

Spring task

B4 Fukuda Shingo

2020.3.3

- 1 Search the word "Lorenz equation", "Rossler equation", "white gaussian noise", "Logistic mapping", "permutation entropy", "mutual information" and "complex network" by book or online

- Lorenz equation

It is a differential equation created to simplify Navier-Stokes equation. It is one of nonlinear differential equation having chaotic solution. It is a system of three differential equations now known as Lorenz equation:[1]

$$\frac{dx}{dt} = -px + py \quad (1)$$

$$\frac{dy}{dt} = -xz + rx - y \quad (2)$$

$$\frac{dz}{dt} = xy - bz \quad (3)$$

The variables are (x, y, z, t) . The constants are (p, r, b) . Because of initial value sensitivity, the result of equation greatly depend on initial value. The initial value requires infinite precision, therefore it is effectively impossible to predict the result.

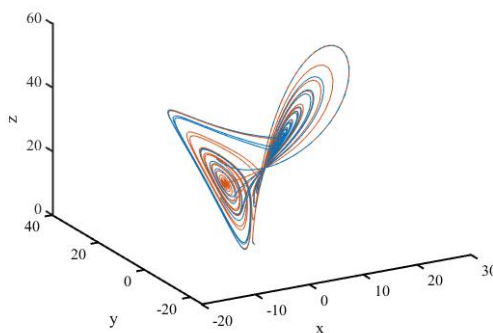


図1 Error amplification of Lorenz equation

figure 1 is a change of solution when one of the initial value is changed by 0.1.

- Rossler equation

It is one of nonlinear differential equation having chaotic solution. It is a system of three differential

equations now known as Ressler equation:[2]

$$\frac{dx}{dt} = -y - z \quad (4)$$

$$\frac{dy}{dt} = x + ay \quad (5)$$

$$\frac{dz}{dt} = b + xz - cz \quad (6)$$

The variables are (x, y, z, t) . The constants are (a, b, c) . Unlike Lorenz equation, Ressler equation doesn't base on real physical model. Like Lorenz equation, Ressler equation have initial value sensitivity.

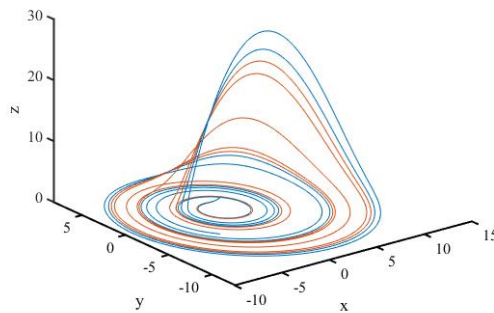


図2 Error amplification of Ressler equation

figure 2 is a change of solution when one of the initial value is changed by 0.1.

- White Gaussian Noise

It is a noise that model a effect of random. White means that the frequency band is uniform. It comes from white light. Gaussian means that its amplitude follows normal distribution, Gaussian distribution.

- Logistic mapping

It is the recurrence formula obtained by discretization of Logistic equation derived to discuss the dynamics of population change in the field of mathematical ecology[1]. It is a system of one differential equations now known as Logistic mapping:

$$x_{n+1} = ax_n(1 - x_n) \quad (7)$$

定数 a によって、値の変化が大きく変わる。 a が十分に小さいときは 0 に収束し、値が大きくなるにつれて、収束値が大きくなり、周期的な振動を示し、最終的にカオス的な挙動を示す。 a ごとの x の値の変化を下記に示す。

