

統計数理

第71巻 第2号

(通巻138号)

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

目次

特集「諸科学における統計数理モデリングの拡がりⅡ」

「特集 諸科学における統計数理モデリングの拡がりⅡ」について	
横山 雅之	99
行動意思決定研究者の読んだ『確率の出現』とその示唆 [研究ノート]	
広田 すみれ	101
生態学におけるモデルと法則に関する科学哲学論考の意義 [研究ノート]	
島谷 健一郎	119
果樹豊凶の決定論的非線形予測	
—アンサンブル再構成と Lorenz 類推法による1年先収量予測— [研究ノート]	
酒井 憲司	129
力学系を組み込んだ変分自己符号化器による顕微鏡動画モデリング：ゼニゴケ精子の運動を例として [原著論文]	
近藤 洋平・南野 尚紀・上田 貴志	149
位相的データ解析による銀河分布の定量化とバリオン音響振動抽出 [原著論文]	
竹内 努・河野 海・クレスチエータ・西澤 淳・村上 広耶・馬 海霞・本武 陽一	159

2023年12月

大学共同利用機関法人 情報・システム研究機構 統計数理研究所

〒190-8562 東京都立川市緑町10-3 電話 050-5533-8500(代)

本号の内容はすべて <https://www.ism.ac.jp/editsec/toukei/> からダウンロードできます

ISSN 0912-6112

統計数理

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

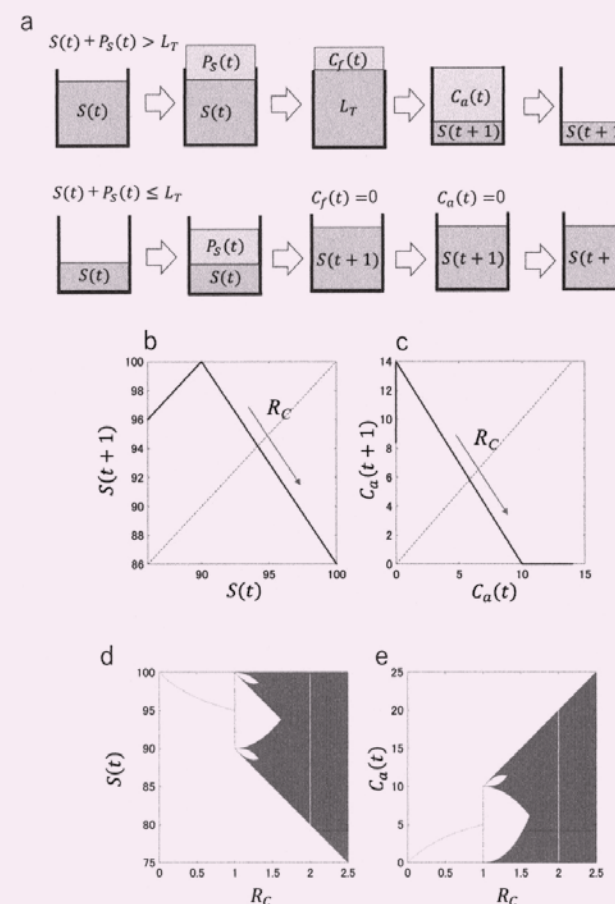
第71巻 第2号

2023

統計数理

Vol. 71, No.2

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS



統計数理研究所

統計数理

(年2回発行)

編集委員長 瀧澤 由美

編集委員 朴 堯星

逸見 昌之

村上 大輔

持橋 大地

吉本 敦

特集担当編集委員 菊地 和平

島谷 健一郎

三分一 史和

編集室

池田 広樹

川合 純華

長嶋 昭子

「統計数理」は、統計数理研究所における研究成果を掲載する統計数理研究所「彙報」として1953年に歴史を始め、1985年に誌名を変更し今の形となりました。現在は、統計数理研究所の研究活動に限らず、広く統計科学に関する投稿論文を掲載し、統計科学の深化と発展、そして統計科学を通じた社会への貢献を目指しています。

投稿を受け付けるのは、次の6種です。

- a. 原著論文
- b. 総合報告
- c. 研究ノート
- d. 研究詳解
- e. 統計ソフトウェア
- f. 研究資料

投稿された原稿は、編集委員会が選定・依頼した査読者の審査を経て、掲載の可否を決定します。投稿規程、執筆要項は、本誌最終頁をご参照ください。

また、上記以外にも統計科学に関して編集委員会が重要と認める内容について、編集委員会が原稿作成を依頼することがあります。

その他、「統計数理」に関するお問い合わせは、各編集委員にお願いします。

All communications relating to this publication should be addressed to associate editors of the Proceedings.

大学共同利用機関法人 情報・システム研究機構

統計数理研究所

〒190-8562 東京都立川市緑町10-3 電話 050-5533-8500(代)

<https://www.ism.ac.jp/>

© The Institute of Statistical Mathematics 2023

印刷：笹氣出版印刷株式会社

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

Vol. 71, No. 2

Contents

Special Topic : Expansion of Statistical-mathematical Modelling in Various Research Field II

On the Special Topic “Expansion of Statistical-mathematical Modelling in Various Research Field II” Masayuki YOKOYAMA	99
A Short Review of Hacking’s <i>The Emergence of Probability</i> and Its Implications from the Perspective of a Researcher for Behavioral Decision Making Sumire HIROTA	101
Does Philosophical Discussion about Ecological Modelling and Laws Contribute to the Developments of Ecological Studies? Ichiro Ken SHIMATANI	119
Deterministic Nonlinear Prediction of Tree Crops —Ensemble Reconstruction of Discrete Dynamics and a Lorenz Method for One-year Forward Prediction of Individual Yield— Kenshi SAKAI	129
Dynamical-system-integrated Variational Autoencoder for Modeling of Time-lapse Microscopy Data : Application to Liverwort Spermatozoid Motion Yohei KONDO, Naoki MINAMINO and Takashi UEDA	149
Quantification of Galaxy Distribution with Topological Data Analysis and Detection of the Baryon Acoustic Oscillation Tsutomu T. TAKEUCHI, Kai T. KONO, Suchetha COORAY, Atsushi J. NISHIZAWA, Koya MURAKAMI, Hai-Xia MA and Yoh-Ichi MOTOTAKE	159

December, 2023

Research Organization of Information and Systems

The Institute of Statistical Mathematics

10-3 Midori-cho, Tachikawa, Tokyo 190-8562, JAPAN

表紙の図は本誌 131 ページを参照

「特集 諸科学における統計数理モデリングの 拡がりⅡ」について

横山 雅之^{1,2} (オーガナイザー)

「統計数理」第71巻第1号(2023年6月)に続く第2号として本特集号を刊行できることをとてもうれしく感じている。第1号巻頭言で記したように、ここ数年間にわたって継続的に開催している共同研究集会では、多様な分野(諸科学)における統計数理モデリングや統計思考に関する数多くの講演・話題提供がなされ、フリーディスカッションなども含めて分野を超えた多様な議論が行われた。第1号をお読みいただいた皆様には、「各分野固有の考え方に捉われない統計数理モデリングの可能性やその科学的解釈について議論することを目的」とした共同研究集会の趣旨も感じ取っていただけたものと期待している。研究集会主宰側の関係者の拡がり、講演を契機とした新たな共同研究、参加者からの講演者提案などによって、「諸科学」の幅が広がってきている。

そこで、この第2号では、共同研究集会題目を「諸科学における統計思考」と発展的に改名している状況を踏まえて、幅広い分野における統計思考に関する論文や研究ノートを執筆いただくこととした。

竹内論文は、我々が直接観測できるバリオン(宇宙論における通常の物質)が形作る「宇宙の大規模構造」に対して位相的データ解析手法を応用することで、その形成要因の一つであるバリオン音響振動信号を実際に検出できることを示している。広田論文は、科学哲学書である「確率の出現 第2版」(イアン・ハッキング著、2006年)に記された確率の出現に至る概念の変遷や認識論的側面・偶然的側面の二元性などを紹介しながら、それらがリスク研究に対して意味することを行動意思決定研究者である著者の視点で考察している。島谷論文は、生態学における科学哲学的問題を提起し、科学哲学誌上で議論が行われてきた論文を取り上げ、哲学誌上の議論が往々にして生態学の現場から切り離された形で展開されはするものの、そのような議論が成熟することで研究現場に恩恵をもたらしていると考察し、議論が不毛に終わることを恐れずに分野間のフィードバックが欠かせないと締めくくっている。酒井論文は、果樹園から得られる収量時系列集合(年1回の計測で時系列の大きさは小さいが、大きな作物個体数から大きな時系列集合となる)に基づいたダイナミクスのアンサンブル再構成、決定論的非線形予測(→1年先の収量予測)に関する方法論および実践例を示しつつ、一般性を有するその定式化から、カオスと農業の親和性の高さを述べている。近藤論文は、顕微鏡による観察で、かつてない高時空間分解能で多くの現象が観測可能となっている反面、複雑化・高次元化が進む顕微鏡動画データからの知見抽出が困難になっているという課題に対して、教師なし特徴抽出を目的とした変分自己符号化器に力学系を組み込んで動画データをモデリングすることで、生物学的に興味のある情報を得る枠組みを提案し、ゼニゴケの精子運動様式の観察で実際にその有用性

¹ 大学共同利用機関法人 自然科学研究機構 核融合科学研究所六ヶ所研究センター：〒039-3212 青森県上北郡六ヶ所村大字尾敷字表館 2-166

² 統計数理研究所 客員：〒190-8562 東京都立川市緑町 10-3

を検証している。

第1号巻頭言の再掲となりますが、共同研究集会の歴代の所内受入教員である、野村俊一先生、矢野恵佑先生、さらに、本特集号のオーガナイザーとしてもご尽力いただいた島谷健一郎先生、三分一史和先生に感謝申し上げます。また、査読を引き受けていただいた皆様にお礼申し上げます。2号にわたる特集の機会を与えていただいた編集委員会、および、編纂進行のお世話になりました編集室の皆様にもお礼申し上げます。

共同研究集会立ち上げや運営、第1号の特別編集委員としても多大な貢献をしていただいた故菊地和平さん(親しみを込めて)も、2号にわたる特集がここに無事に刊行されたこと、さらなる発展への契機となることを天国から期待してくれていると思います。

行動意思決定研究者の読んだ『確率の出現』と その示唆

広田 すみれ[†]

(受付 2022 年 12 月 31 日；改訂 2023 年 5 月 15 日；採択 5 月 16 日)

要 旨

本稿は科学哲学書であるイアン・ハッキング『確率の出現 第2版』(Hacking, 2006)の簡単な紹介である。同書は世界中で話題を呼び、現在でも引用される書籍である。同書では確率の出現について、パスカル以前に注目し、ルネサンス時代において“probability”という語は「是認に値するもの」という意味を持ち、またそれが臆見に属するものであることを示し、その後証拠の概念が変化し内的証拠が現れたことが出現の背景にあることを述べた。そしてその前史から確率は出現の時期から認識論的側面と偶然的側面の二元性があったことを主張した。また確率が明示的に記載されたのは『ポールロワイヤル論理学』の追加された最後の数章であることを示した。さらにホイヘンス、パスカル、ライブニッツ、J. ヒュッデ、グラント、J. ベルヌーイといった人物を確率解釈と帰納推論の観点から論じた。本稿では同書について、パスカル以前の特徴的な部分についてやや詳細に記載し、以降を章ごとに順に紹介した。最後にリスク研究に関して本書の主張が意味することを考察した。

キーワード：認識論的確率、偶然的確率、二元性、パスカル以前、帰納推論、蓋然性。

1. 『確率の出現』の影響と位置付け

この稿は世界的に定評のあるイアン・ハッキング『確率の出現』(原著初版はHacking, 1975, 第2版はHacking, 2006, 邦訳2013)について、邦訳者の1人として概要を紹介することを目的とする¹⁾。これは統数研の島谷氏の強い要請によるもので、また筆者(広田)自身、この著書の主張の紹介が現在でもリスクと意思決定に関する理解を深め、社会的に有益だろうと考えたためである。半世紀近く前の書籍を今さら、と考える向きもあるだろう。しかし原著者自身が第2版(Hacking, 2006)で書いているように、この本は世界的な確率の哲学の研究の流行を生み出し、「確率論の哲学」という一つの領域を生み出すほどの影響力を持ち、さらにBernsteinの“Against the Gods: The Remarkable Story of Risk”(Bernstein, 1998)のような世界的なベストセラーでも重要な典拠とされている。だがその割に、哲学の領域を除き日本での確率に関する議論に現れにくいのは、この本に邦訳がなかったことが大きな理由だと推測する。筆者(広田)の専門であるリスクや意思決定の研究者にとっても確率に関する理解は必須だが、世界的常識となっているその部分の共通理解が本邦で欠けがちなることを残念に思ったのが翻訳の動機である。邦訳は科学哲学者の森元良太氏(北海道医療大学)との共訳で彼に負うところが大きい。

まず、あらかじめ本稿のスタンスを述べる。『確率の出現』は一見歴史書にも見えるが、著者

[†] 東京都市大学 メディア情報学部：〒224-8551 横浜市都筑区牛久保西3-3-1

表 1. イアン・ハッキング『確率の出現』(2013)の章立て.

第 1 章	欠落していた考え
第 2 章	二元性
第 3 章	臆見
第 4 章	証拠
第 5 章	しるし
第 6 章	最初の計算
第 7 章	ロアネーズ・サークル
第 8 章	偉大な意思決定
第 9 章	思考法
第 10 章	確率と法
第 11 章	期待値
第 12 章	政治算術
第 13 章	年金
第 14 章	等可能性
第 15 章	帰納論理
第 16 章	推測法
第 17 章	初めての極限定理
第 18 章	デザイン
第 19 章	帰納
二〇〇六年版序論 確率的推論の考古学	

自身が明言しているようにフーコーの方法を用いた科学哲学書である。したがって専門外の筆者に全てを十全に説明することはそもそも不可能なため、本稿では同書の記述にできるだけ忠実に従い、記述されたことの範囲を踏み越えないよう努めながら、科学哲学者以外の諸賢が本書を読む手がかりになるよう内容を紹介する。また原著出版(1975年)後、半世紀近くの間内外問わず確率の哲学や統計学史に関する文献が多数出版された(一例だが伊藤, 1997; フランクリン, 2018)が、その点に関しては第2版で付け加えられた「2006年版序論」に関わる議論に限定する。ただ、行動意思決定の枠組みでリスクコミュニケーション研究を行ってきた立場から、最後にこの書籍がリスク研究や意思決定研究に示唆するところについていくつか書く。以降、本文中のページ数は邦訳(ハッキング, 2013)のページ数である。なおこの本はハッキングによる19世紀の確率・統計の発展による社会変化を扱ったHacking(1990)(邦訳はハッキング, 1999, 『偶然を飼いならす—統計学と第二次科学革命—』)の以前に出版された姉妹書でもある。またこの文の内容は広田(2021)が統計数理研究所の2021年冬の共同研究集会(通称:「諸科学」集会)で行った講演内容をベースに大幅加筆・修正したものである。

2. 『確率の出現』の構成、キーワードと予備知識

『確率の出現』は19章から成り、加えて2006年の第2版には「2006年版序論 確率的推論の考古学」(以降「序論」と表記)という、第1版への反響や議論に対するコメント・反論が追加されている。章立ては表1の通り(邦訳に基づく。なお「序論」は原著第2版では冒頭に置かれているが、内容的に初版を踏まえるため、邦訳では後ろに置いた)。内容は、確率計算が初めて考案されたパスカルとフェルマーの往復書簡の時期(第6, 7章)を挟んだ時系列で、第2-5章は古代に始まり主に中世・ルネサンスにおける「その後確率になったもの」という前史について述べている。続く第6-17章はパスカルとフェルマーの往復書簡により数的な確率が出現する1660年頃の直前からの約10年間を主に扱い、J. ベルヌーイが極限定理を考案した17章までを、それぞれ主な人物と(論理も含め)確率に関わる貢献や彼らの確率解釈を論じている。第18

章と第 19 章は『確率の出現』がその直後に影響を与えた科学哲学上の事象、すなわち死亡率や出生の男女比に見られるような「統計的安定」を神のデザインとする英国の思潮(第 18 章)を述べ、確率の出現によって可能になった「帰納推論」とそれにより生じたヒュームの帰納の懐疑を論じて終わる。

本書の特徴だが、まず題で想像される歴史書ではなく、数的な確率より、哲学的な確率という概念の変遷やそれに影響した要素、確率で可能になった「新しい論理」帰納推論、そして意思決定の考え方を扱った科学哲学書であることに注意が必要である。実際、通常確率論史で語られる事象は、出てくるものの、場合によっては非常に軽く扱われている。題の『出現(emergence)』という用語には強い含意があるようである。ハッキングの主張は一貫して、1660 年頃以前に確率の元になる考え方はあったものの、確率という網の目のような考え(ideas. 複数形であることに注意)はそこから 10 年間に突然現れ、以降多くの人を知るようになったとしており、それ以前とは断絶があるという考え方である。この点にはその後議論がかなりあったようだが、「序論」からするとハッキングは多くの反論にもかかわらずこの考えを堅持している。とはいえそれは「(科学)革命」ではなく、また進化でもなく(p.325. 以降本文中に記載された頁数はハッキング, 2013 の頁を指す)、侵食や雪崩等の地質学上の変化と同様に(p.326), それまで表に出なかった変化がここで一気に「出現」した、と考えており、それがこの題に反映されている。

