決定論的ジャンプ過程のシステム同定と長期予測に適したサンプリング手法の検討

大塚 陽介[1,a)]　鈴木 智也[1]

受付日 2011年8月18日，再受付日 2011年10月11日，
採録日 2011年11月4日

概要：経済市場の価格変動など，不等時間間隔で変化するシステムを観測する場合，変動ごとに現象をとらえる方法と，物理時間に基づいて等時間間隔で現象をとらえる方法がある．世界情勢やディーラの思惑など，現象を生み出す背景活動は連続的かつ多変量である可能性を考慮すれば，物理時間を無視した前者のサンプリングを安易に適用できない．そこで本研究では，決定論的な背景活動を有するジャンプ過程をモデル化し，2 種類の観測方法によって得られた時系列データから，元の背景活動を特徴づける要因を同定できるのかを実験した．なお，システムの決定論性を保持するサンプリングであれば，優れた時系列予測を実現できると考えられる．しかし長期予測においては，むしろシステムの特徴を破壊してしまう等時間間隔サンプリングの方が高い予測精度を得た．この理由として，観測データの質と予測反復回数に関するトレードオフが考えられ，数理モデルを通じて検証実験を行った．その結果，システム同定と長期予測では適切なサンプリング方法が異なり，その理由を数理モデルの観点から考察する．

キーワード：ジャンプ過程，決定論的システム，サロゲート法，非線形予測法

Data Sampling Strategies for Nonlinear Analyses and Predictions of Deterministic Jump Systems

Yosuke Otsuka[1,a)]　Tomoya Suzuki[1]

Received: August 18, 2011, Revised: October 11, 2011,
Accepted: November 4, 2011

Abstract: Real jump processes such as dealing prices look like discrete, but they are often derived by a continuous background dynamics. To record their movements as time-series data, two sampling methods have been applied: the nonuniform sampling based on discrete jump timings and the uniform sampling based on a continuous background dynamics. Because financial systems are continuously affected by many elements, such as world news and dealers' minds, the nonuniform sampling must be applied carefully. That is why, we examine whether the nonuniform sampling can detect the background dynamics better than the uniform sampling. Next, a good sampling method is generally considered as a good prediction method, but we can see that the uniform sampling method works rather better for long-term predictions in real foreign-exchange markets. To explain the reason why the appropriate sampling for system detection and that for long-term prediction are different, we preform numerical simulations applying our jump models derived by background dynamical systems, and demonstrate that the difference is caused by the trade-off between the quality of sampled data and the iteration number of short-term predictions.

Keywords: jump process, dynamical system, surrogate method, nonlinear prediction

[1] 茨城大学大学院理工学研究科知能システム工学専攻
Graduate School of Science and Engineering, Ibaraki University, Hitachi, Ibaraki 316–8511, Japan
[a)] 11nm911n@hcs.ibaraki.ac.jp

1. はじめに

実世界には，為替価格の変動や神経細胞の発火など，不等時間間隔で変化する現象が多く存在する．その変動の様子は離散的に見えるが，しかし根拠なく唐突に変動が発現しているとは考え難い．つまり変動の背後には，それを生み出すダイナミクスが連続的に駆動しているであろう．たとえば市場価格は，絶えず思考する人間の取引行動によって変化し，その思考は世界中のニュースからリアルタイムに影響を受ける．神経細胞においては，連続的に変化する内部電位が閾値をオーバしたときに発火が起こる．つまり不等時間間隔で変化する現象は，連続的な背景活動が何らかのきっかけで表面化することで引き起こされる．

このような現象を時系列データとして記録し，解析するには 2 つのサンプリング方法が考えられる．まずは連続的な背景活動に着目し，物理時間に従って等時間間隔で変動をサンプリングする方法である．サンプリングの時間間隔を拡大すれば，少ないデータ数でも長期の現象を解析できるという利点がある．しかし，この時間間隔に応じて重複や欠損が発生し，変動の全体像は破壊されてしまう．一方，変動タイミングに着目して不等時間間隔でサンプリングすれば，連続的な背景活動を無視することになり，データ間の物理時間は伸縮してしまう．また，長期の現象を記録するにはデータ量が膨大になりがちである．このように，サンプリングに関する時間スケールには長所と短所が存在する．特に経済データの解析では，全変動を記録した高頻度データの入手が可能になったものの，解析に適切な時間スケールは不明である [1], [2]．そのため現在もなお，分次や日次データのような等時間間隔で観測されたデータを用いて解析する研究事例は少なくない．

そこで我々の先行研究 [3] では，決定論的な背景活動を有するジャンプ過程をモデル化し，2 種類の観測方法によって得られた時系列データから元の決定論性を同定できるのかを実験した．その結果，変動が生起するタイミングでサンプリングする方法が，最も適切に背景活動の特徴を抽出できた．またその理由は，連続的な背景活動から離散的な変動を生成する機構をポアンカレ断面ととらえるならば，背景活動が持つ基本的特徴は，離散的な不等時間間隔変動に引継がれるからだと解釈している．

しかし世の中の現象のすべてが，ポアンカレ断面のように背景活動の特徴を保存しているとは考え難い．そこで本研究では，連続的な背景活動から離散的な現象に写像する際に，ポアンカレ断面のような特徴を保存する機構が不完全である様子を，乱数によって表現する．そしてこの不完全さの増大によって，先行研究で得られた知見がどのように変化するのかを検証する．

さらに先行研究 [3] では，最も単純な非線形予測モデル [4] で検定統計量を算出し，システム同定実験を行っていた．そこで本研究では，実験の信頼性を向上させるべく，より一般的な非線形予測モデル [5] を適用する．このように数理モデルおよび予測モデルを改良して，データサンプリングにおける適切な時間スケールを再検討する．

さらに経済データ解析の工学的応用として，長期予測のために最適なサンプリング方法も検討する．通常，システムの特徴を保持するサンプリングであれば，時系列予測も精度良く行えると推測されるが，しかし長期予測では予測の反復による誤差の拡大を考慮する必要がある．たとえ短時間間隔のサンプリングがシステムの特徴を保持できるとしても，長期予測のためには，1 ステップ予測を繰り返すことで予測経過時間を延長させる必要がある．一方，等時間間隔によるサンプリングではサンプリングごとの時間間隔を可変できるので，少ない反復回数で長期予測を実現できる．しかし先述のとおり，サンプリング間隔を拡大するほど，背景活動の特徴を破壊してしまう．つまり長期予測においても，サンプリングに関する時間スケールには長所と短所が存在し，観測データの質と予測反復回数に関するトレードオフがある．そこで本研究では，実際の為替取引価格データに対しても長期予測を行い，最適なサンプリング方法を検証する．さらに，ジャンプ過程を模擬した先述の数理モデルに対しても長期予測を行い，得られた結果について数理モデルの観点から考察する．

