

自己相似性を用いたフラクタル時系列の予測とその応用

時永 祥三 池田 欽一 竹林 渉
九州大学経済学部経済工学科

あらまし:本報告では、フラクタル時系列に対する予測手法を展開するとともに、予測誤差の検討、フラクタル次元の推定、および応用例について示す。まず、フラクタル時系列をスケール関数の展開形式で表現されたインパルス応答関数と入力信号との畳み込みにより表現するモデルを仮定し、時間軸の伸長に対してインパルス応答の自己相似的な性質が保持されることを用いた予測手法を示す。また、フラクタル時系列のウェーブレット変換係数を用いてフラクタル次元を推定する。計算例として、シミュレーションにより発生させたフラクタル時系列の1ステップ先の予測誤差が極めて小さくなり、 n ステップ先の予測誤差についても一定の範囲内で小さくなることを示す。応用例として、株価の予測とオプション取引への応用、通信トラフィック予測、および需要予測について述べ、有効性を示す。

Prediction of Fractal Time-Series by using the Self-Similarity and its Application

Shozo Tokinaga, Yoshikazu Ikeda, Wataru Takebayashi

Department of Economic Engineering, Faculty of Economics, Kyushu University

This report deals with the prediction method for the time series bearing fractal geometry. The prediction error of the method, the estimation of parameters (fractal dimension et.al.) and several applications are also discussed. At first, the time series is represented by the convolution of the impulse response expanded by a set of scaling function and the input signal. Then, by using the time scale expansion, future values are estimated. By using the simulation study, it is shown that 1-step ahead prediction of the fractal time-series is very well, and n -step ahead prediction where the forecast is used as an observation is limited in a range. The method is applied to the forecast of stock price and option premium, the prediction of input traffic at the ATM, and the forecast of demands. These results show better performance compared to the conventional methods.

1 まえがき

時系列や図形などの一部分を拡大、縮小した場合に同じような形状をもつ相似形が出現する性質(自己相似性)を説明するフラクタル理論が展開されている。フラクタル理論は、物理化学現象など自然現象の解明、信号処理、図形符号化など工学分野をはじめ、さまざまに応用がなされている [1][2]。

本報告では、フラクタル時系列において時間軸のスケール伸長に対してインパルス応答の自己相似的な性質が保持されることを用いた予測手法を示し、また、時系列がフラクタル性を持つかどうかを検証する方法を示す [3],[4]。更に、これらを各種の時系列の予測問題へと適用し、有効性を確認する。特に、トラ

フィックと需要予測で示すように、ある確率的な分布をもったタスクの到着を仮定した場合には、これらを集計(統計多重)した時系列は極めて強いフラクタル性を持つことが示され、長期的な相関の発生メカニズムを解明する手がかりとなる。

以下では、まず2.においてフラクタル性をもつ時系列のインパルス応答関数をスケール関数により近似的に展開した場合に、時間軸方向にインパルス応答を伸長することにより予測が行える原理について説明する。3.では、さまざまなフラクタル次元の時系列に適用した場合の予測誤差などについて整理する。4.では、具体的な応用例として日本の証券市場における日々の株価予測、通信トラフィック予測、および需要予測について述べ、有効性を示す。

2 フラクタル時系列の予測手法

2.1 スケール関数による近似

一般的な線形時変入出力システム

$$y(t) = \int_0^{t_0} h(t, t-\tau)x(\tau)d\tau, t > t_0. \quad (1)$$

を考察する。線形時変システムのインパルス応答関数 $h(t, \tau)$ がスケール関数 $\phi(t)$ を用いて次のように展開されると仮定する。

$$h(t, \tau) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} h_{ij} \phi_{Ni}(t) \phi_{Nj}(\tau). \quad (2)$$

ただし、

$$\phi_{Ni}(t) = \phi(2^{-N}t - i). \quad (3)$$

$$\phi(t) = \begin{cases} 1, & 0 < t \leq 1; \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