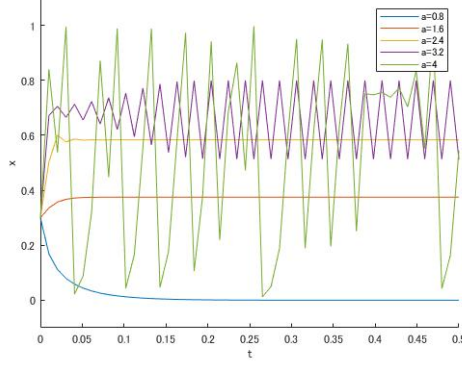


図3 a の変化によるLogistic写像の値の変化

- 順列エントロピー

時系列の乱雑さを定量評価する指標であり、時系列の値ではなく、大小を比較する手法である。長さが L の時系列に対して遅れ時間 τ の間隔があるデータ D 個を D 次元の遅れ時間座標 $\mathbf{X}_j = \{x_j, x_{j+\tau}, \dots, x_{j+(D-1)\tau}\}$ とし、この座標を考える。ただし、 $j = 1, 2, \dots, L - (D-1)\tau$ とする。この D 次元の座標のそれぞれの値の大小をパターン化すると $D!$ 個のランクオーダーパターンに分類される。あるランクオーダーパターン π_i の相対度数 $p(\pi_i)$ は、

$$p(\pi_i) = \frac{\sum_{j=1}^N \chi_i(X_j)}{L - (D-1)\tau} \quad (8)$$

$\chi_i(X_j)$ は指示関数であり、

$$\chi_i(X_j) = \begin{cases} 1 & (\pi_j = \pi_i) \\ 0 & (\pi_j \neq \pi_i) \end{cases} \quad (9)$$

となる。そしてこの相対度数の情報エントロピー(Shannonエントロピー)で計算したものが順列エントロピーである。さらに定量的に様々な時系列を比較するために正規化を行ったものを順列エントロピーと定義する。その式を下記に示す。

$$H_p = \frac{-\sum_{i=1}^{D!} p(\pi_i) \log_2 p(\pi_i)}{\log_2 D!} \quad (10)$$

- 相互情報量

2つの情報源の依存度を表した値である。一方の情報をもっているとき、他方の情報を得たらどの程度の不確かさ(情報エントロピー)が減少するのかわを示している。2つの情報源 X, Y からそれぞれから情報 x, y を得られる確率を $P_X(x), P_Y(y)$ とし、同時に得られる確率を $P_{X,Y}(x, y)$ とすると、相互情報量の式は、

$$H_p = \sum_{x \in X} \sum_{y \in Y} P_{X,Y}(x, y) \log_2 \frac{P_{X,Y}(x, y)}{P_X(x)P_Y(y)} \quad (11)$$

となる[4].

- 複雑ネットワーク[5]

ノード(点)間をリンク(線)で結んで構成されているものをネットワークという。その中で「スケールフリー性」,「スモールワールド性」などの特徴をもつものを複雑ネットワークという。あるノードに結ばれているリンクの数を次数といい,その分布を次数分布という。複雑ネットワークにおいては,次数分布がべき分布となっている。べき分布は確率変数 X の要素 x に対して,確率密度関数が,

$$f(x) = ax^k \quad (12)$$

となる。式(12)の確率変数が定数倍(c 倍)になったとき,確率密度関数は,

$$f(x) = a(cx)^k \quad (13)$$

となり,式(12)の c^k 倍すれば表現できるため,べき分布はスケール不変性を持っている。そのため複雑ネットワークは「スケールフリー性」という特徴を持つ。2つのノードを結ぶ最短のパスの長さの平均を平均ノード間距離,あるノードに対して,隣接ノードの内2つのノードを選んだときのその2つのノードがリンクされてるときの確率の平均を平均クラスター係数とする。複雑ネットワークでは,平均ノード間距離が小さく,平均クラスター係数が大きい。このことを「スモールワールド性」という。

2 相互情報量とLorenz 方程式に関するMATLAB ファイル(mutual.m) を実行し,図を出力する

出力された結果を下記に示す。

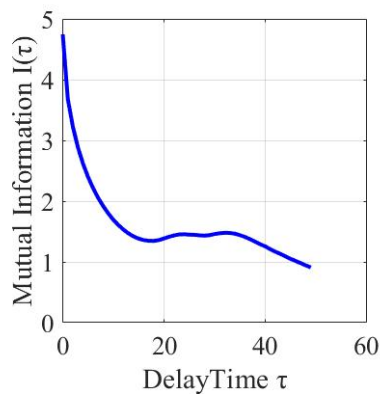


図4 課題2の出力結果

また,時間遅れの極小値は18となった。このことからアトラクタの構築するために最適な時間遅れは18だと考えられる。 x 軸をMATファイルの時系列データ $f(t)$ とし, y 軸をMATファイルの時系列データ $f(t + \tau)$ とし, z 軸をMATファイルの時系列データ $f(t + 2\tau)$ としたときのプロットを下記に示す。

