab} , e-接続係数 (e=exponential) $\overset{e}{\Gamma}_{abc}$, m-接続係数 (m=mixture) $\overset{m}{\Gamma}_{abc}$ はそれぞれ

$$\begin{aligned} g_{ab} &= g_{ab}^\epsilon := \epsilon^2 E_u[\partial_a \ell_\epsilon \partial_b \ell_\epsilon], \\ \overset{e}{\Gamma}_{abc} &= \overset{e}{\Gamma}_{abc}^\epsilon := \epsilon^2 E_u[\partial_a \partial_b \ell_\epsilon \partial_c \ell_\epsilon], \\ \overset{m}{\Gamma}_{abc} &= \overset{m}{\Gamma}_{abc}^\epsilon := \overset{e}{\Gamma}_{abc} + \epsilon^2 E_u[\partial_a \ell_\epsilon \partial_b \ell_\epsilon \partial_c \ell_\epsilon], \end{aligned}$$

for $a, b, c \in \{1, \dots, m\}$, と定義する. 普通の定義に対して因子 ϵ^2 がかかっているのは, これらの量が $O(1)$ となるように調整したためである. このような調整をしても, 双対性 $\partial_a g_{bc} = \overset{\epsilon}{\Gamma}_{abc} + \overset{m}{\Gamma}_{acb}$ は成り立つ.

本稿では, 添え字 $\{a, b, \dots, f\}$ 及び $\{a', b', \dots, f'\}$ は座標系 u に関する量を表す; 例えば ∂_a や g_{ab} などと書く. また, Einstein の縮約規則も用いる; 例えば $\phi^a \psi_a$ は $\sum_{a=1}^m \phi^a \psi_a$ の意味である. $(g_{ab}) = (g_{ab}^\epsilon)$ の逆行列を $(g^{ab}) = (g^{ab\epsilon})$ と記す. また $\overset{\epsilon}{\Gamma}_{ab}^c := \overset{\epsilon}{\Gamma}_{abd} g^{cd}$ とおく. $E[(\epsilon \partial_a \partial_b \ell_\epsilon - \overset{\epsilon}{\Gamma}_{ab}^c \epsilon \partial_c \ell_\epsilon) \epsilon \partial_d \ell_\epsilon] = 0$ が成り立つので, $\overset{\epsilon}{\Gamma}_{ab}^c$ は $\epsilon \partial_a \partial_b \ell_\epsilon$ を $\epsilon \partial_c \ell_\epsilon$ で回帰したときの回帰係数の意味を持つ.

形式的なテイラー展開により, 上記の幾何学的な量を漸近的に求めることができる. 大雑把に言って, ここで扱う統計量は, 伊藤の公式により

$$F_T^\epsilon(X) = F_0^\epsilon + \int_0^T f_0^\epsilon(X_t) dt + \int_0^T f_1^\epsilon(X_t) dW_t$$

の形で表されるので, Yoshida (1992b) の結果が適用できると考えられる. ただし $F_0^\epsilon \in \mathbf{R}^k$ と $f_0^\epsilon, f_1^\epsilon$ は滑らかな $\mathbf{R}^k$ -値関数 ($k \geq 1$) である. しかし, ここではその正当性については議論しないことにする.

以下では, $x \in \mathbf{R}$ と $t \in [0, T]$ の関数 $f(x, t)$ に対して, $[f]_t^\epsilon$ は $f(X_t^\epsilon, \epsilon)$ を意味するものとする. また, $\bar{\mu}_{a_1 \dots a_k}$ は $\sigma^{-1} \partial_{a_1} \dots \partial_{a_k} \mu$ の意味であるとする. $Y = \{Y_t\}$ は決定論的な微分方程式

$$dY_t = [\partial_x \mu]_t^0 Y_t dt, \quad Y_0 = 1,$$

の解とする. ただし $\partial_x := \partial/\partial x$ とおいた.

まず, 次の補題が成り立つ.

補題 2.1 u を真のパラメータとする. 対数尤度のパラメータ微分はそれぞれ以下のよう書ける:

$$\begin{aligned} \partial_a \ell_\epsilon(X^\epsilon, u) &= \epsilon^{-1} \int_0^T [\bar{\mu}_a]_t^\epsilon dW_t, \\ \partial_a \partial_b \ell_\epsilon(X^\epsilon, u) &= -\epsilon^{-2} \int_0^T [\bar{\mu}_a \bar{\mu}_b]_t^\epsilon dt + \epsilon^{-1} \int_0^T [\bar{\mu}_{ab}]_t^\epsilon dW_t, \\ \partial_a \partial_b \partial_c \ell_\epsilon(X^\epsilon, u) &= -3\epsilon^{-2} \int_0^T [\bar{\mu}_{(ab} \bar{\mu}_{c)}]_t^\epsilon dt + \epsilon^{-1} \int_0^T [\bar{\mu}_{abc}]_t^\epsilon dW_t, \end{aligned}$$

ただし $f_{(abc)} = (f_{abc} + f_{bca} + f_{cab})/3$ は f_{abc} を対称化したものである. また, 4 階以上の微分は全て $O_P(\epsilon^{-2})$ である.

証明 (2) を微分した式に (1) 式を代入して得られる. □

補題 2.2 (1) 式の解は

$$X_t^\epsilon = X_t^0 + \epsilon D_t + \frac{\epsilon^2}{2} E_t + o_P(\epsilon^2)$$

と展開される。ただし D と E は

$$\begin{aligned} D_t &= Y_t \int_0^t Y_s^{-1} [\sigma]_s^0 dW_s, \\ E_t &= Y_t \int_0^t Y_s^{-1} ([\partial_x^2 \mu]_s^0 D_s^2 ds + 2[\partial_x \sigma]_s^0 D_s dW_s). \end{aligned}$$

で定義される確率過程である。

証明 Gihman & Skorohod (1972, pp. 58-59) 参照。 $\square$

上記の補題 2.1 と補題 2.2 を用いて、幾何学的量の漸近展開を求めたものが次の定理である。

定理 2.1 1. Fisher 計量は次のように漸近展開される：

$$\begin{aligned} g_{ab}^\epsilon &= \int_0^T E_u [\bar{\mu}_a \bar{\mu}_b]_t^\epsilon dt \\ &= \int_0^T [\bar{\mu}_a \bar{\mu}_b]_t^0 dt + \frac{\epsilon^2}{2} \left\{ \int_0^T [\partial_x^2 (\bar{\mu}_a \bar{\mu}_b)]_t^0 Y_t^2 \left(\int_0^t Y_s^{-2} [\sigma^2]_s^0 ds \right) dt \right. \\ &\quad \left. + \int_0^T [\partial_x (\bar{\mu}_a \bar{\mu}_b)]_t^0 Y_t \left(\int_0^t Y_s [\partial_x^2 \mu]_s^0 \left(\int_0^s Y_v^{-2} [\sigma^2]_v^0 dv \right) ds \right) dt \right\} \\ &\quad + o(\epsilon^2). \end{aligned} \tag{3}$$

