

腸内細菌叢の数理モデルとその相互作用行列の解析

田口 優真, 時田 恵一郎

名古屋大学 大学院情報学研究科

Analysis of mathematical model of gut flora and its interaction matrix

Yuma Taguchi and Kei Tokita

Graduate School of Informatics, Nagoya University

概要

近年、腸内細菌叢が様々な疾患に影響を及ぼす可能性などが示唆されており、このような大規模群集の全容解明は、我々の QOL 向上において重要な役割を果たすことが考えられる。そこで、我々は、多自由度非線形進化力学系のパラメータ推定を行うことで、大規模群集内の相互作用を推察し、全容解明に寄与することを目標とする。本研究では、腸内細菌叢の実例においてパラメータ推定が行われた例のさらなる解析と、多自由度系に対応したパラメータ推定アルゴリズムの精度検証を行った。

1 はじめに

近年、日本では、潰瘍性大腸炎をはじめとする炎症性腸疾患の患者数は増加傾向にある。潰瘍性大腸炎は、その発症に腸内細菌叢の多様性の低下が関連している可能性が示唆されている疾患であり、腸内細菌叢の全容解明がなされると、治療に大きく役立つことが考えられる。腸内細菌叢は、およそ数百種、数十兆匹にもなる巨大な生態系を形成しており、炎症性腸疾患以外にも、腎臓病や自閉症スペクトラム障がいなどの全身性疾患や、肥満などの代謝系にも関連している可能性が示唆されている [1]。腸内細菌叢の複雑な相互作用を同定することで、その全容解明に寄与することができ、これらの疾患の治療や予防に有用であると考えられる。

本研究では、非線形進化力学系のパラメータ推定から、相互作用の同定にアプローチする。例えば、Lotka-Volterra 捕食者被食者系

$$\frac{dx}{dt} = ax - bxy \quad (1)$$

$$\frac{dy}{dt} = cxy - dy \quad (2)$$

において、 a, b, c, d はそれぞれ種の増殖率や密度効果、種間の捕食被食への影響の大きさなどを表しており、これらが推定されることで、相互作用が推定されたと考えることが可能である。

近年の分子生物学分野におけるシーケンス技術の発達により、腸内細菌叢のような、生体内の大規模なデータが、より大量に、より正確に得られるようになってきている。このため、これまで難しかった大規模群集の実データを用いた研究が比較的行いやすくなっているという背景がある。

腸内細菌叢のような大規模群集においては、この非線形進化力学系の連立する数が多くなる。それにより、最小二乗法を用いた一般的なパラメータ推定では、推定に要する時間が増加することが予想される。そこで、このような多自由度系における推定で計算時間を短縮させる方法 [2] が提案されている。本研究では、いくつかの非線形力学系にその手法を適用し、得られた結果の精度の比較を行い、実データへの応用可能性を検証した。

2 腸内細菌叢の相互作用

2.1 先行研究

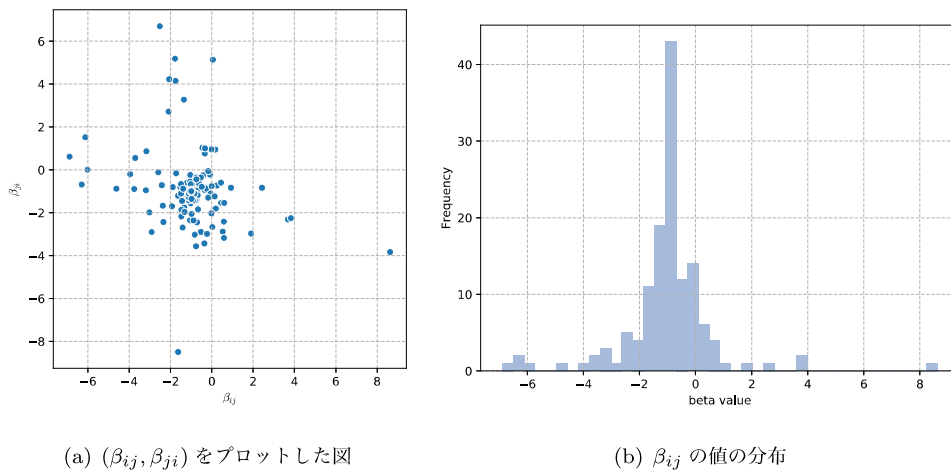
Marino et al.(2014)[3] では、マウス腸内細菌叢のシーケンスデータより得られた存在量データを OTU(operational taxonomic unit) を用いて 17 種に分類し、それらの時系列データをもとに、対応する一般化 Lotka-Volterra モデル

$$\frac{dx_i(t)}{dt} = \alpha_i x_i(t) \left(1 - \frac{x_i(t)}{K}\right) + x_i(t) \left(\sum_{j=1, j \neq i}^n \beta_{ij} x_j(t)\right) \quad (i = 1, 2, \dots, n) \quad (3)$$