特に、以下では入出力が同じ時系列である同定問題 ($y(t) = x(t)$ の場合) を考える [3],[4]。

いま、式 (1) による予測値 $\hat{x}(t)$ と入力時系列 $x(t)$ との差の最小 2 乗近似を考え、これを最小化するようにインパルス応答関数の係数 h_{ij} を決定する。計算にあたり式 (2) で $N=0$ とし、 j の範囲を $j=1 \sim 4$ に限定する。最適化の方法として最急降下法を用いているが本論文で考察する問題の範囲では誤った解への局所的な収束の問題はない。式 (4) を用いて観測値と予測値の乗誤差を最小化する場合のしきい値は 10^{-4} に選び、推定値が得られるまでの繰り返し回数は平均して 20000 ステップである。

2.2 時間軸の伸長による予測

いま、 $T_s \leq t \leq T_e$ は時系列 $x(t)$ が観測される時間区間であり $T_1 = T_e - T_s$ とする。また、 $0 < t \leq T_2$ は $x(t)$ を予測する区間とする。次の量を定義する。

$$b = a^D, a = T_2/T_1, T_2 > T_1. \quad (5)$$

D は時系列 $x(t)$ のフラクタル次元であり、 $1 < D < 2$ である。 $x(t)$ がフラクタル性を持つ場合には、その自己相似性により、 $0 < t \leq T_2$ において、次の式が近

似的に成立する。

$$x(t) = \int_0^{bt_0} h\left(\frac{t}{b}, \frac{t-\tau}{b}\right)x(\tau)d\tau, t > bt_0 \quad (6)$$

すなわち、時間軸が a 倍された領域の中に b 個のフラクタル図形が入ると解釈される [1]。スケール関数で近似された時系列のフラクタル性とは、インパルス応答関数により以下のように説明される。インパルス応答関数 $h(t, \tau)$ の性質はスケール伸長されても時系列の自己相似性から、同じ線形予測の関係式を成立させる性質を保持するので、区間 T_1 で推定された $h(t, \tau)$ を b 倍したものをいれれば、 a 倍に伸長した時間域においても、式 (6) の予測モデルが成立する。

計算においては、 $a = \text{整数}$ となるように T_1, T_2 を選んでいる (例: $T_2 = 2T_1$)。サンプリングされた時系列を仮定し、サンプリング間隔を 1 とするすなわち、式 (1) では $t = t_0 + 1$ であり、式 (6) では $t = bt_0 + b$ となるので、次の予測式となる。

$$x(bt_0 + b) = \int_0^{bt_0} h\left(\frac{t}{b}, \frac{t-\tau}{b}\right)x(\tau)d\tau, t > bt_0 \quad (7)$$

予測を議論する場合には、通常、現在までの観測データをもとにして、1 ステップ先のサンプルの予測値を計算することが行われる。これを、以下では b 時刻先の予測とよぶ。

これに対して、式 (7) に従って、逐次的に予測された値を観測値と見なして、予測を継続していく場合を考え、これにより n ステップ将来の値を予測する場合を nb 時刻先の予測とよんでおく。 nb 時刻先の予測においてもインパルス応答が計算されているので、線形予測をそのまま継続して用いる場合になどと比較して、予測誤差は極めて小さいものとなる。また、現在の時刻から離れた時刻の予測値を得ることができるので、株式投資におけるオプションの設定などに利用できる。

2.3 フラクタル次元の推定

次に、フラクタル時系列のパラメータをウェーブレット変換を用いて推定する方法について述べる。時系列がフラクタル性をもつ場合には、そのスペクトルの時間平均が周波数のべき乗に反比例する性質を持つことが証明されている [5][6]。すなわち、時系

