九州大学学術情報リポジトリ
Kyushu University Institutional Repository

遺伝的プログラミングによる状態方程式近似とモンテカルロフィルタ法を用いた非定常時系列からの状態推定とその応用

儲，梅芬

https://doi.org/10.15017/3000381

出版情報：経済論究. 123, pp.49-61, 2005-11-30. 九州大学大学院経済学会
バージョン：
権利関係：

遺伝的プログラミングによる状態方程式近似とモンテカルロフィルタ法を用いた非定常時系列からの状態推定とその応用

儲　　梅　芬

1　まえがき

株価や為替レート変動などの実際的な時系列生成をより精密にモデル化し，特性解析やパラメータ抽出，あるいは予測に応用する方法が示され，さまざまな非線形モデル（例えばARCHモデル）が提案されている［1］。このような課題とならんで，時系列生成モデルを仮定し，内部状態に相当する変動要因を推定する方法論が展開されている。最近，このような非線形モデリングにおいて，遺伝的アルゴリズム(Genetic Algorithm: GA）と同様な考えで時系列モデルの推定を行う方法が提案され，複数のシミュレーションを実行して，逐次的にモデルの改善を行うことから，モンテカルロフィルタ(Monte Carlo Filter）法と呼ばれている［2］-［6］。しかし，このモンテカルロフィルタによる手法では，時系列生成のモデルを記述する状態方程式が既知であること，システムノイズの性質は正規分布などの既知のプロセスであることが仮定されており，生成過程そのものを推定する問題に適用するには，何らかの拡張が必要である。

本論文では，モンテカルロフィルタ法を適用し非定常時系列から内部状態を推定する課題において，時系列生成（ダイナミックス）を推定する場合に遺伝的プログラミング（Genetic Programming: GP）を適用し，実際的な時系列からの内部状態推定を実行する方法を示す［7］-［14］。本論文で示すGPによる近似手法の基本的な考え方は，状態方程式の形状を直接関数表現により近似することにあり，時系列生成のダイナミックスを推定しながら，同時にモンテカルロフィルタ法による内部状態推定を進めることにある。本論文ではこのようなGP手法とモンテカルロフィルタ法を組み合わせた手法で得られる内部状態推定の方法により，人工的に生成された時系列，および日経平均株価から求められる内部状態推定を実施し，その有効性を確認している。

以下，2.ではモンテカルロフィルタによる状態推定の基本モデルについて述べ，3.ではGP手法の原理について示す。4.では人工的なデータと実際の経済時系列への適用を示す。

2　モンテカルロフィルタによる状態推定の基本モデル

2.1　非線形・非ガウス時系列モデル

経済分野における時系列処理の手法や応用は多様であるが，これを大きく分けると，時系列予測とパラメータ推定に分類できるであろう。時系列予測は文字どおり，現在までの時系列データを用いて将来の値を予測する方法であり，関数論的な手法のほかに，ニューラルネットワーク手法などが提案されている。パラメータ推定は，時系列生成のモデルに含まれる次数や係数を推定する問題や，スペクトルを推定する問題などが含まれる。これに関連して，さまざまな線形・非線形の時系列生成モデルが提案されている。

時系列を生成するモデルに関して，時系列の特性に注目しながら，従来の線形モデルから非線形モデルへと拡張することがなされている。従来は自己回帰モデルを中心とした議論がなされているが，株価のボラティリティー（volatility）推定に関連してARCH（Autoregressive conditional heteroscedasticity）モデルおよび拡張形であるGARCH（Generalized-ARCH）モデルが提案されている［1］。これらのモデルにおいては，ボラティリティーが時間的に変動する時変モデルとなっており，このことにより，例えば株価のオプション価格の推定精度を改善するなどの効果が期待されている。

時系列生成モデルにおけるパラメータ推定から派生した１つの分野として，時系列データをもとにして状態変数記述によるモデルを仮定し，内部状態に相当する変動要因を推定する方法論が展開されている。従来の時系列分析手法に対して内部状態の推定を用いる利点としては，間接的に影響する要因の分析を行うことができることがある。例えば，為替レート変動がどのような原因で発生しているかを推定する場合に用いられている。

なお，状態推定を行うためには関数系によるシステム記述は適切ではなく，時系列生成のシステムを状態方程式で記述することが行われる。この状態方程式による記述を用いると，時系列が定常である場合には，従来のカルマンフィルタ法が適用でき，更に，モデルを非線形の状態方程式で記述する場合においても，その次数が小さい場合には数値積分を実行することにより，状態推定が可能であることが示されている。しかし，次数が増加した場合には数値積分による方法は適用できない［2］。

最近，このような非線形モデリングの課題を解決する方法として，GAと同様な考えで時系列モデルの推定を行う方法が提案され，複数のシミュレーションを実行して，逐次的にモデルの改善を行うことから，モンテカルロフィルタ（Monte Carlo Filter）法と呼ばれている［2］-［6］。モンテカルロフィルタ法は，状態推定における多次元の数値積分計算を数値計算に置き換える方法であり，非線形フィルタの実現方法として極めて有効である［2］-［6］。

本論文では，このモンテカルロフィルタにおける方程式記述をGPにより直接近似する方法を提案する。まず，最初に，GP手法を説明する前に，モンテカルロフィルタ法の概要を以下で整理しておく。

モンテカルロフィルタ法の代表的な適用分野として，為替レート変動からその変動要因を推定する問題がある。従って，以下では，株価時系列の解析あるいは為替レート時系列解析にならって，時系列の対数の変化（対数収益率）を対象とした時系列モデルを考察するが，一般化するのは容易である。