確率論史としてはトドハンター(2002, 原著は Todhunter, 1865)が知られ、ハッキングも冒頭で触れている。だが本書の眼目はむしろ、トドハンターが「わずか数行」で済ませたパスカルとフェルマーの往復書簡の話題(第 6 章)以前、数的に確率が検討され始める前の、第 3-5 章の中世部分である。この部分でハッキングは probability あるいは probable という語が過去にどのように用いられ、その時代にどんな意味を持っていたかに基づいて検討を進め、同書の中心的な主張の一つである「確率の二面性(認識論的, 偶然的)は出現の以前から備わっている」ことの根拠の核をまずここで示している。ただ、ここで probability や probable は現在のような(数的な)確率、あるいは「ありうる」といった意味ではなく、例えば中世では、意味が異なる点に注意が必要である(訳出の際には意味するところが変化することから同じ語でも蓋然性としたり、ルビを振ったりする方法をとった)。哲学における「蓋然的(英語では probable)」は岩波の哲学小辞典によると以下の通り。「確からしいこと。いくつかの、どれも可能な将来の事件が考えられ、しかもそのどれかが必ず起ることが分かっているとき、他のものに比してそのうちのとくに一つが起ることが、他の諸可能性を排除してしまうほどの十分な根拠はないにしても、ある程度の根拠によって予期される場合、その事件の起ることが蓋然的であると言われる。その予期の度合を蓋然量(または確率・公算)という。確率が過去の事実により、または形式的に数学的に測られる場合については→統計, 確率」(栗田・古在, 1979)。よって必ずしも数学的ではなく、こういった不確実な事象の発生の予測に関するものが蓋然性である。ちなみに哲学の中ではキリスト教の護教論での蓋然論(probabilism)に基づく決疑論はパスカルが激しく攻撃したもので、中世ヤルネサンスの語としての「蓋然性」には必ずしも肯定的ではないニュアンスを含んでいた点も知っておく必要がある。

さて本書のもう一つ重要な柱に「新しい種類の論理」がある。哲学での probability は過去の事象から未来の事象を予測する帰納論理や帰納的推論と強く結びついたものであり、ハッキングはこの「新しい論理」に本の中で一貫して関心を向ける。ハッキングは『ポールロワイヤル論理学』(初版は Arnauld and Nicole, 1662; 邦訳はアルノー・ニコル, 2021; 以下、『論理学』と表記)の最後の数章を確率が初めて出現した章とするが、それを「新しい種類の非演繹的な推論の研究(p.128)」として紹介し、ライプニッツは「チャンスの理論は不確実な状況下での推論という別事例に移行できると推量し(p.150, 傍点広田)」、「確率論を偶然的な出来事に関する論理と

みなしていた(p.151)」とする。つまり、確率の出現を帰納的推論の研究の始まりという角度から捉え、それゆえ最後は帰納論理への懐疑を呈したヒュームで終わる。この帰納推論との関連は第2版で追加された「序論」でより明確に強調されている。帰納は哲学上の重大問題で、専門家が多数いることから筆者(広田)は表層に触れるだけに留めるが、本書の最終章(19章)が「帰納」でヒュームの帰納の懐疑で終わるのは本書の企図を示している。ハッキングは帰納論理を、確率が出現し、J. ベルヌーイが極限定理を証明した(第16, 17章)ことで可能になったと考えるが、この点については「序論」で伺えるように、多くの反論があったようである。だが確かに現代の統計手法は入手されるデータから全体像を推定する帰納推論であり、その意味ではハッキングの関心は当然でもある。「確率はヒュームが著したとき完全に出現した(序論, p.340)」。ちなみに、ハッキングは後年教科書として“An introduction to probability and inductive logic”(Hacking, 2001)を書いているが、この書籍はこの点をより詳しく、かつやさしく説明しており、哲学者としての関心の所在の一つを示している。

また不確実な状況での合理的意思決定の形式にも大きな関心が向けられている。パスカルは数学史ではフェルマーと共に確率計算を考案した人物として評価され、確率の生みの親といった位置付けにされる。しかし、ハッキングはパスカルについて、むしろ「意思決定理論の構造を広く理解した(p.323)」という点で評価している。実際、パスカルについて割かれている7-8章では確率計算が案出された部分はややゴシップ的な面に頁が割かれているのに対し、「偉大な意思決定(第8章)」では「パスカルの賭け」の意思決定の枠組みについて詳しく論じ、続く「思考法」(第9章)では『論理学』において、合理的な決定(リスク評価)のために確率が導入されたことを紹介する。

第6-7章、第9-17章では確率論史でお馴染みの人物が扱われるが、その業績の検討も、それぞれの人々の確率解釈(認識論的確率、あるいは偶然的確率として捉えていたのか)や、新しい論理にどのように寄与したのか、といった観点から検討が進められる。そしてハッキングは同時に、確率解釈に関わる現代の議論にも折々に触れる(例えば1回のみの試行での期待値の正当化可能性の議論(p.159)など)。

また本書は特定の誰かによって確率が出現した、という立場を取っていないことにも注意が必要である。ハッキングは第1章や、第2版で加えられた「序論」の「確率[出現]の前後の違い」の中など、繰り返し、特定の人物の寄与によって確率が出現したのではなく、その前後10年間に欧州各地で「確率という出現してきた考えをすべての市民が認識した(p.2)」と考えていることを強調している。その点においても歴史書ではなく、ハッキングは第2版でははっきりとフーコーの分析手法を応用したことを明かしている(p.316)。『確率の出現』は、まず1650年以前の前史において事前からあったものが、「侵食、雪崩」のように(「序論」, p.326)その後どのように眼に見える大きな変化を生み出したのかを述べる。その点から見ると、科学的思考の成立に影響され、またどう影響して確率が出現したかを記述する書籍、という見方で読むべきものである。

以上からこの本は科学哲学史の常識を踏まえて書かれている点を再度注意しておく。次項で述べるように、ハッキングの主張では、ルネサンスとそれ以降で「証拠」の概念が変化してしまったことが現代的な確率の出現と重要な関わりがあるとす。また、出現後(J. ベルヌーイより後)、統計上の死亡率の一定の規則性や、出生についての男女比、といった統計的安定性について、英国を中心に「世界はあらゆる点で明らかにうまく配置されているので、複雑な一個の時計仕掛けのように職人を必要とする(Derham, 1713)(p.287)」という(誤った)主張が流布したというが、この「時計仕掛けの宇宙論」も科学史でよく知られた考え方である。具体的に科学哲学関係の概念を挙げると、「しるし(sign)」「証拠(外的証拠と内的証拠)」「知識と臆見(opinion)の区別」「高級科学と低級科学」がある。

以下、羅列的ではあるが本の流れを追って順に概要を紹介する。ただし、次項部分がおそらく重要な点だと推測するため、この部分に関しては少し説明を加えながら紹介する。

3. 導入部(第1-2章)とし、証拠の概念の変化と確率の前身の二元性(第3-5章)

第1, 2章は導入部である。第1章では確率の出現に関する諸説をざっと眺めるとともに、問題を欧州での確率の問題に絞ることを宣言している。というのは、ここで引用されているインドの『マハーバーラタ』(紀元前2-1世紀。邦訳は上村, 2002)には、驚くことに現在という標本から母集団を推測する統計的推測にあたる物語があり、そのように確率の出現以前に関係する記載が欧州以外にも見受けられるためである。この話題は非常に興味深い、それは書籍にあたっていただきたい。第2章は確率の二元性について、現代の確率解釈のさまざまな説を概観しながら、確率は元(前身)から認識論的確率と偶然的確率という二元性がある、という本書の最も重要な主張を展開している。

その前身である中世における probability について検討しているのが第3-5章である。本書はこの部分が難解なので、ここでは少し詳しく解説する。これらの章では中世やルネサンスでの科学の変化を背景としながら、probability の意味が変化してきたことが論じられる。ハッキングの考えでは、パスカルとフェルマーの往復書簡から J. ベルヌーイによる極限定理までの約10年間で人々は確率をよく知るようになったが、しかしパスカル以前にも確率の元となるものはあり、この「確率ではない前提条件」の中で、現代の人々が抱えている概念の特性が決定された、とする。その「特性」とは何か。すなわち確率はその起源から二元性がある、というものである。よく知られるように、数学的には一意に定まる確率は、哲学的解釈の側面からはさまざまな説が存在する。たとえば Stanford Encyclopedia of Philosophy (n.d.) では「確率の哲学的解釈」として長文の項が設けられており (Hájek, 2019)、哲学的解釈としては大きく分けると認識論的確率と偶然的確率が存在するのは周知の通りである。認識論的確率にはケインズの論理説や主観確率が代表的なものであるが、これを確率として認めない、という立場も依然存在し、ハッキングの記述からすると世界的にも当時そのようであったと推察される。しかし、ハッキングは第1章から旗幟を明確にし、確率の二元性は確率が出現する1660年以前の「確率ではないが、突然変異のようなものを通じて確率へと変化した何か」にすでに備わっていた、とする。それについて記載したのが第3-5章である。まず前章(第2章)で「統計的で、ストカスティックな法則に関連する偶然的確率と、「認識論的で、統計的な背景の全くない命題についての合理的な信念」認識論的確率を区別する。極端な立場では「確率は一種類しかないのでその区別は見せかけだ、という(p.23)」が、本書では確率は元来二元性があったことを全体で示し、そして前史は「ルネサンス」が重要であるとする。

第3章は「臆見」(opinion)である。「臆見」は「ドクサ」のことで、知識と対置されるものである。哲学では古くから知識とは「正当化された真なる信念(p.33)」で論証によって到達するものとされるが、一方臆見は、論証ではなく、反省や議論、討論から得られ、論証できない命題にも適用される。この2つは明白に区別される。ハッキングは、probability は(現代の我々にとっては意外なことに)臆見に含まれ、ラテン語の probabilitas には「是認に値すること」のような意味があるなど、語の使用に基づいて複数の根拠を紹介する。臆見に含まれるということは、論証に必要な証拠による裏付けがあるのではなく、「尊敬されている人による裏付け(p.37)」があるという意味である。つまり、大まかに言えば、この時点では誰か尊敬されている人がそうだと、たとえば probable なのである。この部分に関しては主に Byrne (1968) による研究によるものようだが、ハッキングは Byrne の文を引き、プロバビリティーは「臆見に関する是認可能性」を意味していたとする (p.38)。

そのことから、ルネサンス期の(認識論的な)probability は知識と論証による科学に専念していた物理学者ではなく、論証的である見込みが一切なかった内科医と錬金術師の probable なしるし(sign)から(確率の元となる)発展中の諸概念が見つかる、とした。実際ガリレオやベーコンは言葉として probabilitas(ラ)/probability を使ってはいるものの、それらには「是認に値すること」という古い意味しかないことを示した。その上で「ここで目を向けなければならないのは、天文学、幾何学、力学といった「高級科学」ではない。むしろ目を向けるべきは、臆見に手を出さなければならなかった低級な経験主義者たちなのである(p.46)」とする。しるしも本書の最重要概念の1つであるが、ここでのしるし(sign)とは症候ではなく、予後診断のようなものである。この低級科学の内科医と錬金術師の probable なしるしは、その後変化がみられた。たとえば疾病の病原菌理論を考えたイタリアの科学者フラカストロ(1478-1553)は伝染病について「これらの中にはほぼ常に信じられるものと、信じられることが多いものがある。したがって、それらすべては…プロバビリティーを含んだしるしとしてのみ考えるべき(p.47)」と述べている。つまり、フラカストロのこの言でのプロバビリティーは起こりやすさである。それまでの臆見の「蓋然性」は権威や大昔の書物によって証言されることで是認されるものだったが、フラカストロやその他のルネサンス期の著作では起こりやすさ(probability)を備えたしるしが見受けられる、という。

第4章の「証拠」はこれを受け、証拠の概念の変化について説明する。現代の哲学者は確率とはある仮説とその仮説を支持する証拠との関係、とするが、ハッキングは証拠の概念自体が現代とは異なると説明する。中世における証拠は人によるもので、外的証拠と呼ばれ、証言や証言者、権威による。一方、新しい現代的な証拠は事物によるもので、内的証拠と呼ばれ、ハッキングは内的証拠の例として、ある動物が豚であるかどうかの証拠に、足跡や痕跡、匂いをあげる。そして彼は、「近代的な確率概念は、しるしが証拠になる時に生じた時に多くを負っている(p.57)」とする。続く第5章で彼は確率は低級科学から生じたとし、医師で錬金術師のパラケルスス(1493-1541)を例に挙げる。パラケルススや同時代の者たちは「劣った書物の研究をやめて優れた書物の研究を始めよう」と言ったという。この場合、劣った書物というのは複写により誤りもある古代の文書、「優れた書物」は自然そのものである。ごく簡単にいえば、自然は神が書いた書物で、そこにある事物の名前が「しるし」となって治療方法がわかる、という考え方をとっていたという。例えば水銀で梅毒が治せるというのは、梅毒の罹患場所が市場(mercury)で、商業の神水星(planet Mercury)は市場のしるし、よって水星と同じ名前のついた水銀は梅毒の治療薬である、といった考え方である。現在では到底理解しにくい、しかし当時の人々にとって自然は神の作った「真の書物」であることからこういった名前をしるしとする対応関係が考えられたようである。

ということから「世界がpのしるしを与えたなら、それは立証される(p.70)」。ここでpは命題である。pは正当な権威(この場合は神)から証言されるという意味で蓋然的(probable)、ということになる。といってもすべてのしるしが常に信頼できるわけではなく、「ほぼ常に信頼できるしるしもあれば、多くの場合に信頼できるものもあり」したがってこれらは「確からしさ(probability)を伴うしるし」ということになる。つまり、かつては権威によって証言されることが証拠だったが、その後自然という書物(=証言)が受け入れられることにより、あらゆる権威と同様に読み取られるべきものになった。これが「蓋然性すなわち証言と、安定した法則的な頻度が結びついたのは、内的証拠という新たな概念が生じた経緯の結果(p.72)」だというのがハッキングの主張である(「外的証拠」と「内的証拠」は記述から見るとハッキングのオリジナルの区別のようなものである)。第5章の終わりにはホップズが1650年に出版した、自然のしるしとその正確さの頻度の対応について書いた文章が引用され、そこには「しるしが二十回に一回外れるなら、人はその出来事について二十回に十九回まで賭けをするかもしれない」とあり、確率

は実質上出現したとする。

最後に頻度の話題は出てくるものの、第5章で最も重要な点は、元々 probability は人の証言に基づく臆見に関するものであり、したがって本来的に認識論的なものであったのが、証拠の概念が変化することで人の証言から自然という神の書物を内的証拠とし、そこでの頻度から確からしさという意味合いに変化したという点であろう。そして内的証拠の出現がなければ帰納の懐疑は存在しない(第19章)。ということで、この点は以降何度か記載されている。

4. カルダーノ、パスカルとフェルマーの往復書簡から『論理学』、ライプニッツまで(第6-10章)

この部分の前半は確率論史でよく知られた事象を扱い、第6章はカルダーノ、第7章はパスカルとフェルマーの往復書簡について紹介するが、第8章、第9章は数学的な確率論史には出てこない部分を熱心に論じている。

カルダーノの組合せ計算については、ハッキングは錬金術師のライムンドゥス・ルルス(1234-1315)が組合せ論の創始者であり、世界のすべての要素を真なるしるしで表現し、その組み合わせによって宇宙のすべての可能な化合物のしるしを生み出すことを考えたことを紹介している。この「組合せ」の発想は言うまでもなく後年ライプニッツの『結合論』(ライプニッツ, 2019. なおこの書籍の題はそもそも De Arte Combinatoria, すなわち直訳なら「組合せの方法」)のように、記号の組合せであり、引いてはそれにより普遍言語を志向する発想と共通のものである。これらから見ると、組合せ論自体が認識論的考え方と深く結びついていることを示唆している。その上で計算法に関する吟味をおこなっている。またガリレオが頻度説的な考え方であったのに対し、カルダーノは傾向説的であることも指摘している。

第7章はこの本の中で一番一般的な確率論史的な部分であり、ハッキングは往復書簡で確率計算が生まれたものの、むしろ分配問題の重要性を指摘し、この章はあっさりと終わる。

科学哲学者としてのハッキングにとってより重要であったのはむしろ第8-9章であろう。第8章は「パスカルの賭け」を扱い、第9章は『論理学』第4部終わりの4章(第13-16章)で初めて確率が明白に記載されたことを指摘している。通常数学的な確率論史ではこれらは現れないが、ハッキングはこれを取り上げ、論理の中で「probable な程度」がどのように考え方を変化させたか、について詳細に論じている。「パスカルの賭け」については哲学の中ですでに詳しく論じられているので割愛するが(例えば伊藤, 1997)、ここで注目するのは章題「偉大な意思決定」で、この章は現代まで連なる意思決定理論との関係が詳しく検討され、無差別(indifference)といった意思決定や効用理論で用いられる概念が「パスカルの賭け」という神への信仰に関する論証で現れたことを指摘する。さらに論証自体は護教論的にはすでに意味がないという批判は認めつつも、ここでの論証は「神が存在するか、しないか」という分割に基づいていること、期待値計算であることにも注目する。

続く第9章では「認識論的な色合いをもつある確率的表現が、測定可能なものを示すために体系的に用いられた最初の状況(pp.125-126)」としてパスカルの仲間であるアントワヌ・アルノー、ピエール・ニコルが執筆したと目される『論理学』を取り上げる。ハッキングが「プロバビリティーの数章」と呼ぶ章は後から付け加えられたものである。「新しい種類の非演繹的な推論の研究が始められている(p.128)」, すなわち後日帰納論理と呼ばれるもの、具体的には確率が出てくることで「不確実な状況における推論(p.129)」が始められたというのが注目の理由である。ここはリスクの研究者として非常に興味深く思うので、少し長くなるが本文のうち『論理学』からの引用部分を記載する。「雷鳴を聞くと過度に怯える人はたくさんいる。(…)その人々を異常な恐怖で満たしているのが死の危険だけであるなら、それが不合理なことを示す

