3B-4

等時間間隔サンプリングによって見失う非線形システムの特徴

鈴木 智也 †
† 茨城大学 工学部 知能システム工学科

1 はじめに

経済市場の取引価格など不等時間間隔で変動しているシステムは，実世界において数多く存在する．しかし従来の経済分析においては，分次データのような等時間間隔で観測されたデータを解析対象とする場合が多い．これは物理時間を基準にして解析するためであり，また観測データ数を削減し少数データでも長期の性質を分析できるという利点もある．しかし，システム本来の特徴が破壊される危険性が生じる．サンプリングの時間間隔を大きくすれば観測漏れが生じ，小さくすれば重複して観測してしまう．本研究では，カオス的な数理モデルや実際の為替取引価格データに対して，非線形性を分析するサロゲートデータ法 [1, 2] を適用し，サンプリングの時間間隔の変化によって，どのように非線形性が欠落するのかを分析した．

2 決定論的なジャンプ過程

不等時間間隔で変動するシステムは点過程であり，取引価格の増分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．ここで生起時刻を t_n，変動幅を $z(t_n)$，取引価格の状態値を $x(t)$ とすれば，システムの振舞いはジャンプ過程で記述される (図 1)．

$$x(t) = x(0) + \sum_{n=1}^{T} z(t_n)\ , \quad t \geq T \qquad (1)$$

本研究では式 (1) を決定論的システムに拡張するために，マーク付き点過程 $z(t_n)$ の数理モデルとしてローレンツ方程式の第三変数 z の極大値を用いた．なお $x(t)$ が単調増加にならないように，極大値の系列に対して標準化を施した．得られた $x(t)$ を図 2(a) に示す．なお，ローレンツ方程式は $\dot{x} = -\sigma x + \sigma y, \dot{y} = -xz + rx - y, \dot{z} = xy - bz$ で定義される連続時間システムであるが ($\sigma = 10, b = 8/3, r = 28$ とした)，極大値をとることでポアンカレ断面上の非線形写像として表現できる [3]．

3 等時間間隔サンプリングによる数値実験

3.1 サンプリング方法

システムの特徴を欠落させずに $x(t)$ を観測するには，どのようにサンプリングすれば良いであろうか？

Uniform Temporal Sampling Destroys Property of Observed Nonlinear Systems

† Tomoya Suzuki (tsuzuki@mx.ibaraki.ac.jp)

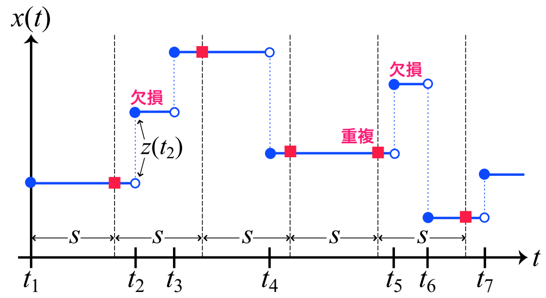

図 1: ジャンプ過程の状態値 $x(t)$ を等時間間隔 s で観測した様子．図中の■が観測されるデータであり，s の大きさに応じて欠損や重複データが発生する．

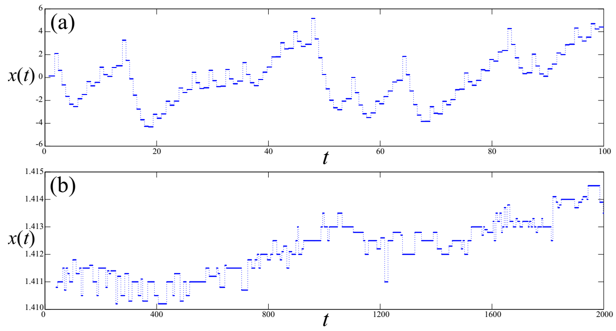

図 2: ジャンプ過程の状態値 $x(t)$．上図 (a) は式 (1) の数理モデル，下図 (b) は実際の為替取引価格 (売値) であり，時間 t の単位は秒である．

おそらく値が変化する時刻 t_n でサンプリングする (経済分析ではティックデータと呼ばれる) のが最適であろうが，データ数 N が膨大になり長期間のデータ分析が困難になる場合も多い．そこで前処理として，分次データのように等間隔サンプリングによってデータ量の削減が施される．本研究では，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．サンプリングされたデータは $x(ns)$ であり，そのデータ数を N とする．$r > 1$ である程欠落するデータ量が増え，逆に $r < 1$ である程同じ状態値が重複して観測される (図 1)．しかし，サンプリング後に欠落したデータを補うのは容易ではないが，重複データは単に削除することで改善できる可能性がある．そこで，重複データを削除しない標準的な方法を “方法 1”，削除する対処法を “方法 2” として区別する．なお $r = 0$ の場合，ティックデータのように理想的な (時刻 t_n で) サンプリングをするものとする．

さらに実データとして，1993/1/1〜1993/12/31 のアメリカドルとドイツマルクの為替取引価格 (売値)[4] も解析した (図 2(b))．本来は全取引が記録されているティックデータ ($N = 1,472,241$) であるが，数値実験において $x(ns)$ を得るために等時間間隔サンプリングを施した．その後，トレンド成分の除去として階差データに変換し ($x(ns) - x((n-1)s) \mapsto x(ns)$)，以後の解析に用いた．この前処理は数理モデル (式 (1)) においても同様に行った．なお市場参加者の減少に伴い取引時間間隔が拡大する場合があるので，$\Delta t_n \geq 300$[sec] の場合 (全体の約 0.01%) は $\Delta t_n = 300$ に補正した．