時刻 n における時系列の値を z_n とした場合に，この対数収益率の時系列を y_n とする。

$$y_n = \log(z_n / z_{n-1}) \tag{1}$$

時系列の変動成分として，さまざまなものが考えられるが，これまでの先行研究を参考にして，以下では，次のような非線形モデルを考察する。

$$y_n = T_n + p_n + \sigma_n \varepsilon_n \tag{2}$$

T_n はトレンドを表し，通常は階差モデルなどの簡単な形で表現される。p_n は周期成分などに対応して，通常はこのような定常的なスペクトルの存在を仮定することができるARモデルなどで表現される。第３目の項目は不規則な変動に対応するものである。この第３番目の項目のモデル化と，そのパラメータ推定に，これまで多くの先行研究がなされている。ε_n は，平均ゼロ分散１の正規分布に従う乱数であり，これにか

けられている σ_nは時間的に変動する時系列の分散である。

トレンド T_nと時変分散 σ_n^2の対数値は，それぞれ次数が k，l である確率的な階差モデルに従って変化すると仮定し，その挙動を次のように表す。

$$\Delta^k T_n = e_{1n},\ e_{1n} \sim N(0,\ \tau_1^2) \tag{3}$$

$$\Delta^l \log \sigma_n^2 = e_{3n},\ e_{3n} \sim N(0,\ \tau_3^2) \tag{4}$$

また，定常時系列成分である p_nについては，次のような自己回帰モデルで記述されると仮定しておく。

$$p_n = \sum_{j=1}^{m} a_j p_{n-j} + e_{2n}, e_{2n} \sim N(0, \tau_2^2) \tag{5}$$

ここで，e_{1n}, e_{2n}, e_{3n}は，それぞれ，トレンド，対数分散，周期成分のシステムノイズの分散である。

システムの簡潔な表現を得るために，次に示す状態ベクトルを導入する。

$$x_n = (T_n, T_{n-1}, \ldots, T_{n-k+1}, p_n, \ldots, p_{n-m+1}, \log \sigma_n^2, \ldots, \log \sigma_{n-l+1}^2)^T \tag{6}$$

この変数ベクトル x_nを用いて式(3)(4)(5)の関係を整理すると，次のような簡潔な表現となる。

$$x_n = F x_{n-1} + G v_n = f(x_{n-1}, v_n) \tag{7}$$

$$v_n = (e_{1n}, e_{2n}, e_{3n}) \tag{8}$$

ただし，v_nは時系列を生成する乱数e_{1n}, e_{2n}, e_{3n}をまとめたものである。ここで，行列 F，G は次のような形で示される。

$$F = \begin{pmatrix} c_1^k & \cdots & c_k^k & & & & & & \\ 1 & & 0 & 0 & & 0 & & & \\ 0 & & 1 & & & & & & \\ & & & a_1 & \cdots & a_m & & & \\ & 0 & & 1 & & 0 & & 0 & \\ & & & & \cdots & & & & \\ & & & 0 & & 1 & & & \\ & & & & & & c_1^l & \cdots & c_l^l \\ & 0 & & 0 & & & 1 & & 0 \\ & & & & & & & \cdots & \\ & & & & & & 0 & & 1 \end{pmatrix} \tag{9}$$

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix} \tag{10}$$

観測方程式の相当するものは，時系列 y_nが状態ベクトル x_nの一部を用いて表現できることから，次のような非線形の関数を導入して求められる。

$$y_n = T_n + P_n + \sigma_n \varepsilon_n = h(x_n, \varepsilon_n) \tag{11}$$

式(7)(8)と式(11)とをあわせることにより，この時系列生成モデルは，非線形の状態方程式により表現することができる。

ここでは説明は省略するが，時系列が定常である場合には，カルマンフィルタによる状態推定が可能であり，式(7)(8)(11)に含まれる行列を用いた逐次計算のアルゴリズムが適用できる。

また，モデルが非線形である場合においても，状態方程式の次数が小さい場合には，次に示すような多重積分を数値計算として実施することにより，状態推定が可能である。観測値の集合を $D_n = [y_i, i = 1, 2, \ldots, n]$とするとき，ベイズの定理により次のようになる。

$$p(x_n | D_n) = \frac{p(y_n | x_n) p(x_n | D_{n-1})}{\int p(y_n | x_n) p(x_n | D_{n-1}) dx_n} \tag{12}$$

この式から解析解を求めることは困難であるので，数値計算を実施することになる。

2.2 モンテカルロフィルタによる時系列からの状態推定

すでに述べたように，時系列が定常である場合にはカルマンフィルタによる状態推定が可能

であり，次元数が小さい場合には，多重積分を数値計算で実施することにより状態推定が可能である。しかし，時系列が非定常でありモデルが非線形となる場合には，この方法は適用できない。

これまでの議論で時系列を生成するシステムの挙動を非線形の状態方程式で書くことができることが示された。従って観測された時系列から，状態およびシステムを記述するパラメータを推定する方法が適用できる。この場合，モデルが非線形であるので，通常のカルマンフィルタ法が適用できないため，これを解決するモンテカルロフィルタ法が提案されている。

