

局所線形化法による確率微分方程式の パラメータ推定

筑波大学* 庄 司 功
統計数理研究所 尾 崎 統

(1996年6月 受付)

1. はじめに

時間と共に変動する確率的な現象をモデル化するために、確率微分方程式を用いることは有効な手段の一つである。この時、対象となる現象を適切に表現するには、確率微分方程式に含まれるパラメータの値が極めて重要な役割を演じる。しかし、パラメータの値が事前に知られていることは、実際上はほとんど期待できない。むしろ、こうしたパラメータは観測されたデータから推定することが普通である。したがって、人々は、手持ちのデータから如何に効率よくパラメータを推定するかに関心を寄せている。ここで注意すべきは、観測されたデータはほとんどすべての場合、ある時間間隔を持った離散時間の時系列データである。したがって、実用の観点からは、確率微分方程式のパラメータ推定といえども、離散データを前提にせざるをえないのである。

離散データから確率微分方程式のパラメータを推定しようとする試みには、Bergstrom (1990), Dembo and Zeitouni (1987), Harvey and Stock (1985), Jones and Tyron (1984), Ozaki (1985), Singer (1993), Yoshida (1992) などが既に知られている。このうち、Ozaki (1985) と Yoshida (1992) を除いた研究では、線形な確率微分方程式を対象としてパラメータの推定を論じるものばかりである。確かに、線形な確率微分方程式は、離散化モデルが容易に導けることや離散データに対応した尤度関数が容易に導けることなど、取り扱いやすさにおいては大変便利な方程式である。しかし、現実の複雑な現象を記述するには能力不足である点は否めない。それに対して Ozaki (1985) や Yoshida (1992) は、線形はもちろん非線形な確率微分方程式までも範疇とする方法を提案している。両者は共に最尤法を用いてパラメータを推定しようとする方法であるが、Yoshida (1992) では、目的関数を構成する連続量をオイラー法を用いて離散近似しているのに対して、Ozaki (1985) では局所線形化法と呼ばれる方法を提案し、離散化モデルの導出に工夫を凝らしている。実際、Ozaki (1992a, 1992b, 1993) には、局所線形化法に基づく推定のパフォーマンスについて数値実験から極めて良好な結果が報告されている。したがって、我々は、Ozaki (1985) が開発した局所線形化法に注目したい。

局所線形化法の基本的なアイデアは、所与の確率微分方程式を局所的に線形な確率微分方程式で近似することである。上に述べたように、線形な確率微分方程式は、極めて扱いやすいために、離散的なサンプルパスからパラメータを推定することが容易である。したがって、例えば、遷移確率密度関数を求めることが困難な非線形な確率微分方程式であっても、線形な確率微分方程式を局所的につなぎ合わせることによってパラメータ推定が可能となる。

局所線形化法は、パラメータ推定において次のような利点がある。まず第1に、局所線形化

* 社会工学系：〒305 つくば市天王台 1-1-1.

法から導かれた離散化モデルは、よく用いられるオイラー法から導かれるそれよりも効率的である。そもそも、オイラー法では、確率微分方程式のドリフト項の係数が局所的に定数であるとして離散化モデルを導いているのに対して、局所線形化法では、線形近似しているために、局所線形化法の離散化モデルがより効率的であることは容易に想像がつく。もちろん、オイラー法よりも効率的に離散化モデルを導く方法として陰的なスキームを用いるルンゲ・クッタ法が知られているが、その陰的なスキームのゆえに、本稿で問題にしているパラメータ推定には用いることができない。第2に、線形な確率微分方程式の遷移確率密度関数は正規であるので、局所線形化法の離散化モデルの尤度関数は容易に求めることができる。したがって、パラメータの推定値は最尤法により求められる。最後に、局所線形化法の離散化モデルは、オイラー法のそれよりも数値的に安定であると Ozaki (1993) によって指摘されている。つまり、所与の確率微分方程式を満たす確率過程が発散しない場合でも、オイラー法の離散化モデルでは数値的に発散してしまう場合があるのに対して、局所線形化法の離散化モデルではそのようなことが起こらないのである。

このような利点があるにも拘わらず、局所線形化法にはまだ解決しなければならない問題点がある。もともと、局所線形化法は、次のような確率微分方程式を、

$$dx = f(x)dt + \sigma dB,$$

次のような線形な確率微分方程式で近似することであった。

$$dx = Lxdt + \sigma dB.$$

したがって、上のような近似が意味を持つには、 $f(0)=0$ であることが必要である。逆に言うと、 $f(0) \neq 0$ であれば、近似が意味を成さなくなってしまうのである。もちろん、上のような線形な確率微分方程式の代わりに次のような確率微分方程式を選択することも可能であるが、

$$dx = (Lx + f(0))dt + \sigma dB$$

今度は、 $f(0) < \infty$ であることが必要になってくる。この点で、局所線形化法の適用範囲が制限されることになる。更に、係数を決めるために所与の確率微分方程式を満たす確率過程の条件付き期待値が、拡散項を除いた常微分方程式から導かれる軌道に等しいという条件が加わり、局所線形化法の適用範囲がますます制限されることになる。

本論文の目的は、局所線形化法の持つ上のような制限を取り除き、より広範囲な確率微分方程式に適用可能な新たな局所線形化法を提案することにある。新たに提案する方法はオリジナルな方法と線形近似する方法が基本的に異なる。具体的には、オリジナルな方法では、ドリフト関数の線形近似を常微分方程式の枠組みで行っており、その結果、プロセスの持つ確率的な振る舞いが除外されている。他方、新たな方法では、伊藤の公式を通じてプロセスの持つ確率的な振る舞いを考慮し、そこからドリフト関数を線形近似している。この線形化方法によって、オリジナルな方法が抱えている上のような制限が結果的に取り除かれることになる。

