

生態／社会系における普遍性の探求

Approach to the Universality in Social Data

島田 尚^{*1*2}Takashi Shimada^{*1*2}^{*1}東京大学 大学院工学系研究科 物理工学専攻

Department of Applied Physics, Graduate School of Engineering, The University of Tokyo

^{*2}JST CREST

JST CREST

Social and economical systems today are one of the most feasible object to obtain large and accurate data. However, analyzing those data involves difficulties which mainly stem from its non-stationarity, its *smallness* comparing to the degree of freedom of the original system, and the impossibility of making experiments for validation. Here I discuss how to obtain better understanding from such data, taking some of our approaches as examples.

1. 問題意識

近年、社会系・経済系は最も大量の客観的データが得られる対象へと急速に変貌している。これらのデータの特徴として、(大量のデータではあるものの)系の自由度に比べて時系列が短く、また非定常性が無視できないことが挙げられる。また、科学データと異なり繰り返しの実験は不可能であることが多い。このため、解析は静的な特徴についてのものになりがちで、また得られた統計的特徴から元の系について何が言えるかという吟味をすることは一般に難しい。このような問題については万能の数理的手法は無く、普遍的な側面を指向しつつ個別の解析をすることが肝要であると考えられる。以下ではこのような挑戦の要素を持つ下記事例を紹介し、本講演での非物理系データへの取り組み方についての議論のもととした。

2. 事例

2.1 非定常性：株式市場構造の遷移ダイナミクス

この研究では、株式市場 (S&P500) における株価変動の相関構造とそのダイナミクスについて調べた [1]。株価変動の相関構造については、基本的には Pearson の相関係数

$$C_{ij} \equiv \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sigma_i \sigma_j}$$

のような標準的な量によって評価する。ここで r_i は株 i の価格変動、 $\langle \cdot \rangle$ はある観測期間における平均を表し、 σ_i はその観測期間における r_i の標準偏差である。実際に過去約 20 年に渡る実データから全期間を通して S&P500 に選定され続けた 300 社余りについて計算した株価変動の相関行列を図 1 に示す。この図から、例えば同一産業セクター内、及びセクター間で異なる強さの相関があることがわかる。

このような相関構造の抽出に際して問題となるのは対象の非定常性である。すなわち、市場の相関構造は時間と共に変化していると考えられるので、観測時間は理想的にはある程度短くとる必要がある。一方で、そのような短い観測時間から得られる相関係数には常に大きなノイズが伴う。

このトレードオフの問題をある程度解決する試みとして、我々は相関行列の構造そのものではなく異なる観測期間にお

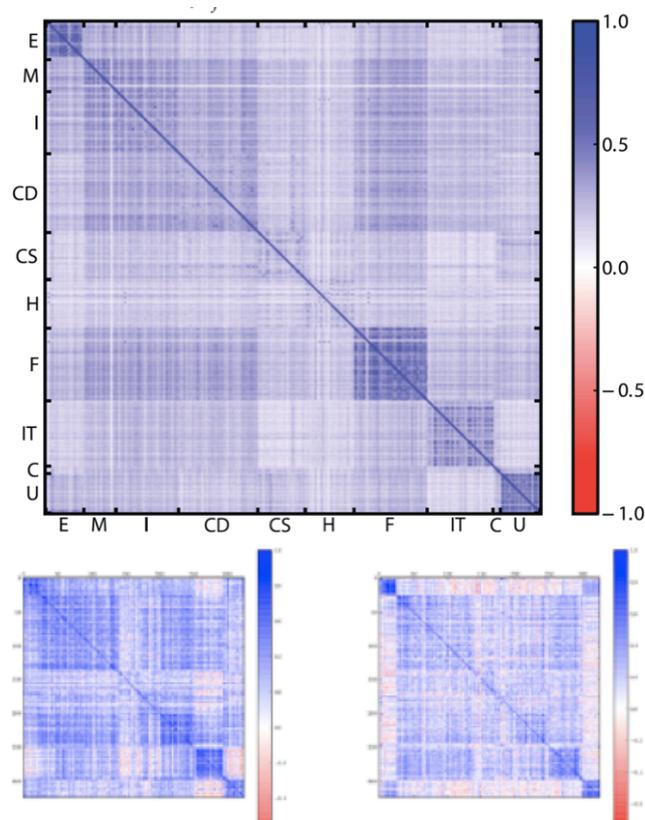


図 1: S&P500 における過去 20 年間の株価変動の相関行列 (上) と、これとは明らかに異なる構造を示す、ある 2 ヶ月間について計算した相関行列の例 (下)。白が無相関、青が正の相関、赤が負の相関を表す。企業は Global Industry Classification Standard (世界産業分類基準) インデックスに従って並べてある。産業区分は以下の通り。

E: エネルギー、**M:** 素材、**I:** 資本財・サービス (資本財 (航空、建設、電気設備、機械、商社・流通など)、商業・専門サービス、運輸)、**CD:** 一般消費財・サービス (自動車・自動車部品、耐久消費財・アパレル、消費者サービス (レジャー、教育など)、メディア、小売り)、**CS:** 食品・生活必需品、**H:** ヘルスケア (ヘルスケア機器・サービス、医薬品など)、**F:** 金融 (銀行、保険、不動産など)、**IT:** 情報技術、**C:** 通信、**U:** 公益事業 (電気水道ガスなど)

連絡先: shimada@ap.t.u-tokyo.ac.jp

〒 113-8656 東京都文京区本郷 7-3-1

る相関行列同士の距離に注目した。行列間の距離としては要素の差の絶対値平均

$$\zeta(C^\alpha, C^\beta) \equiv \langle |C_{ij}^\alpha - C_{ij}^\beta| \rangle,$$