この方法では，観測値を用いて状態の予測値を求め，更に，この予測値を用いて状態の平滑化を進める手順を繰り返す。これをまとめると次のようになる。

(1)初期分布の近似を生成する

乱数を用いて，与えられた初期分布 $p(x_0 \mid Y_0)$ に従うM個の粒子を生成する。

$$\{F_0^1, F_0^2, \ldots, F_0^M\} \sim p(x_0 \mid y_0) \tag{13}$$

以下のステップ(2)から(5)までを，時系列の時刻 n について繰り返す。

(2)システムノイズの近似

あらかじめ与えられた密度関数 $p(v)$ に従う乱数 v_nを生成する。

$$\{v_n^1, v_n^2, \ldots, v_n^M\} \sim p(v) \tag{14}$$

(3)予測分布の近似

$n=1$ の場合にはステップ(1)で求めた粒子を式(7)の右辺に代入し，予測分布を表す粒子 P_n^j の値を求める。

$$P_n^j = f(F_n^j, w_n^j) \tag{15}$$

(4)ベイズ係数の計算

次に示すベイズ係数を計算する。

$$\alpha_n^j = Prob(y_n \mid x_n = P_n^j) \tag{16}$$

このベイズ係数は，それぞれの粒子の重要さを表すので，この大きさに従って，次に示すステップで改善をはかる。

(5)リサンプリングにより分布の近似

P_n^j を $\alpha_n^j / \sum_{k=1}^{M} \alpha_n^k$ の確率で $F_n^{(j)}$ とするリサンプリングを実施する。

この手法を数値計算の方法論から見ると，GAにおける適合度の計算によるエリート保存戦略であると見ることができる。乱数として多くの可能性のある解を初期値として生成し，これらの中でシステムの近似能力の高い組み合わせを選択する確率を，次第に増やしていく方法である。

なお，ステップ(5)におけるリサンプリングにおいて用いた各粒子の尤度 α_n^j を用いて，システムの当てはまりの良さを示す対数尤度は，次のように計算できる。

$$l(\theta) = \sum_{n=1}^{N} \log p(y_n \mid Y_{n-1})$$

$$= \sum_{n=1}^{N} \log\left(\sum_{j=1}^{M} \alpha_n^j\right) - N \log M \tag{17}$$

2.3　モンテカルロフィルタ適用事例

以下では，モンテカルロフィルタを適用して内部状態を推定した事例について引用し，概要を述べる。これらの適用事例では，状態方程式の形式が仮定され，観測されたデータからこの観測値に内部状態が，どのように影響を与えているかを分析している。

[事例１]：日経225データの場合

文献[4]においては，日経225株価平均の1987年１月から1990年８月31日までのデータについて，モンテカルロフィルタ法を適用している。この方法は，基本的に2.2節に示したものと同様であり，トレンド変動と分散変動のモデルの次数 k，l は，それぞれ，$k=2$，$l=2$としている。

このモデルに対して，モンテカルロフィルタによるパラメータ推定を実施すると，推定結果

として $\hat{\tau}_1^2=9000$, $\hat{\tau}_2^2=0.0026$を得ており，この場合のAICはAIC＝13580となっている。なお，このAICの値は，通常の解析パッケージを用いた場合に得られる値AIC＝14190よりも610小さくなっている。図１には，解析の結果として得られるトレンド変動と分散変動の推定時系列を引用している。

［事例２］：損傷推定の場合

建築構造物の維持管理コストを削減するため，構造物の内部状況を構造物を破壊することなく，推定する方法が提案されている。構造物の最大変位応答値（maximum absolute displacement）を観測値として与え，この構造物の速度変位および構造物損傷を内部状態として仮定する。ランダムノイズの分布は正規乱数を仮定している。

このデータに対して，モンテカルロフィルタによる内部状態推定を適用する。ただし，経済時系列の場合と異なり，内部状態の時系列はデータとして収集することができるので，真の値との一致を直接的に検証することができる。図２には，文献［6］に示された結果を引用している。推定値が真の値に，極めて近くなっていることが分かる。

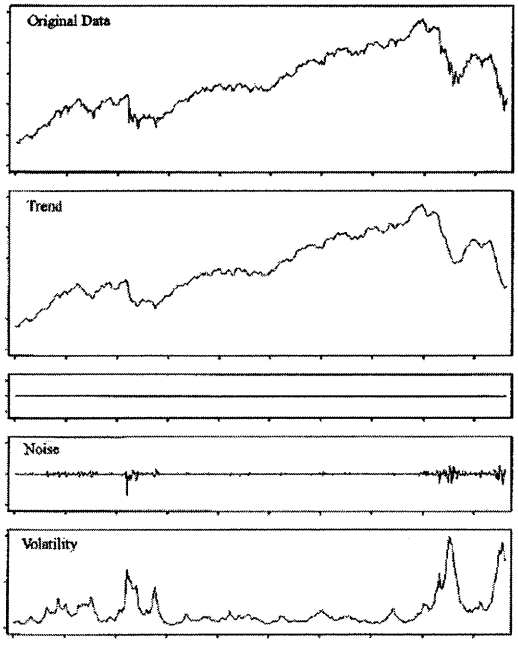


図１：日経225データからの内部状態推定（文献［4］より引用）

2.4 GPによる状態方程式推定への一般化

状態方程式の形が既知である場合，すなわちあらかじめ時系列が生成されるモデルを仮定し，その中に含まれる状態を推定するケースにおいては，これまで述べたモンテカルロフィルタ法を適用することができる。しかしながら，経済時系列を例にとった場合に，必ずしも時系列生成モデルが既知であるケースばかりではない。特に，経済現象は，工学や自然科学の現象と異なり，一般には状態方程式を記述するための物理モデルが存在しない。