また、本稿では、新たな方法によるパラメータ推定がどの程度うまくいくかを数値実験によって評価する。数値実験では、非線形な確率微分方程式を対象にして、オリジナルな局所線形化法とオイラー法とを比較対象として選び、3者の推定のパフォーマンスを比較評価する。更に、サンプリングの時間間隔の違いによって、推定のパフォーマンスがどのような影響を受けるかという点は、実用上重要であるため、幾つかのサンプリング間隔を選んで、推定のパフォーマンスの違いを調べることにした。

2. 新たな局所線形化法

まず最初に、次のような1次元の幾分一般的な確率微分方程式を考察する。

$$(2.1) \quad dx_t = f(x_t, t)dt + g(x_t)dB_t,$$

ここで、 $f(x_t, t)$ は第1の変数について2回連続的微分可能で第2の変数について1回連続的微分可能な関数であり、 $g(x_t)$ は変数について1回連続的微分可能な関数である。また、 B_t は標準的ブラウン運動を表わす。ここで、変数変換 $y_t = \varphi(x_t)$ を考えると、伊藤の公式から y_t に関する確率微分方程式が以下のように得られる。

$$(2.2) \quad dy_t = \left(f \frac{d\varphi}{dx} + \frac{g^2}{2} \frac{d^2\varphi}{dx^2} \right) dt + \sigma g \frac{d\varphi}{dx} dB_t.$$

$\varphi(x)$ が微分方程式 $gd\varphi/dx = \sigma$ を満足すれば、上記の(2.1)の確率微分方程式は、拡散項の係数が定数となる。したがって、もとの確率微分方程式は、上のような変数変換によって、拡散項の係数が定数となるので、今後は、特に断らない限り以下のような確率微分方程式を考察する。

$$(2.3) \quad dx_t = f(x_t, t)dt + \sigma dB_t.$$

さて、局所線形化法とは、局所的に、即ち、時刻 t がある微小時間間隔 $[s, s+\Delta t)$ にある時に、ドリフト関数 $f(x_t, t)$ を x_t の線形関数で近似する方法であったから、まず始めに、 $f(x_t, t)$ の局所的な振る舞いに注意を向ける必要がある。そして、この局所的な振る舞いは、 $f(x_t, t)$ の微分によって明らかにされ、この微分は伊藤の公式によって特徴づけられる。即ち、

$$(2.4) \quad df = \left(\frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x^2} + \frac{\partial f}{\partial t} \right) dt + \frac{\partial f}{\partial x} dx$$

が得られる。 f を線形近似するためには、オイラー法の時と同様に、 $t \in [s, s+\Delta t)$ において dx , dt の係数 $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial t}$, が定数であると仮定すればよい。この仮定の下に、 f の線形近似が次のように求められる。

$$(2.5) \quad f(x_t, t) = f(x_s, s) + \left(\frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x^2} + \frac{\partial f}{\partial t} \right) (t-s) + \frac{\partial f}{\partial x} (x_t - x_s).$$

この近似式を用いて、確率微分方程式(2.3)を以下の確率微分方程式で近似する。

$$(2.6) \quad dx_t = (L_s x_t + M_s t + N_s)dt + \sigma dB_t.$$

但し、

$$\begin{aligned} L_s &= \frac{\partial f}{\partial x}(x_s, s), \\ M_s &= \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x^2}(x_s, s) + \frac{\partial f}{\partial t}(x_s, s), \\ N_s &= f(x_s, s) - \frac{\partial f}{\partial x}(x_s, s)x_s - \left(\frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x^2}(x_s, s) + \frac{\partial f}{\partial t}(x_s, s) \right) s. \end{aligned}$$

上で求められた確率微分方程式は線形なので、その解は直ちに求められて、

$$(2.7) \quad x_t = x_s + \frac{f(x_s, s)}{L_s} (e^{L_s(t-s)} - 1) + \frac{M_s}{L_s^2} (e^{L_s(t-s)} - 1 - L_s(t-s)) \\ + \sigma \int_s^t e^{L_s(t-u)} dB_u,$$

となる。これが、求める離散化モデルとなる。そして、よく知られているように、第4項の確率積分は平均が0で分散が

$$(2.8) \quad \text{Var}_s(x_t) = \sigma^2 \left(\frac{e^{2L_s(t-s)} - 1}{2L_s} \right)$$

の正規分布に従う。したがって、離散化モデル自体も正規分布に従うことになる。

Shoji (1995) によれば、離散時間間隔が限りなく0に近づく時、離散化モデルはローカルエラー、グローバルエラー共に L_2 ノルムで0に収束することが報告されており、この意味で、近似離散化モデルは一致性を持っているといえる。

ここで、上で求められた離散化モデルを、オリジナルな局所線形化法から求められる離散化モデルやオイラー法から求められる離散化モデルと比較してみる。まず、オリジナルな局所線形化法では離散化モデルは以下のようなようになる。

$$(2.9) \quad x_t = x_s + \frac{f(x_s, s)}{L_s} (e^{L_s(t-s)} - 1) + \sigma \int_s^t e^{K_s(t-u)} dB_u,$$

ここで、

$$\text{Var}_s(x_t) = \sigma^2 \left(\frac{e^{2K_s(t-s)} - 1}{2K_s} \right), \\ K_s = \frac{1}{t-s} \log \left(1 + \frac{f(x_s, s)}{x_s L_s} (e^{L_s(t-s)} - 1) \right).$$

次に、オイラー法では以下のようなになる。

$$(2.10) \quad x_t = x_s + f(x_s, s)(t-s) + \sigma \int_s^t dB_u,$$

ここで、

$$\text{Var}_s(x_t) = \sigma^2(t-s).$$

ここで、3つの離散化モデルの違いを分散に注目して見てみる。オイラー法の分散はプロセスの状態によらず一定であるのに対して、両局所線形化法の分散はプロセスの状態に依存して変わる。両局所線形化法の分散の違いは、形式的には、 L_s と K_s の違いとして現れる。この違いは、ドリフト関数の線形近似の方法の違いに由来するものである。即ち、オリジナルな方法では、プロセスの局所的な確率的変動を無視しているのに対して、新たな方法では、伊藤の公式を利用して局所的な確率的変動を考慮しており、これが L_s と K_s の違いとなって現れるのである。3つの離散化モデルの分散の違いは、特に拡散項の係数を推定する場合に顕著に現れることが、後述の数値実験から示される。