のは容易である。200万人に一人が落雷で死亡すると言うのは大きさであろう。あまり一般的でない非業の死はどんな種類であれ、滅多にないのである。危害への恐怖は危害の重さだけでなく、その出来事の起こりやすさ(probability)にも比例するべきである。そして落雷よりもまれな死因は滅多にないので、ほとんどの死因は落雷よりも恐るべきものである。(p.132)」

このように、『論理学』のこの項の著者は自然現象の起こりやすさの測定に頻度を用い、かつ「意思決定の問題では効用だけでなく、確率も含めた期待値計算が必要であることを、著者が十分自覚していた(p.132)」ことを示している、という。

一方『論理学』では奇跡について検討し、討議で「証言の信頼性」に注意が払われていることにも注目している。その上で、改めて内的／外的証拠の区別が生じたことに論を向け、後半はジョン・ウィルキンズ(1614–1672)の著書(Wilkins, 1638)に初めてprobableな証拠が使われていること、さらにこの時期に証拠として論証、証言以外に「混合した証拠」と呼ばれる中間的なものが現れたことを指摘している。そして改めて、20世紀の認識論では内的証拠(事物の証拠)が基礎的であるが、かつては証言による外的証拠が主流であり、この時期になって初めてprobableな証拠という中間的なものが出てきたことを述べている。

第10章はこれ以降何度となく出てくるライブニッツを取り上げている。ライブニッツは法学者でもあることから、証拠は本来法律上の概念であり、ローマ法には証拠の尺度に関するさまざまな記載があると言う。ハッキングは、「条件論」(Leibniz, 1665)での条件付き権利について取り上げ、絶対的であるか、無であるかの間に条件付きを位置付け、その条件を「不確実」または「偶然的」とし、これを分数で表したことを指摘している。のちにライブニッツは“Ad stateram juris de gradibus et probabilitatum”(Leibniz, ca. 1676)という著書で、法学をモデルに、それと対抗する蓋然性(probability)の理論であるジェズイットの決疑論を非難し、この蓋然性と実在の確率(probability)を区別する必要があるとしている。そしてライブニッツは確率の度合いに基づいて「新しい種類の論理」について考え始めた。ハッキングによると、ライブニッツで特徴的なのは最初から数的な確率を認識に関する概念と見なし、論理に関わるものとしている点である。

5. 期待値計算から年金、極限定理まで(第10–17章)

以下は、やや羅列的になるが、各章で扱われている主な内容を記載する。第11章ではホイヘンスの「運任せゲーム」における期待値計算が論じられている。『論理学』で紹介された、効用と確率の両方に基づいて意思決定するよう方向付ける考え方は、言い換えればその時点ではまだ期待値の理解が当然ではなかったことになるが、ハッキングによればその直後のホイヘンス(Huygens, 1657)にはすでに似た概念が現れているという。すなわち、賭け事の賞金に対して公平な代価をどのくらい払うべきか、という問題に対する答えである。ただ、そこでは繰り返しのあるゲームにおける代価であり、1回のゲームに対してはいくらが正当か、に関する答えは現在まで議論があることへも言及する。同時に、ホイヘンスの整理の中には効用理論のいくつかの原理(加法性、複合的なケースの存在、いわゆる独立性の公理)がすでに当然のものとして含まれているという(p.161)。

また期待値は平均余命という考え方に影響し、ハッキングはジョン・グラントの書籍『死亡表における自然のおよび政治的諸観察』(Graunt, 1662)を読んだ弟ルードヴィッヒ・ホイヘンスのコメントに基づいて兄のC.ホイヘンスが「運任せゲームに関する私のルールによると…これが新しく命を宿した子どもの期待値となる」とした記述を紹介している。この平均寿命に関する記述が年金計算、すなわち第12章の「政治算術」および第13章「年金」に繋がる。

「政治算術」(第12章)は主にペティヤグラントの話題で、ここでも証拠の概念の変化の重要

性が指摘されている。具体的に挙げられているのはリスク研究者にとって興味深い疫病の例で、感染理論と瘴気理論(質の悪い大気によって病気が起こるとする)の対立を例に、理論がない限り、証拠は証拠とはならないことを述べる。続章(第13章)ではデ・ウィットとヒュッデによる年金計算が取り上げられるが、ここで問題となったのはホイヘンスが明らかにした数学的期待値、すなわち平均寿命を用いることの正当化であり、デ・ウィットはそれについて書いた一連の論文の冒頭でこれを論証している(p.194)。続いて問題になったのは、死亡率を年齢に関わらず一様にするか否か、という点であった。これは「等可能性」の問題として次の章に引き継がれる(この年金を巡る詳細な議論も非常に興味深いが紙幅の関係でここでは割愛する)。

通常ラプラスが確率を「等可能な場合の総数に対する好都合な場合の数の比」として定義したとされるが、ハッキングはライプニッツが1678年にこれをすでに行なっていたとする。同時にこの「等可能性」は循環論法でしか定義できない馬鹿げたものとする一方で、実はこれは確率の二元性から生じている「可能性の二元性」だとする。ライプニッツは確率を「可能性の度合い」とするが、結果的にライプニッツは確率という用語を二元性のあるものとして使っていることを指摘し(p.213)、また論理に関しても認識論的な使い方と偶然的なもの両方を用いているとする(等可能性は確実性の度合いという意味で認識論的だが、物理的な等可能性という意味でも用いている)。

第15章の「帰納論理」も主にライプニッツの業績を取り上げているが、ここでは帰納の論理的側面に焦点が当てられ詳しく検討されている。ライプニッツは法学から「確率は仮説と証拠の間の関係であることを学んだ(p.237)」のだとすれば彼の考える確率は認識論的な確率に当たるが、一方で「確率は物理的傾向性の問題(p.237)」と考えていたともする。したがってライプニッツの場合、確率の二元性の両義で確率を用いていることになる。またライプニッツが確率によって可能だと考えた「新しい種類の論理」は、カルナップを予言するものとされ、この章ではジェフリーズやケインズ、カルナップの論理を紹介しながらライプニッツが最初に発案したこの論理について説明している。

第16章は「推測法」で、この章題はJ. ベルヌーイの書籍名“*Ars conjectandi*” (Bernoulli, 1713)から取られているが、この題自体が、実は『論理学』のラテン語のタイトル“*Ars cogitandi*”に倣って付けられており、『論理学』の終わったところを引き継ぐものと位置付けられる(p.246)。ハッキングはこの書籍の出版で「確率の出現は完全に終了した(p.243)」とし、数学的には確率の極限定理を初めて示したことが何を意味するか、を論じている。

J. ベルヌーイの確率解釈は初めての主観論だと言われるが、頻度的だとも、カルナップの諸理論を予見するもの、すなわち論理説だとも言われるそうである(p.244)。ただ、ハッキング自体はJ. ベルヌーイは現代人と同じで、これらのいずれにも引き付けられていたと考えている。「推測法」に関しては、ホイヘンスの不明瞭な問題を解決し、「期待値」などに関する諸概念を整理したとされる。特にハッキングが目しているのは第4部で、この部分が『論理学』を引き継いでおり、確率を確実性の度合いとし、その確実性の度合いには「主観的なものと客観的なものの2種類がある(p.247)」とし、ハッキングは「間違いなくベルヌーイが「主観的」という言葉を確率論に取り入れた(p.249)」とも述べている。ところがもう一方で、「この場合の主観主義は個人的主観説や論理説の観点には当てはまらず、物理学者の説に似ている(p.249)」とし、その直前の記載からすると、どうやらJ. ベルヌーイにとっては「主観確率は実験によって実際調べることができる」ものであったらしい。この部分では現代までのさまざまな主観説や論理説と照らし合わせて論じられている。とはいえ、結論的には彼の場合は自然に両方の意味で用いていた、というのがハッキングの考えである。

この後でどのようにして証拠を組み合わせて認識に関する確率を推定するか、というJ. ベルヌーイのモデルに移り、純粋な証拠と混合した証拠の組み合わせによる推測が論じられる。

さらにもう1点、J. ベルヌーイは確実性の度合いとして非加法的な確率を考えていた点も注目されている。

第17章は極限定理自体に関する論である。この中で特に注目され論じられているのは帰納論理との関係である。繰り返し試行のできるチャンス生成装置で、任意の与えられた試行で成功する(S)ことについて、未知だが恒常的なチャンス p が存在し、 n 回試行が行われた時に観察された成功割合を s_n とする。この p の区間推定を行うに当たり、

$$\text{確率}(p \text{ は } s_n \pm \varepsilon \text{ に入っている} \mid s_n)$$

を計算することで、 p が $s_n \pm \varepsilon$ の区間に入っていることの信用可能性の尺度として用いることができ、これは頻度解釈からは引き出せないものの、リチャード・プライスによるとこの計算方法は帰納の問題の解決に本質的であるとされた。同時に、これはベイズの論文で解決されたものである。ただしJ. ベルヌーイの時代には帰納の問題はまだ哲学の中心問題として存在していなかったため、ベルヌーイはこの解法を考案しなかった、というのがハッキングの意見である。

この後、ライブニッツとJ. ベルヌーイの間で交わされた議論が紹介されるが、これもリスクの研究者には興味深いものである。ライブニッツは結果からのアポステリオリな推定では利用できる基礎確率集合がないことで苦悩しており、サイコロと壺の問題についてはそれがあるのは認めるものの、疾病や大気の変動についてもそれがあることには納得できないでいる。それに対してJ. ベルヌーイは両者の違いを「同等に不確かで不確定」という認識論的な回答(p.277)をおこなったが、後日の書簡では「人の身体に置き換え、その身体には壺に石が入っているように病気の素が入っているとしよう」という偶然に依拠する言い方で回答を行なっている。ライブニッツとJ. ベルヌーイのここでの議論はあまり知られていないと推測するが、現在でも素人が感じる、「病気をサイコロやくじで表されるような確率で捉えられるのか」という疑問と合致していて筆者(広田)には非常に興味深かった。

6. 科学哲学上でのその直後の展開(第18–19章)

さて、第17章で基本的には確率出現の物語は終了するが、なお2つの章がある。これらは直接的に確率と関わるものではなく、確率の出現のその後の影響という科学的な意味がより強い章である。すでに述べたが、第18章の「デザイン」は科学哲学での「時計仕掛けの宇宙論」、すなわち統計的に見られる安定性は神のデザインによるものである、という考え方である。具体的にはアーバスノットによる、出生の男女比が安定していることについて、それを「支配しているのは、チャンスではなくわざ(art)」とし、神の摂理によってそれが生じている、という考え方を紹介し、しかしニコラス・ベルヌーイが反対するように、その論は同じ年に出版されたJ. ベルヌーイの極限定理の帰結を知れば神のわざは必要ない、と述べている。ただ、確率の理解について、当時こういった風潮が英国を中心に極めて有力だったこともまた述べられ、ド・モアブルでさえその支持者だったという。第19章の「帰納」は本書全体の論理に関するまとめとして、ヒュームの帰納の懐疑を取り上げる。まず帰納での分析についてはライブニッツも、J. ベルヌーイもホップスも、パスカルも関心を持っていたことを指摘する。ただ、帰納に関する懐疑の問題は、基本的に証言によるものだった証拠の概念に、内的証拠という新しいものが生じることで初めて現れたことを指摘して終わる。ここでは本書全体で見てきたさまざまな変化を振り返る形で論をまとめている。

このように確率の出現に関わるJ. ベルヌーイやライブニッツのような人物も、偶然に依拠する確率ばかりではなく、むしろ自然に確率を認識論的に捉えていることからすると、確かに

確率に関しては完全に独立した形で確率解釈が存在するのではなく、そもそも probability には二元性がある、というハッキングの主張には強い説得力がある。

7. 「2006年版序論 確率的推論の考古学」

本書最後は、第2版で加えられた序論である(原著では冒頭)。「序論」での主張にはすでに何回か触れてきたが、ハッキングは初版後に現れた重要な関係した思潮(例えば Poovey, 1998(未邦訳)など)に触れつつ、初版に対する反響や批判に対して応えている。この部分は本書への評価や反論を伺うことが可能なため、本項ではその点を中心にまとめる。

「序論」冒頭は、フーコーによる影響を明白に述べ、この本が確率の哲学に関する流行を作り出し、また多くの関連書籍の出版に繋がり、本書は二元性という現象を解明する根源的な試みだった、と自負を覗かせる。確かに「序論」全体を読んでも「二元性は出現時から備わっていた」という主張自体はそれほど強い反駁には遭っていないように見える。

一方、多くの頁が割かれているのは「断絶」という主張に関するものである。ハッキングが主張するほどの大きな断絶が1660年の前後にあり、そこが本当に出現の時期なのかに関しては多くの疑問が呈されたようである。具体的には、ルネサンス以前にも原始の確率は存在することから、「ハッキングは誤っている」とまで明言した強い反論(Gerber and Zabell, 1979)や、歴史家イヴォ・シュナイダーによる古代の確率の存在の指摘がある(p.337)。また1660年以前の、古代から中世の弁論術・法学・医学等での蓋然性概念の発達を重視するフランクリン(原著はFranklin, 2001; 邦訳はフランクリン, 2018)に対してはやや長い反駁を行なっている(p.335)。また帰納の懐疑の問題について、ハッキングはヒュームの『人間本性論』から始まるとの見解に立っていたが、この点についてもかなり批判があったと見え(p.338)。「帰納の問題はヒュームに始まるものではない、と論じる批評家たち(p.340)」という記述もある。著者自身は、一つにはヒュームについてそれほど踏み込んだ説明を行なわなかったためもある、と考えているようである。

これらの指摘について、ハッキングは反論をそれぞれに行っている。確率に関しては過去に似たようなものがあったとしても1650年頃に広まった「網の目のような」確率の考え(ideas)とは似て非なるものであるという反論である。帰納の問題に関しても、帰納の分析の問題と、帰納の懐疑の問題の区別を示し、それらの反論で指摘されている帰納の問題や帰納の懐疑の問題はヒュームのものとは異なることを述べる。フランクリンの批判については、その前の法律家や医師などについての記載が本書に不足していた点は認めつつも、「なんでも屋は、多様な背景に基づく考えを、同じ目的をもち、すべてが現在の確率的推論へ向かう傾向として示しがち(p.336)」と強い調子で反論している。以上の論点について筆者(広田)は評価を行う力はないが、推測としてはこの点に関してはまだかなり議論があるものと推測される(実際、断絶に関するハッキング自身の考えも複雑で、彼は一方でクロンビーの連続主義的で進化的なアプローチに共感も示し、半ばは連続的な部分があることを認めてもいるように見える(pp.328-330))。

一方で、彼はこの断絶を(科学)革命とは言わない、とも明言している。I. B. コーエンが題の『出現』という表現を適切としたのを評価するものの、コーエンの「適用における(in application)科学革命」とする考え方は否定する。ハッキングは「社会思想と公衆衛生に確率という考えが適用されたこと」を(科学)革命とは考えず、それらは「ある考えを適用したものではなく」「絶対的に本質的なもの」とする(p.325)。とはいえ政府が多くの相対頻度(統計)を記録するようになったことが大きな社会変化を引き起こしたとし、特に生産や貿易だけでなく、むしろ「自殺、犯罪、売春、狂人の監禁といった『逸脱の公刊』」を第二次科学革命の引き金と考えており、逸脱の公刊が社会科学を発展させ、決定論を侵食し、パースの「偶然の宇宙に住んでいる」という発

言につながる、と述べている。

8. 意思決定やリスク論との関係

最後に、科学哲学の素人ながら意思決定やリスク論の研究者として本書を隅々まで読んで考えたことを述べたい。まず予想外にも、ハッキング自身がリスク論や意思決定を折々に意識していることが読み取れたのが驚きだった。すでに引用したが、『論理学』の中で雷鳴などに関する恐れを確率で評価し、それによって合理的に判断することを勧めているのはリスク論の原点である。期待値、あるいは期待効用を用いて意思決定をする枠組みは「パスカルの賭け」と共通のものであるので、その点はそれほど驚きではなかったが、それ以前に確率自体の出現過程とリスクが関わっているとは考えていなかった。ハッキングは「序論」でフーコーの影響をよりはっきり述べているが、そこで「(フーコーの)考古学とは現在を理解するために過去を組織化するもの(p.318)」と述べており、その「現在」の射程の一つが「リスク社会」だと推測される。「一般市民の不安は、確率を用いて果てしなく議論されている。原子炉の炉心融解、癌、強盗、地震の可能性、核の冬、エイズ、地球の温室効果、次はなんだろう。(中略)統計のおかげで、人々はウルリッヒ・ベックがリスク社会と呼んだものの中で生活している(p.318)」「パスカルという一つのオークの実から、リスク社会と呼ばれる怪しいオークの木が育った(p.323)」。これはその現れである。