このようなことを考慮すると，観測された時系列から状態推定を行う問題においても，状態の記述モデルが確定できないと考えるのが自然である。本論文では，このように状態方程式の推定を，GPにより同時並行的に実施する方法論を与えるものである。

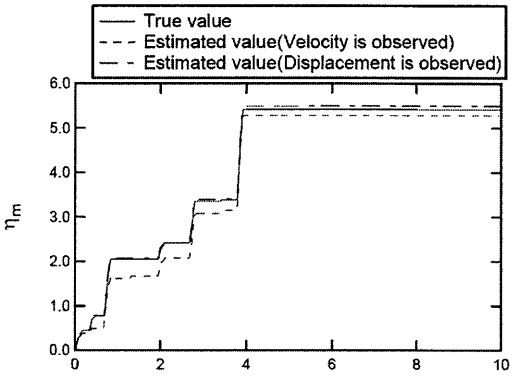


図２：構造物損傷からの内部状態推定（文献［6］より引用）

GP手法は，これまでさまざまな分野に応用されており，その有効性が確認されている。基本的には，関数近似の手法であり，カオス時系列を観測データとして用い，その生成メカニズムダイナミックスを推定する問題，推定されたダイナミックスを利用してカオス制御を実施する方法への適用，あるいは，関数推定を学習的に行うエージェントを仮定した人工市場のシミュレーションへも応用されている [7] -[14]。GP手法による状態推定の概要を図3に示す。

図3においてGPにおける個体(詳細は後述する）のそれぞれが，1つの状態方程式の実現形に対応している[7]-[14]。この個体のそれぞれに対して，これまで述べたモンテカルロフィルタによる状態推定を実施する。式(17)に示す対数尤度を計算して，この値が小さいほど状態推定が適切であることに相当するので，以下の遺伝的操作において，より個体の遺伝的生成や個体プールにおける残存確率が大きくなるように設計する。

モンテカルロフィルタ法におけるGP手法の応用は，式(7)に示された非線形の関数である $f(x_{n-1},v_n)$ を推定する場合に使用する。GP手法の原理は，関数 $f(x_{n-1},v_n)$ に対して多数の関数近似の形を個体の形で与えておいて，この個体の近似度（適合度として定義される）の相対的に高いものを選択して，交差処理などの遺伝的操作を加えて，より近似度の高い個体を生成するものである。

個体の適合度は，個体によるシステムモデル近似の関数近似の度合いを示すが，実際には，個体を用いて近似された状態方程式にモンテカルロフィルタ法を適用して得られた対数尤度の逆数を個体の適合度として与える。すなわち，GPの個体が表現するモデルから観測された時系列の発生確率が高いほど，個体の評価が高くなる。

3　GP手法の原理

3.1　GPによる関数近似

既に述べたように，本論文では状態方程式の関数を推定する方法としてGPを用いる。具体的には，GPにおける個体を関数 $f(x_n, v_n)$ を近似するものとして状態変数，およびシステムノイズの関数として与え，関数の形状を直接近似する。GP手法においては，このような関数近似の候補を複数準備しておいて，これらの近似度に応じてより近似度の高い個体を生成するアルゴリズムを適用する [7]-[16]。

GPにおける1つの予測モデルは，木構造で表現され，その非ターミナルノードには演算子の集合から選択された記号をおき，ターミナルノードには変数と定数をおく。演算子の集合として，次のようなものを仮定する。

＋，－，×，/

exp，　abs，　sqrt，　log

関数を木構造としてそのまま用いるのは効率が悪いので，以下では，前置表現(prefix repre-

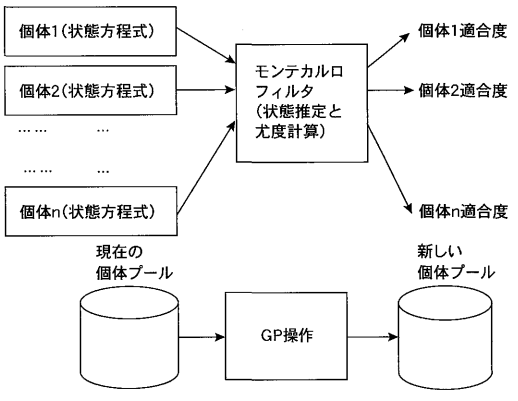


図3：GPによる状態方程式近似とモンテカルロフィルタ法

sentation）を用いる。この前置表現の事例を，次に示す。

$$(5\times x_1-x_2)\times(x_3-11)\rightarrow\times-\times 2x_1x_2-x_3 7 \quad (18)$$

GPを適用する場合の個体の適合度として，モンテカルロフィルタを適用して得られる尤度を用いる。尤度が小さいほど，時系列を生成する場合のシステムノイズのモデルが適切であったことを意味している。

このような適合度の計算と遺伝的操作を十分な回数だけ繰り返す。その結果，自動的に予測の精度の高いモデルは残り，その個体に対して実施されたGPの結果としてもたらされる個体も，新しくプールに加わる。これに対して，適合度の低い個体はプールから除去されていく。

GPによる関数や算術式を近似する方法については，その方法論や近似精度について，すでに過去の研究で述べられているので，ここでは，必要な範囲で概要についてまとめる［7]-[16]。

最初に関数を木構造で表現することとし，1つの個体が1つの関数近似に対応する。木構造の表現として，演算記号と演算子を適切な順序に一列にならべる表現である前置表現を用いる。個体の関数値を計算するには，変数に値を与えて解釈する。