例. 3つの離散化モデルの違いをより具体的に見るために、次のような非線形な確率微分方程式を取り上げる：

$$(2.11) \quad \begin{aligned} dx_t &= f(x_t)dt + \sigma dB_t, \\ f(x_t) &= -x_t^3, \end{aligned}$$

ここで、 σ は定数である。新たな局所線形化法では次のような離散化モデルが得られる。

$$(2.12) \quad \begin{aligned} x_t &= x_s + \frac{x_s}{3}(e^{-3x_s^2(t-s)} - 1) - \frac{\sigma^2}{3x_s^3}((e^{-3x_s^2(t-s)} - 1) + 3x_s^2(t-s)) \\ &\quad + \sigma \int_s^t e^{-3x_s^2(t-u)} dB_u, \end{aligned}$$

ここで、分散は、

$$\text{Var}_s(x_t) = \sigma^2 \left(\frac{e^{-6x_s^2(t-s)} - 1}{-6x_s^2} \right),$$

となる。オリジナルな局所線形化法では次のような離散化モデルが得られる。

$$(2.13) \quad x_t = x_s + \frac{x_s}{3}(e^{-3x_s^2(t-s)} - 1) + \sigma \int_s^t e^{K_s(t-u)} dB_u,$$

分散は、

$$\begin{aligned} \text{Var}_s(x_t) &= \sigma^2 \left(\frac{e^{2K_s(t-s)} - 1}{2K_s} \right), \\ K_s &= \frac{1}{t-s} \log \left(1 + \frac{1}{3}(e^{-3x_s^2(t-s)} - 1) \right), \end{aligned}$$

となる。オイラー法では次のような離散化モデルが得られる。

$$(2.14) \quad x_t = x_s - x_s^3(t-s) + \sigma \int_s^t dB_u,$$

分散は、

$$\text{Var}_s(x_t) = \sigma^2(t-s)$$

となる。

3. 離散化モデルの推定

前節で述べたように、新たな方法から得られる離散化モデルは局所的に正規分布に従うため、離散データに対応する尤度関数は直ちに得られ、したがって、パラメータの推定値は最尤法によって容易に求めることができる。

今、等間隔に N 個の時系列データ $x_1, \dots, x_N$ が与えられたとして、その同時確率密度関数を $p(x_1, \dots, x_N)$ で表す。マルコフ性を用いれば、対数尤度関数は次のようになる。

$$(3.1) \quad \begin{aligned} \log(p(x_1, \dots, x_N)) &= \sum_{n=1}^{N-1} \log(p(x_{n+1} | x_n)) + \log(p(x_1)), \\ &= -\frac{1}{2} \sum_{n=1}^{N-1} \left\{ \frac{(x_{n+1} - E_n)^2}{V_n} + \log(2\pi V_n) \right\} + \log(p(x_1)), \end{aligned}$$

但し、

$$\begin{aligned}
E_n &= x_n + \frac{f(x_n, t_n)}{L_n} (e^{L_n \Delta t} - 1) + \frac{M_n}{L_n^2} \{ (e^{L_n \Delta t} - 1) - L_n \Delta t \}, \\
V_n &= \frac{(e^{2L_n \Delta t} - 1)}{2L_n} \sigma^2, \\
L_n &= \frac{\partial f}{\partial x}(x_n, t_n), \\
M_n &= \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x^2}(x_n, t_n) + \frac{\partial f}{\partial t}(x_n, t_n).
\end{aligned}$$

さて、ここで、もう一度確率微分方程式(2.1)に注目する。前節で述べたように、(2.1)は適当な変数変換を用いることによって拡散項の係数が定数となるような確率微分方程式に変換できる。したがって、このようなより一般的な形をした確率微分方程式に対しても、本稿の局所線形化法が適用できる。但し、変数変換を用いた場合には、尤度関数が若干変更されることに注意する必要がある。今、観測データ x_n に対して、それに変数変換 φ を施したものを y_n とし、それぞれの尤度を、 $p(x_1, \dots, x_N)$, $p(y_1, \dots, y_N)$ とする。確率密度関数の変換規則によって次を得る。

$$(3.2) \quad p(x_1, \dots, x_N) = p(y_1, \dots, y_N) \left| \frac{\partial(y_1, \dots, y_N)}{\partial(x_1, \dots, x_N)} \right|,$$

ここで、

$$\left| \frac{\partial(y_1, \dots, y_N)}{\partial(x_1, \dots, x_N)} \right|$$

はヤコビアンを表わし、これは

$$\prod_{n=1}^N \left| \frac{d\varphi}{dx} \right|_{x=x_n}$$

に等しい。したがって、元の確率過程に対する対数尤度関数は次のように表わすことができる。

$$\begin{aligned}
(3.3) \quad \log(p(x_1, \dots, x_N)) &= -\frac{1}{2} \sum_{n=1}^{N-1} \left\{ \frac{(y_{n+1} - E_n)^2}{V_n} + \log(2\pi V_n) \right\} \\
&\quad + \log(p(y_1)) + \sum_{n=1}^N \log \left(\left| \frac{d\varphi}{dx} \right|_{x=x_n} \right),
\end{aligned}$$

ここで、

$$\begin{aligned}
E_n &= y_n + \frac{\tilde{f}(y_n, t_n)}{L_n} (e^{L_n \Delta t} - 1) + \frac{M_n}{L_n^2} \{ (e^{L_n \Delta t} - 1) - L_n \Delta t \}, \\
V_n &= \frac{(e^{2L_n \Delta t} - 1)}{2L_n} \sigma^2, \\
L_n &= \frac{\partial \tilde{f}}{\partial y}(y_n, t_n), \\
M_n &= \frac{\sigma^2}{2} \frac{\partial^2 \tilde{f}}{\partial y^2}(y_n, t_n) + \frac{\partial \tilde{f}}{\partial t}(y_n, t_n), \\
\tilde{f} &= f \frac{\partial \varphi}{\partial x} + \frac{g^2}{2} \frac{\partial^2 \varphi}{\partial x^2},
\end{aligned}$$