の α_i および β_{ij} を推定した。

2.2 種間相互作用行列の解析

ここで得られた、種間相互作用行列 β の特徴についての解析を行ったところ、 β_{ij} と β_{ji} 間の非対角要素間に強い相関がないこと (図 1(a))、および、値の分布が単峰のベル型となっており、正規分布に近い分布となっていた (図 1(b))。すなわち、この相互作用行列は、非対称ランダム行列に近い行列となっていることがわかった。これは、May が 1972 年に解析したランダム行列 [4] の実例となっていることが考えられる。



(a) (β_{ij}, β_{ji}) をプロットした図

(b) β_{ij} の値の分布

図 1 [3] で得られた種間相互作用行列 β の解析結果

2.3 長時間シミュレーション

式 (3) の数理モデル、および推定された α 、 β を用いて、先行研究 [3] で行ったシミュレーションより長時間のシミュレーションを行った結果、ある初期値からカオス的な振動をしながら 5 種が共存し、12 種が絶滅すると言う結果が得られた (図 2)。

このカオスや大量絶滅と言った結果は、ランダム行列を相互作用にもつ系の一般的な振る舞いとして知られている [5, 6]。しかし、この結果は、「腸内細菌叢は多くの種が複雑な相互作用をしながら生存する」という我々の直感からは外れた結果を示しており、さらに高密度・長時間データを用いて検証することが必要であると考えられる。