一方，個体の初期値および2つの個体を部分的に切断して交差処理により相互に接続する処理の結果として生成される個体が，関数を表現しているかどうかをチェックする目的のため，*StackCount*という変数を導入する。*StackCount*は，スタックの中に存在する演算子の数から，演算記号の数を引いた数値である。従って，前置表現において，その途中まで，この*StackCount*をかぞえて，その数値が同じであれば，この位置で2つの個体を切断し交差処理をしても，意味のある方程式表現を与えていることになる。この概要を図4に示す。

以上のような準備のもので，GPを適用するアルゴリズムを整理すると以下のようになる。

（ステップ1）：個体の初期値生成

最初の個体の集合（プール）を乱数をもとにして発生させる。この場合，すでに述べた*StackCount*を用いて，関数として意味をなすものが得られるまで繰り返す。

（ステップ2）：個体の適応度の計算

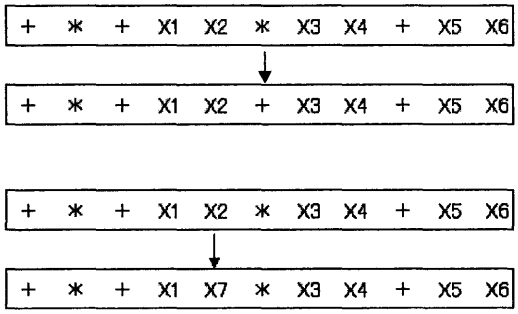
すでに述べた関数近似における予測値と観測値との2乗誤差の逆数を，個体 i の適合度 S_i として定義する。個体を適応度の大きい順に並びかえておく。

（ステップ3）：交差処理

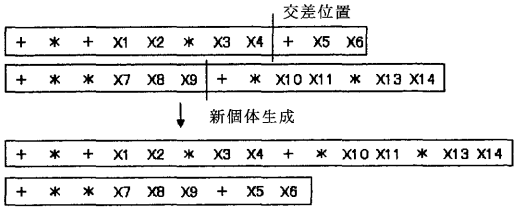
適応度に応じて2つの個体を取り出し，交差処理を実施する。交差処理では，乱数を1個発生させておいて，一方の個体の切断個所とし，ここまでの*StackCount*の値を計算する。こののち，もう一方の個体の*StackCount*を計算しながら，同じ*StackCount*の値となる場所を検出し，これらから任意に1個を選択して切断位置を決める。この2つの個体に対して遺伝的操作を行い，生成された新しい個体を次のステップにおける代替個体のプールであるP-Bに格納しておく。このような新しい個体の生成を，規定回数繰り返す。新規個体の生成が終了したら，プールP-Aの個体の中で，相対的に適合度の低い個体を，プールP-Bの個体により置き換える。

（ステップ4）：突然変異

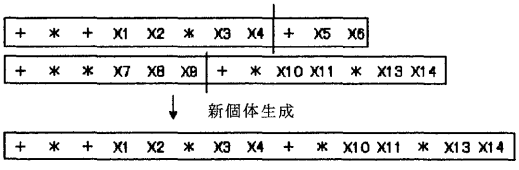
ステップ2からステップ3までの交差処理を，決められた個数の個体に適用し，新しい個体のプールを作成したあとに，次に示す突然変異を実施する。GPにおける突然変異として，次の2つのものを用いることとし，一定の確率で適用する。



(a)　交差処理



(b)　グローバル突然変異



(c)　ローカル突然変異

図 4 ：GPにおける交差処理と突然変異

グローバル突然変異

ある個体(木構造) I_sを生成しておいて，目的とする木構造の葉の部分を，この I_sにより置き換える方法である。もとの木構造は，やや複雑なものへと変化する。

ローカル突然変異

木構造の葉の部分を，別の変数に置き換える操作である。あるいは，木構造における根の部分に相当する原始関数を，別のものに置き換える操作であることもある。

4　応用例

4.1　既知の生成モデルを用いた場合の近似

以下では，数値シミュレーションをもとにして，本論文で提案する非線形時系列生成モデルの推定を実施する。この最初の例として，既知の非線形モデルを仮定し，GP手法によりこの関数の形状を正確に近似できるかを確かめる。以下のようなモデルを仮定する。

$$T_n = 2T_{n-1} - T_{n-2} + e_{1n},\ e_{1n} \sim N(0,\ 0.5^2) \quad (19)$$

$$p_n = 0.8p_{n-1} + 0.28p_{n-2} + e_{2n},\ e_{2n} \sim N(0,\ 0.5^2) \quad (20)$$

$$\log\sigma_n^2 = \log\sigma_{n-1}^2 + e_{3n},\ e_{3n} \sim N(0,\ 0.05^2) \quad (21)$$

$$y_n = T_n + p_n + \sigma_n\varepsilon_n,\ \varepsilon_n \sim N(0,\ 1.0^2) \quad (22)$$

それぞれの内部状態の初期値と，GPによる推定のための条件は，次のようになっている。

$T_0 = 50.0,\ T_{-1} = 49.98$

$p_0 = 5.0,\ p_1 = 4.8,\ \log\sigma_0^2 = 6.0$

内部状態の次元数：6

モンテカルロフィルタにおける粒子数：3000

GPにおける個体数：50

GPによるモデル推定用観測データ：150

システムノイズの平均と分散は，既知であると仮定する。観測データの利用については，GPによる状態方程式推定に150サンプルを用いたのち，このデータに350サンプルを加えた合計500サンプルを用いて，モンテカルロフィルタ法による内部状態推定を実施している。