2. e-, m-接続係数はそれぞれ次式のように展開される：

$$\begin{aligned} \bar{\Gamma}_{abc}^\epsilon &= B_{abc} - A_{abc} + o(1), \\ \bar{\Gamma}_{abc}^m &= B_{abc} + A_{bca} + A_{cab} + o(1), \end{aligned}$$

ただし

$$\begin{aligned} A_{abc} &= \int_0^T [\partial_x (\bar{\mu}_a \bar{\mu}_b)]_t^0 Y_t \left(\int_0^t Y_s^{-1} [\partial_c \mu]_s^0 ds \right) dt, \\ B_{abc} &= \int_0^T [\bar{\mu}_{ab} \bar{\mu}_c]_t^0 dt. \end{aligned}$$

証明 補題 2.1, 2.2 から得られる。詳細は Sei & Komaki (2003) を参照。 $\square$

次に、埋め込み e-曲率に対応するテンソル量 $\bar{H}_{abcd}$ を定義する：

$$\bar{H}_{abcd} = E[\hat{y}_{ab} \hat{y}_{cd}],$$

ただし $\dot{y}_{ab} := \epsilon \partial_a \partial_b \ell + \epsilon^{-1} g_{ab} - \epsilon \dot{\Gamma}_{ab}^c \partial_c \ell$ とおいた. Efron (1975) は統計的曲率と呼ばれる次のスカラー量を定義した.

$$\gamma = \gamma(u) = \dot{H}_{abcd} g^{ac} g^{bd}$$

これは, 独立同一標本のモデルを考えたとき, モデルが指数型分布族からどの程度ずれているかを示す量である. この性質は微小拡散モデルについてもあてはまることが 3 節の高次漸近理論で明らかになる.

命題 2.1 テンソル $\dot{H}_{abcd}$ は

$$\dot{H}_{abcd} = \int_0^T [m_{ab}]_t [m_{cd}]_t dt - \dot{\Gamma}_{ab}^f \dot{\Gamma}_{cdf} + o(1),$$

と展開される. ただし

$$[m_{ab}]_t = [\bar{\mu}_{ab}]_t^0 - \left(\int_t^T [\partial_x (\bar{\mu}_a \bar{\mu}_b)]_s^0 Y_s ds \right) Y_t^{-1} [\sigma]_t^0.$$

証明 補題 2.1, 2.2 による. □

ここでいくつかの例を与える.

例 1. Ornstein-Uhlenbeck 過程モデル (O-U 過程モデル) は

$$dX_t = -uX_t dt + \epsilon dW_t, \quad t \in [0, 1], \quad X_0 = 1, \quad (4)$$

で定義される. ただしここでは u_1 を単に u と書く. このモデルの幾何学的な量はそれぞれ ($\epsilon \rightarrow 0$ の下で)

$$\begin{aligned} g_{11} &= (2u)^{-1}(1 - e^{-2u}), \\ \dot{\Gamma}_{111} &= (2u^2)^{-1}(1 - (1 + 2u)e^{-2u}), \\ \dot{H}_{1111} &= (2(-1 + e^{2u})u^3)^{-1}(1 + e^{-4u} - (2 + 4u^2)e^{-2u}), \\ \gamma &= ((-1 + e^{2u})^3 u)^{-1}(2((-1 + e^{2u})^2 - 4e^{2u}u^2)), \end{aligned}$$

となる.

例 2. 平均回帰型 Ornstein-Uhlenbeck 過程モデルは次式で定義される:

$$dX_t = (-u_1 X_t + u_2)dt + \epsilon dW_t, \quad t \in [0, 1], \quad X_0 = 0. \quad (5)$$

このモデルの Fisher 情報量は $\epsilon \rightarrow 0$ の下で

$$\begin{aligned} g_{11} &= (2u_1^3)^{-1}(-3 + 2u_1 - e^{-2u_1} + 4e^{-u_1})u_2^2, \\ g_{12} &= u_1^{-2}(1 - u_1 - e^{-u_1})u_2, \\ g_{22} &= 1 \end{aligned}$$

となる。その他の幾何学的量の漸近形も解析的に求まるが、煩雑になるので割愛する。

例 3. 次のような確率過程を考える：

$$dX_t = -\sin(X_t - u)dt + \epsilon dW_t, \quad t \in [0, T], \quad X_0 = 0. \quad (6)$$

このモデルのことをここでは円上の拡散モデルと呼ぶことにする。\$\epsilon = 0\$ のとき、(6) 式の解 \$x_t\$ は次式で与えられる。

$$x_t = u + \arccos \left\{ \frac{\tanh(t) + \cos(x_0 - u)}{1 + \tanh(t) \cos(x_0 - u)} \right\}.$$

特に

$$\begin{aligned} \cos(x_t - u) &= \frac{\tanh(t) + \cos(x_0 - u)}{1 + \tanh(t) \cos(x_0 - u)}, \\ \sin(x_t - u) &= \frac{\sin(x_0 - u)}{[1 + \tanh(t) \cos(x_0 - u)] \cosh(t)}. \end{aligned}$$

Fisher 情報量は \$\epsilon \rightarrow 0\$ の下で

$$g_{11} = T - \frac{\sin(x)^2 \sinh(T)}{\cosh(T) + \cos(x) \sinh(T)}$$

となる。その他の幾何学的量の漸近形も解析的に求まるが、煩雑になるので割愛する。

各モデルについて統計的曲率 \$\gamma\$ のグラフを描いたものを図 1 に示す。

3 Fisher 一致推定量の漸近展開と高次漸近有効性

この節では、(1) 式のドリフト係数がある関数 \$\phi^p(\cdot)\$, \$\nu_p(\cdot)\$ for \$p \in \{1, \dots, K\}\$ を用いて

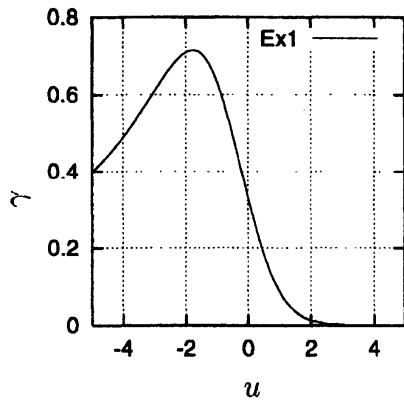
$$\mu(x, u) = \phi^p(u) \nu_p(x) := \sum_{p=1}^K \phi^p(u) \nu_p(x) \quad (7)$$

という形に書けると仮定する。前節で挙げた例は全てこの仮定を満たしている。添え字 \$\{p, q\}\$ は \$\phi, \nu\$ に関係する量を表す。(7) 式が成り立つとき、その対数尤度関数は