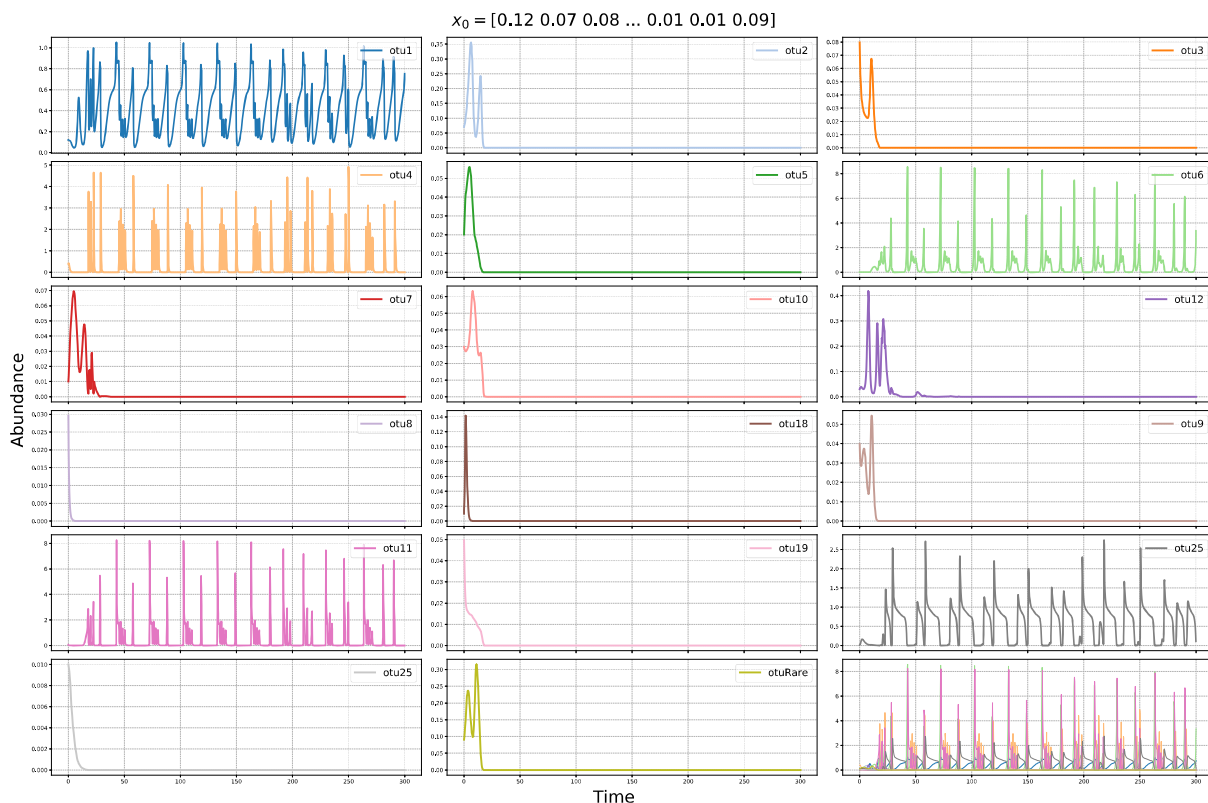


図2 パラメータ α , β の推定値を用いた数理モデルのシミュレーション結果 ($0 \leq t \leq 300$ (日))。右下を除くパネルは各 OTU の個体群動態，右下のパネルは全 OTU の個体群動態を重ね合わせて描いたもの。

2.4 固定点

以下では推定されたパラメータを持つ 17 種 LV 系の固定点を求める。固定点は、系

$$\frac{dx}{dt} = f(x) \quad (4)$$

において、

$$f(x^*) = 0 \quad (5)$$

となる点 x^* のことである。ここで、全ての種が $x^* > 0$ であると仮定すると、各式の両辺をそれぞれ $x_i(t)$ で割ることにより、

$$\alpha_i x_i(t) \left(1 - \frac{x_i(t)}{K}\right) + x_i(t) \left(\sum_{j=1, j \neq i}^n \beta_{ij} x_j(t)\right) = 0 \quad (6)$$

$$\alpha_i \left(1 - \frac{x_i(t)}{K}\right) + \sum_{j=1, j \neq i}^n \beta_{ij} x_j(t) = 0 \quad (7)$$

と、17 元 1 次連立方程式にすることができる。これを数値的に解析したところ、内部平衡点 ($\forall x_i^* > 0$) がないことがわかった。

共存種数を減らした部分系についても調べた。例えば、 $x_1^* = 0$ と仮定し、その他の $x^* > 0$ とすることで、先ほどと同様に 16 元 1 次連立方程式とすることができる。これを 2 種、3 種と順に増やしていくことで、合計 $2^{17} - 1$ パターンを検証したところ、最大でも 10 種が生存する固定点しか存在しないことが判明した。

この結果は、先の 12 種が絶滅 (5 種が生存) するという結果と整合しないが、ここでの 10 種共存解を取りうる範囲が狭く、先のシミュレーションにおいては、初期状態がその中に入らなかったためであると考えられる。ここでも、ランダム行列を相互作用に持つ系の「平衡状態の予測不可能性」[5] という性質が現れていることがわかった。

3 パラメータ推定手法

3.1 概要

従来、パラメータ推定において一般的に用いられている手法として、最小二乗法がある。一般的に、最小二乗法は、

$$\frac{dx_n}{dt} = f_n(x_1, x_2, \dots, x_n | \theta_n) \quad (8)$$

という θ をパラメータとする系について、推定値 $x_n(t)$ と実データ $\hat{x}_n(t)$ について、

$$\sum_n \sum_t \{x_n(t) - \hat{x}_n(t)\}^2 \quad (9)$$

を最小化するパラメータを求めるものである。しかし、この手法では、腸内細菌叢のような、種数の多い多自由度系においては、学習の収束に膨大な時間を要し、かつ推定の精度が下がるという欠点がある。そこで、各変数をそれぞれ独立に考え、それぞれの変数 n ごとに

$$\sum_t \{x_n(t) - \hat{x}_n(t)\}^2 \quad (10)$$

最小化するパラメータを決定するという手法が提案された [2, 3]。この手法では、

$$\frac{dx_n}{dt} = f_n(x_1 | x_2, \dots, x_n, \theta_n) \quad (11)$$

のような、注目する変数以外は実データを補間したものをパラメータとして用いるため、従来のものと比較すると精度が下がることが予想されるが、数十種、数百種という巨大な系についても、各変数を並列化して計算することが可能になるため、有用であると考えられる。そこで、本手法に関して、まずは次元の低い系において、精度を検証した結果を紹介する。