2. 決定論的ジャンプ過程

価格変動やニューロンの発火など，不等時間間隔で変化する現象は点過程に分類される．また，価格変動のように変動幅が一定ではない場合は，一般にマーク付き点過程と呼ばれる．この変動幅を $m(t)$ とすると，

$$m(t) = \begin{cases} 0 & \text{if } t \neq t_n \\ m(t_n) & \text{if } t = t_n \end{cases} \tag{1}$$

ここで t_n は変動の生起時刻，$m(t_n)$ はその変動幅，n は変動の生起番号である．変動 n は離散的に生起されるが，時刻 t は連続的である．変動幅を時間積分すると，現在のシステムの状態値 $x(t)$ が算出される．

$$x(t) = x(0) + \int_0^t m(T)dT \tag{2}$$

この状態値 $x(t)$ は**図 1** のようになり，この振舞いはジャンプ過程と呼ばれる．なお，初期値 $x(0) = 0$ とした．状態値 $x(t)$ を等時間間隔 s でサンプリングすれば，s が大きいほど欠落するデータ量が増え，逆に s が小さいほど同じ状態値が重複して観測される危険性が生じる．

本研究では式 (1)，(2) に決定論的ダイナミクスが内在するように，ローレンツ方程式 [6] および池田写像 [7] を利用した．まず第 1 の数理モデルして，ジャンプ過程を生み出す背景活動に次式のローレンツ方程式をあてはめる．

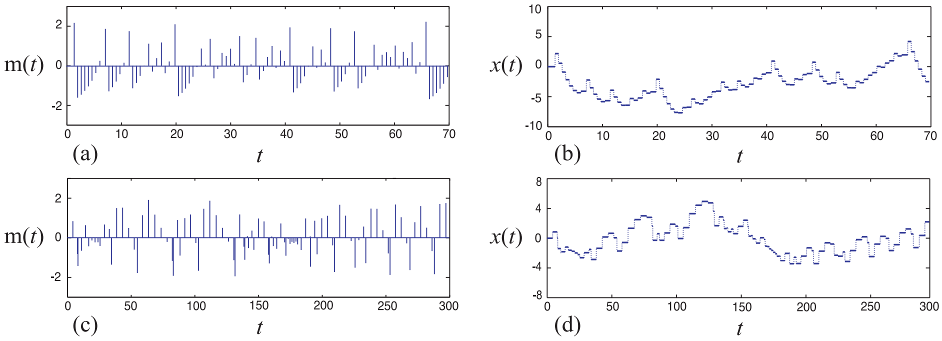

図 2　変動幅を表すマーク付き点過程 $m(t)$ と，変動自体を表すジャンプ過程 $x(t)$．(a)(b) はローレンツ型ジャンプ過程，(c)(d) は池田型ジャンプ過程より生成された

Fig. 2　The marked point process $m(t)$ and the jump process $x(t)$ generated by (a)(b) the Lorenz-type jump model and (c)(d) the Ikeda-type jump model.

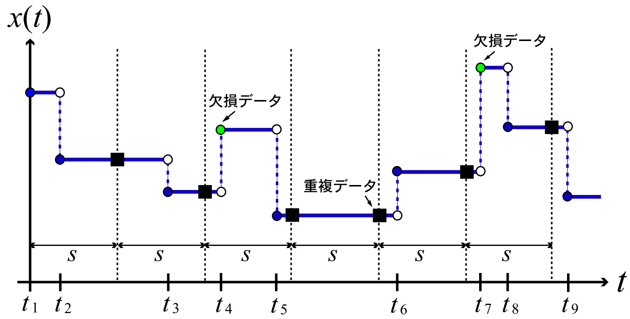

図 1　ジャンプ過程の状態値 $x(t)$ を等時間間隔 s で観測した様子．図中の四角印が観測されるデータであり，s の大きさに応じて欠損や重複データが発生する

Fig. 1　Continuous behavior of the jump process $x(t)$. If this behavior is sampled at every temporal uniform interval s, it causes missing data and overlapping data according to the size of s.

$$\begin{cases} \dot{X}_1 &= -\alpha X_1 + \alpha X_2 \\ \dot{X}_2 &= -X_1 X_3 + \beta X_1 - X_2 \\ \dot{X}_3 &= X_1 X_2 - \gamma X_3 \end{cases} \tag{3}$$

なお，$\alpha = 10$，$\beta = 28$，$\gamma = 8/3$ とした．このような連続的な背景活動が存在し，これをある条件によって離散化することはポアンカレ断面をとることに相当する．特にローレンツ方程式の場合，$X_1 X_2 = \gamma X_3$（つまり $\dot{X}_3 = 0$）の断面をとると，ローレンツ方程式の基本的特徴を保持した離散データを得られることが知られている．その際，主に X_3 の極大値が離散データとして用いられる [4]．そこで本研究でも，第 3 変数 X_3 の極大値 $X_3(t_n)$ を用い，この生起時刻を t_n とすると，**図 2** (a) のマーク付き点過程 $m(t)$ を表現できる．さらに式 (2) の時間積分によって，システムの状態値 $x(t)$ は図 2 (b) のジャンプ過程として実現できる．ただし $x(t)$ が単調増加にならないように前処理として，極大値 $X_3(t_n)$ を平均値 0 かつ分散値 1 に標準化した．

さて，実世界のシステムでは，ポアンカレ断面のような機構によって，背景活動の特徴がうまく保持されているとは限らない．そこで本研究では，背景活動を離散化する際の不完全さをノイズを用いて表現する．つまり，次式のように正規乱数 $\eta(n)$ を極大値 $X_3(t_n)$ に付加することで，$m(t_n)$ に転移する式 (3) の特徴的構造の度合いを調節できるようにする．

$$\eta(n) \sim N(0, \sigma) \tag{4}$$

$$\sigma = q\sigma_0 \tag{5}$$

$$m(t_n) = X_3(t_n) + \eta(n) \tag{6}$$

ここで σ_0 は元データである $\{X_3(t_n)\}$ の標準偏差である．パラメータ q を変えることで，ノイズ量を可変にできる．ノイズを増やせば，背景活動の特徴的構造はうまく転移されないので，$m(t_n)$ はランダムな変動に近づく．本論文では，この数理モデルをローレンツ型ジャンプ過程と呼ぶ．