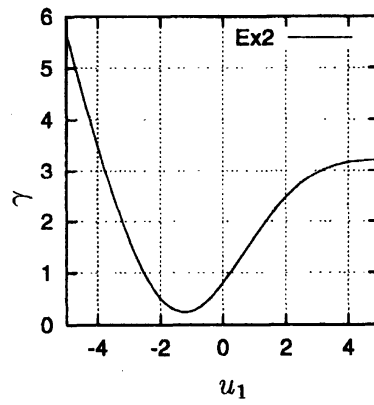
$$\ell_\epsilon(X, u) = \epsilon^{-2} \left(\phi^p(u) \int_0^T \bar{\nu}_p \sigma^{-1}(X_t) dX_t - \frac{1}{2} \phi^p \phi^q(u) \int_0^T \bar{\nu}_p \bar{\nu}_q(X_t) dt \right), \quad (8)$$

となる。ただし、\$\bar{\nu}_p := \nu_p \sigma^{-1}\$ とおいた。よってモデル \$\mathcal{P}\$ は曲指数型分布族である。これを包む指数型分布族 (envelope family) \$\mathcal{E} = \{\ell_\epsilon(\cdot, \theta)\}\$ を

$$\ell_\epsilon(X, \theta) = \epsilon^{-2} (\theta^i s_i(X) - \Psi^\epsilon(\theta)), \quad \theta \in \Theta^\epsilon(\mathcal{E}),$$



例 1



例 2

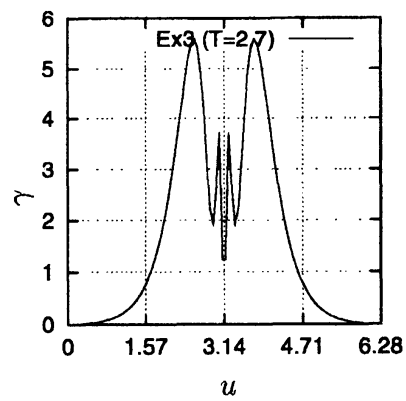
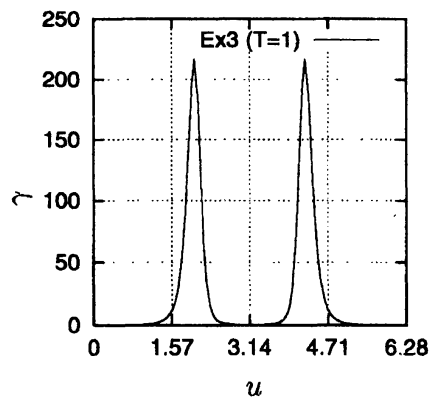
例 3 (左 : $T = 1$, 右 : $T = 2.7$)

Figure 1: 各例について, 統計的曲率を示す. 例 2 の u_2 の値は 1 に固定した. 例 3 では $T = 1$ と $T = 2.7$ の場合についてそれぞれ描いた.

と書く. ここに, $\theta = (\theta^i)$ は N 次元自然パラメータ ($N = \frac{1}{2}K(K+3)$), $s(X) = (s_i(X))$ は θ に対応する十分統計量, $\Psi^\epsilon(\theta)$ はポテンシャル関数, $\Theta^\epsilon(\mathcal{E})$ は $\Psi^\epsilon(\cdot)$ の定義域であり, 以下のように定義する.

$$s(X) = \left(\left(\int_0^T \bar{\nu}_p \sigma^{-1}(X_t) dX_t \right)_{1 \leq p \leq K}, \left(\int_0^T \bar{\nu}_p \bar{\nu}_q(X_t) dt \right)_{1 \leq p \leq q \leq K} \right), \quad (9)$$

$$\Psi^\epsilon(\theta) = \epsilon^2 \log E_0^\epsilon \left[\exp(\epsilon^{-2} \theta^i s_i(X)) \right],$$

ただし E_0^ϵ は $X_t = \epsilon \int_0^t \sigma(X_s) dW_s$ に関する期待値である. また,

$$\Theta(\mathcal{E}) = \Theta^\epsilon(\mathcal{E}) = \{\theta \in \mathbf{R}^n \mid \Psi^\epsilon(\theta) < \infty\}.$$

元のモデル $\mathcal{P}$ は $\mathcal{E}$ の部分多様体である:

$$\theta(u) = ((\phi^p(u))_{1 \leq p \leq K}, (-(1 + \delta_{pq})^{-1} \phi^p(u) \phi^q(u))_{1 \leq p \leq q \leq K}),$$

ただし δ_{pq} はクロネッカーのデルタである. また, $\Theta(\mathcal{P}) = \{\theta(u) \mid u \in U\} = \{\theta \mid \ell_\epsilon(\theta) \in \mathcal{P}\}$ と記す. 曲指数型分布族とその envelope family については Küchler and Sørensen (1996, 1997) を参照せよ.

次の補題より, Ψ^ϵ は $\Theta(\mathcal{P})$ の近傍で $O(1)$ ($\epsilon \rightarrow 0$) である.

補題 3.1 ある定数 $C_0 > 0$ が存在して, 全ての $x \in \mathbf{R}$ に対して $|\nu_p(x)| \leq C_0(1 + |x|)$ と $1/C_0 < |\sigma(x)| < C_0$ が成り立つと仮定する. このとき各 $u \in U$ に対して, ある定数 $\delta > 0$, $C > 0$ が存在し, 任意の $\epsilon > 0$ に対して $\|\theta - \theta(u)\| \leq \delta \implies \Psi^\epsilon(\theta) \leq C$ が成り立つ.

証明 一般性を失うことなく $T = 1$ とおける. $\theta = (\phi^p(u), -(1/2)\phi^p \phi^q(u) + \xi_{pq})$, $\|\xi\| < \delta$ とおく. すると

$$\begin{aligned} \exp(\epsilon^{-2} \Psi^\epsilon(\theta)) &= E_u \left[\exp \left(\epsilon^{-2} \sum_{p,q} (\xi_{pq} \hat{\eta}_{pq}) \right) \right] \\ &\leq \prod_{p,q} E_u \left[\exp \left(\delta \epsilon^{-2} \int_0^1 |\bar{\nu}_p \bar{\nu}_q(X_t)| dt \right) \right] \\ &\leq \int_0^1 E_u \left[\exp \left(C_1 \epsilon^{-2} (1 + |X_t|^2) \right) \right] dt, \end{aligned}$$

ただし $C_1 = 4K^2 C_0^4 \delta$, となるので, 次式を満たす定数 $C > 0$ が存在することを示せば十分である; 十分小さい $C_1 > 0$ に対して

$$E_u \left[\exp \left(C_1 \epsilon^{-2} |X_t|^2 \right) \right] < \exp(C \epsilon^{-2}).$$

すると, Lipster & Shiryaev (2001) の定理 4.7 の証明と同様に, 線形の確率微分方程式

$$dX_t = C_0 X_t dt + \epsilon C_0 dW_t, \quad X_0 = x_0,$$

のケースのみ考えればよく, この場合は直接確認できる. $\square$

自然パラメータに関係する量の添え字には $\{i, j, k\}$ を用いる. $\eta^\epsilon(\theta)$ を期待値パラメータとする:

$$\eta_i^\epsilon(\theta) = E_\theta[s_i(X)] = \partial_i \Psi^\epsilon(\theta),$$