である。

4. 数値実験

本節では、非線形な確率微分方程式のパラメータをシミュレーションデータから推定する実験を試み、本稿で提案する方法 (NLL) がどの程度うまくパラメータを推定できるか、既存の推定方法と比べてどのような差があるのかをそれぞれ評価する。また、データのサンプリング間隔の違いによって、推定のパフォーマンスがどのように変化するかという点も実用上は重要であるので、この点も併せて評価する。

ここでは、既存の方法として、オリジナルな局所線形化法 (LL) とオイラー法 (Euler) を比較対象として選んだ。そして、確率微分方程式が含意する真の離散化モデルに対するそれぞれの近似離散化モデルの当てはまりの良さを情報量規準 AIC で評価する。但し、3つの離散化モデルは共に同数のパラメータを含んでいることを考慮すれば、最大対数尤度の値で評価すれば十分であるので、ここでは、数値的最適化から得られる最大対数尤度の値で各モデルの当てはまり具合を比較評価する。

4.1 実験方法と実験データ

数値実験では次のような3つの非線形な確率微分方程式を選んだ。

例 1. $dx = (\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3)dt + \sigma dB,$

例 2. $dx = (\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \alpha_4 x^4 + \alpha_5 x^5)dt + \sigma dB,$

例 3. $dx = (\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3)dt + \sigma x^\gamma dB,$

ここで $\alpha_1, \dots, \alpha_5, \sigma, \gamma$ は推定すべきパラメータであり、真の値を表1に示した。3つの確率微分方程式は共に、ドリフト関数がプロセスの多項式で表現されており、特に、例2は例1に比べてより非線形性が強い。また、例3は、拡散項の係数が定数ではなく、プロセスの累乗に比例する係数となっており、この累乗の指数もまた推定すべきパラメータとなっている。例3のような拡散項の係数を持つ確率微分方程式は、経済データ、特に、金利の時系列的な変動を表現するモデルとしてしばしば用いられている。もちろん、前節で説明した通り、例3のような確率微分方程式であっても、局所線形化法が適用できることは言うまでもない。

さて、まず最初に、各例のサンプルデータを生成する必要がある。しかし、各例は、共に、真の離散化モデルも遷移確率密度関数も求めることが困難であるために、何らかの近似方法を用いてサンプルデータを生成しなければならない。確率微分方程式の数値的解法は、オイラー法をはじめルンゲ・クッタ法などいろいろな方法が知られているが、どの方法も、離散時間間隔が0に近づけば近似離散化モデルも真のモデルに近づくことが保証されているので、離散時間間隔を十分小さくにとっておけば、近似離散化モデルから生成されるデータ (サンプルパス) を真のデータと看做しても差し支えない。また、大抵の数値実験ではオイラー法が用いられていることから、本論文でも離散時間間隔を十分小さくにとってオイラー法を適用し、サンプルデータを生成する。もちろん、サンプルデータ生成にオイラー法を使用した影響が全く残らないとは言いきれないが、仮にそのような影響があったとしても、それはオイラー法を用いた推定に

表1. パラメータの真値。

	α_0	α_1	α_2	α_3	α_4	α_5	γ	σ^2
例 1.	—	1.0	0.0	-1.0	—	—	—	1.0
例 2.	—	-3.0	0.0	4.0	0.0	-1.0	—	1.0
例 3.	6.0	-11.0	6.0	-1.0	—	—	0.5	1.0

有利に働くことはあっても、局所線形化法を用いた推定に有利に働くことはないと考えられる。例3の確率微分方程式にオイラー法を適用するに際しては、拡散項の係数が定数となるように変数変換を施す必要がある。単純に、元の確率微分方程式にオイラー法を適用した場合、ひとたびプロセスが大きな値を取ると、それは大きな分散となって反映されるために、数値的に発散してしまう虞があるからである。したがって、ここでは $y = x^{1-\gamma}/(1-\gamma)$ なる変数変換を施して定数拡散係数を持つ確率微分方程式に変換して y_t を生成し、しかる後に、逆変換 $x = ((1-\gamma)y)^{1/(1-\gamma)}$ を用いて元のプロセスのサンプルデータを得る。

時間間隔 $\Delta t_0 = 0.005$ でデータを生成する。そして、生成されたデータからサンプル時間間隔 Δt_s なるデータを拾い上げてサンプルデータを作成する。本実験では、推定に対するサンプル時間間隔の影響を見るために、 Δt_s を 0.05 から 0.05 刻みで 0.2 までの4つの時間間隔を設定してサンプルデータを作成する。また、各サンプルデータは2000個のデータから構成されるようにするため、先頭からそれぞれ2000番目のデータまでを拾い上げる。

図1は、例1のサンプルデータをプロットしたものである。上から順にサンプル時間間隔がそれぞれ0.05, 0.1, 0.15, 0.2に相当するデータプロットである。また、例2のサンプルデータは図2に、例3のサンプルデータは図3にそれぞれ描かれている。

確率微分方程式のパラメータは、上のサンプルデータから最尤法に基づく数値的最適化から推定する。但し、例1と例2の対数尤度は(3.1)を用い、例3は(3.3)を用いる。