第 2 の数理モデルでは，ポアンカレ断面上に残存する背景活動の特徴的構造を直接的に表現すべく，一例として次式の池田写像 [7] を用いる．

$$\begin{cases} X_1(n+1) = \alpha + \beta\left[X_1(n)\cos(\theta(n)) - X_2(n)\sin(\theta(n))\right] \\ X_2(n+1) = \quad\ \beta\left[X_1(n)\sin(\theta(n)) + X_2(n)\cos(\theta(n))\right] \\ \qquad \theta(n) = \gamma - \kappa/\left[1 + X_1^2(n) + X_2^2(n)\right] \end{cases} \tag{7}$$

なお，$\alpha = 1.0$，$\beta = 0.9$，$\gamma = 0.4$，$\kappa = 6.0$ とした．$X_1(n)$ を式 (1) の $m(t_n)$，$\theta(n)$ を $\Delta t_n = t_n - t_{n-1}$ と見なすと，図 2 (c) のマーク付き点過程 $m(t)$ を表現できる．ただし，$\Delta t_n < 0$ を避けるために，$\Delta t_n = \theta(t) - \min(\theta(t))$ としている．さらに式 (2) の時間積分によって，システムの状態値 $x(t)$ は図 2 (d) のジャンプ過程となる．なお，式 (7) の $\{X_1(n)\}$ の標準偏差を σ_0 とすると，式 (4)～(6) と同様に，

$$\eta(n) \sim N(0, q\sigma_0) \tag{8}$$

$$m(t_n) = X_1(t_n) + \eta(n) \tag{9}$$

となり，パラメータ q によって背景活動の特徴的構造の残存度を可変にできる．本論文では，この数理モデルを池田型ジャンプ過程と呼ぶ．

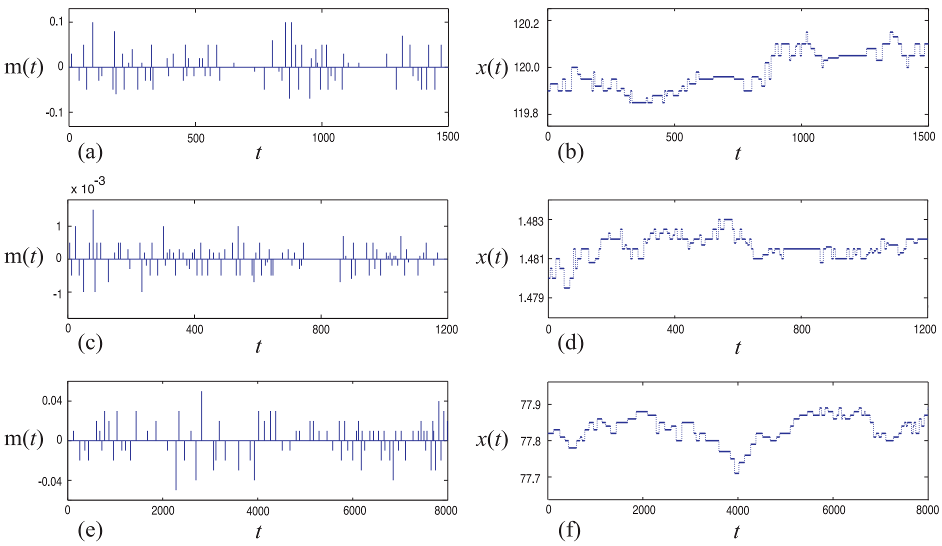

図 3　為替市場における取引価格の変動幅 $m(t)$ と取引価格 $x(t)$．(a)(b) は JPY/USD の為替レート，(c)(d) は USD/DEM の為替レート，(e)(f) は JPY/DEM の為替レート

Fig. 3　Real foreign-exchange prices $x(t)$ and their movements $m(t)$ of (a)(b) the USD/DEM, (c)(d) the USD/DEM, and (e)(f) the JPY/DEM.

表 1　解析に用いた為替取引価格データの特徴

Table 1　Characteristics of real foreign-exchange price data for analyses.

為替レート	データ数	$\langle\Delta t_n\rangle$	異常値の閾値 θ
JPY/USD	442,159	73 [秒]	1,812 [秒]
USD/DEM	1,187,395	27 [秒]	675 [秒]
JPY/DEM	125,426	255 [秒]	6,389 [秒]

これらと類似の振舞いを示す実例として，為替取引価格データ [8] を解析する．用いたデータは次の 3 通貨：(1) 日本円対アメリカドル（JPY/USD），(2) アメリカドル対ドイツマルク（USD/DEM），(3) 日本円対ドイツマルク（JPY/DEM）であり，いずれも 1993/1/1〜1993/12/31 の売値を用いた．前処理として，価格変化のともなわない取引データ，つまり $m(t_n) = 0$ を除去した．さらに，夜間・休日などの市場参加者の減少にともない取引時間間隔 Δt_n が拡大するため，$\Delta t_n \geq \theta$[秒] の場合，これを異常値と見なし時間間隔 $\Delta t_n = \theta$ に補正した．この閾値 θ は Δt_n の上位 0.1%に基づいて設定し，異常値を削除した後の各データの詳細を**表 1** に示す．さらに，前処理を施した価格変動幅 $m(t)$ および価格変動 $x(t)$ を**図 3** に示す．いずれの市場も，図 2 の数理モデルの振舞いと類似している．

3.　システム同定のためのサンプリング実験

3.1　2 種類のサンプリング方法

ジャンプ過程 $x(t)$ の振舞いに対して，システムの特徴を欠落させずに時系列データ $x(ns)$ を観測するには，どのようにサンプリングすればよいであろうか？　もし，ポアンカレ断面によって連続的な背景活動を点過程 $m(t_n)$ として切り取る際に，背景活動の特徴的構造をうまく残存させることができていれば，時刻 t_n でサンプリングするのが効果的であろう．経済データ分析ではこれをティックデータと呼ぶため，本研究では変動ごとにサンプリングする方法をティックサンプリングと呼ぶ．しかし実データの場合，ポアンカレ断面の不完全さをノイズで表現したように，ポアンカレ断面上によって特徴的構造を的確にとらえられるとは限らない．また時間間隔が短いサンプリングでは，データ数が膨大になり長期間のデータ解析が困難になるので，実際の解析では毎分や毎時データのように等時間間隔サンプリングによってデータ量の削減が施される場合が少なくない．そこで本研究では，このティックサンプリングと等時間間隔サンプリングの 2 手法について比較実験を行う．

