F_021

ブートストラップ法による少数時系列データのリアプノフ指数の精度向上

Improving the Lyapunov Exponent of Short Time Series by Bootstrap Method

今野 良彦 † 森 康久仁 ‡ 松葉 育雄 ‡
Yoshihiko Imano Yasukuni Mori Ikuo Matsuba

1. はじめに

現在時系列解析の分野はコンピューターの発達と共に飛躍的な進歩を遂げている．特にその統計的側面の解析は計算量の多さを気にせずに行える．本稿で扱う非線形時系列，つまりカオス時系列においてもその発展は目覚しい．

非線形時系列解析では時系列がカオスか判断するのにしばしばリアプノフ指数が用いられる．リアプノフ指数とは時系列の広がりを定量化した値であり，その値が正ならばカオスであると判断できる．しかしリアプノフ指数はノイズに不安定な値で，データにより上手く求められないことが多い．例えば，互いに独立な白色ノイズのデータでは本来リアプノフ指数は負の値になるべきだが，正の値を算出してしまうことがある．本稿の最終的な目的はこのリアプノフ指数の信頼性を高めることである．

リアプノフ指数に限らず時系列解析を行ううえで，これまでには統計的な手法に基づいたサロゲート法がよく使われている．サロゲート法はカオス解析において有効に作用するが，データ数が十分にあることが前提であった．ここで問題となるのが実際はデータ数が十分にあるとは限らないことである．例えば，本研究で用いた算出方法では，ローレンツモデルなら 1000 程度のデータがあれば信頼できる値が得られる．しかし経済時系列データの日次日経平均などは一日一データ（土日は含まない）なので，年間で高々 250 データほどである．このような問題に対してジャックナイフ法，ブートストラップ法が提案され，様々に応用されてきた．特に時系列データに対してはブロックワイズブートストラップ法などがよく用いられている [1]．

しかしながら，現在までリアプノフ指数の算出におけるブートストラップ法の適用は十分に検討されているとは言えない．そこで本稿では，データの少なさを補う手法としてブートストラップ法を用いて，少数の時系列データから精度よくリアプノフ指数を求めることを目的としている．今回のシミュレーションでは時系列の時間相関を考慮に入れることでリアプノフ指数の計算精度をあげられる可能性があることを示唆できた．

2. ブートストラップ法によるリアプノフ指数の推定

ブートストラップ法は，Efron により提唱された代表的なリサンプリング法の一つである．標本から単純無作為抽出を元の標本の長さになるまで繰り返し新しい標本（ブートストラップ標本）を構成し，推定量を求め，多数回繰り返すことで推定量のバラツキを計算する手法である．

時系列データを扱う場合には，リサンプリングする際にデータを一個ずつ抽出するのではなく，時系列としての性質を失わないために隣り合うデータを複数個ずつ抽出する方法が一般的である．この同時に抽出するデータの個数 (l) のことをブートストラップ長と呼ぶ．時系列データを $\boldsymbol{X} = (X_1, ..., X_N)$ としたとき，k 番目の抽出 ($k = 1, 2, ..., s$) に対して，乱数 $r_k (1 \leq r \leq N - l + 1$ の一様乱数) を用いてブートストラップ長 l のデータを $\boldsymbol{X}^{*k} = (X_{r_k}, ..., X_{r_k+l-1})$ とすると，一回のリサンプリングで $\boldsymbol{X}^* = (\boldsymbol{X}^{*1}, \cdots, \boldsymbol{X}^{*s})$ となるように抽出する．これをブートストラップ標本と呼ぶ．s は一般的には $ls = N$ であるが，ここでは $ls = 1000$ となるように s を設定した (端数は切り捨てないので正しくは 1000 以上となる)．このブートストラップ標本の構成法の概念図を図 1 に示す．これを $\boldsymbol{X}_1^*$ から $\boldsymbol{X}_B^*$ まで B 個発生させる．リアプノフ指数を $\lambda = \lambda(\boldsymbol{X})$ で求められるとすると，