図5には状態方程式推定と内部状態推定に用いる観測時系列を示している。図6，7，8には，それぞれ，推定されたトレンド周期成分および分散変動の推定値を，観測値（モデルを仮定して生成した時系列）との比較を示している。GPによる内部状態推定の結果は，次のようにまとめることができる。

(1)状態方程式の推定

GPによる推定改善の結果として，約7世代のGP繰り返しにおいて，あらかじめ与えていた状態方程式と同じ形の方程式が推定されており，GPによる関数近似の能力が優れていることが示されている。

(2)トレンドの推定

内部状態の中では大きな数値を占める成分となっており，この推定が推定の精度全体を規定すると言える。GP手法により状態方程式が推定されたあとのモンテカルロフィルタ法による内部状態トレンドの推定は，図を見る限り極めて正確であり，GP手法の有効性が確認できる結果となっている。

(3)周期成分と分散成分の推定

周期成分と分散成分の推定についてはこれらの成分の絶対値がトレンド成分に比べて小さいことおよびどちらもランダムな変動を代表しており相互に分離できない可能性があることなどが説明できる。モンテカルロフィルタにより推定された周期成分は観測値よりも小さくなる傾向がある。これとは反対に分散成分は推定された値が観測値よりも大きくなる傾向になっている。なお式(16)に定義された対数尤度は，このシミュレーション例では-29.84となっている。

以上のように，GP手法による状態方程式の推定に関しては，ほぼ満足できる結果となっている。しかし，絶対値が相対的に小さく，ランダムな性質が類似している周期成分と分散成分を分離することには，やや問題がある。この問題の理由の1つとして，GPによる過学習がある。GPの第7世代において，すでに正確な状態方程式を推定しているが，このあとも学習を継続することにより，最適な状態から乖離する現象が見られる。これを制御する方法が必要であろう。

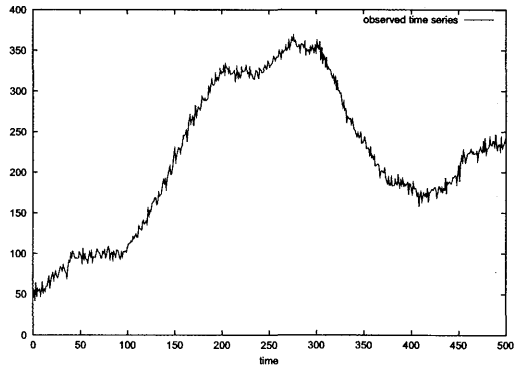


図5：人工的に生成された時系列

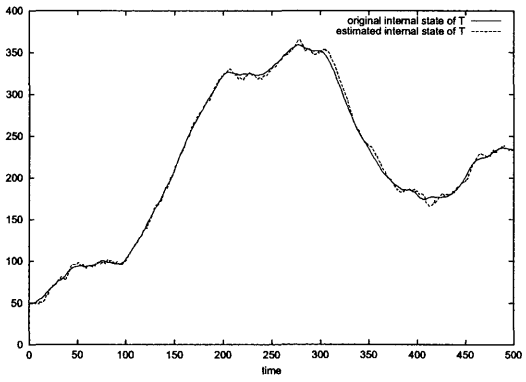


図6：内部状態 T_nの推定時系列

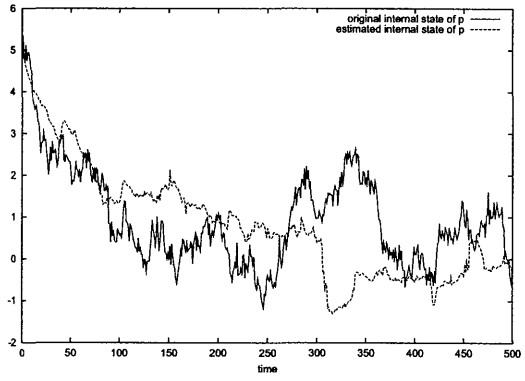


図7：内部状態 p_nの推定時系列

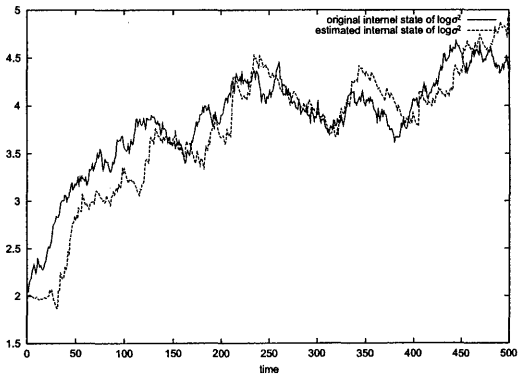


図 8 ：内部状態 $\log\sigma_n^2$ の推定時系列

4.2　実際の時系列への適用

次に，本論文の手法を実際の経済時系列にあてはめ，その結果から時系列の特徴を検証する。事例として日経平均株価を観測データとして用いた場合の結果を示す。

用いるデータは，1987年1月5日から1988年10月27日までの日経平均株価である。最初に，1987年1月5日から1987年9月22日までの200サンプルのデータを用いて，GPにより状態方程式を推定し，そのあとに，更に，1988年10月27日までの300サンプルを追加して内部状態の推定を行う。