まず，サンプリング間隔 $s = r\langle\Delta t_n\rangle$ によってシステムの振舞い $x(t)$ を観測することにする．ここで $\langle\Delta t_n\rangle$ は Δt_n の平均値を表し，r をサンプリング比率と呼ぶ．なお，ローレンツ型ジャンプ過程では $\langle\Delta t_n\rangle = 0.8$，池田型ジャンプ過程では $\langle\Delta t_n\rangle = 2.9$ である．サンプリングされた観測データは $x(ns)$ であるが，サンプリング後は離散データになるので，以後 $x(ns)$ を x_n（$n = 1, 2, \cdots, N$）と表記する．ここで n はサンプリング番号である．比率 r は自由に調節できるパラメータであり，$r > 1$ であるほど欠落するデータ量が増え，逆に $r < 1$ であるほど同じ状態値が重複して観測される（図 1）．なお便宜上，$r = 0$ の場合はティックデータのように時刻 t_n で $x(t)$ をサンプリングするものとする．

その後，時系列解析における一般的な前処理として主要なトレンド成分を除去すべく，階差データ $\dot{x}_n = x_{n+1} - x_n$ に変換して以後の解析に用いる [9]．なお，この処理は変動幅 $m(t)$ をサンプリング間隔で時間積分することに対応し

ている．

$$\dot{x}_n = \int_n^{n+1} m(t)dt \tag{10}$$

欠損が発生するほど，サンプリング間隔 s 内で多くの $m(t_n)$ が生起し，それらが 1 つの積分値として加算されてしまう．また重複とは，まったく $m(t_n)$ が生起しないうちにサンプリング間隔 s が経過して $\dot{x}_n = 0$ となることに相当する．

3.2　サロゲートデータ法

次に，解析対象データ $\dot{x}_n$ 内にシステム本来の特徴が残存しているかを確かめるべく，サロゲートデータ法 [10], [11], [12], [13], [14] による統計的仮説検定を行う．設定する帰無仮説の違いによって，サロゲートデータの作成方法は異なる．

まず，RS（random shuffled）サロゲート法 [10] では，解析対象データは時間構造を有することを仮定する．それゆえ，解析対象データをランダムにシャッフルし，時間構造を完全に破壊したサロゲートデータを生成する．もしオリジナルデータとサロゲートデータに有意差が確認されれば，時間構造は解析対象のシステムにおいて必須の特性であるといえる．

次に，SS（small shuffled）サロゲート法 [11] では，解析対象データは短期間の時間構造を有することを仮定する．解析対象データ内に含まれる短期間の成分どうしをランダムにシャッフルすることで，時間的な局所構造を破壊する．シャッフルさせる時間幅は文献 [11] に従った．このように作成されたサロゲートデータとオリジナルデータに有意差が確認されれば，短期の時間構造は解析対象のシステムにおいて必須の特性であるといえる．

最後に，IAAFT（iterated amplitude adjusted Fourier transformed）サロゲート法 [14] では，解析対象データは非線形構造を有することを仮定する．解析対象データのフーリエ変換によって得られた位相をランダム化した後，逆フーリエ変換することでサロゲートデータを生成する．さらに，オリジナルデータの頻度分布を保存するように補正を施す．位相のランダム化と頻度分布の補正を繰り返すことで，オリジナルデータの線形構造をうまく保存し，非線形構造のみを破壊する．オリジナルデータとの有意差が確認されれば，非線形構造は解析対象のシステムにおいて必須の特性であるといえる．

いずれのサロゲートデータ法においても，オリジナルの解析対象データ $\dot{x}_n$ から，各サロゲートデータ $\tilde{x}_{in}$ を 60 本作成し（$i = 1 \sim 60$），統計的仮説検定を行う．

3.3　検定統計量としての非線形予測精度

次に，サロゲートデータ法の検定統計量として予測精度を算出する．時系列データ $\dot{x}_n$ または $\tilde{x}_{in}$ の前半部を学習データ $L(n)$（$n = 1 \sim N/2$）として用い，後半部 $T(n)$（$n = N/2+1 \sim N$）を予測する．もし，オリジナルデータに時間構造や非線形構造が内在しているならば，局所線形近似法 [4], [5], [17], [18], [19] によってその構造を学習できるので，特徴的構造を破壊したサロゲートデータと比べて予測精度において有意差が発生する．なお，先行研究 [3] では局所線形近似法の中でも最も単純なローレンツ類推法 [4] を用いているが，本研究ではより詳細な時間構造を学習できる高次の予測モデルとして，非線形自己回帰予測モデル [5] を採用する．なお後述するが，非線形自己回帰予測モデルはローレンツ類推法を内包する予測モデルである．

まず前処理として，1 次元時系列データ $L(n)$ から多次元アトラクタ $\boldsymbol{v}_L$ を再構成する [15], [16].

$$\boldsymbol{v}_L(n) = [L(n), L(n-\tau), \cdots, L(n-(d-1)\tau)] \tag{11}$$

ここで，τ は遅れ時間，d は埋め込み次元であり，$\boldsymbol{v}$ を再構成アトラクタと呼ぶ．本研究ではトレンドを除去した階差データ $\dot{x}_n$ を解析するので $\tau = 1$ でよい．次に，システムのダイナミクスを関数 $\boldsymbol{F}$ とすると，システムの時間発展を

$$\boldsymbol{v}_L(n+1) = \boldsymbol{F}[\boldsymbol{v}_L(n)] \tag{12}$$

と表現できる．学習データを用いてこの関数 $\boldsymbol{F}$ を近似できれば，その近似関数 $\hat{\boldsymbol{F}}$ を評価データ $T(n)$ に適用することで予測値 $\hat{T}(n+1)$ を算出できる．つまり，

$$\boldsymbol{v}_T(n) = [T(n), T(n-\tau), \cdots, T(n-(d-1)\tau)] \tag{13}$$

$$\hat{\boldsymbol{v}}_T(n+1) = \hat{\boldsymbol{F}}[\boldsymbol{v}_T(n)] \tag{14}$$

