非線形で確率論的な対象を時系列データから同定する: 数理モデル化基礎論

Identifying if the underlying system is nonlinear and stochastic based on a time series: Fundamental theory on mathematical modelling

平田 祥人
Yoshito Hirata
筑波大学システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba
Email: hirata@cs.tsukuba.ac.jp

ここでは、数理モデルを立てる上での前提となる決定論性・確率論性、線形性・非線形性の２つの検定に関して、できるだけ平易な言葉を使って解説を試みようとする。可能な限り、数式・図は、使わない。この解説記事で日本語で大体の概要を把握して頂き、詳しくは、それぞれの原論文を当たって頂こうという趣旨である。

数理モデル化は、今興味のある対象を理解し、その将来を予測したり、制御したり、最適化したりするのに役に立つ。数理モデル化する上で、重要なのは、正しい前提の数理モデルを選ぶことである。そうしないと、間違った判断をしたり、対策が非効率になったりすることがある。

動的な対象は、次の３つの要素で大きく、分けることができる：(i) まず、対象が、定常か、それとも、非定常か。定常な対象の場合、その性質は、時間とともに変化しない。一方で、非定常な対象の場合、時間とともに対象の性質が変化する。(ii) 対象が、決定論的であるか、それとも、確率論的であるか。決定論的である時には、現在や過去の状態を完全に把握できれば、その後の状態が完全に一意に決まる。一方、確率論的な場合には、現在や過去の状態が決まっても、将来の状態が一意には決まらない。(iii) 対象が、線形であるか、それとも、非線形であるか。線形の場合、将来の状態は、現在や過去の状態の重ね合わせで書くことができる。一方で、非線形の場合には、将来の状態が、現在や過去の状態の重ね合わせで書くことができず、いくつかの効果の相乗効果を考えなければ、ならなくなる(効果が純粋な足し合わせよりも弱くなることもある)。

対象が定常か、非定常かに関しては、Kennel (1997)などの検定の先行研究がある。一方で、対象が、決定論的か、または、確率論的か、線形か、または、非線形かに関しては、２つの性質の検定が、完全には分離されてこなかったという歴史がある。

話をわかりやすくするために、対象の数理モデルを次のような式で書くことにする。

$$x_{t+1} = f(x_t, p_t).$$

この場合、現在の状態x_tと現在の入力p_tが決まることで、次の値x_{t+1}が決まる。p_tが一定の場合は、定常で決定論的になる。p_tの時間変化に時間構造がある場合には、非定常ということになる。一方、p_tの時間変化に時間構造がない場合は、定常で確率論的ということになる。fがx_tの線形関数で書ける時線形、そうでないとき非線形である。このように話を整理すると、上記の式のような skew product を考えれば、当面の話を進めることができそうである。よって、ちょっと言葉に拡大解釈があるが、p_tが一定の時に決定論的、そうでない時に確率論的、fの関数が線形の時に線形、そうでない時に非線形と呼ぶことにすると、図 1 のような 2 つの軸を考えて、対象を整理すれば良い。