列の分散, フラクタル次元を, それぞれ, σ^2, D とするとスペクトルは次のようになる。

$$S(\omega) = \sigma^2 \omega^{-\gamma} \quad (8)$$

$$\gamma = 5 - 2D \quad (9)$$

与えられた時系列 $x(t)$ をウェーブレット変換する。

$$x(t) = \sum_n \sum_m x_n^m \psi_n^m(t). \quad (10)$$

$$x_n^m = \int_{-\infty}^{\infty} x(t) \psi_n^m(t) dt. \quad (11)$$

ここで, $\psi_n^m(t)$ は基本関数 $\psi(t)$ に対する次のスケール, シフト変換とにより構成される。

$$\psi_n^m = 2^{m/2} \psi(2^m t - n). \quad (12)$$

ここで, m, n は, スケール変換, シフト変換のインデックスである。 $x(t)$ がフラクタル性をもつことから, ウェーブレット係数 x_n^m の満たすべき条件として, 次の関係式が得られる。

$$\text{var} x_n^m = \sigma^2 2^{-\gamma m}. \quad (13)$$

式 (13) の両辺の対数をとると, m についての線形の直線となるので, 左辺により計算されるデータに対して回帰直線を当てはめ, この直線との 2 乗平方誤差 $rmse$ の大きさによりフラクタル性を判定できる。

$$rmse = \left[\sum_m (\log(\text{var} x_n^m) - c_0 - c_1 m)^2 \right]^{1/2} / M \quad (14)$$

M は添字 m の取り得る個数である。

ウェーブレット係数 x_n^m を用いてフラクタル時系列の分散とフラクタル次元を推定する方法が示されている (詳細は省略する)。

3 シミュレーションによる予測誤差の検討

以下では, 式 (8) を用いた b 時刻先の予測, nb 時刻先の予測の誤差について, シミュレーションによる検討結果をまとめておく。フラクタル時系列データは fBm (fractional Brownian Motion) として発生させておく ($a = 2$, 時系列データサンプル数=1000)。

まず, 表 1 には b 時刻先の予測について, いくつ

かのフラクタル次元の時系列の平均予測誤差を示している。予測誤差の定義として, 期間内の時系列の最大値と最小値の差 (振幅) に対する予測誤差の割合 (%) をとっている。これより分かるように, $a = 2$ の場合には 0.3% 程度の数値となっており, 極めて小さな値である。

次に, 表 2 に nb 時刻先の予測について予測誤差の平均値を示している。この結果が示すように nb 時刻先の予測についても誤差は 6% 程度におさまっている。

線形予測を行なった結果について表は省略しているが, 本手法による予測と比較して誤差は極めて大きく, 例えば, $nb = 45$ において平均予測誤差は 20% 程度となっている。

表 1. 本手法の b 時刻先の予測誤差

次元 D	1.80	1.50	1.25
予測誤差	0.32	0.19	0.13

表 2. 本手法の nb 時刻先の予測誤差 ($D = 1.5$)

	nb=30	nb=60	nb=90	nb=120
予測誤差	3.7	4.9	5.2	5.6

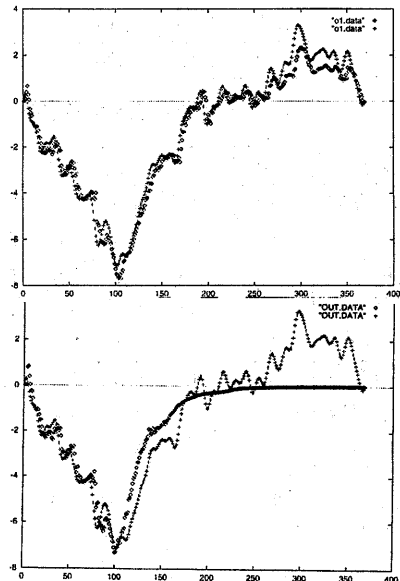


図 1. fBm の 1-step 先, n-step 先予測

4 応用例

4.1 株価時系列への適用

以下では、現実の株価についてフラクタル性の検証と次元推定、および予測の基本的な性質を調べる。用いたデータは日本の株式市場における電子化された日々の株価であり、概要は次のようである。ファイル:東洋経済新報社 CD 株価, 種類:一日の終値, 収録期間:1970年から1982年, サンプル:11業種からランダムに10銘柄を選択, データ長さ:500および2000