によって $\hat{\boldsymbol{v}}_T(n+1)$ を得れば，この第 1 成分が予測値 $\hat{T}(n+1)$ である．

さて，式 (12) の関数 $\boldsymbol{F}$ を以下のようにモデル化する．

$$L(n+1) = \boldsymbol{a} \cdot \boldsymbol{v}_L(n) + a_0 \tag{15}$$

ここで，$\boldsymbol{a} = [a_1, a_2, \cdots, a_d]$，$a_0$ は定数である．この係数推定に用いる学習データの選択として，$\boldsymbol{v}_L(n)$ を中心とした半径 ϵ 内に含まれる近傍集合 $\{\boldsymbol{v}_L(n_k)\}$ と 1 ステップ後の点集合 $\{\boldsymbol{v}_L(n_k+1)\}$ を利用すれば，システムの時間発展ダイナミクス $\boldsymbol{F}$ を最小二乗法により推定することができる．ここで近傍集合の要素数を K とし，

$$\boldsymbol{V}(n+1) = [L(n_1+1), L(n_2+1), \cdots, L(n_K+1)]^{\mathrm{t}},$$

$$\boldsymbol{V}(n) = \begin{bmatrix} \boldsymbol{v}_L(n_1) & 1 \\ \boldsymbol{v}_L(n_2) & 1 \\ \vdots & \vdots \\ \boldsymbol{v}_L(n_K) & 1 \end{bmatrix}, \quad \boldsymbol{F} = [a_1, a_2, \cdots, a_d, a_0]^{\mathrm{t}}$$

と書くと，

$$\boldsymbol{V}(n+1) = \boldsymbol{V}(n)\boldsymbol{F}$$

$$\hat{\boldsymbol{F}} = \left[\boldsymbol{V}(n)^{\mathrm{t}}\boldsymbol{V}(n)\right]^{-1}\boldsymbol{V}(n)^{\mathrm{t}}\boldsymbol{V}(n+1) \tag{16}$$

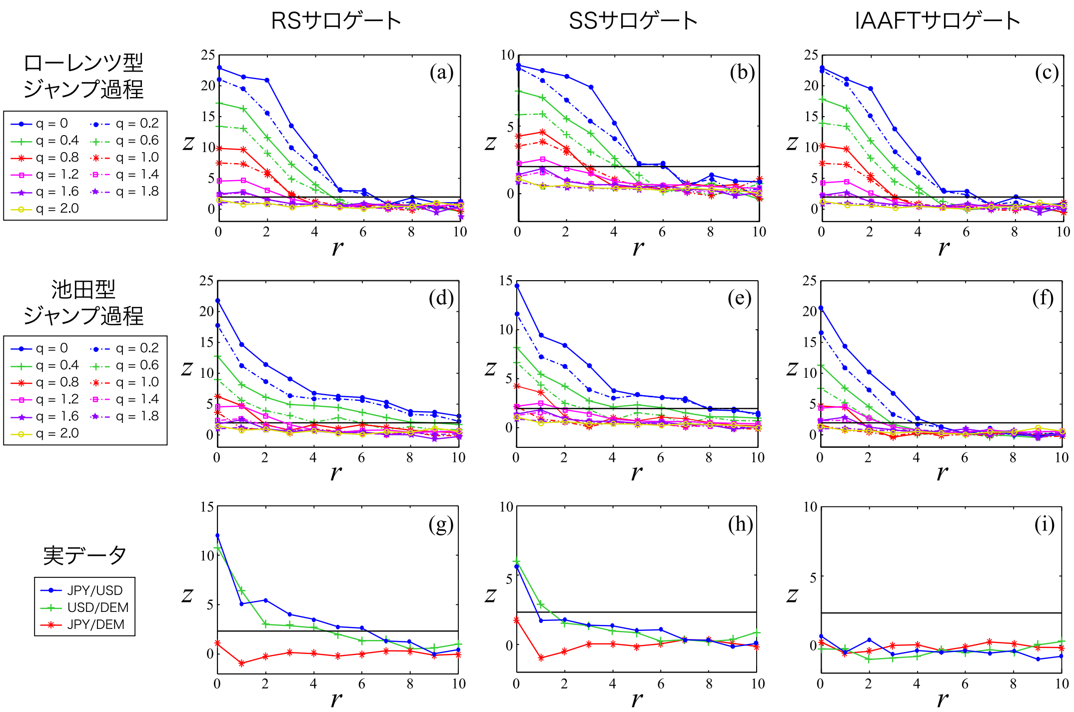

図 4　サロゲートデータ法により算出された z 値（式 (18)）の結果．各図の実線は $z > 2.33$ の棄却域の境界である

Fig. 4　Results of the surrogate data tests, where each solid line indicates the rejection region of $z > 2.33$.

によって推定値 $\hat{\boldsymbol{F}}$ を得る．その後，推定された $\hat{\boldsymbol{a}}$ と $\hat{a}_0$ を用いて予測対象点 $\boldsymbol{v}_T(n)$ の時間発展を

$$\hat{T}(n+1) = \hat{\boldsymbol{a}} \cdot \boldsymbol{v}_T(n) + \hat{a}_0 \tag{17}$$

によって予測し，予測値 $\hat{T}(n+1)$ を得る．しかし式 (16) において，$\boldsymbol{V}(n)$ を構成する近傍集合の要素数 K が少ない場合，または近傍どうしが類似している場合，行列 $\boldsymbol{W} = \boldsymbol{V}(n)^{\mathrm{t}}\boldsymbol{V}(n)$ が非正則行列になり，逆行列 $\boldsymbol{W}^{-1}$ の計算が不安定になる場合がある．その対策として，特異値分解によって得られた疑似逆行列を利用することができる [20]．

このように非線形自己回帰予測モデルは，$\boldsymbol{F}$ を局所的に線形近似するため，大域的には曲面近似を行う非線形近似法である．もし近似する領域 ϵ を拡大すれば，線形自己回帰モデルに相当し，また $\boldsymbol{a}$ の要素をすべて 0 にすれば，ローレンツ類推法 [4] と等価になる．つまり非線形自己回帰モデルは，これらの予測モデルを一般化したものである．なお，予測モデルの尤度を最大化すべく，最尤法の一種である交差確認法 [21], [22] を用いて，モデルパラメータである d と近傍集合の要素数 K を最適化した．

次に，予測値 $\{\hat{T}(n)\}$ と真値 $\{T(n)\}$ の相関係数によって予測精度 C を算出した．サロゲートデータとオリジナルデータの予測精度をそれぞれ $\tilde{C}_i$，C と書くと，以下の z 値を満たす場合に有意水準 1%の棄却域において，サロゲートデータとの有意差を確認できる．

$$z = \frac{C - \left\langle \tilde{C}_i \right\rangle}{\sigma_{\tilde{C}_i}} > 2.33 \tag{18}$$

ここで $\langle\cdot\rangle$ は平均値，$\sigma_{\cdot}$ は標準偏差を意味する．この z 値は，点 C と分布 $\{\tilde{C}_i\}$ のマハラノビス距離であり，z 値が大きいほどオリジナルデータとサロゲートデータの乖離も大きい．決定論的なカオス方程式より構成した数理モデルでは，いずれのサロゲートテストにおいても有意差が現れるはずである．しかし，サンプリングの時点でオリジナルデータ $\dot{x}_n$ の特徴的構造を破壊した場合は，有意差は検出されずに z 値 $\simeq 0$ となる．経済データでは特性が未知であるため，解析結果によってその特性が明らかとなる．