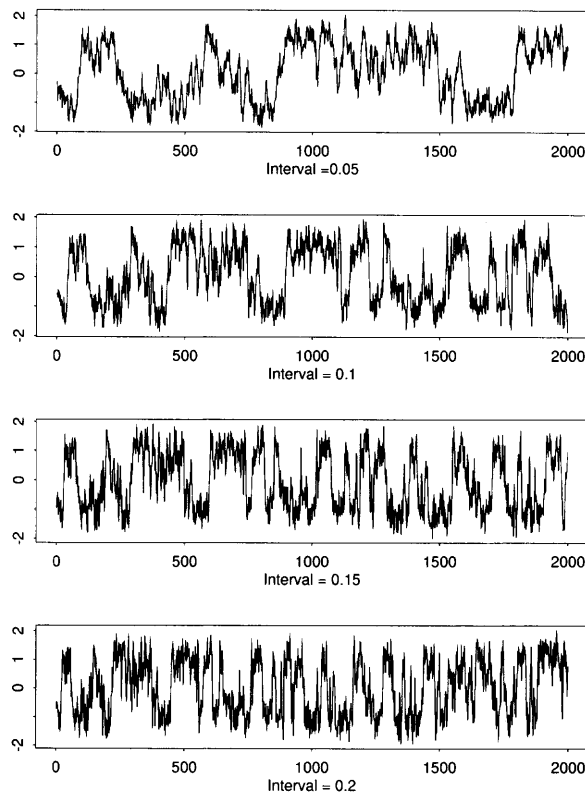


図1. 例1の時系列プロット。

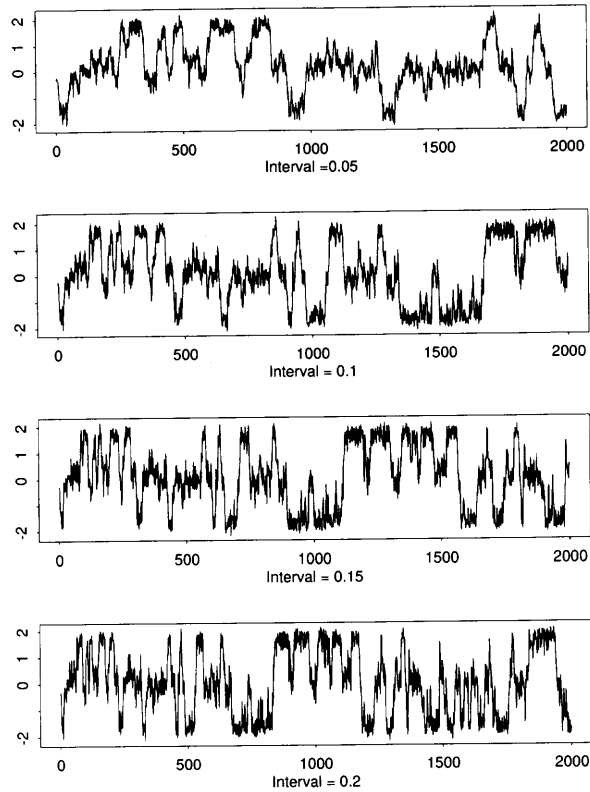


図 2. 例 2 の時系列プロット。

4.2 パラメータの推定

例 1 に対する推定結果は表 2 に示されている。最大対数尤度の値を見ると、すべてのサンプル時間間隔で NLL が最も良いパフォーマンスを示している。それとは対照的に Euler が最も悪いパフォーマンスを示している。オイラー法と両局所線形化法の差は、サンプル時間間隔が小さい時にはそれほど大きくはないが、間隔が大きくなるにつれてその差が顕著になっている。NLL と LL の差は、全般的にそれほど大きくはない。次に、個別の推定値に目を向けると、分散の推定では、NLL が LL や Euler に比べてより正確な値を出している。この傾向はサンプル時間間隔が大きくなるにつれて次第に顕著になっている。ドリフトの推定を見る場合には、各パラメータの推定値を個別に見るよりも、観測された x の範囲内でのドリフト関数の形状を見る方が適切である。実際、数学では、関数と言えば係数の値よりもそのグラフによって特徴づけられている。そして、このグラフとは、簡単に言えば、座標平面上の関数の形状を指している。そこで、横軸にプロセスの値、縦軸にドリフト関数の値を描いたものが、図 4 であり、そこにはサンプル時間間隔が 0.05 から 0.2 に対応するグラフが描かれている。また、それぞれには、真のドリフト関数の形状と 3 種類の方法によって推定されたドリフト関数の形状が描かれている。推定されたドリフト関数の形状は真のそれとほぼ似たような形状を示している。しかし、より細かく見ていくと幾つかの違いが明らかになってくる。LL によって推定されたドリフト関数は他の 2 つに比べて傾斜がきつく、逆に、Euler によって推定されたドリフト関数は最も平坦な形状を示している。そして、この傾向はサンプル時間間隔が大きくなるにつれて顕著になっている。このような形状の違いは、真のドリフト関数を持つ性質を記述する際にも重大な

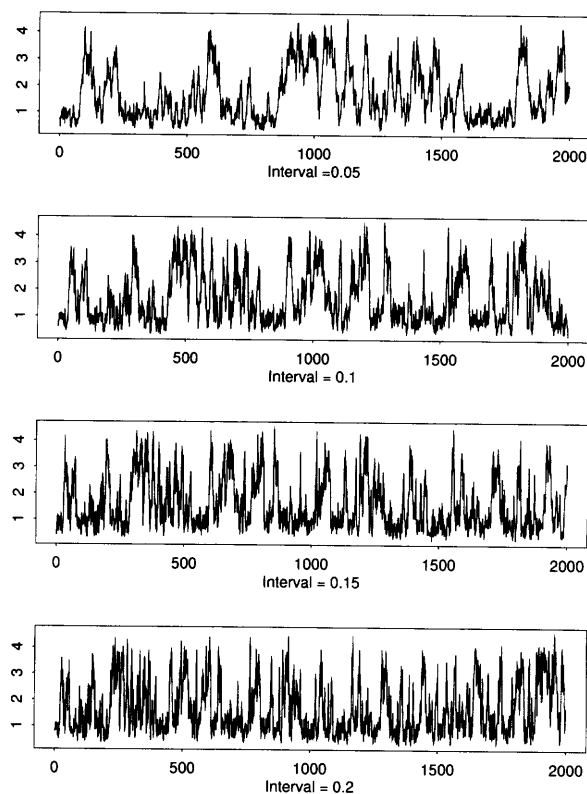


図3. 例3の時系列プロット.