や差の二乗和の平均

$$\zeta'(C^\alpha, C^\beta) \equiv \sqrt{\langle (C_{ij}^\alpha - C_{ij}^\beta)^2 \rangle},$$

のような自然な量が自然でこのような平均量に関しては、相関行列の要素数 (\sim 銘柄数)²/2) が多いことから良い精度の値が得られるためである。

実際に二ヶ月程度の互いに重ならない観測期間についてそれぞれ計算した相関行列 $C(t)$ に対して、各期間間の距離を過去 20 年に渡って 2 次元表示すると、タータンチェック状の非常に綺麗な異時間相関マップが得られる。この異時間相関マップの構造は、「市場の相関行列は特徴的ないくつかの構造 (準安定状態のようなもの) を持ち、その間を間欠的に遷移している」ことを示唆している。そこで、この予想を裏付けるためにクラスター解析によって特徴的な状態の抽出を試みた。この結果、複数個の「典型的な市場状態」を再構成することに成功し、更には市場がこれらの状態間を遷移している様子を捉える事ができた。

状態遷移図からはまず、一度訪れると長く安定に滞在されやすい状態や、時間的に疎に現れる状態など典型的なダイナミクスの異なる状態があることが伺える。特に時間的に疎に埋め込まれた典型的状態の発見はクラスター解析の成果であると言える。また、時間の経過と対応する大域的な状態遷移の方向性が伺える一方で過去に訪れた状態を再訪する事も珍しく無いことも分かる。更には状態遷移のダイナミクスの観点から特徴的と見出される時点がドットコムバブルの始まりと終わりや 2008 年秋の金融危機による株価急落の直前期等に対応する等、興味深い知見が得られた。

以上の結果は、例えば金融市場におけるリスクの予想と回避という実際上の問題についても有用である。同時に、非定常な時間発展をする大自由度系のダイナミクスの特徴を捉えた例としても物理的に興味深いものであると考える。

2.2 普遍性と個別性の吟味：日台動画サイトの解析

この研究においてはまずニコニコ動画における再生数分布の解析を行い、動画再生数のランクサイズ分布においてべき則からべき則へのクロスオーバーという興味深い分布を見出した。しかしながらこの系においてもデータの非定常性などから帰無仮説的分布が明らかでないという問題がある。このため、台湾における類似サイト (Yam) において同様の解析を行い、分布の比較を行った。この結果、べき則については普遍的であるもののクロスオーバーするという特徴は共有されないこと等が分かり、両サイトの違いがどのように分布の差と対応しているかについて推察することができた。

2.3 統計性の起源の理解：共存進化系における種の寿命分布

進化論的時間における生態系の性質は、多様化し続ける (もしくは少なくとも多様性を維持し続ける) 開放系の代表として物理学も含む幅広い分野の注目を集めてきた対象である。観測データから絶滅や種の寿命分布、食物連鎖網や系統樹の構造などにおける特徴的なパターンが報告されており、我々はこれらの異なった側面を多種が共存・相互作用しあう系が示す普遍的性質として理解しようと研究を進めてきた [2, 3, 4]。

この研究ではまず、非常に簡単なモデルが複数の異なった統計則を同時に再現しうる事を示した。その上で次に、特に特徴的な分布である種の寿命分布について互いに大きく異なる数理モデル群を用いてその再現の条件を探った。この両方向の解析から、対象にしてきた生態系に見られる統計性の多くはランダムな新規要素の導入を受ける競争進化系の自然な性質として理解できる事が分かって来た。

このような抽象化・簡素化された理解からは、研究対象にしてきた普遍的な側面が社会系や経済系等の他の“生態系”にも共有される可能性が予想される。実際に小売り店における商品陳列の寿命分布において生態系における種の寿命分布と同様の分布則が見出されており、普遍的様相についての明確な理解の重要性が確認できる。

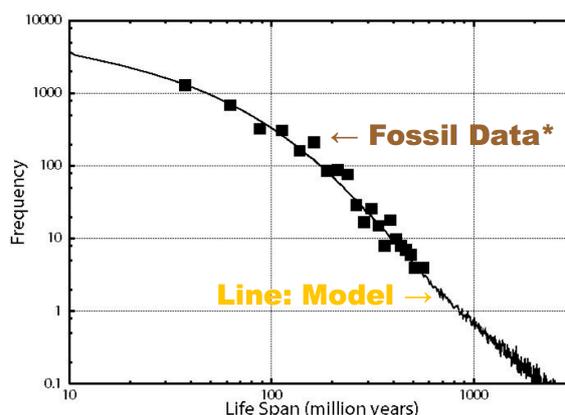


図 2: 化石データの示す種の寿命分布 (四角) と、系の自由度の増減を許す簡単な競争進化系モデルにおける自由度の寿命分布 (実線)。

参考文献

- [1] Michael C. Münnix, Takashi Shimada, Rudi Schäfer, Francois Leyvraz, Thomas H. Seligman, Thomas Guhr, and H. Eugene Stanley: “Identifying states of a financial market” SCIENTIFIC REPORTS Vol. 2, (2012) 644.
- [2] T. Shimada, S. Yukawa and N. Ito “Self-organization in an ecosystem” Artificial Life and Robotics, Vol. 6 (2002) 78-81.
- [3] T. Shimada, S. Yukawa and N. Ito “Life-span of Families in Fossil Data Forms q-exponential Distribution” Int. J. of Mod. Phys. C, Vol. 14 (2003) 1267-1271.
- [4] Y. Murase, T. Shimada, and N. Ito “A simple model for skewed species-lifetime distributions” New Journal of Physics Vol. 12, (2010) 063021.