表3にはフラクタル性を検査する $rmse$ および次元 D の推定について、平均値と5種類の銘柄に対する結果を示している。 R_{500}, D_{500} はそれぞれサンプル数が500の場合の $rmse$ と次元 D を意味する。これらの結果より分かるように、全体的に $rmse$ は0.2程度となっている。この広がりについては、銘柄による差は大きくない。次元については、ほぼ、1.5と仮定してよいことが分かる。

表3. $rmse$, 次元の観測長さによる変化

銘柄	R_{500}	D_{500}	R_{2000}	D_{2000}
平均	0.17	1.49	0.14	1.51
三洋電気	0.22	1.34	0.15	1.51
ハウスイ	0.29	1.37	0.17	1.50
KDD	0.15	1.58	0.12	1.56
ダイソー	0.20	1.51	0.13	1.57
キャノン	0.17	1.49	0.14	1.51
定常波	0.47	1.12	0.34	1.25

表4. nb 時刻先の予測誤差 (%)

銘柄	$nb=28$	56	84	112	140
富士電機	4.0	4.7	4.3	5.4	6.2
日立工機	4.4	4.1	5.5	5.6	6.6
極洋	4.5	4.6	5.2	5.6	6.6
ラサ工業	5.9	4.6	6.7	6.2	7.0
平均	4.2	4.5	5.1	5.5	6.5
富士電機	4.1	4.3	6.5	7.4	7.6
日立工機	4.7	5.6	5.7	6.4	7.4
平均	4.3	5.3	5.4	5.4	6.5

表4には、株価予測の nb 時刻先の予測誤差について、平均値と、いくつかの個別銘柄に対する結果を示している。予測誤差は期間内の最大振幅に対する誤差の割合(パーセント)としている。サンプル数

=2000の場合を考察している。予測期間については、 nb 時刻先について、 $nb=28$ から $nb=140$ までをとり、この場合の予測誤差を示している。

これらの結果より分かるように、全体的に nb 時刻先の予測誤差については、平均値4%ないし6%は程度である。ここで考察するオプション取引は次のような方法で行なわれると仮定する。いま、例としてコールオプションを購入する場合を考察する。一定の期日(時刻) t あとでの原株の取引について、オプション価格 C_m が市場に示されているとする。一方、本報告で示した時系列予測の方法により、 t 時刻あとの予測株価 S_t からオプション価格の予測値 C_f を求めておく。すなわち、

$$C_f = \max[S_t - K, 0] \quad (15)$$

ここで、 K は行使価格である。

オプション取引引きを有利にすすめるには

$$C_f > C_m \quad (16)$$

のときに限ってオプションを購入すればよい。これ以外のケースでは、市場でのオプション価格は、予測値より高く割りだかな投資となるからである。

データは日経225の記録データを用いる。オプション取引のある期日についてオプションを購入することを仮定し限月が存在する1ヶ月、2ヶ月、3ヶ月オプションのみを購入することにする。モデルには時系列の分散を含んでいるので、統計的な推定方法を用いて計算する。

結果について表5にまとめている。結果から分かるように、本報告の予測手法を用いた場合には、20ないし55の利益となっている。これは、平均して投資を開始した時点での株価に対して約2%利益をあげていることになる。

表5. オプション取引の評価

	1カ月前	2カ月前	3カ月前
本手法による利益	20.8	36.3	54.9

4.2 通信トラヒックの予測

ATM交換機におけるトラヒックを表現するモデルとしてIPP(Interrupted Poisson Process)やMMPP(

Markov Modulated Poisson Process) が用いられる。MMPP は IPP を多重化したものと解釈されている。

N 個の局 i の IPP の多重化を仮定し、それぞれの局 i の IPP の *on* 期間におけるセルの発生数を平均値 λ_i の指数分布、*on* 期間と *off* 期間の長さを、それぞれ、平均 a_i, b_i の指数分布と仮定する。流体近似として、入力トラフィックを一定時間間隔 t ごとに観測する到着セル数で表現する。流体近似のもとでは、局 i の IPP のプロセスが継続して *on* 状態にある場合、あるいは、時間間隔 t の間に新たに局 i が *on* 状態になった場合、セル到着率 λ_i のトラフィックが継続して加わる。一方、*off* 状態に移行すればこのトラフィックレベルはセル到着率 λ_i だけ減少することになる。