$$\bar{\lambda} = \frac{1}{B}\sum_{i=1}^{B}\lambda(\boldsymbol{X}_i^*) \tag{1}$$

を真のリアプノフ指数と比較する．リアプノフ指数の求め方は，システムを再構成しアトラクタの起動の広がりを平均する A.Wolf らによる手法を用いた [2]．ここで l が短ければ時間の流れが無視されてしまい，無相関な標本になってしまう．しかし，長すぎれば同じようなデータばかりがリサンプリングされてしまい，これではブートストラップ法の意味がない．したがって適切なブートストラップ長を与えるための指標が必要になる．その指標として時系列データの時間相関の長さに注目した．ここでいう時間相関の長さとはデータ間の相関が残存するような時間（ラグ）のことである．すなわち，時間相関の長さが 1 ならば 1 時刻先のデータとは相関があり，2 時刻以降のデータとは相関がない時系列である．

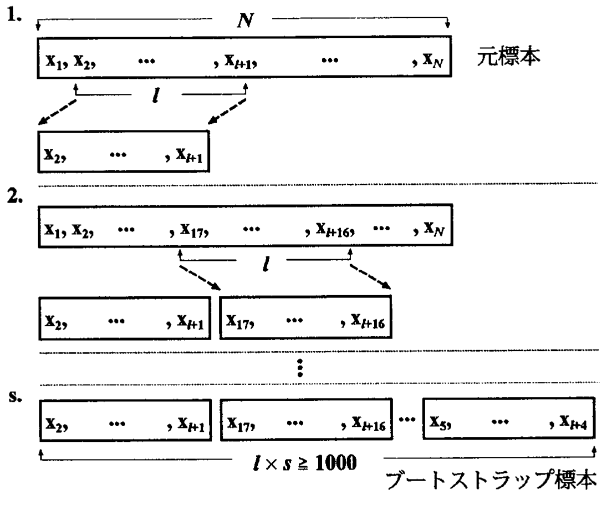


図 1: ブートストラップ標本の抽出の例：1. 元の標本から無作為に l 個抽出　2. 再度 l 個抽出　s. データ数が 1000 を越えるまで s 回繰り返す．

そこで，時系列の時間相関が保たれる長さを l にとればよいとの予測の下，シミュレーションを行い，精度のよいリアプノフ指数について算出について考察する．本研究の第一段階とし

† 千葉大学大学院自然科学研究科
‡ 千葉大学工学部情報画像工学科

て，真の値との比較を行うために既知のシステムより発生させたデータを用いる以下のようなシミュレーションを行った．

3. シミュレーション実験

シミュレーションには3種類の時系列データを用いた．最初のデータは，代表的なカオスシステムであるローレンツモデル，

$$\begin{aligned} \frac{dx(t)}{dt} &= -10x + 10y \\ \frac{dy(t)}{dt} &= 28x - y - xz \\ \frac{dz(t)}{dt} &= xy - \frac{8}{3}z \end{aligned} \tag{2}$$

であり，刻み時間を 0.01 として発生させた時系列のうち x のデータを使用した．残りの2種類のデータはロジスティック写像，

$$x_n = f(x_{n-1}) = ax_{n-1}(1 - x_{n-1}) \tag{3}$$

を用いた．時間相関が長いデータを得るために工夫し3回再帰させたシステム，$x_n = f(f(f(x_{n-1})))$ からパラメータを変え ($a = 4.0, 3.858$)，時間相関の長さの異なる時系列データを使用した．少数データに対する実装とし3種類の時系列それぞれにおいてデータ数は $N = 10, 20, 30, 40, 50, 100, 200, 300, 400, 500,$ に設定して計30パターンのシミュレーションを行った．ブートストラップ長 $l = 1 \sim N$ それぞれに対しブートストラップ標本からリアプノフ指数を算出し，100回ずつ繰り返すこと ($B = 100$) で平均と分散を求めた．また十分なデータ数（$N = 10000$）でシステムから真のリアプノフ指数を求め，シミュレーションの結果と比較検討を行った．表1にそれぞれの真のリアプノフ指数と時間相関の長さを示す．時間相関の計算には自己相関関数を用いて，自己相関がはじめて0になる一つ手前のラグをとった．ただし $\pm 1.96/\sqrt{N}$ 以内を0と判断した．また，今回は時間相関の計算は $N = 10000$ の場合で求めた．