ただし $\partial_i := \partial/\partial\theta^i$ とした. 簡単のため $\eta_i^\epsilon(u) = \eta_i^\epsilon(\theta(u))$ と書く. また $S^\epsilon(\mathcal{P}) = \{\eta^\epsilon(u) \mid u \in U\}$ とおく.

envelope family $\mathcal{E}$ 上の Fisher 情報行列, e-接続, m-接続も, $\mathcal{P}$ 上と同様に定義される:

$$\begin{aligned} g_{ij} &= g_{ij}^\epsilon := \epsilon^2 E_\theta[\partial_i \ell_\epsilon \partial_j \ell_\epsilon] = \partial_i \partial_j \Psi^\epsilon(\theta), \\ \overset{\circ}{\Gamma}_{ijk} &= \overset{\circ}{\Gamma}_{ijk}^\epsilon := \epsilon^2 E_\theta[\partial_i \partial_j \ell_\epsilon \partial_k \ell_\epsilon] = 0, \\ \overset{\text{m}}{\Gamma}_{ijk} &= \overset{\text{m}}{\Gamma}_{ijk}^\epsilon := \overset{\circ}{\Gamma}_{ijk} + \epsilon^2 E_\theta[\partial_i \ell_\epsilon \partial_j \ell_\epsilon \partial_k \ell_\epsilon] = \partial_i \partial_j \partial_k \Psi^\epsilon(\theta). \end{aligned}$$

以下の議論においては, これらの量は $\mathcal{P}$ 上のみで値を評価する. 補題 3.1 から次の関係が導かれる.

補題 3.2 $\theta \in \Theta(\mathcal{P})$ とする. このとき $\eta_i, g_{ij}, \overset{\text{m}}{\Gamma}_{ijk}$ は $O(1)$ である.

次に, 以下の 3 つの条件が成り立つことを示す.

[A1] 十分統計量 s_i は

$$s_i(X) = \hat{\eta}_i + \epsilon^2 \xi_i^0 + o(\epsilon^2),$$

と書ける. ただし ξ_i^0 は決定論的な量であり, $\hat{\eta}_i$ は n 次元部分多様体 $S(\mathcal{M})$ ($n \leq N$) 上を動く確率変数である. g_{ij} の $S(\mathcal{M})$ への制限は正則である.

[A2] 統計量 $\hat{\eta}_i$ は漸近正規性を持つ:

$$\tilde{\eta}_i := \epsilon^{-1}(\hat{\eta}_i - \eta_i^0(u)) \rightarrow N(0, g_{ij}).$$

[A3] 全ての u に対して, $\Psi^\epsilon(\theta(u))$ は ϵ に依存しない.

Remark 1. $S(\mathcal{M})$ のことを**観測多様体**と呼ぶことにする. $n < N$ のとき, envelope family $\mathcal{E}$ 上では Fisher 情報量が退化する. このとき, 例えば $\theta \mapsto \eta^0(\theta)$ は一対一対応でないので, 注意を要する.

Remark 2. 条件 [A3] より強く, もし全ての $\theta \in \Theta(\mathcal{E})$ に対して $\Psi^\epsilon(\theta)$ が ϵ に依存しない場合, モデルは exponential dispersion model となる. しかし, これは一般には成り立たない.

補題 3.3 $\theta \in \Theta(\mathcal{P})$ とする. $\nu_p(\cdot)$ は一次独立とし, $\phi^p \nu_p(x_0) \neq 0$, $\partial_x(\bar{\nu}_p \bar{\nu}_q)(\cdot)$ ($p \leq q$) は一次独立とする. このとき条件 [A1]-[A3] は満たされる. 観測多様体は

$$S(\mathcal{M}) := \{(\eta_p(x), \eta_{pq}) \mid x \in \mathbf{R}, (\eta_{pq}) \in \mathbf{R}^{N-K}\}, \quad (10)$$

となる. ただし $\eta_p(x) := \int_{x_0}^x \bar{\nu}_p(z) dz$, $n := N - K + 1$ である. 特に, $K \geq 2$ ならば $n < N$ である.

証明 付録参照. □

例 3 (続き). $\phi(u) = (\cos(u), \sin(u))$, $\nu(x) = (-\sin(x), \cos(x))$ より, 十分統計量は

$$\begin{aligned} s_1(X) &= -\int_0^1 \sin(X_t) dX_t = \cos(X_t) + (\epsilon^2/2) \int_0^1 \cos(X_t) dt, \\ s_2(X) &= \int_0^1 \cos(X_t) dX_t = \sin(X_t) + (\epsilon^2/2) \int_0^1 \sin(X_t) dt, \\ s_{11}(X) &= \int_0^1 \sin^2(X_t) dt, \\ s_{12}(X) &= -\int_0^1 \sin(X_t) \cos(X_t) dt \end{aligned}$$

である. $s_{22}(X) = \int_0^1 \cos^2(X_t) dt = 1 - s_{11}(X)$ は不要である. 観測多様体は

$$S(\mathcal{M}) = \{(\eta_1, \eta_2, \eta_{11}, \eta_{12}) \in \mathbf{R}^4 \mid (\eta_1)^2 + (\eta_2)^2 = 1\}$$

(3次元部分多様体) で与えられる. □

以下の議論は微小拡散モデルに限らず, 条件 [A1]-[A3] を満たす任意のモデルについて成り立つ.

次の形の推定量を考える:

$$\hat{u}_\epsilon^a = u_\epsilon^a(s) = u_0^a(s) + \epsilon^2 \bar{u}^a(s) + o(\epsilon^2), \quad \text{for } s \in \mathbf{R}^N,$$

ただし u_0 と $\bar{u}$ は $\mathbf{R}^N$ から U への滑らかな写像とする. また, u_0 は Fisher 一致性を持つ, すなわち全ての $u \in U$ に対して

$$u^a = u_0^a(\eta^0(u))$$

が成り立つと仮定する。条件 [A1] より、この推定量は

$$\hat{u}_\epsilon^a = u_0^a(\hat{\eta}) + \epsilon^2((\partial^i u_0^a) \xi_i^0 + \bar{u}^a) + o(\epsilon^2), \quad (11)$$

と書ける。ここで、 $\partial^i := \partial/\partial\eta_i$, $\bar{u} := \bar{u}(\eta^0(u))$ とおいた。

推定多様体 $\mathcal{A}_u$ を

$$\mathcal{A}_u = \{\eta \in S(\mathcal{M}) \mid u_0(\eta) = u\},$$

によって定義する。これは $S(\mathcal{M})$ の $(n-m)$ 次元部分多様体である。各 u に対して $\mathcal{A}_u$ に座標 $v = (v^{m+1}, \dots, v^n)$ を導入し、 $w = (u, v)$ とおく。 w と $\eta \in S(\mathcal{M})$ が (少なくとも $\mathcal{P}$ の近くで) 一対一対応であるので、その対応を $w = w(\eta)$, あるいは $\eta = \eta(w)$ で表す。一般性を失うことなく全ての $u \in U$ に対して $\eta((u, 0)) = \eta^0(u)$ が成り立つと仮定することができる。また、 w の推定量 $\hat{w} = (\hat{u}, \hat{v})$ を $\eta(\hat{w}) = \hat{\eta}$ で定義する。 $\hat{w}$ は最尤推定量になる。 $\hat{w}$ の u -部分である $\hat{u}$ は、以上の定義から $u_0(\hat{\eta})$ に他ならない。

u に関する添え字を $\{a, \dots, f, a', \dots, f'\}$, v に関する添え字を $\{\kappa, \dots, \nu, \kappa', \dots, \nu'\}$, w に関する添え字を $\{\alpha, \beta, \dots, \delta, \alpha', \beta', \dots, \delta'\}$ とおく。 $S(\mathcal{M})$ 上の計量 $g_{\alpha\beta}$ は

$$g_{\alpha\beta} = (\partial_\alpha \theta^i) g_{ij} (\partial_\beta \theta^j)$$