また、確率出現の初期の歴史の中に、いわゆるハザードや疫病といったリスクに関わる事象が事例として頻繁に出てくることを改めて強く認識した。中世から近世の初めであることからキリスト教の信仰に関わる事例(奇跡や神の存在)が多いのは当然としても、パスカル以前にも、本書では前出の『論理学』の雷に対する確率を踏まえた意思決定や、すでに述べたようにフラカストロによる、伝染病に関する著作からの記述がある。グラントの死亡表がペストの流行がなければ人はどのくらい生きられるかに関する動機から作成されたのは有名だが、本書では人の死にチャンスの術を適用することに疑問を抱くライブニッツとJ.ベルヌーイとの間のやりとりが記載されている(p.277)。ライブニッツはサイコロと壺には基礎確率集合があるが、疾病や大気の変動にそれがあることに納得できず、J.ベルヌーイに問いかけたと言う。この辺りは、現代人でラプラス型の「ツボとサイコロ」で確率を学校で学び、だがあまり慣れていない人には共通する疑問なのではないか。しかし、たとえ基礎確率集合がない認識論的な確率であっても「両者の数—疾病の[種類の]数と壺の中の小石の数—は同等に不確かで、不確定である」(ベルヌーイの返信)という考えは、認識論的確率と偶然的確率がもともと同根であるとするなら少しは納得が得られるかもしれない。

このように、哲学的な確率解釈が多数ある中で、それらが完全に個々に分離したものだったのではなく、元来 probability にはそういった複数の意味性が混在していることを知ったのには蒙が拓かれた。本書で紹介している時期の人々の多くは、語の probability に偶然的確率と認識論的確率を同時に認識し、用いているが、それは認識論的確率が後から「発明」されたのではなく、本来的に両方を備えていたためであろう。現代においてリスクは確率で表現されることが一般的であるものの、確率の理解が人により異なることや、主観的確率も含まれるような複数の計算法、さらに場合によって計算結果が異なってしまうことへの違和感や疑問はそこでもしばしば提起される。だが、この事実を知ってそれが緩和された。計算方法が異なることや結果が異なることの問題はさておき、確率解釈が一意でないことは出現経緯からすると特段おかしなことではなく、むしろ自然だと言っても良い。このような確率解釈とリスク論に関わる問題意識も、実は「序論」を読むとハッキング自身が本書執筆の段階ですでに持っていたものと推測される。「その「現在」は、哲学者を悩ます概念、すなわち二面性を持つ概念としての確率に関

する物である。一面は事実に向けられたもので、ハンガリーの自殺者であれ坑道に置かれたガイガーカウンターのクリック音であれ、異なる種類の出来事が生起する相対頻度に向けられたものである。もう一面は、確信がもてない事物に対するあなたの信念の度合いに向けられたものである。(p.319)」そして、これに対してただ一つの確率の考えだけが正しく、有益で、科学的と主張する教条主義者の存在に触れ、この問題は「1975年の時点でそうだったように、2005年も同様である。予想では2035年もそうだろう。(p.320)」と悲観的な見通しを述べる。彼はこのような哲学的な問題が「どのように生じ、またなぜそれほど解決が難しそうなのかを示そうとした」とする。とすれば残念ながら、解決の見込みは立たなそうである。

現実場面でリスクコミュニケーションでは確率を伝達する場面があり、多くの領域でこの点に苦慮している。筆者は確率の表現による効果や地震予測の確率を題材に確率をどう伝えればわかりやすいか、またどのように認知されているかを検討していた時期がある(広田, 2011, 2015a, 2015b; Hirota and Oki, 2016)。世界的にも確率の伝達は極めて多数の研究があり、例えば心理学研究の中ではKahnemanが意思決定のプロスペクト理論の中で、非常に小さい確率は過大視されるが、そこから上はおおよそ過小視という確率の重みづけ関数を発表していることはよく知られる(Kahneman and Tversky, 1979)。また直接的に確率を言語的に伝達する場面を扱ったものとして、マックスプランク研究所のGigerenzerら(Gigerenzer et al., 2005)は5つの地域で「明日は30%の確率で雨」をどう理解しているかを尋ねたところ、正しく「明日のようないちいちのうちに」と理解している人は最も高い地域で65%、低いと20%程度に過ぎないことを明らかにした(他の選択肢は「その地域の30%で」「その時間帯の30%で」)。また別の研究では、乳がん検診の説明に関して、条件付き確率だと医師でさえ正しい理解ができないことも明らかにされている(Gigerenzer, 2015)。ちなみにこの著書は『確率の出現』が執筆のきっかけの一つだった、とあとがきに書かれている)。

確率伝達のレビューの結果から見ると(Visschers et al., 2008)、頻度による伝達の有効性が一部示された程度で、数値計算された確率を正確に伝達するのは非常に困難であり、多くの場合文脈依存であることも明らかになっている。ハッキング自身も、「序論」の中で、心理学者で行動意思決定の研究者Tverskyが「自身の研究結果で人間は誤った帰納判断をすることが示されていると論じた(p.336)」ことを紹介しており、悲観的に見える。「臨床医学と、エビデンス・ベースト医学の間の、百出するだけで進展のない議論をみてみよう。エビデンス・ベースト医学は[過去の症例の]頻度やランダム化した試験[に基づくもの]を意味する。一方、臨床医学は、整合的な信念の度合いの形成[に基づくもの]を意味する。またもやおなじみの二元性である。(p.348)」つまり、こういったずれは根源的に生ずるのだろう。

実生活でも、そういった例を耳にしたことがある。福島県で東日本大震災後にリスクコミュニケーションに携わった専門家は低線量の放射線による発ガンリスクについて確率を用いて説明した場面もあったようである。当然ながらそれは過去のデータに基づき、頻度論に基づくリスク計算である。それを聞くと、子供を持つ母親は大概このように聞き返すそうである。「…で、うちの子供は大丈夫でしょうか?」。筆者(広田)は、これを客観的に計算されるリスクの三人称性と、自分や家族のリスクを捉える際の二人称性の問題(これは野家, 2004のアイデアによる)と考えていたが、本書の内容から確率自体の二元性の問題である可能性も考えるようになった。もしそれが根源的であるなら、この問題は確率を伝えるフォーマットの改善のような表層的なやり方ではまったく解決されない可能性が高い。

今回『確率の出現』を読み直して最も筆者(広田)の印象に残ったのは、確率を使ったリスク評価それ自体が、『論理学』が示すように、合理的決定を志向するという背景や意思決定の枠組みと強く結びついているものだという点である。リスク論で扱われるリスクは、領域によって多少異なるものの、通常(さらされる確率) \times (結果の大きさ)という形で基本的には評価される(日

本リスク研究学会, 2019)。これを見ればわかるように、期待値計算に則したもので、源はパスカルの時代の合理的意思決定であると推測される。この場合の確率は、例えば10万人あたりの死亡者数のような統計を用いて計算され、これを相対評価して合理的意思決定を行うことは社会的には重要であろう。しかし、個人にとっては必ずしもそうではないのではないか。『論理学』が規範として勧めるように、リスクを合理的に判断するため確率を用いることは可能である。しかし、個人が生きていくときの指針は合理性ばかりではない。日常的には、合理的に生きていくことなど意識しない人は多数存在し、それは個人の選択である。確率の性質からしても、個人が繰り返して意思決定を行うのではなく1回の場合には、個人にとって結果は「当たり／はずれ」の2値であり、それを頻度論的な確率に基づいて判断することが個人にとって合理的と断言できるか、という疑問もある。巨視的に社会的観点では合理的だとしても、そもそも日常的に合理的かどうかを考えたことのない個人に、頻度的に計算されたリスクを提示して、こちらがより合理的な選択肢である、というのを理解してもらうこと自体に無理があるのではないか。リスクコミュニケーションを行う時に、確率自体の伝達可能性や受け手の統計的リテラシー以前に、その(厳密には社会的な)合理的意思決定の枠組みが共有されなければ、説明がわからないのは当然であることにあらためて気付かされた。ハッキングもこの件について、終わり部分にこのように書いている。「エビデンス・ベースト医学は勝利を収めるだろうが、それはよい[帰納的]推論ゆえではない。それは、国民健康保険の必要性と結びついた、ますます高額になる医療技術と薬学の成功ゆえである。医学の基礎を大規模な統計的規則性におけば、各症例を臨床的に細かく見るよりもはるかに費用が安い。(p.348)」この部分はまさに、リスク社会の考え方は大数の法則を利用した保険計算に基づく「集団全体での合理性」でしかなく、結果として個人は細かく診てもらえない、という個人にとっての不合理を示している。

確率の社会応用が進んだ現在、しかし現実には昨今の伝染病の感染モデルへの不信感でも明らかのように、社会応用する専門家側とそれを受け止める市民の間には大きなギャップが生じている。科学哲学の専門家だけの理解ではなく、確率あるいは確率モデルを用いる専門家の中でもこういった確率の広義や出現の背景についての理解が広まることで、社会的な合意形成の適切な形を考える一助になれば、というのが本書を訳し、この文を書いた著者(広田)の希望である。

謝 辞

本稿の素稿は椎名乾平先生(早稲田大学)及び森元良太先生(北海道医療大学)にお読み頂き、コメントや修正をいただいた。深く感謝する。むろん最終稿の誤り等の文責は著者である広田にある。

注.

- 1) イアン・ハッキング(Ian MacDougall Hacking)は本稿の執筆中に亡くなった(2023年5月10日)。ご冥福をお祈りする。ついに警咳に接することができなかったのは大変残念である。

参 考 文 献

- Arnauld, A. and Nicole, P. (1662). *La Logique ou l'art de penser*, Chez Guillaume Desprez, Paris. (アールノー, アントワース, ニコル, ピエール (2021). 『ポール・ロワイヤル論理学』(山田弘明, 小沢

- 明也 訳), 法政大学出版局, 東京.)
- 栗田賢三, 古在由重 (1979). 蓋然性, 『岩波哲学小辞典』, 岩波書店, 東京.
- Bernoulli, J. (1713). *Ars Conjectandi*, Thurnisorum, Basle.
- Bernstein, P. L. (1998). *Against the Gods: The Remarkable Story of Risk*, Wiley, New York. (ピーター, バーンスタイン (2001). 『リスク：神々への反逆』 (青山護 訳), 日本経済新聞社, 東京.)
- Byrne, E. F. (1968). *Probability and Opinion: A Study in the Medieval Presuppositions of Post-medieval Theories of Probability*, Springer, Hague.
- Derham, W. (1713). *Physico-Theology: or a Demonstration of the Being and Attributes of God from a Survey of Heavens*, W. and J. Innys, London.
- Franklin, J. (2001). *The Science of Conjecture: Evidence and Probability before Pascal*, The Johns Hopkins University Press, Baltimore. (フランクリン, ジェームズ (2018). 『「蓋然性」の探求 古代の推論術から確率論の誕生まで』 (南條郁子 訳), みすず書房, 東京.)
- Gerber, D. and Zabell, S. (1979). On the emergence of probability, *Archive for the History of Exact Science*, **21**, 33–53.
- Gigerenzer, G. (2015). *Calculated Risks: How to Know When Numbers Deceive You*, Simon and Schuster, New York.
- Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B. and Katsikopoulos, K. V. (2005). “A 30% chance of rain tomorrow”: How does the public understand probabilistic weather forecasts?, *Risk Analysis: An International Journal*, **25**(3), 623–629.
- Graunt, J. (1662). *Natural and Political Observations Mentioned in a Following Index, and Made upon the Bills of Mortality*, John Martyn, London. (ジョン, グラント (1941). 『死亡表に関する自然的及政治的諸観察』 (久留間鮫造 訳), 栗田書店, 東京.)
- Hacking, I. (1975). *The Emergence of Probability*, Cambridge University Press, Cambridge.
- Hacking, I. (1990). *The Taming of Chance*, Cambridge University Press, Cambridge. (ハッキング, イアン (1999). 『偶然を飼いなす—統計学と第二次科学革命』 (石原英樹, 重田園枝 訳), 木鐸舎, 東京.)
- Hacking, I. (2001). *An Introduction to Probability and Inductive Logic*, Cambridge University Press, Cambridge.
- Hacking, I. (2006). *The Emergence of Probability*, 2nd ed., Cambridge University Press, Cambridge. (ハッキング, イアン (2013). 『確率の出現』 (広田すみれ, 森元良太 訳), 慶應義塾大学出版会, 東京.)
- Hájek, A. (2019). Interpretations of probability, *The Stanford Encyclopedia of Philosophy Archives*, Fall 2019 Edition (ed. Edward N. Zalta), <https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>.
- 広田すみれ (2011). リスクコミュニケーションにおける確率を用いた不確実性伝達の心理学的課題, *心理学評論*, **54**(2), 153–167.
- 広田すみれ (2015a). 事故・災害生起確率の集団・時間表現によるリスク認知の違い, *社会心理学研究*, **30**(2), 121–131.
- 広田すみれ (2015b). 地震予測「n年にm%の確率」はどう認知されているか—極限法を用いた長期予測に対する怖さの閾値測定—, *日本心理学会第79回発表論文集*, p.198.
- 広田すみれ (2021). 心理学者が読んだ『確率の出現』と確率解釈の問題—リスク心理学研究の視点から—, 統計数理研究所共同研究集会 (通称: 「諸科学」集会) 2021年12月13日 (オンライン).
- Hirota, S. and Oki, S. (2016). Measurement of the thresholds of fear using methods of limits and effectiveness of colors in contrast of numbers in communicating probabilistic earthquake forecasting, *31th International Congress of Psychology*, p.169.
- Huygens, C. (1657). *Ratiociniis in aleae ludo*, *Exercitium Mathematicorum*, (ed. F. van Schooten), Lodewijk Elsevier, Amsterdam.

- 伊藤邦武 (1997). 『人間的な合理性の哲学 パスカルから現代まで』, 勁草書房, 東京.
- Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decisions under risk, *Econometrica*, **47**, 263–292.
- 上村勝彦 訳 (2002). 『マハーバーラタ：原典訳 1-7』, 筑摩書房, 東京.
- Leibniz, G. W. F. (1665). De conditionibus, Dissertation paper. (reprinted, 2005. Vrin, Paris.)
- Leibniz, G. W. F. (Ca. 1676). *Ad stateram juris de gradibus et probabilitatum*. (In Couturat, L. (1903). *Opuscles et Fragments Inédits de Leibniz: Extraits des Manuscrits de la Bibliothèque Royale de Hanovre*, 210–214, F. Alcan, Paris.)
- ライプニッツ, ゴットフリートヴィルヘルム (2019). 『ライプニッツ著作集』 (下村寅太郎, 山本信, 中村幸四郎, 原亨吉 監修, 澤口昭聿 訳), 工作舎, 東京.
- 日本リスク研究学会(編) (2019). 『リスク学事典』, 丸善, 東京.
- 野家啓一 (2004). 「二人称の科学」の可能性, 聖路加看護学会誌, **8**(1), 50–51.
- Poovey, M. (1998). *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*, 2nd ed., University of Chicago Press, Chicago.
- トドハンター, アイザック (2002). 『確率論史』 (安藤洋美 訳), 現代数学社, 京都. (Todhunter, I. (1865). *A History of the Mathematical Theories of Attraction and the Figure of the Earth from the Time of Newton to that of Laplace, Vol. 1*, Macmillan and Company, UK.)
- Visschers, V. H., Meertens, R. M., Passchier, W. F. and de Vries, N. K. (2008). Audiovisual risk communication unravelled: effects on gut feelings and cognitive processes, *Journal of Risk Research*, **11**(1-2), 207–221.
- Wilkins, J. (1638). *The Discovery of a World in the Moone, or, a Discourse tending to Prove that 'tis probable there may be another habitable world in that planet*, printed by E. G. for Michael Sparke and Edward Forrest, London.

A Short Review of Hacking’s *The Emergence of Probability* and
 Its Implications from the Perspective of a Researcher
 for Behavioral Decision Making

Sumire Hirota

Faculty of Informatics, Tokyo City University

This paper is a brief introduction to Ian Hacking’s (2006) *The Emergence of Probability* (2nd ed.), a book on the philosophy of science. The book has been a hot topic worldwide and is still quoted even today. It focuses on the emergence of probability before the work of Pascal. Hacking notes that the word “probability” in the Renaissance age meant “something worthy of approval” and that word categorized as “opinion(opinio)” in the medieval technical sense of this term. The meaning changed because the concept of evidence changed and “internal evidence” appeared. Given this prehistory, Hacking argues that probability has had both an epistemological aspect and an aleatory aspect from the time of its emergence. He also shows that probabilities were explicitly mentioned in the added last chapters of *The Port Royal Logic* and discusses the works of Huygens, Pascal, Leibniz, J. Hudde, Grant, and J. Bernoulli from the perspective of the implication of probability and inductive reasoning. The present paper presents a detailed description of the characteristic parts of *The Emergence of Probability* that address the concept of probability before the work of Pascal; the subsequent chapters are explained in order. In the last part, the implications of this book’s claims for risk research are discussed.