なお、以下では、セル到着が指数分布ではなく、一定値となるケースも、同時に考察している。これは、交換機における挙動をマクロ的に見た場合に相当しており、理論解析でもモデルの簡単化と一定の計算精度が保証されることが示されている。

数値シミュレーションを条件を変えながら実行して検討した結果、多重化された IPP の場合によるトラフィック発生モデルにおいては、トラフィックがフラクタル時系列として近似できる条件として、次のようなことが推定される。

- 1) 局数が 20 以上であり、*on, off* 期間の長さについては一定の範囲にあること。具体的にはセル長が指数分布のとき *on* 期間の長さが少なくとも 50 以上、固定長のとき 20 以上であること更に *on* 期間が *off* より極端に長くない、あるいは短くない (0.2 倍以上かつ 4 倍程度以下) であることである。
- 2) 局数 N が 20 以上であれば、局数を増加させることにより、特に改善される傾向は見られない。
- 3) *on, off* 期間の長さの分布の分散 σ_a, σ_b の大きさは、極端な値を取らない限り (0.2 から 5.0 の範囲) フラクタル性にほとんど影響を与えていない。
- 4) セル長の分布について平均値の大きさはあまりフラクタル性には影響しないが、分散が大きな場合には結果が変化するので、分散は 0.2 から 5.0 の範囲にあることが望ましい。
- 5) 局からのセルの発生が一定 (固定長) である場合には、時系列のフラクタル性は顕著である。しかし、指数分布の場合には、1) から 4) までの条件が境界値に近づくと、フラクタルではないと見なされるケースが増加する。
- 6) 以上のことから入力トラフィックのセル長の分布が

固定長であるとして近似するか、あるいは位相が程度以上の分布により近似することが望ましい。この場合には、後で述べるトラフィック予測の誤差は小さくできるが、指数分布のままセル長の分布を仮定した場合には誤差が拡大するケースが増加する。

表 6. 多重化された IPP の rmse(セル長指数分布)

	N=5	20	50	100	200
$a/b = 0.25$	0.13	0.13	0.14	0.14	0.14
$a/b = 1.00$	0.19	0.16	0.18	0.18	0.18
$a/b = 0.40$	0.19	0.20	0.21	0.27	0.24

表 7. 多重化された IPP の rmse(セル長固定長)

	N=5	20	50	100	200
$a/b = 0.25$	0.11	0.11	0.11	0.11	0.12
$a/b = 1.00$	0.09	0.05	0.09	0.10	0.10
$a/b = 0.4.0$	0.11	0.11	0.08	0.09	0.10

交換機へのデータトラフィックを流体近似したモデルにおける待ち行列は、以下のようになる。

$$\frac{dQ(t)}{dt} = \begin{cases} J(t) - C, & \text{if } Q(t) > 0, \text{ or } J(t) > C; \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

ここで、 $Q(t)$ は待ち行列長さ、 $J(t)$ は入力トラフィック、 C は交換機の出線スピード、 t_0 はバッファが空となった最近の時刻である。時刻 t における待ち行列の長さは次の式で与えられる。

$$Q(t) = \int_{t_0}^t J(u) du - C(t - t_0). \quad (18)$$

従って、フラクタル性を用いた予測が可能であれば、あらかじめ時刻の待ち行列の長さを予測し、入力規制や伝送品質の低下について警告、制御などを行うことが可能となる。

以下では、入力トラフィックとして観測データと予測データとを用いた場合に、式 (21) より計算される待ち時間の差がどのようになるかをシミュレーションにより検討する。シミュレーションでは、 $J(t)$ の最大値が 40 程度になるように正規化している。なおバッファ容量は無量大と仮定し、出線スピードは有限の値を設定する。

シミュレーションでは、 nb 時刻先の待ち時間の誤差を、観測された入力トラフィックを式 (18) の $J(t)$ として用いた場合の待ち時間と、 nb 時刻先の予測入力