3.4　検定結果

検定結果を図 **4** に示す．なお $0 < r < 1$ においては，重複データの発生により式 (11)，(13) のアトラクタ $\boldsymbol{v}$ を適切に再構成できず，誤った検定結果を引き起こすという問題がすでに報告されている [3], [23]．よって本研究では，ティックサンプリング（$r = 0$）および $r \geq 1$ の等時間間隔サンプリングの結果を示す．また，汎用的な結果を示すために，30 個の z 値を算出しその平均値をプロットした．数理モデルであれば，初期値を変えて $N = 2048$ のオリジナルデータを多数作成し，それぞれについてサロゲートデータ法によって z 値を算出した．為替データであれば $N = 2048$ ごとにオリジナルデータを分割し，それぞれに

ついて同様に z 値を算出した．

決定論的ダイナミクスを有する数理モデルにおいては，サロゲートデータとの有意差を確認でき，統計的仮説検定の妥当性を確認できる．また，いずれのジャンプ過程およびサロゲートデータ法においても，ティックサンプリング（$r = 0$）が最も高い z 値を示している．特に，数理モデルにノイズが混入しない場合，背景活動の特徴をポアンカレ断面によって抽出できるので，ティックサンプリングの優位性は妥当である．つまり，サロゲートデータ法は最適なサンプリングを調べる手法として妥当といえる．

一方，短期の等時間間隔サンプリングであれば z 値は高い値を示し，システムの構造を抽出できているといえる．しかし r の増加にともない，z 値は低下する．つまり，サンプリング比率 r を拡大する恩恵として，データ数 N を一定のままデータ観測期間（$0 \leq t \leq Nr$）を延長できるが，それは逆効果である．むしろ r の拡大にともなって欠損値が増加し，つまり s 期間内に生起する $m(t_n)$ が式 (10) によって統合され，元のシステムの特徴が破壊されてしまう．たとえば図 4 (g) の USD/DEM では，$r = 5$ 程度でもはや有意差を確認できない．このときのサンプリング間隔は $s \simeq 2.5$ [分] であるので，たかだか 3 分程度の短期の振舞いでも，経済システムを特徴づけるには欠くことができない．つまり，これは高頻度データの必要性を主張するものであり，高頻度データを積極的に取り扱う経済物理学 [1], [2] のアプローチを支持する結果である．また，価格変動を単なるランダムウォークと結論づけてきた過去の多くの研究に対して，それは解析に用いた時間スケールが大きすぎたからだと主張できる．

さらに，数理モデルに付加するノイズ量を増やすほど背景活動の特徴は破壊され，サロゲートデータとの乖離が縮小するので z 値は低下する．それでもなお，等時間間隔サンプリングが勝ることはなく，ティックサンプリングの有意性は頑健である．なお，IAAFT サロゲート以外の実データにおいても，ティックサンプリングが最も高い z 値を示している．IAAFT サロゲートでは，すべての実データにおいて非線形構造の存在を確認できなかったが，JPY/USD と USD/DEM については，RS と SS サロゲートによって時間構造を有することを確認できる．一方，JPY/DEM はどのサロゲートデータに対しても有意差がなく，数理モデルに高いノイズを付加した状態と類似している．つまり，背景活動がポアンカレ断面によって離散変動に変換された時点で，元の背景活動を特徴づける構造が破壊されたと解釈できる．

4. 時系列予測のためのサンプリング実験

時系列解析によってシステムのメカニズムを同定し，その知見を時系列予測などの工学的応用に活かすことが期待されるが，果たしてシステム同定と時系列予測は同じサンプリング方法で良いのであろうか？　前章の結果のように，システムの特徴を破壊しないティックサンプリングの利点は重要であるが，観測されたデータ間の時間間隔が短いため，長期の予測を行うには何度も予測を繰り返す必要がある．その反復回数に応じて，予測誤差は拡大するであろう．一方，等時間間隔サンプリングは r が大きいほど元システムの特徴を破壊してしまうが，1 ステップで予測できる期間を長くすることができる．その結果，長期予測においては，等時間間隔サンプリングの方が有利である可能性がある．そこで実際に，数理モデルと実データに対して長期予測を行う．

長期予測の方法は以下のとおりである．まず 3.3 節で述べたように，予測開始点 $\boldsymbol{v}_T(n)$ から 1 ステップの後の将来値を式 (14) によって予測する．次に，得られた予測値 $\hat{\boldsymbol{v}}_T(n+1)$ を，再び予測開始点と見なして同様の 1 ステップ予測を行うと，2 ステップ後の予測値 $\hat{\boldsymbol{v}}_T(n+2)$ を得る．このように 1 ステップ予測を p 回繰り返すと，p ステップ後の予測値 $\hat{\boldsymbol{v}}_T(n+p)$ を得る．これを p ステップ反復予測と呼ぶ．その後，$\hat{\boldsymbol{v}}_T(n+p)$ の第 1 成分である $\hat{T}(n+p)$ と真値 $T(n+p)$ の相関係数によって，予測精度 C を算出する．

1 ティックあたりの平均経過時間は $\langle \Delta t_n \rangle$ であるので，ティックサンプリングにおける p ステップ反復予測では，$p \langle \Delta t_n \rangle$ 秒後の将来値を予測することになる．等時間間隔サンプリングでは，サンプリング間隔 $s = r \langle \Delta t_n \rangle$ を拡大することで，1 ステップで s 秒後の将来値を予測できる．これを 1 ステップ直接予測と呼ぶ．ここで $p = r$ とすれば，p ステップ反復予測と 1 ステップ直接予測による予測経過時間は等しくなる．

ローレンツ型ジャンプ過程，池田型ジャンプ過程，実データの結果をそれぞれ**図 5**，**図 6**，**図 7** に示す．いずれの場合でも，図 4 と同様に汎用的な結果を得るために，30 回のシミュレーションを行い，その平均値をプロットした．さらに標準偏差をエラーバーによって示した．まず p が小さい短期予測であれば，ティックサンプリングでも予測反復回数が少ないので，等時間間隔サンプリングよりも予測精度は高い．この場合はデータの観測において，元のシステムの特徴を保存することを優先すべきである．しかし p を拡大するにつれて，むしろシステムの特徴を破壊する等時間間隔サンプリングが優位となる．前章で確認したように，サンプリング間隔を拡大するほどシステムの特徴を破壊してしまうが，その恩恵として，1 ステップで長期の予測を実現できる．つまり，p が小さい場合は欠損や重複が少ないサンプリングが良く，p が大きい場合は予測反復回数が少ないサンプリングが良い．