生態学におけるモデルと法則に関する 科学哲学論考の意義

島谷 健一郎[†]

(受付 2023 年 1 月 10 日；改訂 4 月 7 日；採択 5 月 11 日)

要 旨

生態学における科学哲学的問題を提起した論文で、科学哲学誌上で繰り返し議論された2つについて、主な論点と哲学論争の経緯を振り返り、現在の統計モデルを活用する生態研究における科学方法論的指針の布石を打つ。「集団生物学におけるモデル構築法」と題する論文では、集団生物モデルにおける一般性・現実性・正確性の間のトレードオフが主張された。用語の定義に曖昧さを含み命題として証明されていないなどの哲学批判を受けたが、そもそも論文の目的は集団生物モデリングの実践論だった。「生物群集に一般法則はあるか?」と題する論文は、そもそも生態学における法則とは何かという論争を招いた。いずれの論文についても、哲学誌上での議論は、一般性や法則などの言葉の定義に割かれがちだった。また、最近公表された科学哲学論文においても、情報量規準やベイズ統計などを駆使する生態研究に関する言及はほとんどない。それでも、何でも込みの複雑モデルと理想化された単純モデルのどちらが適切かのような、データの質・量や計算機の発展を経ても変わらない論点もあり、科学哲学的論考は、議論による成熟を経て研究現場へ恩恵をもたらしていると考えられる。

キーワード：生態学，一般性，法則，科学哲学，モデリング。

1. はじめに

統計学は予測や推定など、実用・応用目的でその価値や役割を議論されがちだが、仮説検証に代表される基礎科学目的でも重要な役割を担っている。とりわけ、野外環境における生き物を扱う生態学では、野外調査で得られるデータは、計測誤差に加え、個体差、1日の中の時間変動、季節変動、環境変動、地域差、地震や津波、山火事などの偶発的イベント等々、研究対象である生態系自体も多様な不確実性を含む。それらを確率分布で表現する統計モデルは、今日、生態データに基づく科学的推論では不可欠になっている。実際、和文誌においても日本生態学会誌や本統計数理誌で特集が組まれており(飯島, 2016; 久保, 2016)、その注目度が伺える。

生態学は、地球温暖化に伴う種の分布域の変遷や生物多様性消失など、環境問題に対峙する科学として、社会の期待感が高まっている。そうした社会的需要である応用・実用問題に貢献するには、種の分布域を制限する要因や種の共存機構を促すメカニズムの解明など、基礎科学としての成熟度を高めていくことも望まれる。すると、不確実さに富む研究対象に関する不確実さに富むデータに基づく科学方法論や推論形式など、科学哲学的な課題や葛藤に見舞われる。科学哲学の視点から生態学の歴史や現状を分析し、さらに、今日の大規模時空間データに

[†] 統計数理研究所：〒190-8562 東京都立川市緑町10-3

基づく推論と統計学の役割について、科学方法論を開拓・整備・確立していく必要がある。

生態学の専門学術誌には、古くから科学哲学的視点を含む論文が掲載され、生態研究者間で議論の俎上に上がっている。ただ、その多くはクーンやポパーなど、有名哲学者による古典的科学哲学書の引用にとどまる。しかし、哲学誌上で引用され科学哲学者により繰り返し議論されるに至った論文も、数は少ないが存在する。

本稿では、生態学における数理モデルの役割や意義について、哲学誌でも盛んに議論された生態学論文の事例を2つ取り上げる。そこでの議論の大半は数理生物学、特に偏微分方程式を用いる数理モデルに重点が置かれており、情報量規準やベイズ統計、機械学習など21世紀以降の主要な統計モデルを活用する生態学への言及はないに等しい(島谷, 2021; Shimatani, 2021)が、その萌芽は見え隠れしている。最終節では、統計モデルを駆使する今日の生態学研究における科学哲学的論考の役割について、筆者の見解を述べる。なお、現在の標準的科学論文の構成は「序>材料と方法>結果>考察」であり、本研究もこれに準じた流れで遂行した。それで、付録に、「序>材料と方法>結果>考察」の形式だった場合の要旨を付けた。

2. 数理モデルの一般性・現実性・正確性とその間のトレードオフ

「集団生物学におけるモデル構築法」と題する Levins (1966)は、生態学者のみならず科学哲学者からも繰り返し引用され議論の俎上に挙げられている。ここで集団生物学(population biology)とは、大きく集団遺伝学(population genetics, 特定種の集団内の遺伝子頻度の変動)と個体群生態学(population ecology, 特定種の集団の個体数の変動)から成る(後の Odenbaugh, 2003では、行動生態学や生物地理学も含めている)。Levins (1966)で革新的だったのは、一般性(generality), 現実性(realism), 正確性(precision)の3者にはトレードオフがあり、3者すべてを最大にする数理モデルを作ることはできない、という主張である。

一般性：その数理モデル(方程式)で表される法則性は、広く様々な現象を説明する。

現実性：現実と著しく乖離した仮定などは置かれておらず、現実に近いものになっている。

正確性：正確な予測を導き出せる。

トレードオフがあるため、数理モデルの構築では、このいずれかを犠牲にせざるを得ない。

1. 一般性を犠牲にする。

ある生態系や個体群などに特化した、例えば種や地域を限定したモデルにする。

2. 現実性を犠牲にする。

古典物理学における摩擦のない斜面や、理想気体などのように、非現実的な仮定を置く(理想化を行う)。現実と乖離していても、それらが互いにキャンセルするといった予測が得られることもある。また、モデル予測と現実の乖離からどのような複雑化が必要かが見えてきたりもし、より現実的なものに修正していくこともできる。

3. 正確性を犠牲にする。

定量的予測をあきらめ、定性的なパターン、例えば安定平衡解の存在、周期解の存在、相転移の発生、縞模様や渦巻きなど空間パターンの形成、あるいは変量間の正の相関や不等式などを数理モデルで予測し、現実と比較検証する。実際のところ、多くの数理生態学モデルはこれを選んでいく。

この論文は数理生態学に大きな影響を与え、その後の数理生態学の多くが定性的パターンに関する考察を主題にしてきた。Orzack (2012)によると Web of Science(2011年10月)で490回以上引用され(2023年1月9日の時点の Google Scholar では引用元数は2560)、その多くが Levins

の3分類の3つめに属することに言及していた。実際、定量的正確性を捨て定性的パターンに焦点を当てることで様々な非線形微分方程式の性質が明らかにされ、生態学を含む諸科学で効果を上げた。

今日、数理モデルを状態空間モデルの中のシステムモデルに用い、観察モデルを通して実データと統計的に適合させる研究が盛んである(飯島, 2016)。システムモデルに一般的かつ現実的な数理モデルを用い、システム・観察の両モデルに確率分布を入れることでデータに適合させる。しかし、両方に確率分布による不確実性が入ると、モデルがデータによく合っている、データはモデルを説明できる、と言われても、それが「正確な予測」であるとは直観的にも受け入れがたい。今日に至るも、正確性を犠牲にする研究が主流と言えるのかもしれない。

数理モデルを用いる集団生物学の方法論として繰り返し引用され参照されている Levins (1966)であるが、Orzack and Sober (1993)で科学哲学の視点で検証を受けると、その議論に多くの問題が指摘された。

Orzack and Sober (1993)が最初に取り上げた問題は、一般性・現実性・正確性の定義である。科学哲学では、言葉の定義が正確でないと議論が始まらないし、それが原因でかみ合わない議論が進行することはしばしば見られる。確かに、上にあげた3者の説明は、筆者による和文表現であり、用語の定義としては不完全である。ニュートンの運動方程式はリングの落下から太陽と地球の動きまで説明する「一般性」を有する、摩擦のない斜面は「現実性」を欠いている、など、3つの概念は直観的に納得いくものであるだけに、特に問題視されてこなかったであろう。

さらに、「トレードオフがあり3者を同時に最大にはできない」と主張するには、前提としてモデル間で一般性や現実性の大小を比較できないと困る。しかし、具体例を取って考えれば、モデル間の相対比較が容易には解決しない問題であることが見て取れる。

重回帰モデルを例に取り上げる。よりたくさんの変数を含む重回帰モデルの方が、より広範な現象に適用できる点で一般性は高いと考えるとする。また、どの変数も、程度の差はあれど目的変数に影響を与えうるので、変数はなるべくたくさん含める方が現実性も高いと考えることにする。さて、データが一つ与えられたとき、どの変数まで含めた重回帰モデルがもっとも予測力が高いか、これは、赤池情報量規準(AIC)などの情報量規準が普及した今日、あまりに少ない変数では予測力は低く、変数を増やすに連れて予測力は増していくが、増やしすぎると逆に予測力は落ちていくというモデル評価が多く見られる。後者は、まさしく一般性・現実性と正確性のトレードオフである。しかし、前者では、変数の増加とともに3者すべてが増加しており、トレードオフは働いていない(Orzack and Sober, 1993では情報量規準についての言及はないが、のちにSoberはAICについての哲学的考察を深めていく(Sober, 2008, 2015))。

つまり、3者のトレードオフはいつでも存在するわけではないか、あるいは、それ以前に3者の概念、特に一般性と現実性における大小関係の定義に問題がある。

ここには、それまでLevins (1966)を引用・参照してきた論文にない視点が2つ入っている。ひとつは、モデルの数式の中のパラメータを固定したときと未知(自由)な状態のときの区別である(Orzack and Sober, 1993はこれらをinstantiated-uninstantiatedと呼んだが、その後、この用語はあまり普及していない)。これは、重回帰モデルのようなネスト構造を有するモデル群について、部分的にパラメータを(0に)固定したモデル(残りのパラメータは自由)とそうでないモデルについて、前者のほうが一般性が高いと断言する、あるいは、ネストしているモデル群について一般性の比較に意味があるか、という問題も含有する。

今一つは、Levins (1966)を参照する多くの論文がひとつの数理モデルを取り上げ、それが3者のどれを犠牲にしているかを論じていたのに対し、Orzack and Sober (1993)は複数の数理モデルについて3者を相対比較するという視点である。

Levins (1966)は、さらに頑健性(robustness)、十分パラメータ(sufficient parameter)という概念も提唱している。この頑健性は、今日の統計学で使われる頑健性とは別の概念で、大雑把には、一つの生物学的仮説について、その数理モデル化(用いる数式、単純化のための仮定、など)によらず同じような結果がもたらされることを指し、Levins (1966)は「真実は独立な嘘の交わりにある」(P423)と書き表した。十分パラメータも十分統計量とは別概念で、個体群増殖率のように生残・成長・繁殖などをひとまとめに表現するパラメータを意味し、重力定数や摩擦係数などと対比させられる概念ととらえている。

こうした批判に対し、Levinsは同じ学術誌上のLevins (1993)で応えている。筆者から見た要点を単純に言ってしまうと、Levins (1966)は集団生物学のモデリング現場における指針であり、3者のトレードオフの存在という哲学の命題を証明する論文ではない。だからこの命題に根拠がないとか真でないという批判は意味がない、とだろう。確かに、一般性や現実性という概念を正確な定義なしに導入することに問題はある。しかし、そもそも論文の目的はモデル論という科学哲学における命題の証明ではなかった。

Odenbaugh (2003)は、科学哲学の立場からLevins (1966)を擁護する。「実用論的(pragmatic)」、「形式的(formal)」、「意味論(semantics)」などの哲学用語で論文の科学哲学における位置付けを行い、目的の違いを説明する。トレードオフを考慮せざるをえない背景として、生態系のような複雑な動態を偏微分方程式系による数理モデルで分析するには、(1)大量のパラメータを推定し、(2)数学として解き、(3)結果を意味付け解釈しないといけない。あまりにつまらないモデルについて、それより3者全てで優るモデルを作ることは可能である(トレードオフは働かない)。しかし、あるレベルを越すと、3者を同時に最大化することは上の(1)-(3)という需要を満たさないといけないためはや無理で、いずれかを犠牲にせざるをえない(P1504のFigure 1はこれらを図で説明している)。

Orzack (2005)はこれに対する反論だが、Odenbaugh (2003)が「OrzackとSoberの議論は失敗している」と書いた(P1496, 1497)のに呼応してか、「Odenbaugh (2003)の主張は集団生物学のモデリングに対し、何の有益な洞察ももたらさない」(P479, 484)という強い表現で始めている。ただ、集団生物学のモデリング現場に位置する立場にとって、議論は科学哲学としての言葉の使い方など末梢的な方向へ向かい、集団生物学の研究現場から遠ざかっていく印象しかもたらさなかったため、これ以上言及しない。

Levins (1966)は、2022年に*American Naturalist*という伝統ある学術誌で公表された「実証研究者のための理論生態学使い方ガイド」と題する論文(Grainger et al., 2022)や、「生態学と進化学において数理モデルについて肝に銘じておくべき9つの事項」と題する論文Joshi (2022)でも取り上げられている。Orzack and Sober (1993)が指摘するような用語の定義の曖昧さなどの問題はあっても、初学者が数理生態モデルを学ぶ時、一般性・現実性・正確性という概念は直観的でわかりやすく、2022年になっても、個別の現象に特化したモデルと根本原理から始まるモデルの間で揺れる現場の研究者、特に初学者には、有益な指針を与えている。

Levins (1966)の時代は、何でも込みの複雑モデル(原文はbrute-force approach)を扱える体制になっておらず、理想化された単純モデル(idealization approach)の方が有効と考えられがちだった。しかし、21世紀に入り、マルコフ連鎖モンテカルロ法と計算ベイズ統計や機械学習の発展により、何でも込みの複雑モデルも扱えるようになってきた。

そうした時代の変化を鑑み、Odenbaugh (2006)では、パラメータ推定以外は解決が見えてきており、Levins (1966)はもはや「あまりに悲観的だった」と結論している(P620)。

「40年後のLevins」と題するWeisberg (2006)は、Levins (1966)がモデリングの意義を「自然を理解し、予測し、上手く付き合う(原文はmodifyで、人の都合のいいように改変していくという意味にとれるが、現在の感覚からはこのような訳語が適切と判断した)」という記述は、「科

学とは説明, 予測, 統制(control)である」という科学哲学界で普通に言われる見解と対応するなど, 科学哲学から見て時代を先取りした視点が多数見られるという指摘から始まる(P625). 一般性・現実性・正確性や何でも込みの複雑モデル・理想化された単純モデルについても, Levens自身の記述より深い洞察を伴う解釈を与える. そして, 査読付き科学哲学誌に掲載された論文ではないため, 記述の不完全さについて科学哲学者から批判を受けてしまったが, Levens (1966)の主張は, 科学哲学の立場から, とりわけ実用(practice)視点で, 40年を経た今でも色褪せることなく読み込める. 願わくば, 自分の学生の学生に「80年後のLevens」を書いて欲しい, そのときは科学哲学も変貌しているだろうが, そこにはLevens (1966)に内在していた様々な科学哲学的発想が吸収されているだろう, と締めくくる(P644).

一方, Orzackは「モデリングの科学哲学, または生物学の科学哲学は何らかの使い道があるか?」と題する論文を哲学誌に出す(Orzack, 2012). そこでは, 「生物学者は哲学に注意を向けるべきか?生物学者は生物学の哲学に注意を向けるべきか?生物学者はモデルに関する生物学の哲学論文に注意を向けるべきか?」という3つの疑問から始め, 3番目についてはYesと断定する. Orzack and Sober (1993)からここに至る論争には, 科学研究先端に位置する研究者と科学哲学者の異分野交流の難しさと, しかし年月を経た後に収穫をもたらす期待を感じさせる.

Orzack and McLoone (2019)では, Breiman (2001)が「統計モデルには, データモデリング文化とアルゴリズムモデリング文化の2つがある」と論じた論文を引用し, そこへさらに「因果モデリング文化」を導入する(P83). データモデリング文化は情報量規準やベイズ推定など, アルゴリズムモデリング文化は機械学習に対応すると考えてよく, ようやく現在の統計モデルを用いる生物研究に言及されるようになった. しかし, 引用は18年前の論文であり, 「アルゴリズム文化(機械学習)に携わる統計研究者は全体の2%である」(P199)といった2023年の感覚と程遠い記述も見られる. 一方Levens (1966)については, 「モデルに対する擬哲学的(pseudo-philosophical)支援」と表現したり(P79), 総合科学誌に掲載されたにも関わらず引用の大半は進化学と生態学に偏ってきた事実を論じる(P79). そして, 集団生物学でなく, 細胞生物や分子生物も含めた生物学全般のモデリング, 特に生物学内でのモデル文化の多様性や違いについても言及しながら, (古典)物理学と生物学における「複雑さ」の捉え方の違いなども議論する.

3. 生物群集の生態的法則

生態学誌 *OIKOS* 誌上で公表された「生態学に一般法則はあるか?」と題する論文 Lawton (1999)では, 群集生態学ではいわゆる一般法則は存在しないと主張された. その根拠は, 群集生態学では山火事や洪水など, 偶発的な(contingent)イベントが群集動態に与える影響が大きい, というものである. 一方, より大きなスケールで考えるマクロ生態学では諸々の偶発的イベントの影響が互いにキャンセルし, 大域的法則性が発見できるかもしれないし, 逆に種を限定する個体群生態学でも対象に限られるため法則性を期待できる. 群集生態学はその中間ゆえ, 法則性は期待できない, というものである.