と書ける。ただし、 $\partial_\kappa \theta^i$ は、次の方程式の一意的でない解の一つとする：

$$\partial_\kappa \eta_j = (\partial_\kappa \theta^i) g_{ij}. \quad (12)$$

計量 $g_{\alpha\beta}$ は well-defined である。なぜならば、(12) 式の一般解は適当な $r_\kappa^i \in \ker(g_{ij})$ を用いて $\partial_\kappa \theta^i + r_\kappa^i$ と書けるからである。条件 [A1] より $(g_{\alpha\beta})$ は正則となる。 $(g_{\alpha\beta})$ の逆行列を $(g^{\alpha\beta})$ と書く。

次の補題から、 $\hat{u}_\epsilon$ の平均二乗誤差を漸近展開したときの 1 次の項が求まる。

補題 3.4 条件 [A1]-[A3] を仮定する。このとき、 $\hat{u}_\epsilon$ の平均二乗誤差は

$$E[(\hat{u}_\epsilon^a - u^a)(\hat{u}_\epsilon^b - u^b)] = \epsilon^2 g^{ab} + o(\epsilon^2),$$

と展開される。ただし、 g^{ab} は $g^{\alpha\beta}$ の u -成分である。さらに、 g^{ab} が正定行列の意味で最小になるためには、 $g_{a\kappa} = 0$ at $\epsilon = 0$ であることが必要十分である。この条件が成り立つとき、 (g^{ab}) は (g_{ab}) の逆行列に一致する。

証明. $w = (u, 0)$, $\tilde{w} = \epsilon^{-1}(\hat{w} - w)$, $\tilde{\eta} = \epsilon^{-1}(\hat{\eta} - \eta)$ とおく。 $\theta \leftrightarrow \eta$ の対応が一対一でないことに注意しながら

$$\begin{aligned} g_{\alpha\beta} &= (\partial_\alpha \theta^i) (\partial_\beta \theta^j) g_{ij} \\ &= (\partial_\alpha \theta^i) (\partial_\beta \theta^j) E[\tilde{\eta}_i \tilde{\eta}_j] \\ &= (\partial_\alpha \theta^i) (\partial_\beta \theta^j) (\partial_\gamma \eta_i) (\partial_\delta \eta_j) E[\tilde{w}^\gamma \tilde{w}^\delta] \\ &= g_{\alpha\gamma} g_{\beta\delta} E[\tilde{w}^\gamma \tilde{w}^\delta]. \end{aligned}$$

が得られる。よって $E[\hat{w}^\gamma \hat{w}^\delta] = g^{\gamma\delta}$ となり、後は通常の方法 (例えば Amari, 1985) によって結論が得られる。 $\square$

ある Fisher 一致推定量の平均二乗誤差の主要項が、全ての Fisher 一致推定量の中で最小となるとき、その推定量は 1 次漸近有効、あるいは単に 1 次有効であるという。以後、 $\hat{u}_\epsilon^a$ は 1 次有効な推定量であると仮定する。

e-接続と m-接続の一部の成分は

$$\begin{aligned}\overset{e}{\Gamma}_{bc\alpha} &= (\partial_b \partial_c \theta^i)(\partial_\alpha \eta_i), \\ \overset{m}{\Gamma}_{\beta\gamma a} &= (\partial_\beta \partial_\gamma \eta_i)(\partial_a \theta^i),\end{aligned}$$

と表される。他の成分は、 $\partial_\kappa \theta^i$ が一意でないために決定できない。しかし、これらの成分の間では接続の双対性が満たされる：

$$\partial_a g_{b\gamma} = \overset{e}{\Gamma}_{ab\gamma} + \overset{m}{\Gamma}_{a\gamma b}.$$

$\overset{e}{\Gamma}_{ab\kappa}$ と $\overset{m}{H}_{\kappa\lambda a}$ は座標系の取り方に依存しない量になるので、特に $\overset{e}{H}_{ab\kappa} = \overset{e}{\Gamma}_{ab\kappa}$, $\overset{m}{H}_{\kappa\lambda a} = \overset{m}{\Gamma}_{\kappa\lambda a}$ と表す。

次の補題は、幾何学的な量を用いて 1 次有効推定量 $\hat{u}_\epsilon$ の漸近展開を求めたものである。

補題 3.5 [A1]-[A3] を仮定する。このとき、1 次有効推定量 $\hat{u}_\epsilon$ は

$$\hat{u}_\epsilon^a = u^a + \epsilon^2 g^{ab} \partial_b \ell_\epsilon(u) - \frac{\epsilon^2}{2} \overset{m}{\Gamma}_{\beta\gamma}{}^a \tilde{w}^\beta \tilde{w}^\gamma + \epsilon^2 (h^{ai} \xi_i^0 + \bar{u}^a) + o_P(\epsilon^2),$$

と漸近展開される。ただし $\tilde{w}^\alpha := \epsilon^{-1}(\hat{w}^\alpha - w^\alpha)$, $h^{ai} := \partial^i u_0^a - g^{ab} \partial_b \theta^i$ とする。

証明 スコア関数は次のように展開される：

$$\begin{aligned}\epsilon \partial_a \ell_\epsilon(u) &= \epsilon^{-1} \partial_a \theta^i (\hat{s}_i(X) - \eta_i^\epsilon(u)) \\ &= \epsilon^{-1} \partial_a \theta^i (\eta_i(\hat{w}) - \eta_i^\epsilon(u)) + \epsilon \partial_a \theta^i \xi_i^0 \\ &= \epsilon^{-1} \partial_a \theta^i \left(\eta_i^0(u) - \eta_i^\epsilon(u) + \epsilon (\partial_\alpha \eta_i) \tilde{w}^\alpha + \frac{\epsilon^2}{2} (\partial_\beta \partial_\gamma \eta_i) \tilde{w}^\beta \tilde{w}^\gamma \right) + \epsilon \partial_a \theta^i \xi_i^0 + o(\epsilon).\end{aligned}$$

条件 [A3] より $(\partial_a \theta^i) \eta_i^\epsilon(u)$ は ϵ に依存しない。よって (11) 式より結果が得られる。 $\square$

この補題を使って、バイアス補正した推定量の平均二乗誤差を漸近展開する。もし、あるバイアス補正した 1 次有効推定量が、そのようなクラスの推定量の中で平均二乗誤差の 2 次項を最小にするならば、その推定量を 2 次漸近有効、あるいは単に 2 次有効であるという。この 2 次漸近有効性の定義は Rao (1962), Efron (1975), Amari (1985) に従っている。

定理 3.1 $\hat{u}_\epsilon$ は 1 次漸近有効推定量とする. 補題 3.3 の条件を仮定する.