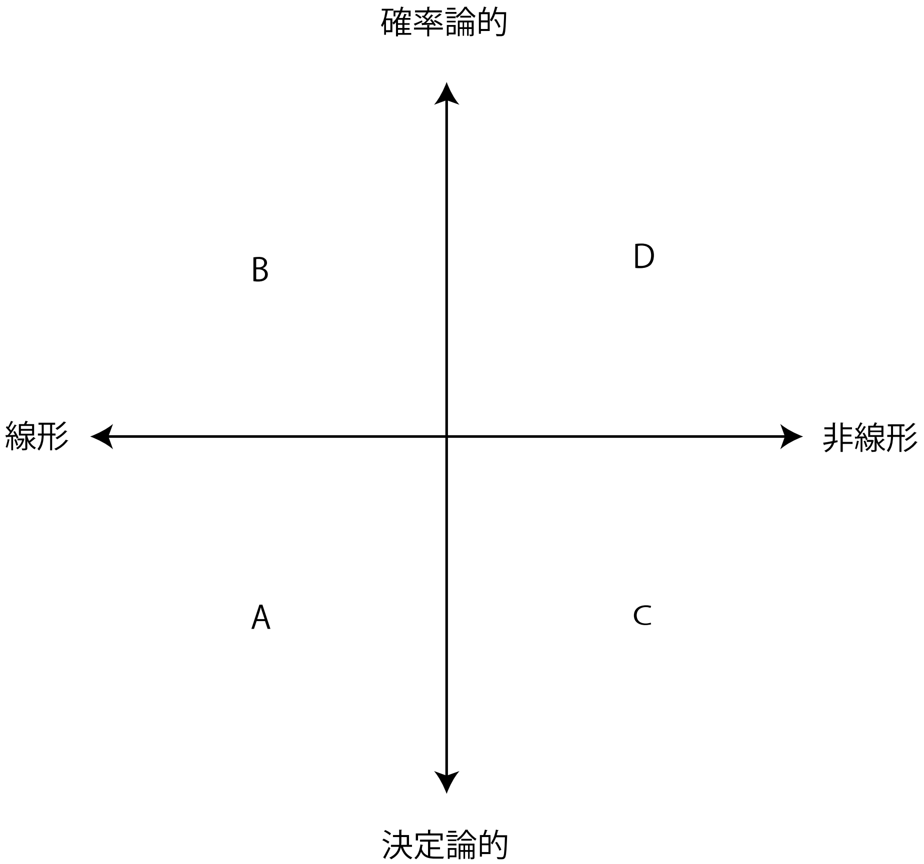

図 1 動的対象の分類

例えば、線形か、非線形かを検定する方法として、フーリエ変換に基礎をおくサロゲートデータ解析がある(Theiler et al., 1992; Schreiber & Schmitz, 1996; Nakamura et al., 2006)。サロゲートデータ解析の主な方法では、与えられた時系列データのパワースペクトラムは、大体保存するようにして、フーリエ変換の位相情報をランダム化することで、線形構造は保存されているが、非線形構造は壊されているランダムなデータを大量に生成する。よって、帰無仮説は、与えられた時系列データは、線形ノイズ(の単調な非線形変換)であるというもの

である。そして、検定統計量を使って、サロゲートデータと、元の時系列データの間に有意な違いがあるかどうかを検定する。検定統計量としては、予測誤差や軌道の平行度など、決定論性の指標が使用されてきた。そのため、決定論的か確率論的かという軸と、線形か非線形かという軸がごちゃ混ぜになっていて、検定として綺麗に整理されて来なかった。

線形・非線形の検定と、決定論性・確率論性の検定は、Hirata & Shiro (2019)によって、初めて分離された。まず、Nakamura et al. (2006)のサロゲートデータ解析を行う際に、検定統計量として、(時刻 t での観測量の 2 乗)*(時刻(t+1)での観測量の 2 乗)という値の時間平均を用いることにした(Hirata & Shiro, 2019)。この量自体は、決定論性や確率論性とは全く関係しない量であるので、線形・非線形の検定と、決定論性・確率論性の性質が、綺麗に分離できた。

一方で、決定論性・確率論性の検定では、順列(permutations; ordinal patterns)(Bandt & Pompe, 2002)の種類の数が、順列の長さを長くして行ったときに、どのように変化するかを利用することにした(Hirata & Shiro, 2019)。順列とは、時系列データの連続する部分の 1 次元の値の大小関係である。例えば、-0.5, 2, 1.5 というふうに時系列データに数字が並んでいれば、-0.5 が 1 番小さく、1.5 が 2 番目、2 が 3 番目なので、(0,2,1)のような記号を割り当てる。これが、順列である。決定論的な対象で順列の種類の数が指数関数的に増えるというのは、次の Amigo and Kennel (2007)の定理による：決定論的で、かつ、拡大的対象では、出現する順列は、指数関数的に増大し、その増加速度は、トポロジカルエントロピーと一致する。ここで、拡大的とは、初期値にちょっとの差がある場合には、必ず、その差が、将来、ある一定以上の差になるという性質である。

実際に使うのは、この定理の対偶である(Hirata & Shiro, 2019)：つまり、出現する順列の種類の数が、指数関数的に増加しないならば、確率論的であるか、拡大的ではない。対象が拡大的であるかどうかは、後に紹介するリカレンスプロットと呼ばれる時系列解析手法を用いれば、一目瞭然である。そのため、出現する順列の種類の数が、指数関数的には増大しないときで、かつ、拡大的であるような場合には、確率論的である可能性のみが残る。