これに対し, 生態学者から, 賛同や, 賛同する面もあるがここまで言い切れないなど, いろいろな反響があった. *Biology and Philosophy* 誌で公表された Roughgarden (2009)は, 進化学のような一般理論が群集生態学には見られないという意見から始まる. 群集生態学では, 古くからニッチ理論(種間相互作用により種には固有の環境条件に生息し, それらは重なり合ってもいい)が一般理論の候補だが, 未だ一般理論に至っていない. そこで, 進化学が突然変異で始まる多様性生成(production)の理論であるのに対し, 群集生態学は多様性形成(formation)の理論を目指すべきと主張する. 進化学における突然変異と自然選択という, 多様性の供給と相

相互作用に関する進化学の概念に対応する群集生態学の概念として、散布や移動により群集に達する新種の加入(移入)と、群集内での種間相互作用(による多様性維持機構)を提唱する。ちなみに論文の結びは、「現在、群集生態学に一般理論はあるか? いいえ」、「あり得るか? はい」、「あるべきか? わからん」である。なお、Roughgarden (2009)は「群集生態学の一般理論は存在するか?」という Lawton (1999)と似たような題で Lawton (1999)も引用しているが、「共通点を積み上げていけば群集生態学の一般論になるかと問われると、大半の人はそう考えていない」という文脈での先行研究例のひとつという引用にとどまっている(P525-526)。

Linquist は一連の論文の中で Lawton (1999)を批判的に検証している。その中のひとつである Linquist et al. (2016)では、まず法則・法則性とは何かについて科学哲学的に整理する。法則が有する一般性は、普遍性、説明能力及び予測能力を特徴として持つ。さらに、法則は様々な場面で安定的に成り立つというとき、生態学では、不変性(invariant)と弾力性(resilient)に分けることを提唱する。そして、弾力性には3つの条件に沿った弾力性があると主張する。

分類ランク間の弾力性(taxonomically resilient)：ある法則が、様々な種、属、科、目を通して成り立つ。

生息地間の弾力性(habitat resilient)：森林、草原、湖沼、海洋、様々な生息環境で成り立つ。

空間スケールの弾力性(spatially resilient)：地域レベル、局所群集レベル、個体レベル、細胞レベル、分子レベルなど、空間スケールを問わず成り立つ。

その上で、公表された生態論文についてメタ解析を行い、個体群生態学やマクロ生態学と比べて群集生態学が弾力性について数的に劣っているかを比較したところ、特に有意な差は見られなかったと報告する。そして、そこで発見された法則として、

- 生息機の分断は送粉と植物の繁殖に負の影響を与える。
- 草食動物が増えると1次生産されるバイオマスは減る。

といった例を挙げている。そして、Lawton の論文の演題「生態学に一般法則はあるか?」に呼応して、「はい、生態学には弾力的一般化(法則)があります」という演題で答えている。

ただ、上記の例を見て「法則」と思う人は生態学界でも少ないのではないか? 法則と聞いて連想するのは、物理学でならニュートンの万有引力の法則や電磁気学におけるクーロンの法則などである。これらと比べるとはるかに矮小で、Linquist et al. (2016)は(弾力的という修飾語が伴うにせよ)法則というより単なる傾向でしかないものを列挙したに過ぎないように思えてくる。

Lean (2019)は、生態学ではニッチ理論や多様性-安定性仮説(多様な生態系ほど安定的である)といったものが法則に対応し、law-like, local law などの用語を持ち出しながら Linquist の事例は低レベル(lower-level)な法則の例である(P139)などを論じ、生態学に欠けているのは一般統一理論(general unificatory theory)であると主張する。

Lorenzano and Díaz (2020)では、法則とモデルと理論について、個体群動態を例に論じる。法則について、数式はそれぞれが特別な法則を表し、基本的法則・ガイド原理とは、生誕・死亡・移出入という基本デモグラフィーに環境など様々な要素を組み合わせていく指針を提供するものとしている。ただ、個体群動態を論じると明記していても、飯島(2016)にあるような最近の状態空間モデルについては触れていない。

なお、Levins (1966)が提起した一般性と、Lawton (1999)が提起した一般法則は、哲学誌上では別に議論されている。本節で取り上げた論文の中では唯一、Lorenzano and Díaz (2020)が Levins (1966)に1段落設けて言及しているが、その段落は「他の著者…」、次の段落は「…の話題に戻ると」で始まり、言及しないわけにいかないから差し込んだという印象を否めない。

4. 科学哲学論考は生態学に恩恵をもたらすか？

生態学の科学哲学を論じた2つの流れを紹介したが、いくつか共通してみられる点がある。

まず、科学哲学界で議論の遡上に上るのに、10年以上の時間遅れが見られる。Levins (1966)にOrzackとSoberが議論をもちかけたのも、コンピュータが身近になり、数理生態学の教科書も多数出版されるようになった1990年代のことである。

次に、10年遅れで哲学議論が始まっても、科学哲学として噛み合う議論に成熟するまで、新たに10年を要する。その間には、哲学者間による言葉の定義を争うような、研究現場から乖離した議論も挟まる。

そして、さらに10年を経て、科学哲学的論考は研究現場へフィードバックされ始める。実際、本稿で引用した論文に限っても、Odenbaugh (2006)やWeisberg (2006)、Orzack (2012)などは、生態学はじめ各専門分野の学術誌において、モデルの役割や限界などの文脈で引用されている。

ただ、こうした哲学系論文の生態誌における引用のされ方を見てみると、その多くはLevins (1966)を引用するとき、その理解を助けてくれる論文の例としてこれらを列挙しているに過ぎない。Evans et al. (2012)では、どういうモデルを使うかは哲学的スタンスに基づいて決められるという文脈でOrzack (2012)は引用されているに過ぎない(P164)。一方、Amano (2012)では、生態学のモデルはほぼすべて(分子や細胞などから個体に至るまで、などの)詳細を省いたブラックボックスを含んでいるが、ブラックボックスの存在自体が問題ではなく、その活用の仕方が重要であるという主張を補佐する先行研究としてOrzack (2012)を引用している(P501)(本稿では特に取り上げなかったが、Orzack, 2012ではブラックボックスを含むモデルについても言及している)。

残念ながら、今回調べられた範囲では、ほぼすべて哲学者が書いた主要な主張を外した引用と言わざるを得ない。ただ、この理由のひとつに、哲学を専門としていない研究者に哲学者の主張の主要な部分の大半は理解できない、という現実がある。実際、本稿で哲学者による論文を言及するところでは、「…について論じている」という書き方にせざるを得なかった場合も多く、その最大の理由は、「…」についてその哲学者が何を主張したのか、日本語で要約できるほどの理解に筆者が至らなかったからである。

哲学者による論文の生態学者による引用を見る限り、生態研究者の中で科学哲学論文を読んで直接自身の研究思考に活用している人は少ないと言わざるを得ない。しかし、引用回数だけからでも、哲学者の論文を読んでいる生態研究者が決して少なくないことはわかる。そして、そこで得た発想を議論や講演の場で発信したり論文の中で言及するなどして、科学哲学界の議論を拡散している。哲学者による議論は生態学者の中だけの議論とは異なる次元の思考を含み、生態学の中の哲学的問題を成熟させる。こうして、科学の業界は哲学論考の恩恵を間接的に受けていると考えられる。

たとえば今日、生物多様性消失や復元という実社会の問題を前に、現実に即応できる何でも込みのモデルか、多様性に関する基本法則を表現する理想化されたモデルか、今でも研究業界は2分されがちである。こうした2者択一論はわかりやすいが、単純化され過ぎるなどの落とし穴に陥りやすい。一方、Levins (1966)が唱えたのは3者とその間のトレードオフであり、何でも込みモデルと理想化モデルの対立はこの中の1断面でしかない。こんなモデリング思考が科学哲学的に精査を繰り返されたため、(科学的証拠を提示できるわけではないが)生態学の業界において安易な2分論は抑制されているのではないか。

問題提起から20年以上を経て扱うモデルやデータが大きく変貌を遂げている時代になってようやく、哲学的論考は成熟し、科学哲学論文はモデルを使う生態研究者に参照され、そのモ

デリング思考を助けるなど、生態学に貢献するに至る。Levins (1966)が提起したモデリング論は、哲学論争を経たからこそ40年を経た今なお有益な指針を与え続けられるようになった。Lawton (1999)はまだそこまで成熟していないが、もう10–20年の哲学論争を経れば、同様な期待を持てる。

統計モデルを駆使する現在の生態研究については、例えば主要な統計手法のひとつである情報量規準に関する哲学でさえ、Sober (2008, 2015)など論文や書籍こそ公表されているものの、統計学研究現場へのフィードバックを実感できない。それは、哲学論考がまだ成熟に至っていないからと考えられる。機械学習や生態学における状態空間モデルに関する哲学的考察は、それぞれ島谷 (2021), Shimatani (2021)などのように、始まったばかりである。これらが成熟期を迎える頃、機械学習も状態空間モデルも、現在とは異なる様相を示していることだろう。それでも今の哲学的発想のいくつかは科学哲学界で議論されることで成熟していき、その恩恵は20年後40年後の研究者にもたらされる。それには、議論が不毛に終わることを極度に恐れることなく、分野間の積極的なフィードバックが欠かせない。

付録 序>材料と方法>結果>考察 の形式だった場合の要旨

生態モデリングにおける科学哲学的問題を取り上げた科学哲学論文は、生態モデリングの発展に貢献しているか。

材料と方法

1. 生態モデリングにおける科学哲学的問題を提起した論文で哲学誌でも繰り返し引用された論文を探し適切と思われる論文を2つ以上選ぶ。
2. 議論された内容の流れを追い、共通する傾向の有無を探す。
3. 2で取り上げた哲学論文の中で最近の生態学論文で引用されているものを検索し、引用のされ方を調べることで科学哲学の生態学への貢献(フィードバック)を判断する。

結果

1. 生態学から提起された哲学的問題が哲学誌で議論されるまでにタイムラグがある。
2. 哲学者が議論を始めても生態学者との議論が噛み合うまでに新たなタイムラグがある。
3. それでもそうした科学哲学論文は生態学者に引用されるようになる。ただし、主要な部分でなく参考資料の例程度の引用が目立つ。

考察

生態学術誌における哲学論文の引用のされ方はマイナーなものが目立つが、引用されたという事実は生態学界に哲学論文を読む研究者が少なからずいることを示す。哲学論文を読むことでそこでの哲学思考の感化を受けているはずで、間接的ながら科学哲学者による議論は現在の生態学研究に正のフィードバックをもたらししていると考えられる。

参 考 文 献

- Amano, T. (2012). Unravelling the dynamics of organisms in a changing world using ecological modelling, *Ecological Research*, **27**, 495–507.
- Breiman, L. (2001). Statistical modeling: The two cultures, *Statistical Science*, **16**(3), 199–231.
- Evans, M. R., Norris, K. J. and Benton, T. G. (2012). Predictive ecology: Systems approaches, *Philosophical Transactions of the Royal Society B*, **367**, 163–169.
- Grainger, T. N., Senthilnathan, A., Ke, P.-J., Barbour, M. A., Jones, N. T., DeLong, J. P., Otto, S. P., O'connor, M. I., Coblenz, K. E., Goel, N., Sakarchi, J., Szojka, M. C., Levine, J. M.

- and Germain, R. M. (2022). An empiricist's guide to using ecological theory, *The American Naturalist*, **199**, 1–20.
- 飯島勇人 (2016). 生態学分野における状態空間モデルの利用：趣旨説明, *日本生態学会誌*, **66**(2), 337–338.
- Joshi, A. (2022). Nine things to keep in mind about mathematical modelling in ecology and evolution, *Journal of Biosciences*, **47**, 19.
- 久保拓弥 (2016). 「特集 生態学における統計モデリング」について, *統計数理*, **64**(1), 1–2.
- Lawton, J. H. (1999). Are there general laws in ecology?, *Oikos*, **84**, 177–192.
- Lean, C. H. (2019). General unificatory theories in community ecology, *Philosophical Topics*, **47**(1), 125–142.
- Levins, R. (1966). The strategy of model building in population biology, *American Scientist*, **54**, 421–431.
- Levins, R. (1993). A response to Orzack and Sober: Formal analysis and the fluidity of science, *The Quarterly Review of Biology*, **68**, 547–555.
- Linquist, S., Gregory, T. R., Elliott, T. A., Saylor, B., Kremer, S. C. and Cottenie, K. (2016). Yes! There are resilient generalizations (or “laws”) in ecology, *The Quarterly Review of Biology*, **91**(2), 1–13.
- Lorenzano, P. and Díaz, M. A. (2020). *Laws, Models, and Theories in Biology: A Unifying Interpretation. Life and Evolution: Latin American Essays on the History and Philosophy of Biology* (eds. L. Baravalle and L. Zaterka), 163–207, Springer Nature, Switzerland.
- Odenbaugh, J. (2003). Complex systems, trade-offs and mathematical modeling: Richard Levins. ‘Strategy of model building in population biology’ revisited, *Philosophy of Science*, **70**, 1496–1507.
- Odenbaugh, J. (2006). The strategy of “The strategy of model building in population biology”, *Biology and Philosophy*, **21**, 607–621.
- Orzack, S. H. (2005). Discussion: What, if anything, is “The strategy of model building in population biology?” A comment on Levins (1966) and Odenbaugh (2003), *Philosophy of Science*, **72**, 479–485.
- Orzack, S. H. (2012). The philosophy of modelling or does the philosophy of biology have any use?, *Philosophical transactions of the Royal Society of London Series B*, **367**, 170–180.
- Orzack, S. H. and McLoone, B. (2019). Modeling in biology: Looking backward and looking forward, *Studia Metodologiczne*, **39**, 73–98.
- Orzack, S. H. and Sober, E. (1993). A critical assessment at Levins’s The strategy of model building in population biology (1966), *The Quarterly Review of Biology*, **68**, 533–546.
- Roughgarden, J. (2009). Is there a general theory of community ecology?, *Biology and Philosophy*, **24**, 521–529.
- Shimatani, K. (2021). Philosophy of statistical sciences: The roles of mathematics and statistical models in estimation and other inductive inferences, *Annals of the Japan Association for Philosophy of Science*, **30**, 5–22.
- 島谷健一郎 (2021). 推定目的の統計モデルを用いる帰納推論における数学・演繹の位置付けと役割, *Linkage: Studies in Applied Philosophy of Science*, **1**, 32–41.
- Sober, E. (2008). *Evidence and Evolution: The Logic Behind the Science*, Cambridge University Press, U.K. (松王政浩 抄訳 (2012). 『科学と証拠：統計の哲学入門』, 名古屋大学出版会, 名古屋.)
- Sober, E. (2015). *Ockham’s Razors: A User’s Manual*, Cambridge University Press, U.K. (森元良太 訳 (2021). 『オッカムのかみそり：最節約性と統計学の哲学』, 勁草書房, 東京.)
- Weisberg, M. (2006). Forty years of ‘The strategy’: Levins on model building and idealization, *Biology and Philosophy*, **21**(5), 623–645.

Does Philosophical Discussion about Ecological Modelling and Laws Contribute to the Developments of Ecological Studies?

Ichiro Ken Shimatani

The Institute of Statistical Mathematics

The literature includes ecological papers that have examined philosophical issues in ecology and have been repeatedly cited in both ecological and philosophical journals. This paper discusses two examples. One example is the tradeoff among generality, realism, and precision in population biology modeling. This proposed tradeoff has been criticized by philosophers because of a lack of proof for the existence of a tradeoff relationship and because of unclear usage of the three terms, although the aim of the original paper was to discuss modeling strategies for population biologists. The second example is a paper that asked if there are general laws in ecology. The philosophical literature tends to focus on what the terms “general” and “law” indicate, and some of the discussions diverge from the concerns of ecologists. In addition, very few philosophers have mentioned recent ecological studies that use up-to-date statistical modeling. Even though some philosophical issues have been invariant despite rapid developments in data and computer technologies, philosophical discussions are expected to make ecological issues mature and then feedback to ecological research.

果樹豊凶の決定論的非線形予測

—アンサンブル再構成と Lorenz 類推法による1年先収量予測—

酒井 憲司[†]

(受付 2023 年 3 月 9 日 ; 改訂 7 月 18 日 ; 採択 7 月 20 日)

要 旨

決定論的カオスは軌道不安定性を有する低次元のダイナミクスから生成される。観測時系列からカオス性を判別することが非線形時系列解析(カオス時系列解析)の第1の目的である。その中でも、決定論的非線形予測はよく利用される手法の一つであるが、大きな時系列サイズが必要である。一方、作物の収量は年に1回しか計測できないため、取得できる収量時系列の大きさは非常に小さい。その反面、圃場や果樹園の作物個体数は大きいため、時系列集合のサイズは大きくなる。また、多くの果樹の繁殖様式が低次元の非線形ダイナミクスを有していると仮定できる。本稿では、果樹園から得られた収量時系列集合に対して行ったダイナミクスのアンサンブル再構成、決定論的非線形予測そして収量の1年先予測の実践例をウンシュウミカンとピスタチオについて紹介する。

キーワード：カオス同期、カオス集団力学、決定論的非線形予測、隔年結果、アンサンブル再構成、共通ノイズ同期。

1. はじめに

果樹や森林樹木の果実および堅果(ドングリ)の生産量は個体ごとに大きく変動する。さらに、これらの樹木個体群は種子・果実生産において豊凶同期現象を呈する。ある山全体のドングリが大豊作であると翌年は実をつけない。豊凶の周期は2年から数年、同期の空間的範囲は数kmから数百kmまで樹種によってマस्टィングと呼ばれる多彩な同期現象を示す(Koenig and Knopes, 2005)。柑橘類、ナッツ類、広葉果樹では豊凶が隔年交互に現れ(Sakai et al., 2008)、これを隔年結果と呼ぶ。樹木集団の同期メカニズムとしては、花粉による間接結合、共通ノイズそして生理融合による直接結合の3者が知られており、樹種と生育環境によって、これらが複合する。Isagi et al. (1997)は物質収支モデル(Resource Budget Model, 以下RBM)によって個体の豊凶メカニズムを一次元写像のカオス振動子としてモデル化し、個体群の豊凶同期(マस्टィング)を花粉結合による大域結合写像として定式化した。樹木豊凶同期は、個体間の相互作用の形態、RBMの制御パラメータおよび共通ノイズによって、同期強度・周期・空間分布などに多彩な挙動を呈する。植物種を超えた多様な同期現象を考察するにあたって、RBMは強力な理論的基盤を与えてくれている(Satake and Iwasa, 2000)。さらに、RBMパラメータの同定についても様々なアプローチが精力的に試みられている(Rees et al., 2002; Crone et al., 2005; 佐竹, 2007; Abe et al., 2016)。また、農業や林業への応用を目指したカオス農学の立場

[†] 東京農工大学 農学研究院：〒183-8509 東京都府中市幸町3-5-8

から (酒井, 1997; Sakai, 2001), ウンシュウミカン (Sakai et al., 2008), ケヤキ (Sakai et al., 2019), ピスタチオ (Sakai et al., 2022), スギ・ヒノキ (Ishibashi et al., 2019) などの多様な同期のメカニズムが, カオス解析・同期解析および RBM に基づいた数理モデルによって解明されてきた (酒井, 2023).