表1．真のリアプノフ指数 λ と時間相関の長さ

		真の λ	時間相関の長さ
ロジスティック写像	$a = 4.0$	2.079	0
	$a = 3.858$	0.957	22
ローレンツモデル		1.02	152

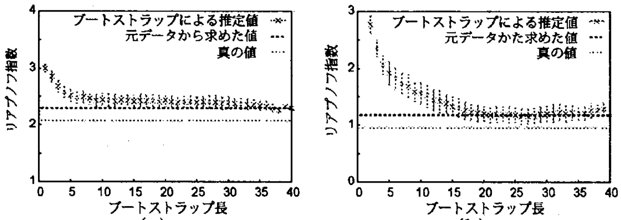
4. シミュレーション結果

時間相関の違いによる影響を検討するためにロジスティック写像による2つの時系列データの結果を比較する．図2に $N = 40$ の場合を示す．またデータ数による違いをみるために図3に $N = 300, 400$ のローレンツモデルによる結果を示す．各図において横軸に l，縦軸に l ごとのブートストラップ法によって求めたリアプノフ指数の平均値を ± 標準偏差一つ分を加えてエラーバー付きで表す (今回の結果ではブートストラップ法でのリアプノフ指数の分布はガウス分布ととらえても問題ないと判断し，標準偏差をエラーバーに使用した)．システムから求めた真のリアプノフ指数と，時系列からのみで求めたリアプノフ指数も図示した．

5. 考察

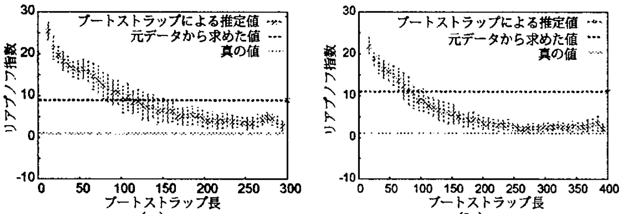
図2より時間相関の長さが長いデータのほうが l をより大きく取らなければ安定したリアプノフ指数が求められないことがわかる．これにより適切なブートストラップ長の決定に時間相関が影響することがわかる．また，ローレンツモデルの結果よりデータ数の違いによっても適切な l の値が変わってくることがいえる．

また図2(b),3(a),3(b)をみると時間相関の長さあたりからリアプノフ指数の平均値が横ばいのグラフになっていることがわかる．つまりブートストラップ長を時間相関の長さ以上にとればよいことがいえる．しかし，ここでブートストラップ長を大きくとりすぎたことによる悪影響はないのかという疑問が残る．



(a) (b)

図 2: ロジスティック写像実験結果 N=40 (a)a=4.0, (b)a=3.858



(a) (b)

図 3: ローレンツモデルの実験結果 (a)N=300, (b)N=400

6. まとめ

本稿ではデータ数が少ない場合のリアプノフ指数算出についてブートストラップ法を用いたシミュレーションを行った．l を時間相関の長さ以上にとることで安定した値が得られ，真の値に近付くことを示した．しかしながら適切な l は時間相関だけでなく標本のデータ数にも影響することが示唆された．また図で示してはいないが，データ数が少なすぎるとデータの偏りが大きすぎ，ブートストラップ法を用いてもうまくリアプノフ指数が求められないことがわかった．データ数が時間相関の長さより短ければこの傾向が見られた．そこで今後の課題として，データ数がいくつあたりからブートストラップ法が適用でき，どのくらい効果があるのかを検討していく必要がある．また，実データ応用にむけて少ないデータから時間相関をどう求めるかを考え，実際に実データでのシミュレーションを行う必要がある．

参考文献

[1] Hans R.Kunsch, "The Jackknife and the Bootstrap for General Stationary Observations", *The Annals of Statistics*, **17**, No.3, pp.1217-1241, 1989.

[2] A.Wolf, J.Swift, H.Swinney, and J.Vastano, "Determining lyapunov exponents from a time-series", *Physica*, D, **16**, pp.285-317, 1985.