状態方程式の形は未知であると仮定するが，その次数は3であることを，あらかじめ仮定しておく。また，内部状態の変数の数は，これまでの議論を参考にして3変数であると仮定しておく。従って，内部状態の次元数は9となる。これを形式的に示すと，次のようになる。

$$x1_n = f1(x_{n-1},\ e_{1n}) \tag{23}$$

$$x2_n = f2(x_{n-1},\ e_{2n}) \tag{24}$$

$$x3_n = f3(x_{n-1},\ e_{3n}) \tag{25}$$

ここで変数ベクトル x_n は次のように定義されている。

$$x_n = (x1_n,\ x1_{n-1},\ x1_{n-2},\ x2_n,\ x2_{n-1},\ x2_{n-2},\ x3_n,\ x3_{n-1},\ x3_{n-2}) \tag{26}$$

また，観測される時系列のモデルは，次で与えられると仮定する。

$$y_n = x1_n + x2_n + x3_n\varepsilon_n,\ \varepsilon_n \sim N(0,\ 1.0^2) \tag{27}$$

推定された状態方程式は，次のようになる。

$$x1_n = x1_{n-2} + 27.59 + e1_n \tag{28}$$

$$x2_n = x2_{n-3} - 35.01 + e3_n \tag{29}$$

$$x3_n = 42.29 \times x3_{n-1} \times x3_{n-2}^2 + e3_n \tag{30}$$

$$e1_n \sim N(0,\ 94.9^2),\ e2_n \sim N(0,\ 11.7^2)$$
$$e3_n \sim N(0,\ 0.05^2) \tag{31}$$

図9，10，11，12には観測値として与えられている日経平均株価と，推定された3つの内部状態の時系列を示している。これらの内部状態を，その形状から類推すると，第1番目の時系列はトレンドに相当すると考えられ，1次の階差モデルとなっている。その構成要素としては，定数に相当する部分と，確率的なシステムノイズを階差の成分としていることが分かる。この時系列変化に関しては，トレンドの性質である，時系列の概要をなぞるように変化している。しかし，時刻が210ステップを越えるあたりから，変動する様相を呈している。この大きな理由は，ブラックマンデーに起因する激しい変動が内部的に発生していることに対応している。

他の変動要因であるについては，同様に1次の階差モデルで記述され，定数部分とシステム

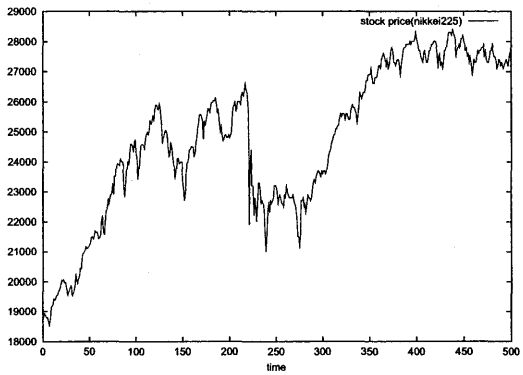


図 9 ：日経平均株価の時系列

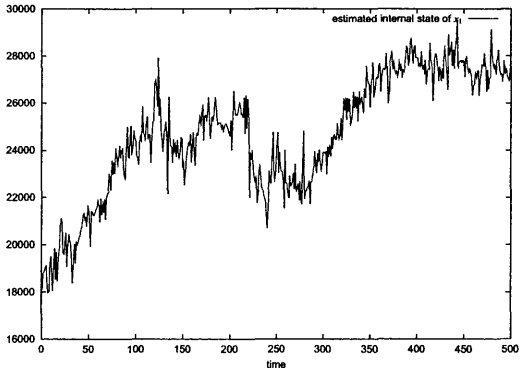


図10：内部状態 $x1_n$の推定時系列

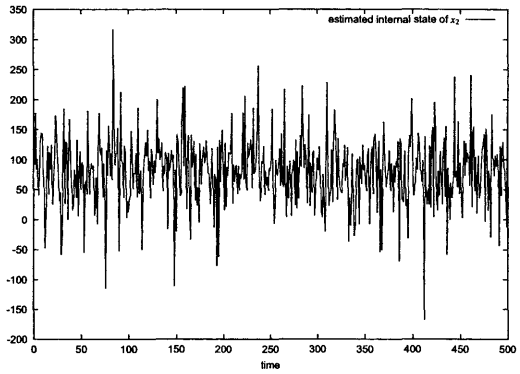


図11：内部状態 $x2_n$の推定時系列

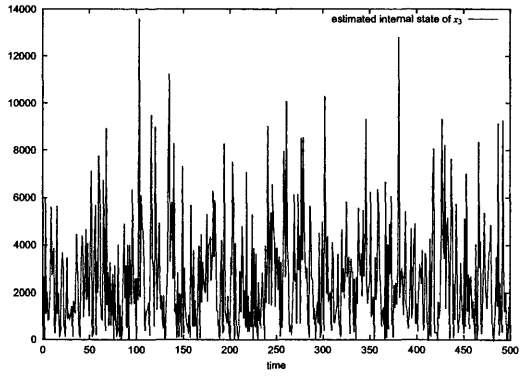


図12：内部状態 $x3_n$の推定時系列

ノイズを，その階差成分として含んでいる。これらの成分の中で第2成分である $x2_n$の解釈としては，第1成分である $x1_n$がブラックマンデーの影響を受けて激しく変動しており，この変動成分をキャンセルしながら進行する成分であると解釈される。また，第3成分は，ゼロの付近で変動する微小な成分であり，時系列の全体の特性には大きく寄与するものではない。

以上のように，状態方程式を仮定しない場合においても，トレンドや分散変動，およびランダムな成分を分離できることは，GP手法によるアプローチが有効であることを示している。