(i) $\hat{u}_\epsilon$ のバイアスは

$$\begin{aligned} b_\epsilon^a(u) &= E_u[\hat{u}_\epsilon^a] - u^a \\ &= -\frac{\epsilon^2}{2}(\bar{\Gamma}_{cd}^a g^{cd} + \bar{H}_{\kappa\lambda}^a g^{\kappa\lambda}) + \epsilon^2(h^{ai}\xi_i^0 + \bar{u}^a) + o(\epsilon^2), \end{aligned} \quad (13)$$

と漸近展開される. ただし h^{ai} は補題 3.5 で定義したものである.

(ii) バイアス補正推定量 $\hat{u}_\epsilon^{*a} = \hat{u}_\epsilon^a - b_\epsilon^a(\hat{u}_\epsilon)$ の平均二乗誤差は

$$E_u[(\hat{u}_\epsilon^{*a} - u^a)(\hat{u}_\epsilon^{*b} - u^b)] = \epsilon^2 g^{0ab} + \frac{\epsilon^4}{2} C^{2ab} + o(\epsilon^4) \quad (14)$$

と漸近展開される. ただし

$$\begin{aligned} C^{2ab} &= -g^{0ac} g^{0bd} g_{cd}^{(2)} + (\bar{\Gamma})^{2ab} + 2(\bar{H}_P)^{2ab} + (\bar{H}_A)^{2ab}, \\ (\bar{\Gamma})^{2ab} &= \bar{\Gamma}_{cd}^a \bar{\Gamma}_{ef}^b g^{ce} g^{df}, \\ (\bar{H}_P)^{2ab} &= \bar{H}_{cd\kappa} \bar{H}_{ef\lambda} g^{ac} g^{be} g^{df} g^{\kappa\lambda}, \\ (\bar{H}_A)^{2ab} &= \bar{H}_{\kappa\lambda}^a \bar{H}_{\mu\nu}^b g^{\kappa\mu} g^{\lambda\nu}, \end{aligned}$$

とし, $g_{ab}^{(2)}$ は g_{ab}^ϵ の 2 次項を表すものとする (定理 2.1 参照). さらに, $\hat{u}_\epsilon^*$ が 2 次有効であるための必要十分条件は $(\bar{H}_A)^{2ab} = 0$ at $\epsilon = 0$ となることである.

証明 (i) 補題 3.5 より成り立つ.

(ii) バイアス補正推定量 $\hat{u}_\epsilon^*$ は

$$\begin{aligned} \hat{u}_\epsilon^{*a} &= u^a + \epsilon^2 g^{ab} \partial_b \ell_\epsilon - \frac{\epsilon^2}{2} \left(\bar{\Gamma}_{cd}^a g^{cc'} g^{dd'} (\tilde{u}_{c'} \tilde{u}_{d'} - g_{c'd'}) + 2g^{ab} \bar{H}_{bc\kappa} g^{cc'} g^{\kappa\kappa'} \tilde{u}_{c'} \tilde{u}_{\kappa'} \right. \\ &\quad \left. + \bar{H}_{\kappa\lambda}^a g^{\kappa\kappa'} g^{\lambda\lambda'} (\tilde{v}_{\kappa'} \tilde{v}_{\lambda'} - g_{\kappa'\lambda'}) \right) + o_P(\epsilon^2). \end{aligned}$$

と漸近展開される. 公式

$$\begin{aligned} E_u[(\hat{u}_\epsilon^{*a} - u^a)(\hat{u}_\epsilon^{*b} - u^b)] &= \epsilon^2 g^{ab} + E_u[(\hat{u}_\epsilon^{*a} - u^a - g^{ac} \partial_c \ell_\epsilon)(\hat{u}_\epsilon^{*b} - u^b - g^{bd} \partial_d \ell_\epsilon)], \end{aligned}$$

が成り立つ. $\tilde{w}_\alpha$ の漸近正規性 (条件 [A2]) から

$$E_u[\tilde{w}_\alpha \tilde{w}_\beta \tilde{w}_\gamma \tilde{w}_\delta] = g_{\alpha\beta} g_{\gamma\delta} + g_{\alpha\gamma} g_{\beta\delta} + g_{\alpha\delta} g_{\beta\gamma} + o(1),$$

が成り立つ. 以上を用いて

$$E_u[(\hat{u}_\epsilon^{*a} - u^a)(\hat{u}_\epsilon^{*b} - u^b)] = \epsilon^2 g^{ab} + \frac{\epsilon^4}{2} \left((\bar{\Gamma})^{2ab} + 2(\bar{H}_P)^{2ab} + (\bar{H}_A)^{2ab} \right) + o(\epsilon^4) \quad (15)$$

となる. (14) 式は (15) 式から直ちに得られる. 3 つの項 $(\bar{\Gamma})^{2ab}$, $(\bar{H}_P)^{2ab}$, $(\bar{H}_A)^{2ab}$ は定義より非負定値である. C^{2ab} の 4 つの項のうち, $(\bar{H}_A)^{2ab}$ だけが推定量の取り方に依存するので, $\hat{u}_\epsilon^*$ が 2 次有効であるためには $(\bar{H}_A)^{2ab} = 0$ が必要十分である. $\square$

系 3.1 $\hat{u}_\epsilon$ を最尤推定量 (MLE) とする. このとき $\bar{u}^a = 0$, $h^{ai} = 0$, $\bar{H}_{\kappa\lambda a} = 0$ となる. 特に, バイアス補正 MLE は 2 次有効である.

証明 MLE は尤度方程式 $\partial_a \theta^i(\hat{u})(\hat{\eta}_i - \eta_i^\epsilon(\hat{u})) = 0$ の解である. $(\partial_a \theta^i) \eta_i^\epsilon$ は ϵ に依存しないので $\bar{u}^a = 0$ となる. 2 番目の式 $h^{ai} = 0$ は直接計算して得られる. 3 番目の式 $\bar{H}_{\kappa\lambda a} = (\partial_\kappa \partial_\lambda \eta_i)(\partial_a \theta^i) = 0$ は, $\mathcal{A}_u = S(\mathcal{M}) \cap \ker(\partial_a \theta^i) \subset \ker(\partial_a \theta^i)$ であることから成り立つ. $\square$

統計量 $\hat{t}$ の情報量損失は $\Delta g_{ab} := \epsilon^{-2} g_{ab} - E[\partial_a \ell_\epsilon^t \partial_b \ell_\epsilon^t]$ によって定義される. ここで ℓ_ϵ^t は $\hat{t}$ の周辺分布である.