果樹の豊凶データは一個体ごとに年 1 点しか得られないため, その時系列サイズは極めて小さい. しかし, 幸いなことに, 果樹園や林分内の個体数は数十から数万にも及ぶため, 時系列集合のサイズを十分大きくすることが可能である. また, RBM によって多くの果樹の繁殖様式が低次元の非線形ダイナミクスを有していると仮定できることから, 時系列集合からのダイナミクス再構成が可能となる. 本稿では, 最もシンプルな決定論的非線形予測法である Lorenz 類推法によって, 数点という極小サイズの時系列に対して, 統計的手法に拠らず, 決定論的非線形予測を応用した 1 年先収量予測の実践例を示す.

2. 樹木豊凶の決定論的ダイナミクス

2.1 物質収支モデル(RBM)としてのテントマップ

Isagi et al. (1997) の RBM の模式図を図 1a に示した. 樹体内には, なんらかの物質 S を貯蔵するプールがある. 環境が一定であれば毎年 P_S がプールに供給される. $S + P_S$ がプールの容量 L_T を超えると, あふれた C_f によって開花数が定まる (図 1a 上段). L_T を超えなければ着花しない (図 1a 下段). 開花すると果実の結実のために C_a が消費される. この時, C_a/C_f をコスト比 R_C と呼ぶ. これを個体 i のサイズ M の個体群に拡張して (2.1)–(2.3) に定式化した. ここで, i は個体番号で $i = 1, 2, \dots, M$ である.

$$(2.1) \quad S_i(t+1) = \begin{cases} L_T - C_{a_i}(t) & S_i(t) + P_S > L_T \\ S_i(t) + P_S & S_i(t) + P_S \leq L_T \end{cases}$$

但し,

$$(2.2) \quad C_{a_i}(t) = R_C C_{f_i}(t)$$

$$(2.3) \quad C_{f_i}(t) = S_i(t) + P_S - L_T.$$

このダイナミクス $S_i(t) \rightarrow S_i(t+1)$ は図 1b のリターンマップに示されるように 1 次元写像である. これは区分線形のテントマップで左側と右側の傾きは, それぞれ 1 と $-R_C$ である. R_C を制御パラメータとする S の分岐図を図 1d に示した. 一方, 現実には S を観測することはできず, C_a を樹木個体の果実生産量として観測する. C_a のリターンマップと分岐図を図 1c と図 1e に示した. C_a リターンマップから明らかなように, $C_a(t) = 0$ の場合には $C_a(t+1)$ は一意に定まらない (図 1c).

2.2 同期モデル

2.2.1 花粉結合モデル

(2.1)–(2.3) の RBM モデルでは個体間の相互作用がなく個体はそれぞれに独立に振る舞うため同期は起らない. 自家不和合性の樹種は他の個体の花粉でないと受精できないため, 開花・結実に至らない. Isagi et al. (1997) は花粉結合による平均場を (2.4) で構成し, (2.2) を (2.5) で置き換えることによって同期モデルとして定式化した. $Y(t)$ は年ごとの集団全体の受粉成功率であり, β が大きいと受粉成功率は低くなり, 同期は強くなる.

$$(2.4) \quad Y(t) = \left[\frac{1}{(M-1)P_0} \sum_{j=1, j \neq i}^M C_{f_i}(t)(t) \right]^\beta$$

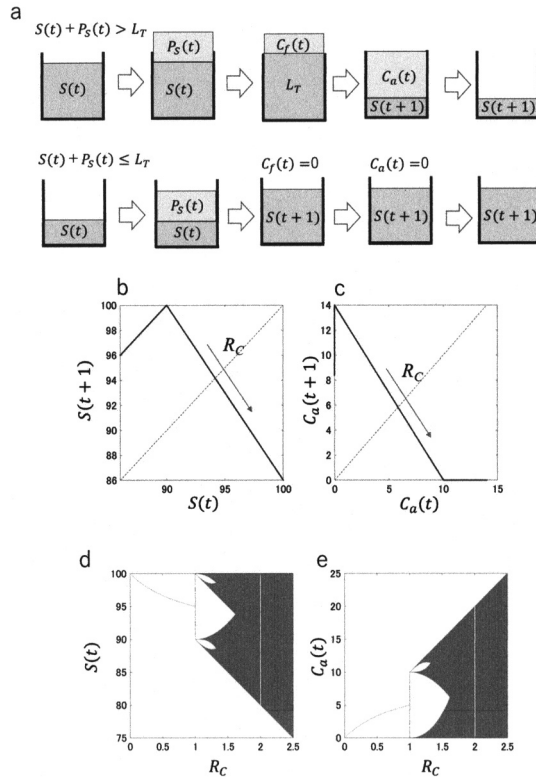


図 1. 物資収支モデル(RBM)の構造. a. 井鷲の物資収支モデル(RBM)の模式図, b. S のリターンマップ($R_C = 1.4$, $P_0 = 10$, $L_T = 100$), c. C_a のリターンマップ($R_C = 1.4$, $P_0 = 10$, $L_T = 100$), d. S の分岐図, e. C_a の分岐図.

$$(2.5) \quad C_{a_i}(t) = R_C C_f^i(t) Y(t)$$

(2.4)の P_0 は秋にプールに供給される物質の内的ポテンシャルであり, (2.1)–(2.3)式の RBM では供給量は毎年一定としているため, (2.1)および(2.3)式において $P_S = P_0$ である.

花粉の到達距離内での同期は大域結合写像 (Kaneko, 1990)として, それより広い範囲では局所結合写像としてモデル化されることになる. β が大きくなり完全同相同期になると, 接線分岐と境界衝突分岐を交互に繰り返す周期増加分岐となる (Sakai et al., 2019). 周期増加分岐では制御パラメータである R_C の増加に従って卓越する周期が1つずつ増加する. これに対して, 周期倍分岐では周期は偶数毎に増加する. 隔年結果やマステイングでは奇数年の周期も現れるため, RBMが周期増加分岐のダイナミクスを内包していることの重要性を強調しておきたい. なお, 不連続点を有しない単峰型の1次元写像とした場合には (Ye and Sakai, 2016; Esmaili et al., 2021), 周期増加分岐ではなく周期倍分岐となる.

2.2.2 共通ノイズモデル

ピスタチオなどの雌雄異株の樹種では, 雌株に対して十分な花粉を提供できるように果樹園内に雄株を配置する. そのため, 花粉結合による同期メカニズムを仮定できないので, ピスタチオの同期は共通ノイズ同期であると考えてきた (Lyles et al., 2009; Rosenstock et al., 2010; Esmaili et al., 2022; Sakai et al., 2022). 微気象変動などの外乱が果樹園内の各個体に同一に

作用すると仮定し(2.6)によって共通ノイズ $P_S(t)$ を導入する. $P_S(t)$ は t 年毎に変動するが, 常にすべての個体に同一に作用する.

(2.1)–(2.3)の P_S を(2.6)式で置き換えてRBMの共通ノイズ同期モデルを定式化した (Sakai et al., 2022).

$$(2.6) \quad P_S(t) = P_0\{1 + e_C\sigma(t)\}$$

ここで, $\sigma(t)$ は平均値 0, 標準偏差 1 の正規乱数であり, e_C が共通ノイズの強度である.

共通ノイズが印加された場合のリターンマップの幾何学的形状は図 1b のテント写像と変わらず, t 年毎の共通ノイズの大きさ $e_C\sigma(t)$ に応じて左右に平行移動する (Sakai et al., 2022).

2.2.3 拡散結合モデル

ピスタチオ果樹園における収量の空間分布に $1/3$ べき乗則をとまなう顕著な長距離相関の存在が明らかになり (Nobel et al., 2015, 2018), 統計物理の Ising モデルとのアナロジーから, 近接する個体間における相互作用の存在が示唆された. そこで筆者らは, 根系ネットワークによる個体間の物質交換 (Klein et al., 2016; Lev-Yadun, 2011; Simard et al., 1997; Gaion et al., 2018) を仮定して, (2.7)によって, 距離 $\gamma = 11$ m にある近傍個体間の拡散結合による結合写像格子 (Kaneko, 1985) として定式化した (Sakai et al., 2022).

$$(2.7) \quad SA_i(t) = S_i(t) + \frac{1}{NE(i)} \sum_{\substack{i \neq j \\ d(i,j) < \gamma}} \varepsilon_{j,i} [S_j(t) - S_i(t)]$$

ただし, $\varepsilon_{j,i}$ を樹木 i, j 間の拡散結合, $d(i, j)$ を樹木 i, j 間の距離, $NE(i)$ を樹木 i の γ 近傍の樹木の数, $SA_i(t)$ を着花前の i 個体内の貯蔵物質としている. (2.3)を(2.8)で置き換え(2.9)が得られる.

$$(2.8) \quad C_{f_i}(t) = \begin{cases} SA_i(t) + P_S(t) - L_T & SA_i(t) + P_S(t) > L_T \\ 0 & SA_i(t) + P_S(t) \leq L_T \end{cases}$$

$$(2.9) \quad S_i(t+1) = SA_i(t) + P_S(t) - C_{f_i}(t) - C_{a_i}(t)$$

ここで示した拡散結合を有するRBMのリターンマップは, 結合に起因する揺らぎがテントマップ上に現れるが, 骨格構造は変わらない (Sakai et al., 2022). (2.7)–(2.9)を追加することで, カオス振動子RBMの結合写像格子 (Kaneko, 1985)の共通ノイズ同期として定式化した. なお, (2.7)–(2.9)の結合形式では同相同期となるが, 拡散結合がおこる生活史上のタイミングが異なれば逆相同期にもなる (Prasad et al., 2017).

2.2.4 RBM 制御パラメータの空間勾配

RBMの制御パラメータは R_C であるが, ピスタチオにおいては受粉成功率 Z の空間分布を考慮する必要があった (Sakai et al., 2022). 果樹園の西端からの距離 l における受粉成功率 $Z(l)$ と, 新たな制御パラメータとしての作物係数 (Cropping coefficient) m_C を以下のように定める.

$$(2.10) \quad m_C = R_C Z(l)$$

(2.2)式を(2.11)式に置き換え, RBM制御パラメータ m_C を定式化した.

$$(2.11) \quad C_{a_i}(t) = m_C C_{f_i}(t)$$

なお,

$$(2.12) \quad Z(l) = \left(\frac{1}{R_C} + \alpha \frac{l}{WE} \right)$$

$$(2.13) \quad \alpha = 1 - \frac{1}{R_C}$$

(2.12)中の WE は果樹園の東西距離 (777m) である。空間勾配を α で表し、西端では $m_C = 1$ 、東端では $m_C = R_C$ となるように (2.12) と (2.13) を定めた。

3. 決定論的非線形予測

3.1 決定論的非線形時系列解析とその農業技術上の意義

決定論的カオス (以下、カオス) は低次元の決定論的ダイナミクスによって生成される一見複雑で不規則な変動である。カオスの軌道不安定性 (初期値への鋭敏な依存性) によって誤差が指数関数的に拡大するため、初期値への鋭敏な依存性を示し、往々にして実験における再現性が困難となる。変動の主要部分がカオスであれば、短期予測や制御の可能性があるので、観測時系列からカオスか否かを判別することが重要になる。その際、カオスの属性である決定性と軌道不安定性を利用するのであるが、カオスは十分条件が定義できないため、決定論的非線形予測、Lyapunov 指数、相関次元、推移誤差、Wayland テスト、False Nearest Neighbor 法、サロゲート法などからカオスの必要条件を多面的に調べ、現象に関する専門的知見と照合しながら、カオス性の判定をおこなう (池口 他, 2000)。これを非線形時系列解析もしくはカオス時系列解析 (池口・合原, 1997) と呼んでいる。農業現場は不規則で複雑な変動や再現性が困難な現象に囲まれている。にもかかわらず、伝承・蓄積されてきた農業技術や農業者の経験と技によってこれらをうまく取り扱いながら作物生産を行ってきた。変動現象にカオスが関与していれば、予測手法や制御手法の導入可能性が生まれ、農業技術への寄与が期待できる。これが、農業におけるカオス理論の意義でありカオス農学の目的でもある。

本稿では果樹収量の 1 年先予測手法の実践を紹介するが、まず 3 章では 1 本の観測時系列からのダイナミクスの再構成法と Lorenz 類推法による決定論的非線形予測手法を紹介した上で、4 章でそれを果樹収量の将来予測手法に拡張する。

3.2 時間遅れ埋込によるダイナミクスの再構成

農業現場で見られる複雑な変動現象の微分方程式や差分方程式が不明であることは一般である。しかし、「時間遅れ埋込法」が発明されたことで、1 つの観測時系列のみを用いて多次元のダイナミクスを再構成することが可能となり、カオス理論の現場応用に扉が開かれた。Takens の埋め込み定理は、得られた観測時系列 $x(t)$ からその時間遅れを成分とした m 次元ベクトル $\mathbf{x}(t) = \{x(t), x(t+\tau), \dots, x(t+(m-1)\tau), t = 1, 2, \dots, N\}$ によって描いた相軌道が、元のダイナミクスとトポロジカルに等価なダイナミクスを再構成していることを保証している (池口・合原, 1977)。実際の解析においては、時間遅れ τ は卓越する周期の 1/4 や自己相関が最初にゼロになる時刻などが経験的に用いられている。埋め込みの次元 m をダイナミクスの次元より大きく設定しないと誤った最近傍ベクトルを選択するため予測精度が大きく低下する。

3.3 Lorenz 類推法による決定論的非線形予測

$\{x(t), 1 \leq t \leq N\}$ が与えられたとき、 m 次元の相空間に予測器と予測対象をそれぞれ (3.1) と (3.2) で構成する。

$$(3.1) \quad \mathbf{X}(i) = \{x(i), x(i+\tau), \dots, x(i+(m-1)\tau), 1 \leq i \leq K\}$$

$$(3.2) \quad \mathbf{X}(j) = \{x(j), x(j+\tau), \dots, x(j+(m-1)\tau), K \leq j \leq N-p-(m-1)\tau\}$$

$K = N/2 - (m-1)\tau$ であれば時系列の前半分と後半分で予測器と予測対象になる。ここで、(3.2)の予測対象 $\mathbf{X}(j)$ の最近傍ベクトル $\mathbf{X}(k)$ を予測器(3.1)から探索する。そうすると、予測対象 $\mathbf{X}(j)$ の p ステップ先の予測ベクトルは $\hat{\mathbf{X}}(j+p) = \mathbf{X}(k+p)$ として算出される。また、 $j+p$ における実ベクトルは $\mathbf{X}(j+p)$ である。予測ベクトルと実ベクトルの第 m 成分はそれぞれ、 $\{\hat{x}(j+p+(m-1)\tau)\}$ および $\{x(j+p+(m-1)\tau)\}$ であり、これらを $K \leq j \leq N-p-(m-1)\tau$ に対して求めると、予測区間 $[K+p+(m-1)\tau, N]$ における予測値が得られたことになる。予測ステップ p の関数として x と \hat{x} の相関係数 $r(p)$ や正規化平均自乗誤差(Relative Root Mean Square of Error) $RRMSE(p)$ などを求める。

$$(3.3) \quad RRMSE(p) = \sqrt{\frac{\sum_{j=K}^{N-p-(m-1)\tau} \{x(j+p+(m-1)\tau) - \hat{x}(j+p+(m-1)\tau)\}^2}{\sum_{j=K}^{N-p-(m-1)\tau} \{x(j+p+(m-1)\tau) - \bar{x}\}^2}}$$

但し、

$$(3.4) \quad \bar{x} = \frac{1}{N-p-(m-1)\tau-K+1} \sum_{j=K}^{N-p-(m-1)\tau} x(j+p+(m-1)\tau)$$

ここで、 $RRMSE(p) = 0$ は完全な予測を示し、 $RRMSE(p) = 1$ は平均値以外全く予測できなかったことを意味する。sin 波、sin 波 + ホワイトノイズ、Lorenz63(カオス)、カラードノイズ、ホワイトノイズの決定論的非線形予測結果を相関係数(図 2a)と $RRMSE(p)$ (図 2b)で示した。そこで短期予測可能性と長期予測不能性を示し、決定性と軌道不安定性を確認したとしてカオスであることを主張するのである。例えば、sin の時系列では長期予測も完全に可能であり、逆に、一様乱数では短期予測も不能であるので当然ながらどちらもカオスではない。Lorenz63 とは、(3.5)式の Lorenz 方程式のパラメータを、 $\sigma = 10$, $r = 28$, $b = 8/3$ として求めた時系列であり、典型的なカオスとして知られている (Lorenz, 1963)。

$$(3.5) \quad \begin{aligned} \dot{x} &= \sigma(y-x) \\ \dot{y} &= -xz + rx - y \\ \dot{z} &= xy - bz \end{aligned}$$

Lorenz63 では短期予測は可能で、長期予測は不能である。これによってカオスの必要条件が確認できる。しかし、一様乱数を積分した有色ノイズでも短期予測可能性と長期予測不能性が示されており、このままではカオスと誤判定してしまう。これを回避するために、推移誤差、サロゲート法、正規化決定論的非線形予測等 (Sakai et al., 2019) の様々な手法が提案されている (池口 他, 2000)。

決定論的非線形予測では対象とする変動が定常であり、ダイナミクスのパラメータも変化しないと仮定して、決定性を検出することが目的である。そのため、予測器と予測対象の定め方は多様に可能である。よって、前半と後半を逆にして、(3.2)を予測器として(3.1)を予測対象としても差し支えない。しかし、本稿では果実収量の1年先予測への応用を意図しているので、現在およびできるだけ近い過去のデータセットのみから未来の収量を予測しなければならない。次節ではそのことを踏まえて予測器と予測対象の設定を工夫する。

4. 果樹収量予測のための決定論的非線形予測

4.1 ダイナミクスのアンサンブル再構成

1年生の作物(イネ、ムギなど)は収穫後には耕起されて更地になるので、年次間の収量変動

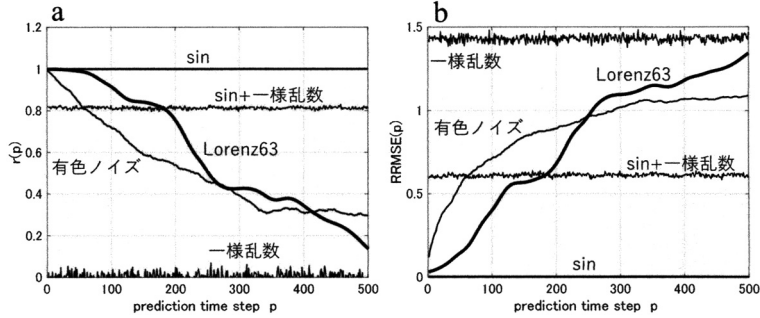


図 2. 決定論的非線形予測. a. 相関係数 $r(p)$ による評価, b. $RRMSE(p)$ による評価.