この知見は，一般の個人投資家にとっては朗報であろう．近年，高頻度のティックデータが入手可能になったとはいえ高価である．むしろ長期予測においては，Web から無料で入手できる毎分や毎時データの方が有用であるため，高

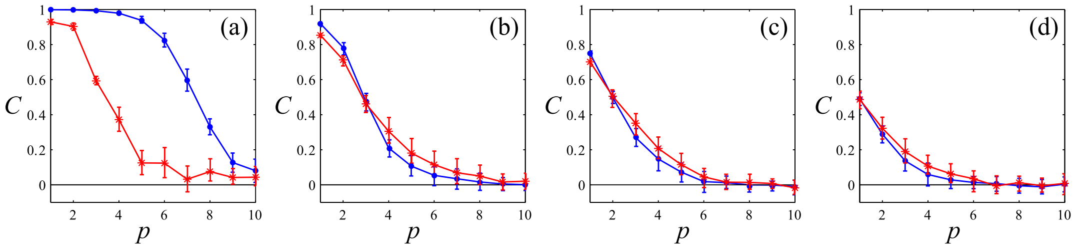

図 5　ローレンツ型ジャンプ過程における長期予測の結果：(a) $q=0$（ノイズなし），(b) $q=0.25$，(c) $q=0.5$，(d) $q=0.75$．実点線はティックサンプリングによる p ステップ反復予測，アスタリスクは等時間間隔サンプリングによる 1 ステップ直接予測の結果である．予測経過時間は $p\,\langle\Delta t_n\rangle$ である．なお，等時間間隔サンプリングでは p はサンプリング比率 r に相当する

Fig. 5　Results of long-term predictions for the Lorenz-type jump model: (a) $q=0$ (noiseless), (b) $q=0.25$, (c) $q=0.5$, and (d) $q=0.75$. Each asterisk shows the case of the nonuniform sampling, where they period of each long-term prediction is total $p\,\langle\Delta t_n\rangle$ each dot shows that of the uniform sampling, where p can be rewritten by r.

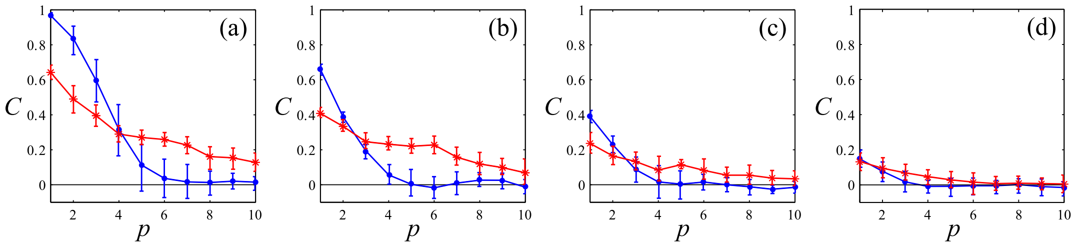

図 6　図 5 と同様．ただし，池田型ジャンプ過程：(a) $q=0$（ノイズなし），(b) $q=0.3$，(c) $q=0.6$，(d) $q=0.9$ の場合

Fig. 6　The same as Fig. 5, but for the Ikeda-type jump model: (a) $q=0$ (noiseless), (b) $q=0.3$, (c) $q=0.6$ and (d) $q=0.9$.

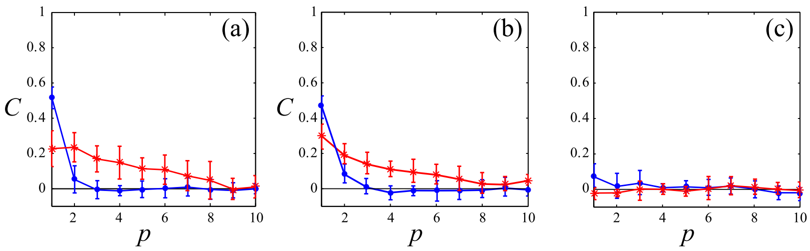

図 7　図 5 と同様．ただし，実データ：(a) JPY/USD 為替レート，(b) USD/DEM 為替レート，(c) JPY/DEM 為替レートの結果

Fig. 7　The same as Fig. 5, but for the real system: (a) the JPY/USD, (b) the JPY/USD and the (c) JPY/DEM exchange rates.

価なティックデータを購入する必要はない．また予測回数を大幅に削減できるので，計算コストの面でも有利である．

この逆転現象が起こる臨界点 p^* は，ノイズ量 q によって変化している．ノイズ量 q が大きいほど再構成アトラクタ $\boldsymbol{v}_L$ は乱れ，近傍点 $\{\boldsymbol{v}_L(n_k)\}$ の選択を誤り，その結果 $\boldsymbol{F}$ の近似精度が低下する．このような未熟な予測を反復すれば，ノイズ量 q が大きいほど予測誤差の拡大が早まるため，逆転現象が起こる臨界点 p^* は小さくなる．等時間間隔サンプリングでも予測精度は減少するが，予測の反復を行わないため，ティックサンプリングに比べて減少量は非常に小さい．

さらに，池田型ジャンプ過程の $q=0.6$ の結果（図 6 (c)）は，図 7 (a)，(b) の実データの結果と類似している．つまり実システムにおいては，ポアンカレ断面は理想的に機能せずにある種の不完全さをともなうため，q を大きくすることで実データの結果に近づけられたと考えられる．この観点から，図 7 (c) の JPY/DEM は q が極度に大きいケースだと考えられる．図 4 (g)～(i) のサロゲートデータ法で

も確認したように，JPY/DEM の為替データにはノイズ以外の特徴を見出せなかった．それゆえ，反復予測や直接予測にかかわらず予測精度はほぼ 0 となったと考えられる．また表 1 によれば，JPY/DEM の市場では取引数が少なく，変動ごとの時間間隔 $\langle \Delta t_n \rangle$ が大きい．そこで複雑系の知見によれば，この市場は流動性が低く，市場参加者どうしの相互作用が小さいため，特徴的な構造が自己組織化されるに至らなかったと示唆される．

5. まとめ

本研究では，市場価格変動などのジャンプ過程をサンプリングする時間スケールを議論すべく，実際の為替取引価格データに対して，システム同定実験と長期予測実験を行った．さらに，決定論性を有するジャンプ過程を数理モデル化することで，そのモデルの観点から得られた実験結果に対して考察を行った．