系 3.2 1 次有効推定量 $\hat{u}_\epsilon$ の情報量損失は

$$\Delta g_{ab} = \frac{1}{2} \left(2(\bar{H}_{\mathcal{P}})_{ab}^2 + (\bar{H}_{\mathcal{A}})_{ab}^2 \right) + o(\epsilon^2),$$

で与えられる. ここで, $(\bar{H}_{\mathcal{P}})_{ab}^2 := \bar{H}_{ac\kappa} \bar{H}_{bd\lambda} g^{cd} g^{\kappa\lambda}$, $(\bar{H}_{\mathcal{A}})_{ab}^2 := \bar{H}_{\kappa\lambda a} \bar{H}_{\mu\nu b} g^{\kappa\mu} g^{\lambda\nu}$ である.

証明 情報量損失は次式から求まる (Amari 1985):

$$\Delta g_{ab} = E[\text{Cov}[\partial_a \ell_\epsilon, \partial_b \ell_\epsilon | \hat{u}_\epsilon]].$$

結果は補題 3.5 から得られる. $\square$

定理 3.1, 系 3.2 より, 統計的曲率 $(g^{ab}(\bar{H}_{\mathcal{P}})_{ab}^2)$ の大きさが推定量の性質に影響を与えることが分かる.

例 1 (続き). X は (4) 式で定義される Ornstein-Uhlenbeck 過程とする. envelope family の対数尤度関数は

$$\ell_\epsilon(X, \theta) = \frac{1}{\epsilon^2} \left(\theta_1 \left(-\int_0^1 X_t dX_t \right) + \theta_2 \left(\int_0^1 X_t^2 dt \right) - \Psi^\epsilon(\theta) \right),$$

である (元のモデルは $\theta_1 = u$, $\theta_2 = -u^2/2$). ここでポテンシャル関数は

$$\Psi^\epsilon(\theta) = \frac{(\theta_1)^2 - z^2}{2(z \coth z + \theta_1)} - \frac{\epsilon^2}{2} \log \left(\cosh z + \frac{\theta_1 \sinh z}{z} \right) + \frac{\epsilon^2}{2} \theta_1,$$

$z := \sqrt{-2\theta_2}$, で与えられる (See Küchler & Sørensen, 1996). Ψ の定義域は

$$\Theta = \{(\theta_1, \theta_2) \mid \theta_2 < \frac{\pi^2}{2} \text{ and } \theta_1 + z \coth z > 0\}.$$

である. このモデルでは, 観測多様体 $S(\mathcal{M})$ は $\mathbf{R}^2$ 全体であり, $n = N$ となる. 統計モデルは, $\theta_1(u) = u$, $\theta_2(u) = -u^2/2$ で与えられる部分多様体となる. Ψ^ϵ の主要項 Ψ^0 は

$$\Psi^0(\theta) = \frac{(\theta_1)^2 - z^2}{2f},$$

$f = f(\theta_1, z) := z \coth z + \theta_1$, である. $\epsilon = 0$ における期待値パラメータは

$$\begin{aligned}\eta_1(\theta) &= \frac{\theta_1}{f} - \frac{\theta_1^2 - z^2}{2f^2}, \\ \eta_2(\theta) &= -\frac{1}{z} \frac{\partial \Psi^0}{\partial z} = \frac{1}{f} - \frac{(\theta_1^2 - z^2)(z^{-1} \coth z - 1/\sinh^2 z)}{f^2}.\end{aligned}$$

$\theta = \theta(u)$ では, これらは

$$\begin{aligned}\eta_1(u) &= (1 - e^{-2u})/2, \\ \eta_2(u) &= (1 - e^{-2u})/(2u).\end{aligned}$$

となる. Fisher 情報行列 (g_{ij}) は, $\theta = \theta(u)$ において

$$\begin{aligned}g_{11} &= \frac{-1 + \coth(u)}{u(1 + \coth(u))^2}, \\ g_{12} &= \frac{-1 + \coth(u) - u \operatorname{csch}^2(u)}{u^2(1 + \coth^2(u))}, \\ g_{22} &= \frac{2(\coth(u) - u/\operatorname{csch}^2(u))}{u^3(1 + \coth(u))^2}.\end{aligned}$$

で与えられる. 差テンソル $T_{ijk} = \partial_i \partial_j \partial_k \Psi^0$ は, $\theta = \theta(u)$ において

$$\begin{aligned}T_{111} &= [-3(-1 + \coth(u))]/[u^2(1 + \coth(u))^3], \\ T_{112} &= [2 + \coth^2(u) + 3u \operatorname{csch}^2(u) - \coth(u)(3 + u \operatorname{csch}^2(u))]/[u^3(1 + \coth(u))^3], \\ T_{122} &= [u \operatorname{csch}^2(u)(5 + 2u \operatorname{csch}^2(u)) + \coth^2(u)(3 - 2u^2 \operatorname{csch}^2(u)) \\ &\quad - \coth(u)(3 + u(3 + 2u) \operatorname{csch}^2(u))]/[u^4(1 + \coth(u))^3], \\ T_{222} &= [3 \operatorname{csch}^3(u)(-((1 + 12u + 8u^2) \cosh(u)) + \cosh(3u) \\ &\quad + 2(3 + 2u - 4u^2 + 3 \cosh(2u)) \sinh(u))]/[4u^5(1 + \coth(u))^3].\end{aligned}$$

で与えられる.

ここで, 次式で定義される推定量 $\hat{u}_e$ を考える:

$$\hat{u}_e := \operatorname{argmin}_u E_{\theta(u)}[\ell_\epsilon(X, \theta(u)) - \ell_\epsilon(X, \hat{\theta})] = \operatorname{argmin}_u \operatorname{KL}(\theta(u) \parallel \hat{\theta})$$

ただし $\hat{\theta}$ は $\eta(\hat{\theta}) = \hat{\eta}$ で定義する. 推定量 $\hat{u}_e$ を e-射影推定量 (e-projection estimator; EPE) と呼ぶことにし, MLE と比較する. EPE の推定方程式は

$$\partial_a \eta_i(\hat{u}_e) (\theta^i(\hat{u}_e) - \hat{\theta}^i) = 0,$$

である ($a = 1$). この式から EPE が Fisher 一貫性を持つことが分かる. 係数は

$$\begin{aligned}\partial_a \eta_1(u) &= e^{-2u}, \\ \partial_a \eta_2(u) &= \frac{e^{-2u}}{u} - \frac{1 - e^{-2u}}{2u^2}.\end{aligned}$$

となる。EPE の推定多様体は自然パラメータ表示では

$$\Theta(\mathcal{A}_u) = \{(\theta_1, \theta_2) \mid \theta_1 = u - (\partial_a \eta_2(u))v, \theta_2 = -u^2/2 + (\partial_a \eta_1(u))v\}.$$

となる。 $\mathcal{A}_u$ の埋め込み e-曲率はゼロになるので、 $\mathcal{A}_u$ の埋め込み m-曲率は

$$\overset{\text{m}}{H}_{\kappa\lambda a} = B_{\kappa}^i B_{\lambda}^j B_a^k T_{ijk},$$