表2. 例1の推定結果: $dx = (\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3)dt + \sigma dB$.

	α_1	α_2	α_3	σ^2	$\log l$
Parameter	1.0	0.0	-1.0	1.0	
$\Delta t = 0.05$					
NLL	1.05	0.08	-0.97	1.01	2064.40
LL	1.00	0.07	-1.00	0.93	2060.38
Euler	0.80	0.06	-0.81	0.94	2057.12
$\Delta t = 0.10$					
NLL	1.03	0.10	-0.99	1.04	1423.08
LL	1.23	0.12	-1.27	0.87	1421.04
Euler	0.81	0.08	-0.84	0.90	1411.01
$\Delta t = 0.15$					
NLL	0.91	0.01	-0.86	1.03	1079.45
LL	1.22	0.01	-1.28	0.82	1078.04
Euler	0.66	0.02	-0.71	0.85	1060.86
$\Delta t = 0.20$					
NLL	0.79	0.02	-0.79	1.05	813.67
LL	1.18	0.05	-1.37	0.80	809.66
Euler	0.50	0.02	-0.61	0.84	783.09

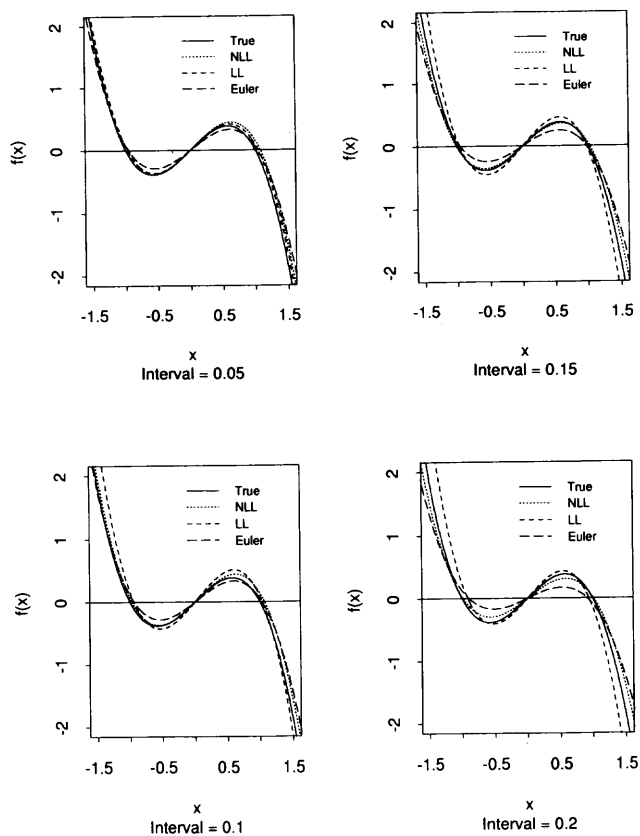


図4. 例1のドリフト関数の形状。

表3. 例2の推定結果: $dx = (\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \alpha_4 x^4 + \alpha_5 x^5)dt + \sigma dB$.

Parameter	α_1	α_2	α_3	α_4	α_5	σ^2	$\log l$
$\Delta t = 0.05$							
NLL	-3.17	0.24	3.45	-0.07	-0.79	1.04	2088.39
LL	-3.07	0.04	3.84	0.00	-0.99	1.00	2074.24
Euler	-2.80	0.00	2.98	0.01	-0.68	0.92	2073.20
$\Delta t = 0.10$							
NLL	-2.18	-0.01	2.84	0.01	-0.69	0.98	1569.92
LL	-2.08	-0.22	3.39	0.10	-1.00	0.89	1505.86
Euler	-1.93	-0.11	2.21	0.05	-0.54	0.82	1501.74
$\Delta t = 0.15$							
NLL	-1.58	-0.04	2.06	0.01	-0.49	0.90	1262.22
LL	-1.57	-0.16	2.70	0.06	-0.84	0.82	1188.72
Euler	-1.59	-0.05	1.74	0.02	-0.41	0.75	1187.23
$\Delta t = 0.20$							
NLL	-0.91	0.02	1.31	-0.01	-0.33	0.81	1079.60
LL	-1.06	-0.28	2.05	0.13	-0.68	0.72	1021.10
Euler	-1.19	-0.02	1.35	0.01	-0.33	0.66	1019.63

違いとなって現れてくる。即ち、真のドリフト関数は図から明らかなように3つの零点を持っている。NLLによって推定されたドリフト関数は、サンプル時間間隔が最大の0.2になっても、この3つの零点を維持しているが、LLやEulerによって推定されたドリフト関数では維持できていないことがわかる。

例2に対する推定結果は表3に示されている。最大対数尤度の値は、例1の結果と同様に、NLLのパフォーマンスが最も良く、Eulerが最も悪いことを示している。特に、NLLとEulerの違いもさる事ながら、NLLとLLの違いも顕著になっている。逆に、LLとEulerの差があまり無いことがわかる。分散の推定に目を向けると、NLLが比較的正確な推定値を出してくるのに対して、その他の方法でかなり不正確であることがわかる。特に、サンプル時間間隔が大きくなると、LLやEulerでは急激に悪化する。図5の推定されたドリフト関数の形状を見ると、NLLではサンプル時間間隔が大きくなっても零点を維持できているのに対して、その他の方法では、維持できなくなっていることが、例1に比べてより鮮明にわかる。このように、ドリフト関数がより高い非線形性を示す場合には、NLLとその他の方法の間に大きな違いが見られ、NLLのパフォーマンスの良さが際立つことになる。

例3に対する推定結果は表4に示されている。最大対数尤度の値は、NLLのパフォーマンスが最も良く、LLが最も悪いことを示している。例2の結果と同様に、LLとEulerの差があまり無く、NLLとこの両者との差はサンプル時間間隔が大きくなるにつれて顕著になる。分散の推定では、NLLでは比較的正確な推定値であるのに対して、LLやEulerではサンプル時間間