4.3 非ガウス型システムノイズ近似への拡張

これまでの議論では，システムノイズの生成は正規乱数によりなれると仮定したが，これまでの手法を拡張して，この関数形状をGP手法により関数近似することも可能である。非ガウス型ノイズの有効性に関しては，例えばバブル期における株価や突発的な株価変動など，まれにしか発生しないが，大きな影響力をもつ現象も含めたモデルにする場合に用いられる。

これまでの研究では，あらかじめ非ガウス型の関数の形状を仮定しており，現実のデータを説明するのみ適切であるかどうかの点に関して検討がなされてきた。株価生成などのモデル分析では，$v_n=(e_{1n}, e_{2n}, e_{3n})$ で記述されているシステムノイズの生成部分において，コーシー分布，ピアソン型分布あるいは2つの分布を結合した混合分布などの非ガウス型分布を導入した解析が試みられている。

その中で，特に注目されているのが，稀にしか起こらないが，その影響が大きな現象を表現する近似である。株価の急落や，インターネットにおけるトラヒックの急増などがこれに相当する。しかし，複雑な現象をモデル化するには十分ではないことも示されている。

本論文では，関数近似の基本方法としてGPを用いて関数形状を推定することが可能であるこ

とだけを示す。その実際的な適用は今後の議論としたい。すなわち，システムノイズを非ガウス型にするには，特に大幅な手法の変更は必要としない。これまでのGP手法による密度関数の近似は予測モデルにおける関数近似に限定されていたが，これに加えてシステムノイズの生成過程に，この関数近似を導入すればよい。

ただし，GP手法による関数近似を2段階にわたって実施することになるので，近似に必要な計算時間が膨大となる可能性がある。そのため，システムノイズを生成する確率密度関数を近似する場合の手順を，大幅に簡素化するなどの改善策が必要である。

5　むすび

本論文では，モンテカルロフィルタ法を適用し，内部状態を推定する場合に未知である状態方程式近似にGPを適用する方法を提案した。更に，この手法を既知のシステムの場合と，実際的な時系列へと適用した。

今後の課題として，計算時間を短縮するアルゴリズムの開発などがあり，検討を進めていく予定である。

参 考 文 献

[1] R. Engel, "Autoregressive conditional heterskedasticity with estimates of the variance of united kingdom inflation", Econometrica, Vol.50, No.4, pp.987-1007, 1982.

[2] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models", Journal of Computational and Graphical Statistics, Vol.5, No.1, pp.1-25, 1996.

[3] 北川源四郎，佐藤整尚，永原裕一，"非ガウス型状態空間モデルによる確率的ボラティリイティ・モデルの推定"，日本銀行金融研究所，Discussion Paper, No.98-J-12, 1998.

[4] 北川源四郎，佐藤整尚，"一般化状態空間モデルによる分散変動時系列の解析"，日本銀行金融研究所，Discussion Paper, No.98-J-22, 1998.

[5] N. Ikoma, "Estimation of time varying peak of power spectrum based on non-Gaussian nonlinear state space modeling," Signal Processing, Vol.49, pp.85-95, 1996.

[6] 荒木時彦，"モンテカルロ・フィルタを用いた損傷推定に関する研究"，京都大学21世紀講演会予稿，No.A13, 2004.

[7] Y. Ikeda and S. Tokinaga, "Approximation of chaotic dynamics by using smaller number of data based upon the genetic programming", IEICE Trans.Fundamentals, Vol.E83-A, No.8, pp.1599-1607, 2000.

[8] Y. Ikeda and S. Tokinaga, "Controlling the chaotic dynamics by using approximated system equations obtained by the genetic programming", IEICE Trans.Fundamentals, Vol.E84-A, No.9, pp.2118-2127, 2001.

[9] X. Chen and S. Tokinaga, "Approximation of chaotic dynamics for input pricing at service facilities based on the GP and the control of chaos", IEICE Trans.Fundamentals, Vol.E85-A, No.9, pp.2107-2117, 2002.

[10] 矢加部 正幸，時永祥三，"遺伝的プログラミグを用いたCNNによる拡散モデルの近似と同期化への応用"，電子情報通信学会論文誌，Vol.E85-A, No.5, pp.548-559, 2002.

[11] 池田 欽一，"共進化によるカオス常微分システムの推定"，電子情報通信学会論文誌，Vol.E85-A, No.4, pp.424-433, 2002.

[12] 陳 暁栄，時永祥三，"共進化GPを用いたマルチエージェントシステムの構成とその人工市場分析への応用"，電子情報通信学会論文誌，Vol.E86-A, No.10, pp.1038-1048, 2003.

[13] 池田欽一，時永祥三，"共進化遺伝的プログラミングによる社会学習を考慮したマルチエージェントシステムの構成とその人工株式市場分析への応用"，経営情報学会論文誌，Vol.13, No.3, pp.17-35, 2003.

[14] Y. Ikeda and S. Tokinaga, "Chaoticity and fractality analysis of an artificial stock market by the multi-agent systems based on the co-evulutionary genetic programming", IEICE

Trans.Fundamentals, Vol.E87-A, No.9, pp. 2387-2394, 2004.

[15] J. R. Koza, Genetic Programming, MIT Press, 1992.

[16] M. J. Keith and M. C. Martin, "Genetic programming in C++: Implementation issues", in (ed) K. E. Kinnerar, Jr., Advance in Genetic Programming MIT Press, 1994.