この解析は、実際にデータ解析してみると、Hirata & Shiro(2019)にあるように、おもちゃモデルでうまく動く。しかし、同時に問題点もある。特に、順列の種類の数が指数関数的に増えるかどうかを検定する部分で、長さ 8 までの順列を考慮に入れると、長さ 1000000 ぐらいの時系列データがなければ、確率論的な場合をうまく同定できない。この長さは、実用的な時系列解析を用意するという立場に立った時、致命的である。

そこで、Hirata (2021)では、順列の代わりに、リカレンスプロットを用いることで、この問

題を克服した。リカレンスプロット(Eckmann et al., 1987; Marwan et al., 2007)は、時系列データを視覚化する道具として、提案された。縦軸・横軸は、同じ時間軸である。対応する２つの時刻の状態間の距離を計算し、もし、距離がしきい値の値よりも大きければ、対応する場所に点を打ち、そうでなければ、点を打たないとして求めることができる。

リカレンスプロットの研究では、打たれている点のパターンを定量化して対象の決定論性に関して議論するという研究が進んできた(Marwan et al., 2007)。特に重要なのは、斜めの線分である。決定論的カオスの対象では、偶然 2 つの点が近傍になると、決定論的性質のため、しばらくは近傍点に留まり続ける。しかし、初期値鋭敏性があるため、2 つの点の距離は次第に拡大していき、やがて、しきい値よりも大きな値になる。この 2 つの点の軌道が寄り添う時間が、斜めに線分となってリカレンスプロット上に現れる。一方、ホワイトノイズの場合には、点が不規則に一様に広がる。そのため、リカレンスプロットの観察に熟練してくると、与えられた時系列データが、決定論的な対象から生成されたものであるか、確率論的な対象から生成されたものであるか、大体わかる。また、これらの定性的な違いがあるため、斜めの線の平均長が、決定論性の指標として広く使われている(Marwan et al., 2007)。

Hirata (2021)では、この打たれている点のパターンの定量化を一歩進めて、決定論性・確率論性が判別できるようにした。Hirata (2021)では、リカレンスプロット内の任意の正方形の領域の左上半分の三角形の領域の点の打たれ方に着目した。この三角形の領域のことを、再帰三角形と呼ぶことにする(Hirata, 2021)。もし、対象か決定論的で、拡大的で、位相推移的な場合、再帰三角形の種類の数は、再帰三角形の大きさを大きくしていくときに、順列同様に、指数関数的に増大する(Hirata, 2021)。ここで、位相推移的とは、対象がとりうる値で構成する任意の開集合を 2 つ用意した時、一方の時間発展を観察すると、必ず、もう一方と重なりを持つ時間が訪れるという性質である。

再帰三角形を使う場合でも、順列と同様に、定理の対偶を用いて検定を構成する。しかし、順列の場合と比較すると、必要となる時系列データが短くて済むという利点がある：順列の場合、時系列データの長さを N とすると、順列は、O(N)個しか抜き出せない；一方、再帰三角形を使う場合には、リカレンスプロットは 2 次元的な広がりを持つので、$O(N^2)$個選ぶことができる。そのため、大体 5000 点以上あれば、決定論性・確率論性の検定を行うことができる。

また、Hirata (2021)には記述されていないが、リカレンスプロット上で決定論性・確率論性が検定できるようになったことで、別のメリットもある。元々、私は、リカレンスプロットを用いた解析手法を、脳神経のスパイク発火のようなイベント系列を扱うために発展させてきた。よって、例えば、イベント系列間の距離(Suzuki et al., 2010; Hirata & Sukegawa,

2019)を使ってリカレンスプロットを求めることによって、リカレンスプロットを用いた決定論性・確率論性の検定は、何の工夫を施すことなく、イベント系列の解析にそのまま用いることもできる。この部分は、将来の研究テーマとして温めている。

以上により、数理モデル化の前提を検定で同定するという研究は、大体見通しがついた。今後は、非線形で確率論的な対象をどう数理モデル化していくかという、本丸の研究が残っている。

そこで、もう一度、図 1 に戻る。今までの時系列データからの数理モデル化の研究では、決定論的で非線形な領域の決定論的カオス、その周辺から少し非線形性が弱いところ、少し確率論性が導入されたところなどが数理モデル化されてきた(Mandic et al., 2008)。一方、統計学の文脈では、王道として、AR モデルや ARMA モデルに代表される線形で確率論的な数理モデル化が研究されてきた。また、そこに少し非線形性が導入された NARMA モデル(Mandic et al., 2008)や GARCH モデル(Lamoureux & Lastrapes, 1990)等も研究されてきている。しかし、この平面上の多くの部分の数理モデル化研究は、まだ、手をつけられていない領域であるように思う。