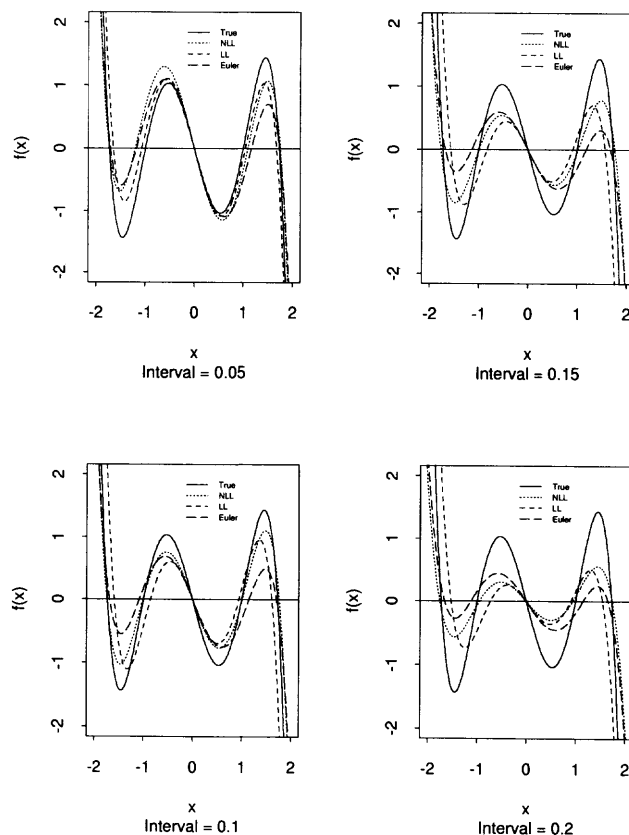


図5. 例2のドリフト関数の形状。

表 4. 例 3 の推定結果: $dx=(\alpha_0+\alpha_1x+\alpha_2x^2+\alpha_3x^3)dt+\sigma x'dB$.

	α_0	α_1	α_2	α_3	γ	σ^2	$\log l$
Parameter	6.0	-11.0	6.0	-1.0	0.5	1.0	
$\Delta t = 0.05$							
NLL	6.25	-11.05	5.83	-0.94	0.47	1.04	1713.43
LL	6.80	-12.17	6.52	-1.07	0.52	0.91	1709.72
Euler	5.81	-10.29	5.44	-0.88	0.50	0.92	1711.91
$\Delta t = 0.10$							
NLL	5.67	-9.93	5.25	-0.86	0.54	1.01	1096.94
LL	7.30	-12.96	6.94	-1.15	0.62	0.79	1085.87
Euler	4.90	-8.46	4.42	-0.72	0.60	0.81	1088.05
$\Delta t = 0.15$							
NLL	4.81	-8.39	4.35	-0.70	0.52	1.02	838.21
LL	6.86	-12.37	6.72	-1.14	0.67	0.73	826.42
Euler	3.98	-6.84	3.51	-0.57	0.64	0.74	831.68
$\Delta t = 0.20$							
NLL	4.25	-7.30	3.86	-0.63	0.56	0.99	536.83
LL	7.40	-13.06	7.09	-1.21	0.70	0.68	507.33
Euler	3.22	-5.16	2.59	-0.42	0.66	0.71	508.55

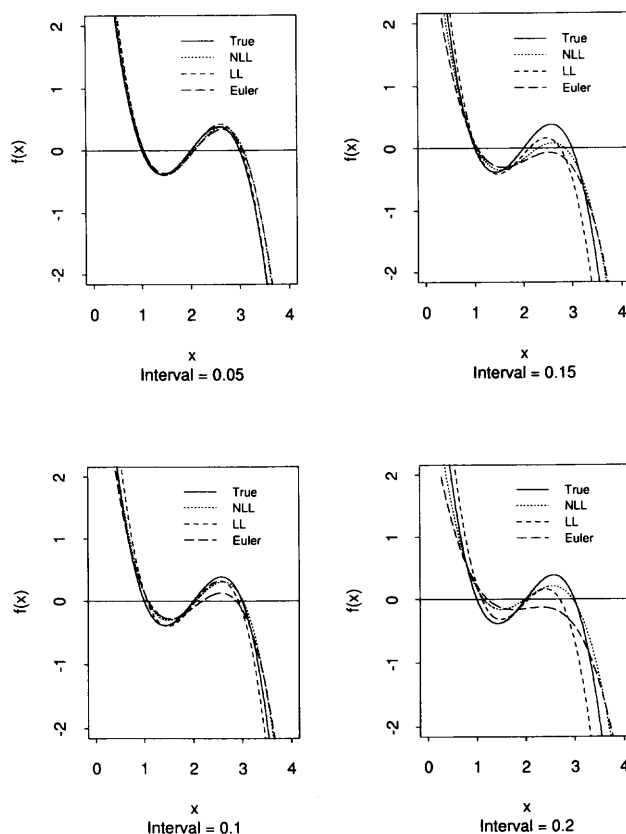


図 6. 例 3 のドリフト関数の形状。

隔が0.05を超えるとかなり過小推定になる。このような傾向は、累乗の指数の推定についても見られ、NLLでは比較的正確であるのに対して、LLやEulerではサンプル時間間隔が0.05を超えるとかなり過大推定になる。図6の推定されたドリフト関数の形状を見ると、サンプル時間間隔が0.1を境に、それ以上ではLLやEulerのドリフト関数の形状は真の形状とはかなり異なっているが、NLLのドリフト関数の形状は全般に真の形状に近いことがわかる。これは零点の観測からも裏付けられることで、サンプル時間間隔が0.1を境に、それ以上ではLLやEulerのドリフト関数は零点を維持できていないのに対して、NLLでは、基本的に維持できていることがわかる。