3.2 検証結果

本研究では、検証のため、あるパラメータ値による解の挙動が既知であるモデルとして、Lotka-Volterra 捕食者被食者系や Lotka-Volterra1 捕食者 2 被食者系などのモデルについて、パラメータ値を設定し、そこから生成したデータを用いて、推定を行った。その結果、多くの場合で、精度の良い結果が得られることが確認された。

さらに、精度を低下させないために留意する点として、データの取り方に注意することもわかった。例えば、Lotka-Volterra1 捕食者 2 被食者系において、振動や収束、カオスなどの固有の状態に乗る前の過渡状態のデータのみでの推定は精度がよくない結果を示したが、定常状態も含むデータを含めたデータでの推定結果は、改善された結果を示した。

3.3 実データへの応用

生成した例題のみでなく、実際の生態系におけるデータに適応した例を紹介する。Lotka-Volterra 捕食者被食者系の実例として有名な、カンジキウサギとオオヤマネコの個体数変動データを用いて、推定を行った (図

3). その結果, 比較的良好な精度で推定することが可能であった。よって, この手法は, 実際の生態系データにも対応可能であることが確認された。

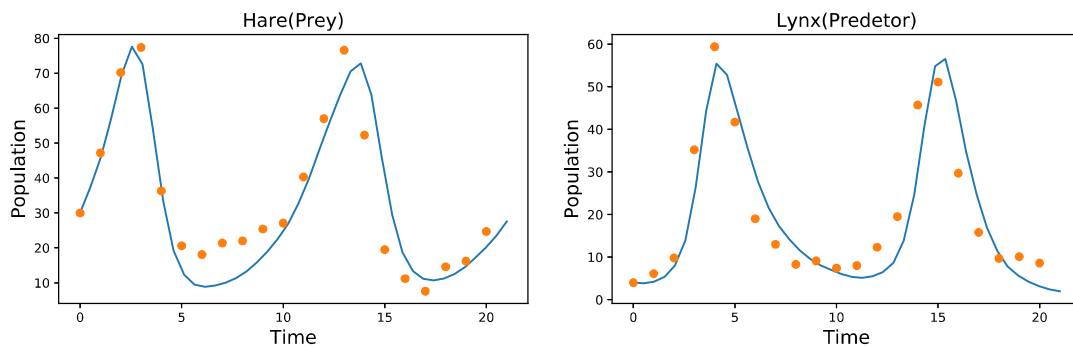


図3 カンジキウサギーオオヤマネコの個体数データを用いた推定結果

4 今後の展望

今後の最重要課題として, ノイズの影響の最小化が挙げられる。生態系などの実データに含まれるノイズは, 推定の精度を下げてしまう。そこで, 推定を行う前にデータの前処理としてノイズを最小化する作業を行うことを検討する。Marino ら [2] では, ノイズの最小化手法として, 信号などのノイズ除去に用いられることの多い, Savitzky-Golay フィルタや Whitaker フィルタ [7] を用いることが提案されている。しかし, 適切にフィルタを用いないと, 元のデータに固有な振動まで除去してしまう可能性があるため, この平滑化手法の適切な適用方法をさらに検討していく必要がある。

さらに, 今回の研究で用いた 3 種 Lotka-Volterra 系よりもさらに多種の非線形力学系を用いた推定精度の検証を行い, 今回, 例として用いた腸内細菌層に限らないマクロな生態系や感染症動態などの多自由度実データを用いた理論モデルのパラメータ推定につなげる予定である。

参考文献

- [1] 福田真嗣 and 佐野ひとみ, Hum. Biol **40**, 463 (2013).
- [2] S. Marino and E. O. Voit, Journal of bioinformatics and computational biology **4**, 665 (2006).
- [3] S. Marino, N. T. Baxter, G. B. Huffnagle, J. F. Petrosino, and P. D. Schloss, Proceedings of the National Academy of Sciences **111**, 439 (2014).
- [4] R. M. May, Nature **238**, 413 (1972).
- [5] K. Tokita and A. Yasutomi, Physical Review E **60**, 842 (1999).
- [6] K. Tokita, 理工研報告特集号: ASTE: advances in science, technology and environmentology: special issue **11**, 59 (2015).
- [7] P. H. C. Eilers, Analytical Chemistry **75**, 3631 (2003).