トラフィックを式(18)の $J(t)$ として用いた場合の待ち時間との差を数値として示している。予測の範囲は、 nb 時刻先の入力トラフィックの予測精度を考慮して 20,40 時刻先に限定している。待ち時間を数値で示しているが、入力トラフィックの最大振幅が 40 程度であることを考慮すると、その相対的な大きさが分かる。シミュレーションではいくつかの出線スピードについての結果を示している。

表の結果においては出線スピードが小さくなるに従って待ち時間も増加するので誤差も拡大している。しかし、出線スピードが比較的大きな場合には待ち時間の予測も良好である。

表 8. 入力トラフィックの待ち時間予測の誤差

C	nb=20			nb=40		
	25	30	35	25	30	35
	25.2	11.3	2.1	27.2	12.3	3.2

4.3 需要予測

実際の商品の需要を示す時系列に対して、本手法を適用した場合の結果を整理する。以下に処理手順を示す。

まず、トレンドを除去する。数年にわたる需要時系列は緩やかなトレンドを含むのでこれを除去した時系列を求め。次に、得られた時系列に対してフラクタル性があるかどうかを $rmse$ を用いて検証する。これらの準備のあと、フラクタル性を持つ時系列に対して予測手法を適用して、その誤差を検討する。結果について以下にまとめている。

(1) 記録周期による違い

表 9 には、需要データの記録周期の違いによる $rmse$ の状況を示している。これらの結果より分かるように、記録の周期が短い周期 1 の場合には時系列そのものがスパイク状であり、フラクタルであるとは言えない。これに対して記録期間の長いケースではフラクタル性の顕著な時系列となる。この場合、全体的に $rmse$ は 0.2 程度となっている。この広がりについては、商品による差が見られる。

記録周期によりフラクタル性に差が見られるのは、商品の発注がまとめてなされること、従って、これが時間を隔てて相関(長期的な相関)を発生している可能性がある。

表 9. $rmse$ の記録周期による変化

周期 1	周期 2	周期 3	周期 4
0.318	0.185	0.102	0.078

(2) 予測誤差

表 10 には予測誤差についてまとめている。フラクタル性が見られるのは周期 1, 2, 3 のデータであるので、これらについて平均予測誤差を示す(誤差の定義は σ_B の場合と同様)。また、フラクタル次元についても約 1.5 の近辺である。

表 10. 記録周期と nb 時刻先の予測誤差

記録周期	$nb = 30$	$nb = 60$	$nb = 120$
周期 2	5.8	6.7	8.9
周期 3	5.8	7.6	8.8
周期 4	4.5	7.7	9.1

5 むすび

今後の課題としては、さまざまな多次元信号をフラクタルで近似するための効率的な方法の開発、長期的な相関だけではなく短期的な相関も考慮に入れた場合の予測理論があり、検討を加えていきたい。

文 献

- [1] Mandelbrot, B. and Van. N.: The Fractional Brownian motions, fractional noises and applications, *SIAM Rev.*, 10, pp.422-436 (Oct. 1968).
- [2] 高安秀樹: フラクタル, 朝倉書店, 1986.
- [3] 時永祥三, 森保洋, 宮崎明雄, 島津宣之: "時系列のフラクタル性を用いた予測手法とその応用", 信学論 (A), J79-A, 11, pp.1793-1800 (1996-11).
- [4] 時永祥三, 森保洋, 宮崎明雄, 島津宣之: "スケール伸長変換およびウェーブレット変換によるパラメータ推定を用いた時系列予測", 信学論 (A), J79-A, 12, pp.2054-2062 (1996-12).
- [5] Wornell, G. W. and Oppenheim, A. V.: "Estimation of fractal signals from noisy measurement using Wavelets", *IEEE Trans. Signal Processing*, Vol. 40, No. 3, pp.611-623 (March 1992).
- [6] Wornell, G. W.: "Wavelet-based representation for the $1/f$ family of fractal processes", *Proc IEEE*, Vol. 81, No. 10, pp.1428-1450 (Oct. 1993)