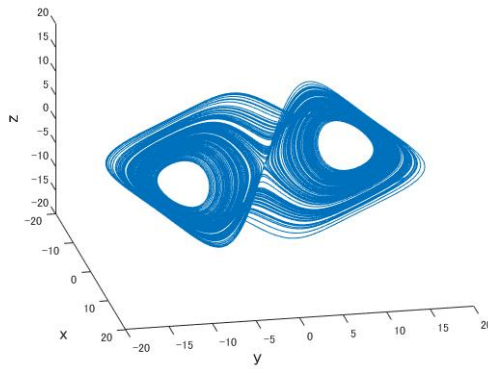


図5 最適な時間遅れのアトラクタ

また，これに対して不適切な時間遅れである $\tau = 2, 1000$ のパターンも下記に示す．

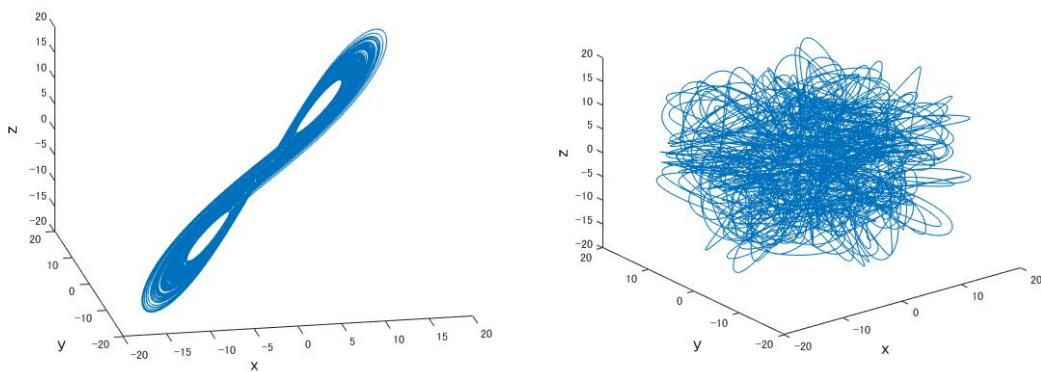


図6 不適切な時間遅れのアトラクタ(左が $\tau = 2$ ，右が $\tau = 1000$)

これは，時間遅れが小さいと時系列と遅れ時系列の相関が高くなり，傾き 45° の直線近傍のプロットとなってしまふ．時間遅れが大きいと無相関となり，不安定な軌道を示してしまう．

3 ホワイトガウスノイズと順列エントロピーに関するMATLABファイル(B4kadai_2020.m)を実行し，図を出力し考察を加える．

まず出力された結果を下記に示す．このときのホワイトガウスノイズの生成数は10000個で次元数は5，遅れ時間は1である．

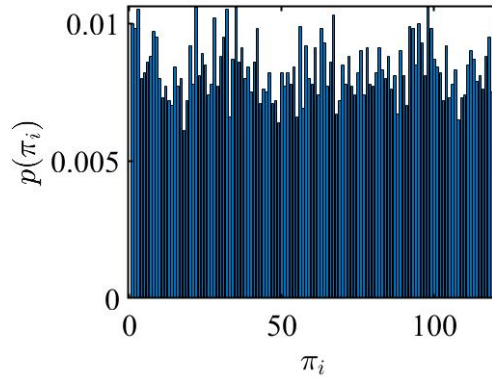


図7 課題3の出力結果(生成数は10000個で次元数は5, 遅れ時間は1)