3.2 サロゲートデータ法

次に，等時間間隔サンプリングされた $x(ns)$ 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，small-shuffle surrogate (SSS) サロゲート検定 [2] を行った．まず時系列データ $x(ns)$ 内に含まれる短期間の成分同士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したサロゲートデータ $\tilde{x}_i(ns)$ を i 本作成する (本研究では $i = 60$)．次に $\tilde{x}_i(ns)$ および $x(ns)$ に対して非線形予測 [3] を行い，真値と予測値の相関係数によって予測精度を算出した．各々を $\tilde{c}_i$, c とし，$\tilde{c}_i$ が正規分布に従うと仮定すれば，z 値 $= \frac{(c - \langle\tilde{c}_i\rangle_i)}{\sigma_{\tilde{c}_i}} > 1.645$ を満たす場合に有意水準 5%の棄却域においてオリジナルデータ $x(ns)$ は特徴的性質を有すると判断できる．もしサンプリングを行った時点で $x(ns)$ 内にこの性質が欠落してしまう場合は，z 値 $\simeq 0$ となる．

図 3 によれば，まずデータ長 N を増やせばデータの欠損や重複に対する頑健性を強化できる．さらに方法 2 として重複データを削除することは効果的である．しかしデータ期間 t を長くするために，N を一定のままサンプリング比率 r を広げても効果は無い．特に経済データの場合は，等時間間隔サンプリングによる特徴量の欠落が顕著である．つまり短期の非線形構造が非常に重要であるため，データ期間 t が短い場合でもティックデータを解析に用いるのが良い．

3.3 リカレンスプロット

さらに定性的に重複データを削除する対処法 (方法 2) の有用性を確認すべく，リカレンスプロット [5] を描画した．まず，1 次元データ x からアトラクタ $\boldsymbol{v}(n) = \{x(ns), x((n-\tau)s), \cdots, x((n-(d-1)\tau)s)\}$ を再構成し [6]，図 4 のように $D(i,j) = |\boldsymbol{v}(i) - \boldsymbol{v}(j)| < \theta$ となる第 (i,j) 画素を着色した．ここで $\theta = 0.1 \max\{D(i,j)\}$ とし，最適な τ と d の設定法として k-fold CV 法 [7] を用いた．得られた知見として，重複データが生じた場合 (図 4(b)(e)) はそれらを削除することで (図 4(c)(f))，理想的なサンプリング状態 (図 4(a)(d)) に近づけることができる．

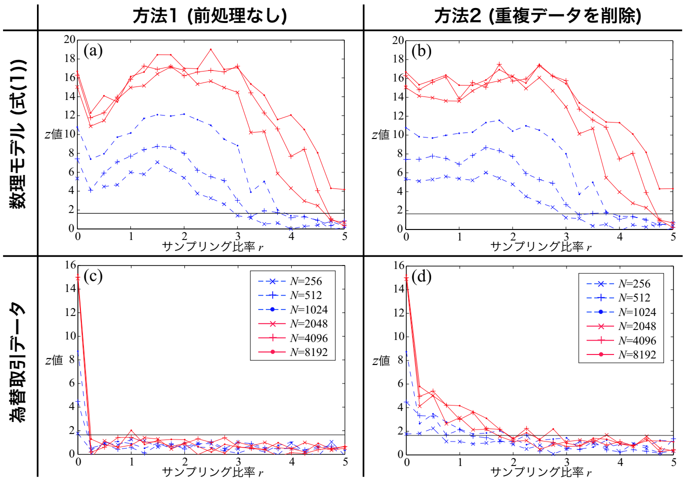

図 3: SSS サロゲート検定の結果．各図の実線は $z > 1.645$ の棄却域の境界である．

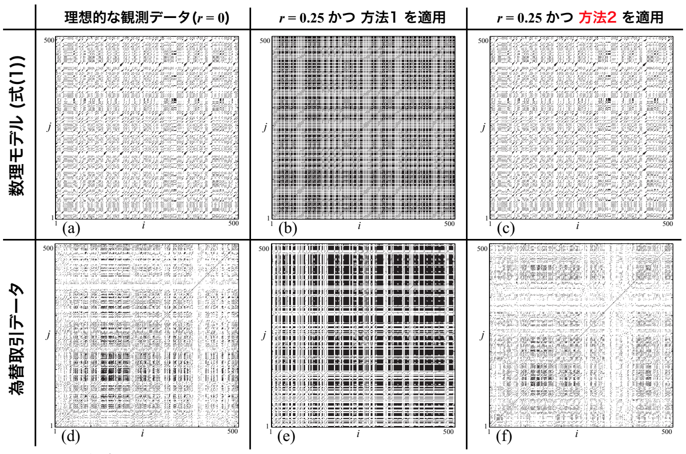

図 4: リカレンスプロット $D(i,j)$ の描画．

4 まとめ

本研究により得られた知見は以下のとおりである．

- データ期間を長くするためにサンプリング間隔を大きくしても効果は無い．
- 重複データの削除は非常に効果的である．
- データ長を増やせば，欠損に対する頑健性を増強できる．
- 等時間間隔サンプリングはデータの特徴を破壊するので，ティックデータを解析するのが最適である．

本研究の一部は，日本学術振興会科学補助金若手研究 (B)(No.20700217) の援助により行われました．

参考文献

[1] T.Schreiber, et al., PRL **77**(4), 635, 1996.
[2] T.Nakamura and M.Small, PRE **72**, 056216, 2005.
[3] E.L.Lorenz, J. of Atmos, Sci. **20**(6), 130, 1963.
[4] OlsenData 社 (http://www.olsendata.com/) より購入.
[5] J.P.Eckmann, et al., Europhysics Lett. **4**(9), 973, 1987.
[6] F.Takens, Lec. Notes in Math. **898**, 366, 1981.
[7] T.Suzuki, et al., PRE **80**(6), 066208, 2009.