3つの例を通して、分散の推定においてNLLは最も良好なパフォーマンスを示し、特に、サンプル時間間隔が大きい時やドリフト関数の非線形性が著しい時にそれは顕著であった。NLLのこのパフォーマンスは、離散化モデルの分散式(2.8)の妥当性に由来するものと考えられる。事実、プロセスの状態に依存しないというEulerの分散式では満足な推定値が得られないこと、更に、LLに対するNLLの優位性を考慮するならば、NLLの分散式がより適切な表現であることが窺える。特に、LLとNLLのパフォーマンスの違いに注目すると、プロセスの確率的な変動を考慮して離散化モデルを導いたNLLがそれを考慮しないLLよりも妥当な離散化手法であることが分散の推定結果から示唆される。

また、NLLはドリフト関数の推定においても良好なパフォーマンスを示した。これは、NLLがドリフト関数の1階及び2階の導関数を用いていることから、真のドリフト関数の情報をより効率的に引き出した結果であると考えられる。事実、1階の導関数を用いるLLに対するNLLの優位性から、2階の導関数の重要性が認識される。一方、理論的な観点からは、関数の変動を考慮する時、その2階の導関数が無視できないことが伊藤の公式により示されている。この事は、関数の微分を基に、それを局所的に線形近似する場合にも、2階の導関数を考慮すべきであることを理論的な立場からも示唆していることに他ならない。更に、数値的な観点からは、2階の導関数の影響がサンプル時間間隔の2乗に比例して大きくなることに注目すると、NLLがサンプル時間間隔が大きい時に特にパフォーマンスが良いことが理解できる。

5. ま と め

本稿では、既存の局所線形化法を改良した新たな局所線形化法を提案した。この方法によれば、オリジナルな方法が確率微分方程式について課す制約、即ち、プロセスの状態が0の時、ドリフト項の関数が0に等しいという制約や離散化モデルの条件付き期待値が、拡散項を除いた常微分方程式から導かれる解の軌道と一致するという制約、を取り除くことができる。それゆえ、本稿で提案する方法は、より広範囲な確率微分方程式に適用可能である。実際、本方法によれば、非線形な確率微分方程式でも、容易にしかも効率的に離散化モデルを導くことができる。しかもその離散化モデルは、離散時間間隔が0に近づく時、真のモデルに収束するという意味で一致性が保証されている。

また、本稿では、離散化モデルに基づいて離散データから確率微分方程式のパラメータをどの程度うまく推定できるかを、オイラー法とオリジナルな局所線形化法とを比較対象として数値実験を行って評価した。その結果、提案する方法は他の方法に比べてかなり良好なパフォーマンスを示し、パフォーマンスの差はサンプル時間間隔が大きくなるにつれて顕著になった。これは、提案する方法が、ドリフト関数の2階微分まで用いていることから、より効率的にプロセスの変動情報を引き出した結果によるものと考えられる。

謝 辞

本研究の過程で有益な助言を下された統計数理研究所の北川教授，田村助教授に感謝いたします。また，本論文に対し貴重なコメントを下された査読委員，編集委員に感謝いたします。

参 考 文 献

- Bergstrom, A. R. (1990). *Continuous Time Econometric Modelling*, Oxford University Press, New York.
- Dembo, A. and Zeitouni, O. (1987). On the parameter estimation of continuous-time ARMA processes from noisy observations, *IEEE Trans. Automat. Control*, **32**, 361-364.
- Harvey, A. C. and Stock, J. H. (1985). The estimation of higher-order continuous time autoregressive models, *Econometric Theory*, **1**, 97-117.
- Jones, R. H. and Tyron, P. V. (1984). Continuous time series models for unequally spaced data applied to modeling atomic clocks, *SIAM J. Sci. Statist. Comput.*, **8**, 71-81.
- Ozaki, T. (1985). Statistical identification of storage models with application to stochastic hydrology, *Water Resources Bulletin*, **21**, 663-675.
- Ozaki, T. (1992a). A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach, *Statist. Sinica*, **2**, 113-135.
- Ozaki, T. (1992b). Identification of nonlinearities and non-gaussianities in time series, *New Directions in Time Series Analysis*, Part I (eds. D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. S. Tqqu), 227-264, Springer, New York.
- Ozaki, T. (1993). A local linearization approach to nonlinear filtering, *Internat. J. Control*, **57**, 75-96.
- Shoji, I. (1995). Approximation of a continuous time stochastic process by a new local linearization method, *Research Memo.*, No. 530, The Institute of Statistical Mathematics, Tokyo.
- Singer, H. (1993). Continuous-time dynamical systems with sampled data, errors of measurement and unobserved components, *J. Time Ser. Anal.*, **14**, 527-545.
- Yoshida, N. (1992). Estimation for diffusion processes from discrete observation, *J. Multivariate Anal.*, **41**, 220-242.

Estimation for Continuous Time Stochastic Processes by a Local Linearization Method

Isao Shoji

(Institute of Policy and Planning Sciences, University of Tsukuba)

Tohru Ozaki

(The Institute of Statistical Mathematics)

A stochastic differential equation (SDE) is often used to model the time evolution of dynamic phenomena. To make the model work in practice, it is necessary to know the parameters of the SDE. Of course, we can not expect to know the parameters accurately, rather we must estimate them from real data. Considering that available data are almost always sampled with regular time interval, we require a method which enables us to estimate the parameters from discrete observation. The objective of this paper is to present a new method which meets this requirement. The local linearization method originally developed by Ozaki (1985) is one of the methods. This method, however, has some problems in application to SDEs. Hence, this paper proposes a new local linearization method which solves these problems. The difference between the new method and the original method lies in the way of linear approximation. In the new method, we derive the linear approximation of a drift function of a SDE with taking account of the stochastic behavior of the process. By contrast, in the original method, its stochastic behavior is neglected. In this paper, numerical experiments are also carried out to compare the performance of the new method with the original local linearization method and the Euler method in estimating some nonlinear SDEs from discrete observation. Furthermore, we evaluate how much the performance of estimation will be changed as varying discrete time interval. The results of the experiments indicate that the new method shows better performance than the others and performs considerably better particularly when discrete time interval is large.