ホワイトガウスノイズは周波数帯域が均一のため理論的には、ランクオーダーパターンも均一になると考えられる。つまりランクオーダーパターンのエントロピーである順列エントロピーは理論的には、1となると考えられる。実際に順列エントロピーの計算すると0.9989となった。予想の結果との誤差は0.11[%]となるので今回の計算結果は妥当であると考えられる。

結果からホワイトガウスノイズのランクオーダーパターンにはバラつきがあることがわかる。今回のシミュレーションでは、次元数5($5! = 120$ 個のパターン)に対して、ノイズの生成数が10000となっているので平坦に見えるにはデータ数が少ない。実際に $p(\pi_i)$ の平均値は約0.083($= 1/120$)となっており、各ランクオーダーパターンの相対度数のヒストグラムを制作すると約0.083($= 1/120$)を中心にガウス分布に近いものとなっている。ノイズの生成を増やせば、各ランクオーダーパターンの相対度数は均一に近づいていくと考えられる。下記にノイズの生成数が100000とノイズの生成数が1000000に増やしたときの出力結果を示す。

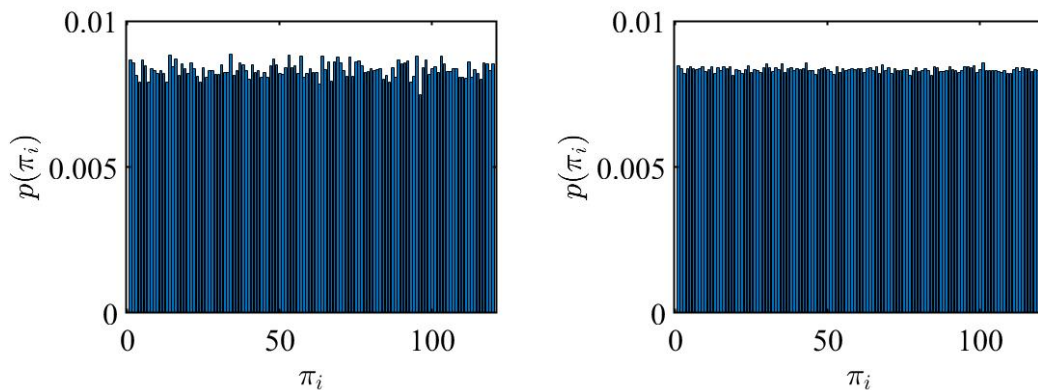


図8 課題3の出力結果(左が生成数は100000, 右が生成数は1000000)

この結果からデータ数を増やすことでバラつきが小さくなることがわかった。また、ノイズの生成数が1000000のときのヒストグラムも下記に示す。

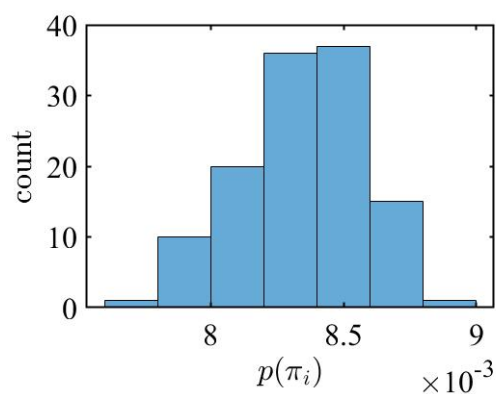


図9 課題3のヒストグラム(生成数は100000)

参考文献

- [1] カオス時系列解析の基礎と応用, 合原一幸, 2005年6月23日
- [2] カオス・フラクタル 講義ノート #8, <https://ocw.hokudai.ac.jp/wp-content/uploads/2016/01/ChaosFractal-2011-Note-08.pdf>
- [3] ホワイトガウスノイズサンプルの生成-MATLAB wgn, <https://jp.mathworks.com/help/comm/ref/wgn.html>
- [4] 相互情報量の意味とエントロピーとの関係 — 高校数学の美しい物語, <https://mathtrain.jp/mutualinfo>
- [5] 複雑ネットワーク: 統計物理学の視点, <http://mercury.yukawa.kyoto-u.ac.jp/~bussei.kenkyu/pdf/03/1/9999-031210.pdf>