構築した数理モデルの特徴として，ジャンプ過程の背後には変化を生み出す連続的な背景活動が存在し，そのポアンカレ写像によって離散的なジャンプを表現した．背景活動の決定論性を陽に想定したモデルがローレンツ型ジャンプ過程であり，ポアンカレ断面によって抽出された背景活動の決定論性を，直接的に表現したモデルが池田型ジャンプ過程である．さらに，ポアンカレ断面の不完全さをノイズとしてモデルに導入した．

システム同定実験では，ジャンプ過程を生起時刻で観測するティックサンプリングが最適であった．数理モデルにノイズが混入しない場合，背景活動の特徴をポアンカレ断面によって抽出できるので妥当な結果である．つまり，サロゲートデータ法は最適なサンプリングを調べる手法として妥当だといえる．また，ノイズ量が多い場合は背景活動の特徴は低減するが，等時間間隔サンプリングが勝ることはなかった．つまり，ティックサンプリングの有意性はロバストである．一方，等時間間隔サンプリングでは欠損データの発生により，ポアンカレ断面によって抽出された特徴は破壊されてしまう．それゆえ，データの観測期間を長くするためにサンプリング間隔を拡大することは逆効果である．以上については，短期予測実験でも同様であった．

しかし長期予測実験では，たとえシステムの特徴を破壊しても，予測反復回数が少ない等時間間隔サンプリングの方が予測精度が良い．さらに，ノイズを増大させるほど毎回の予測誤差が拡大するので，反復予測を行うティックサンプリングは不利である．つまり短期予測から長期予測にかけて，両サンプリングの優劣が逆転する臨界点が存在し，この位置はノイズ量に依存することを示した．また，数理モデルにノイズを導入することで，実システムの結果に近づけることができた．つまり実システムにはノイズが存在し，それゆえ長期予測において等時間間隔サンプリングが優位になると考えられる．

なお本研究の一部は，文科省科研費若手研究（B）（No.22700227）のご支援により行われました．

参考文献

[1] Goodhart, C.A.E. and O'Hara, M.: High frequency data in financial markets: Issues and applications, *J. Empirical Finance*, Vol.4, pp.73–114 (1997).

[2] Mantegna, R.N. and Stanley, H.E.: *An Introduction of Econophysics: Correlations and Complexity in Finance*, Cambridge University Press (2000).

[3] Suzuki, T.: Appropriate Time Scales for Nonlinear Analyses of Deterministic Jump Systems, *Phys. Rev. E*, Vol.83, No.6, 066203 (2011).

[4] Lorenz, E.N.: Atmospheric predictability as revealed by naturally occurring analogues, *Journal of Atmospheric Sciences*, Vol.26, pp.636–646 (1969).

[5] Farmer, J.D. and Sidorowich, J.J.: Predicting Chaotic Time Series, *Phys. Rev. Lett.*, Vol.59, pp.845–848 (1987).

[6] Lorenz, E.N.: Deterministic nonperiodic flow, *Journal of Atmospheric Sciences*, Vol.20, pp.130–141 (1963).

[7] Ikeda, K.: Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system, *Optics Communications*, Vol.30, pp.257–261 (1979).

[8] OLSEN 社（http://www.olsendata.com/）から購入．

[9] Nelson, C. and Plosser, C.: Trends and Random Walks in Macroeconomic Time Series: some evidence and implications, *J. Monetary Economy*, Vol.10, pp.139–162 (1982).

[10] Schreibor, T. and Schmitz, A.: Surrogate time series, *Physica D*, Vol.142, pp.346–382 (2000).

[11] Nakamura, T. and Small, M.: Small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends, *Phys. Rev. E*, Vol.72, 056216 (2005).

[12] Chang, T. et al.: Tests for nonlinearity time series, *Chaos*, Vol.5, No.1, pp.118–126 (1995).

[13] Theiler, J. et al.: Testing for nonlinearity in time series: The method of surrogate data, *Physica D*, Vol.58, pp.77–94 (1992).

[14] Schreiber, T. and Schmitz, A.: Improved surrogate data for nonlinearity tests, *Phys. Rev. Lett.*, Vol.77, pp.635–638 (1996).

[15] Takens, F.: Detecting strange attractors in turbulence, *Lecture Notes in Mathematics*, Vol.898, pp.366–381, Springer-Verlag (1981).

[16] Sauer, T., Yorke, J.A. and Casdagli, M.: Embedology, *Journal of Statistical Physics*, Vol.65, No.3/4, pp.579–616 (1991).

[17] Casdagli, M.: Nonlinear prediction of chaotic time series, *Physica D*, Vol.35, No.3, pp.335–356 (1989).

[18] Sugihara, G. and May, R.M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature*, Vol.334, pp.734–741 (1990).

[19] Kugiumtzis, D.: Regularized Local Linear Prediction of Chaotic Time Series, *Physica D*, Vol.112, No.3–4, p.344 (1998).

[20] Haraki, D., Suzuki, T., Hashiguchi, H. and Ikeguchi, T.: Bootstrap Nonlinear Prediction, *Phys. Rev. E*, Vol.75, 056212 (2007).

[21] Allen, D.M.: Mean Square Error of Prediction as a Criterion for Selecting Variables, *Technometrics*, Vol.13, No.3, pp.469–475 (1971).

[22] Suzuki, T., Ueoka, Y. and Sato, H.: Estimating Structure of Multivariate Systems with Genetic Algorithms

for Nonlinear Prediction, *Phys. Rev. E*, Vol.80, 066208 (2009).
[23] Suzuki, T., Ikeguchi, T. and Suzuki, M.: Algorithms for generating surrogate data for sparsely quantized time series, *Physica D*, Vol.231, pp.108–115 (2007).

大塚 陽介　（学生会員）

昭和 62 年生．平成 23 年茨城大学工学部知能システム工学科卒業．同年 4 月茨城大学大学院理工学研究科知能システム専攻博士前期課程に進学．カオス時系列解析に関する研究に従事．

鈴木 智也　（正会員）

昭和 51 年生．平成 17 年東京理科大学大学院理学研究科物理学専攻博士課程修了．理学博士．同年東京電機大学工学部電子工学科助手，平成 18 年同志社大学工学部情報システムデザイン学科専任講師を経て，平成 21 年より茨城大学工学部知能システム工学科准教授，現在に至る．非線形時系列解析，複雑系，金融工学に関する研究に従事．電子情報通信学会，日本物理学会，人工知能学会，数理社会学会各会員．