で求まる。ここで $B_a^i = \partial_a \theta^i$, $B_{\kappa}^i = \partial_{\kappa} \theta^i$ とする。MLE と EPE に対応する推定多様体を図 2 に示す (期待値パラメータ表示)。

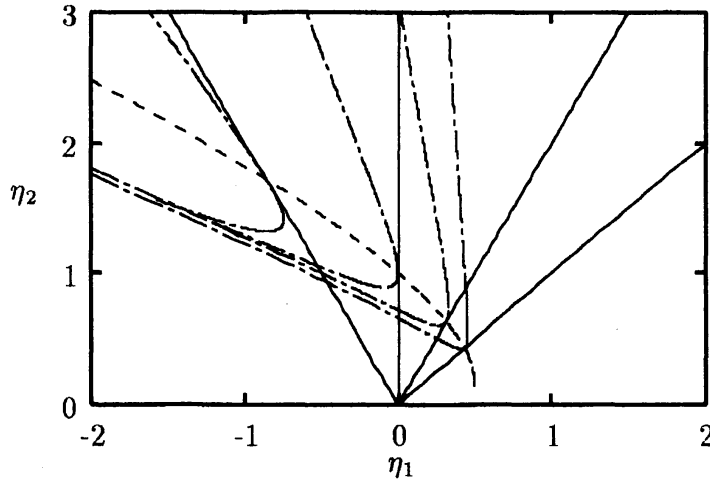


Figure 2: 例 1 に関して、推定多様体を期待値パラメータ空間で描いた図である。実線、1 点鎖線はそれぞれ MLE, EPE に対応する推定多様体を示す。破線は $\epsilon = 0$ におけるモデル $S(\mathcal{P})$ を表す。

MLE と EPE の平均二乗誤差をモンテカルロ法で比較する。真のパラメータは $u = 1.5$ とする。拡散係数のパラメータは $\epsilon = 0.1$ で既知とする。観測回数 (パス X の生成回数) は $\text{OBS} = 10000$ 回とし、観測時点の個数は $\text{STEP} = 65536$ とする。データは、公式 $X_t = e^{-2ut}(x_0 + \epsilon B_{(e^{2ut}-1)/2u})$ (B は標準ブラウン運動) によって生成した。結果を表 1 に示す。漸近展開による理論的な結果と実験結果は、少なくとも MLE が EPE を優越している点では一致している。 $\square$

Table 1: MLE と EPE の平均二乗誤差の理論値と実験値を示す。真のパラメータは $u = 1.5$ とし、微小拡散パラメータは $\epsilon = 0.1$ とした。数値計算のための各パラメータは $\text{OBS} = 10000$ (観測回数), $\text{STEP} = 65536$ (観測時点の個数), 信頼区間は正規分布の 95% 区間である。

	MLE	EPE
誤差 (理論値; 1 次のオーダーまで)	0.03157	0.03157
誤差 (理論値; 2 次のオーダーまで)	0.03043	0.03225
誤差 (実験値)	0.03335 ± 0.00122	0.03657 ± 0.00154

付録

補題 3.3 の証明

補題 3.3 の証明 対数尤度関数は (8) と表されるから、条件 [A3] は明らかに成り立つ。もし $\theta = \theta(u)$ ならば、(9) 式の最初の K 成分は

$$s_p(X) = \int_{x_0}^{X_T} \bar{\nu}_p \sigma^{-1}(z) dz - \frac{\epsilon^2}{2} \int_0^T [\partial_x (\bar{\nu}_p \sigma^{-1}) \sigma^2]_t^\epsilon dt \quad (16)$$

for $p = 1, \dots, K$, となる。ただし伊藤の公式を用いた。(9) 式の後半の $K(K+1)/2$ 成分は

$$s_{pq}(X) = \int_0^T [\bar{\nu}_p \bar{\nu}_q]_t^0 dt + \epsilon \int_0^T \int_t^T [\partial_x (\bar{\nu}_p \bar{\nu}_q)]_s^0 Y_s ds Y_t^{-1} [\sigma]_t^0 dW_t + o_P(\epsilon) \quad (17)$$

for $1 \leq p \leq q \leq K$, となる。ただし補題 2.2 を用いた。(16) 式と (17) 式から [A2] が成り立ち、観測多様体は (10) となる。後は、計量の $S(\mathcal{M})$ への制限が正則であることを示せばよい。 λ , $(\lambda^{pq})_{1 \leq p, q \leq K}$ をそれぞれ、次式を満たすような任意の実数、任意の実対称行列とする：

$$\text{Var} [\epsilon^{-1} (\lambda X_T + \lambda^{pq} s_{pq})] \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0,$$

ここで (s_{pq}) は (17) 式から定まる対称行列である。すると

$$\lambda Y_T + \lambda^{pq} \int_t^T [\partial_x (\bar{\nu}_p \bar{\nu}_q)]_s^0 Y_s ds = 0 \quad (18)$$

for any $t \in [0, T]$, が成り立つ。両辺を t で微分すると

$$\lambda^{pq} [\partial_x (\bar{\nu}_p \bar{\nu}_q)]_t^0 = 0$$

となる。 $([\partial_x (\bar{\nu}_p \bar{\nu}_q)]_t^0)_{1 \leq p \leq q \leq K}$ は一次独立だから、 $\lambda^{pq} = 0$ for all p, q が成り立つ。よって方程式 (18) は

$$\lambda Y_T = 0,$$

となり、 $\lambda = 0$ が示された。 $\square$

References

- [1] Amari, S. (1985) *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics 28, Springer-Verlag, Berlin.
- [2] Efron, B. (1975) Defining the curvature of a statistical problem (with applications to second order efficiency), *Ann. Statist.*, **3**, 1189-1242.
- [3] Gihman, I. I. and Skorohod, A. V. (1972) *Stochastic Differential Equations*, Springer-Verlag, Berlin.
- [4] Küchler, U. and Sørensen, M. (1996) Curved exponential families of stochastic processes and their envelope families, *Ann. Inst. Statist. Math.*, **48**, pp. 61-74.
- [5] Küchler, U. and Sørensen, M. (1997) *Exponential Families of Stochastic Processes*, Springer Series in Statistics, Springer-Verlag, New York.
- [6] Kutoyants, Yu. (1994) *Identification of Dynamical Systems with Small Noise*, Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht; Boston.
- [7] Lipster, R. S. and Shiryaev, A. N. (2001) *Statistics of Random Processes I, General Theory*, Springer-Verlag, Berlin.
- [8] Rao, C. R. (1962) Efficient estimates and optimum inference procedures in large samples, *J. Roy. Statist. Soc. Ser. B*, **24**, 46-72.
- [9] Sei, T. and Komaki, F. (2003) Information geometry of small diffusions, submitted.
- [10] Taniguchi, M. and Watanabe, Y. (1994) Statistical analysis of curved probability densities, *J. Multivariate Anal.*, **48**, 228-248.
- [11] Yoshida, N. (1992a) Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe, *Probab. Theory Related Fields*, **92**, 275-311.
- [12] Yoshida, N. (1992b) Asymptotic expansion for statistics related to small diffusions, *J. Japan Statist. Soc.*, **22**, 139-159.