その観点に立つ時、上記の順列と再帰三角形を用いた研究は、非線形で決定論的な対象、非線形で確率論的な対象から生成された時系列データを扱うための統一的な手法になり得る。というのも、どちらの場合も、それぞれの前提条件を満たせば、順列や再帰三角形のサイズ無限大の極限で、それぞれのモチーフと、状態と入力の組の系列を 1 対 1 に大体対応づけすることができる(Hirata, Sato & Faranda, 2020; Hirata, 2021)。つまり、順列や再帰三角形を用いて、確率論的な場合でも、筋の良い"記号力学系"を構成できる可能性が示されたことになる。記号力学を用いた非線形時系列解析の過去の研究はたくさんあるが、それらを順列や再帰三角形を用いて再度実装し直すことで、様々な解析・応用が広がっていくであろう。

短く要約すると、数理モデル化の前提となる決定論性・確率論性、線形性・非線形性の 2 つの組の検定が、ここ数年の研究で整った。また、非線形な時、決定論的な場合と確率論的な場合をつなぐ統一的な記述方法が発見された。ここでいう確率論的な場合は、通常の意味での非定常な場合も含むので、非常に大きな研究の可能性が、現在広がっている状態だと思われる。この機会に、1990 年代にブームであった複雑系研究の再評価が進むことを願う。

謝辞
この研究の主な部分は、JSPS 科研費 JP18K11461 の助成で行ったものであった。また、京都大学数理解析研究所での発表や本講究禄をまとめるにあって、the Research Institute for Mathematical Sciences, an International Joint Usage/Research Center located in Kyoto

University の助成も受けた。

参考文献

J. M. Amigó & M. B. Kennel (2007), Topological permutation entropy, Physica D 231, 137-142.
C. Bandt & B. Pompe (2002), Permutation entropy: A natural complexity measure for time series, Phys. Rev. Lett. 88, 174102.
J.-P. Eckmann, S. O. Kamphorst, & D. Ruelle (1987), Recurrence plots of dynamical systems, Europhys. Lett. 9, 973-977.
Y. Hirata (2021), Recurrence plots for characterizing random dynamical systems, Commun. Nonlinear Sci. Numer. Simulat. 94, 105552.
Y. Hirata, Y. Sato, & D. Faranda (2020), Permutations uniquely identify states and unknown external forces in non-autonomous dynamical systems, Chaos 30, 103103.
Y. Hirata & M. Shiro (2019), Detecting nonlinear stochastic systems using two independent hypothesis tests, Phys. Rev. E 100, 022203.
Y. Hirata & N. Sukegawa (2019), Two efficient calculations of edit distance between marked point processes, Chaos 29, 101107.
M. B. Kennel (1997), Statistical test for dynamical nonstationarity in observed time-series data, Phys. Rev. E 56, 316-321.
C. G. Lamoureux & W. D. Lastrapes (1990), Persistence in variance, structural change, and the GARCH model, J. Bus. Econ. Stat. 8, 225-234.
D. P. Mandic, M. Chen, T. Gautama, M. M. Van Hulle, & A. Constantinides (2008), On the characterization of the deterministic/stochastic and linear/nonlinear nature of time series, Proc. R. Soc. A 464, 114-1160.
N. Marwan, M. C. Romano, M. Thiel, & J. Kurths (2007), Recurrence plots for the analysis of complex systems, Phys. Rep. 438, 237-329.
T. Nakamura, M. Small, & Y. Hirata (2006), Testing for nonlinearity in irregular fluctuations with long-term trends, Phys. Rev. E 74, 026205.
T. Schreiber & A. Schmitz (1996), Improved surrogate data for nonlinearity tests, Phys. Rev. Lett. 77, 635-638.
S. Suzuki, Y. Hirata, & K. Aihara (2010), Definition of distance for marked point process data and its application to recurrence plot-based analysis of exchange tick data of foreign currencies, Int. J. Bifurcat. Chaos 20, 3699-3708.
J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, & J. D. Farmer (1992), Testing for nonlinearity in time series: the method of surrogate data, Physica D 58, 77-94.