にダイナミクスの存在を仮定できないので本稿の考察対象としない. 一方, 果樹作物は図 1 の RBM で示したように, 現在までの収量と翌年の収量にダイナミクスの存在が仮定できる. 更に, 果樹園の各個体はクローンとして同一のダイナミクスを有すると考えてよいので, 果樹園の個体数と同じサイズの時系列集合を用いてダイナミクスの再構成(アンサンブル再構成)が以下のように可能となる.

個体数 M の果樹園で N 年間の収量を観測した場合, 時系列 $\mathbf{x}_i = \{x_i(t), t = 1, 2, \dots, N\}$ を要素とする集合 $\mathbf{A} = \{\mathbf{x}_i, i = 1, 2, \dots, M\}$ が得られる. 3.1 節で述べた時間遅れを $\tau = 1$ とする.

$$(4.1) \quad \mathbf{X}_i(t, m) = \{x_i(t), x_i(t+1), \dots, x_i(t+(m-1)), t = 1, 2, \dots, N-m+1\}$$

(4.1) で示した M 個の相軌道によってダイナミクスを m 次元相空間内に $(N-m+1) \times M$ のベクトル点で再構成する. これをダイナミクスのアンサンブル再構成と呼んでいる (Sakai et al., 2008).

4.2 果樹収量予測のための予測器構成法

アンサンブル再構成されたダイナミクスに対して Lorenz 類推法による決定論的非線形予測を試みる. 時間遅れ τ は 1 (年) であり, 果樹収量の将来予測を行うため, 予測器 L は予測を行う時点での現在年とその過去のみから構成する.

予測を行う対象年 t_N において p 年先予測を施して予測年 $t_N + p$ の収量を予測する. m 次元のダイナミクスであるとして, 対象年における対象ベクトル $\mathbf{X}_j(t_N, m)$ を m 年間の収量を成分として (4.2) のように構成する.

$$(4.2) \quad \mathbf{X}_j(t_N, m) = \{x_j(t_N - m + 1), x_j(t_N - m + 2), \dots, x_j(t_N)\}$$

なお, (4.1) においてはベクトルの時間表示 t を第 1 成分の時間表示と一致させていたが, (4.2) 以降ではこれを第 m 成分の時間表示 t に一致させている. この表記の方が予測年 $t_N + p$ と対象年 t_N , 予測値および実値の関係が分かりやすくなるためである.

次に, 予測器 $L(t_N, p, m)$ を設定する. 対象 t_N における各個体の相軌道を (4.3) で構成する.

$$(4.3) \quad \mathbf{x}_i(t_N) = \{x_i(1), \dots, x_i(t_N), i = 1, 2, \dots, M\}$$

これを要素として, 予測器 $L(t_N, p, m)$ を (4.4) で設定する.

$$(4.4) \quad L(t_N, p, m) = \{\mathbf{x}_i(t_N), i = 1, 2, \dots, M\}$$

予測器 $L(t_N, p, m)$ において, 対象ベクトル $\mathbf{X}_j(t_N, m)$ の最近傍ベクトル $\mathbf{X}_k(t_N, j, k, m)$ を (4.5)

として探索する.

$$(4.5) \quad \mathbf{X}_k(t_{N,j,k}, m) = \{x_k(t_{N,j,k} - m + 1), x_k(t_{N,j,k} - m + 2), \dots, x_k(t_{N,j,k})\}$$

但し, $m \leq t_{N,j,k} \leq t_N - p$. p 年先予測ベクトル $\hat{\mathbf{X}}_j(t_N + p, m)$ は次式となる.

$$(4.6) \quad \hat{\mathbf{X}}_j(t_N + p, m) = \mathbf{X}_k(t_{N,j,k} + p, m) \\ = \{x_k(t_{N,j,k} + p - m + 1), x_k(t_{N,j,k} + p - m + 2), \dots, x_k(t_{N,j,k} + p)\}$$

(4.7) によって予測値 $\hat{x}_j(t_N + p)$ は, $\mathbf{X}_k(t_{N,j,k} + p, m)$ の第 m 成分 $x_k(t_{N,j,k} + p)$ となる.

$$(4.7) \quad \hat{x}_j(t_N + p) = x_k(t_{N,j,k} + p)$$

なお, 個体 j の $t_N + p$ 年の収量の実値は $x_j(t_N + p)$ である.

この操作を M 個の個体を実施し, 予測年 $t_N + p$ において M 個の実値 $\{x_j(t_N + p), j = 1, 2, \dots, M\}$ と予測値 $\{\hat{x}_j(t_N + p), j = 1, 2, \dots, M\}$ とから相関係数 $r(t_N + p, p)$ と正規化平均自乗誤差 $RRMSE(t_N + p, p)$ を計算する. さらにこれを, $m + p \leq t_N \leq N - p$ における対象年 t_N に対して実施し求めた予測値と実値によって相関係数 $r(p)$ と $RRMSE(p)$ を求め p 年先予測の精度を評価する.

4.3 RBM モデルを用いた共通ノイズの影響評価

決定論非線形予測では時系列に定常性を前提としている. しかし, 実データにおいては厳密な定常性は期待できない. よって, 実データ解析の立場からは, 予測精度が目的に応じて満たされているか否かで妥当性を判断することになる.

果樹個体群の同期現象としての隔年結果は花粉結合と共通ノイズの2者が大きく, 根系ネットワークによる拡散結合もこれに加わる. 共通ノイズの印加によって定常性が失われるし, 大域結合写像や結合写像格子も厳密には大規模な自由度を持つため, 低次元の決定論的カオスであるとする仮定を逸脱している. しかし, この仮定が実用上許容できるかどうかは, 実データへの適用結果で判断すればよい.

5章の実践例に先立って, 共通ノイズの影響を RBM 時系列集合のモデルによって見積もっておく. RBM は図 1b に示すように貯蔵物質 S の 1 次元写像であるので, $m = 1$ として決定論的非線形予測は可能である. しかし, これまでのところ S を測定することは困難で間接的な推定が試みられているが (Rees et al., 2002; Crone et al., 2005), 直接かつ簡便に観測できるのは個体の果実収量 C_a である. 図 1c に示すように, C_a のリターンマップでは, $C_a(t) = 0$ において $C_a(t + 1)$ が一意に定まらない. よって, 少なくとも $m = 2$ 以上としなければならない.

制御パラメータ $RC = 1.9$ について決定論的非線形予測を実施した. ここでは 100 個体について初期値から 180 回計算を繰り返し, 181~210 までの計算値でサイズ $N = 30$ の時系列集合を得た. 初期値を $[L_T - P_0 RC, L_T]$ の一様乱数で与えている. それぞれの個体が全くばらばらに不規則に振る舞うことに留意してほしい. わかりやすいように 100 個体の内の 4 個体を図 3a に示した. 決定論的非線形予測は図 3b のように高い短期予測可能性と長期予測不可能性が明瞭である. 図 3c に短期予測 ($p = 1$) における予測値 (中空) と実値 (中実) を $t = 3 - 30$ において比較した. ほとんどの予測値は実値と重なっている. 図 3d に実値 vs. 予測値プロットを示した. $r(1) = 1.0$ および $RRMSE(1) = 0.150$ とほとんど完璧な予測が得られている. このように, 100 個の RBM のカオス振動子は個体毎に独立に振る舞うが, 全個体のダイナミクスは同一で, ノイズもないため極めて高い短期予測が実現できている. 図 3e に長期予測 ($p = 20$) における予測値 (中空) と実値 (中実) を $t = 21 - 30$ において比較した. ほとんど予測値と実値は重ならない. 図 3f に実値 vs. 予測値プロットを示した. $r(20) = 0.0717$ および $RRMSE(1) = 1.35$ と

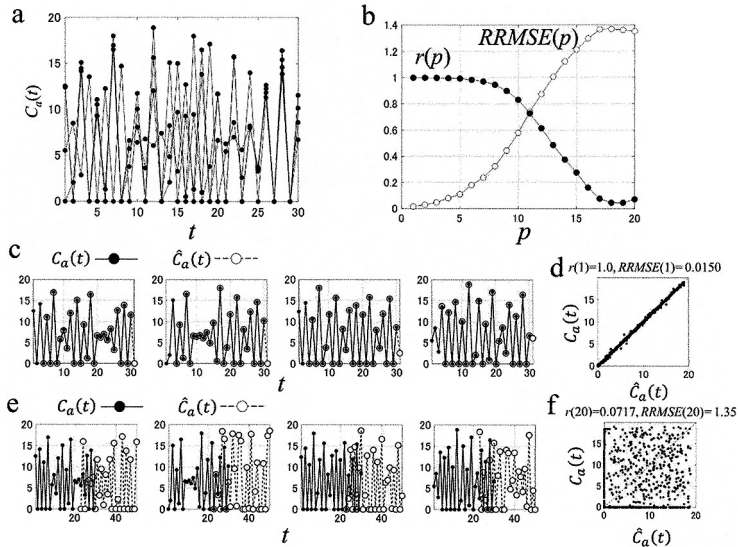


図 3. 共通ノイズが無い場合の RBM 集合の決定論的非線形予測. a. $R_C = 1.9$ における 100 個体のうち 4 個体の収量の時系列を表示 ($t = 1 - 30$). b. 決定論的非線形予測の精度を相関係数 $r(p)$ と $RRMSE(p)$ で表示. c. 短期予測 ($p = 1$) における実値 (中実) および予測値 (中空) 時系列 ($t = 3 - 30$). 100 個体中 4 個体を例示し, $t = 31$ では予測値 (中空) のみプロット. d. 短期予測 ($p = 1$) における実値対予測値プロット ($t = 3 - 30$). e. 短期予測 ($p = 1$) における実値 (中実) および予測値 (中空) 時系列 ($t = 21 - 30$). 100 個体中 4 個体を例示し, $t = 31 - 50$ では予測値 (中空) のみプロット. f. 長期予測 ($p = 20$) における実値対予測値プロット ($t = 21 - 30$). RBM のすべての個体に対して $P_0 = 10$, $L_T = 100$, $e_C = 0$ とした.

カオスの軌道不安定性によって長期予測不能となっている. なお, ここでは共通ノイズが無い場合であるので, (2.6) における共通ノイズの強度は $e_C = 0$ として計算した.

5 章で紹介するウンシュウミカンやピスタチオの実データでは, 共通ノイズが外生的ノイズとして印加されているため定常性が確保できているとは言えない. そこで, RBM 集合の短期予測に対する共通ノイズの影響を調べておく. 制御パラメータは $R_C = 1.4$, (2.6) における共通ノイズの強度は $e_C = 0.2$ として $N = 30$ の時系列 $M = 100$ 個によって時系列集合を構成した. 予測器は予測対象の 1 年以上過去のデータしか用いない. 共通ノイズは年ごとに変動するダイナミカルノイズなので予測精度の低下は当然であるが, その程度が実用上の関心事となる. 図 4a のように収量予測では相関係数 $r = 0.700$, $RRMSE = 0.826$ であり, 図 4b の予測と実値の時系列においても, 共通ノイズによって収量の短期予測精度が低下している. 一方, 図 4c に示す収量差分の相関係数は $r(1) = 0.921$, $RRMSE(1) = 0.438$ と高い予測精度を示し, 図 4d の収量差分時系列の実値と予測値比較からも 1 年先予測もおおよそ成功している. このように, 共通ノイズ $e_C = 0.2$ を印加した場合の収量については共通ノイズ無し ($e_C = 0$; 図 3d) に比較して予測精度は低下したが, $RRMSE(1) = 0.826$ と 1 よりも小さい値に収まっている. さらに, 差分収量の短期予測精度は $r(1) = 0.912$, $RRMSE(1) = 0.438$ と十分高い. 以上から, この程度の共通ノイズの場合には, 果樹栽培の実用上許容できる予測が期待できる. 本節での 1 年先予測手法を実証するために, ウンシュウミカンとピスタチオの実データに対して 5 章で実践例を解説する.

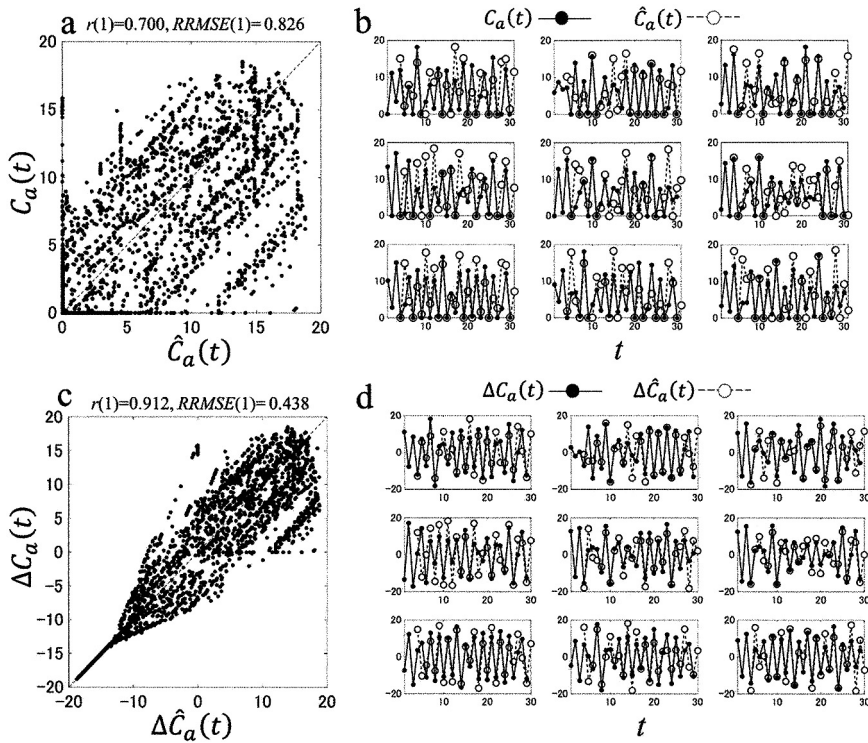


図 4. 共通ノイズが印加された RBM 集合の決定論的非線形予測 ($p = 1$ の短期予測). a. 短期予測 ($p = 1$) における収量の実値 (C_a) 対予測値 (\hat{C}_a) プロット. b. 短期予測 ($p = 1$) における収量の実値 (中実) および予測値 (中空) 時系列 ($t = 3 - 30$). 100 個体中の 9 個体を例示. $t = 31$ では予測値 (中空) のみプロット. c. 短期予測 ($p = 1$) における収量差分の実値 (ΔC_a) 対予測値 ($\Delta \hat{C}_a$) プロット. d. 短期予測 ($p = 1$) における収量差分の実値 (中実) および予測値時系列 (中空). 100 個体中の 9 個体を例示. $t = 31$ では予測値 (中空) のみプロット. RBM のすべての個体に対して $R_C = 1.4$, $P_0 = 10$, $L_T = 100$, $e_C = 0.2$.

5. 果樹の決定論的収量予測

5.1 ウンシュウミカン

ウンシュウミカンの観測データには個体ごとの変異や外乱, さらに比較的弱い共通ノイズ同期が想定される. ウンシュウミカンは単為結実のため, 受粉する必要がなく花粉結合同期を起こすことはないが, 共通ノイズと考えられる弱い同相同期を示す.

図 5 のようにウンシュウミカンの 48 個体 7 年間の収量データ (個体毎の収穫果実数) に対して 1 年先予測を以下のように実施する. なお, 1998 から 2004 年はそれぞれ $t = 1$ から $t = 7$ に対応させている. 1 年先予測であるから予測ステップ $p = 1$ で次元 $m = 2$ である. 個体 j において予測の対象年 $t_N = 3$ として, 予測年 $t_{N+1} = 4$ の収量の予測値 $\hat{x}_j(4)$ を求めよう. 予測の対象ベクトル $\mathbf{x}_j(3) = \{x_j(2), x_j(3)\}$ の最近傍ベクトル $\mathbf{x}_k(2) = \{x_k(1), x_k(2)\}$ を予測器 $L(3, 1, 2)$ から得たとする. (4.5) から $t_{N,j,k} = 2$ である. (4.7) から $\hat{x}_j(4) = x_k(3)$ となるので, 個体 j の予測年 4 の収量の予測値は $\hat{x}_j(4) = x_k(3)$ となる. 即ち, 2000 年 ($t = 3$) より過去の収量データのみから 1 年先 2001 年 ($t = 4$) の収量予測を行ったことになる. 図 5a で示したように

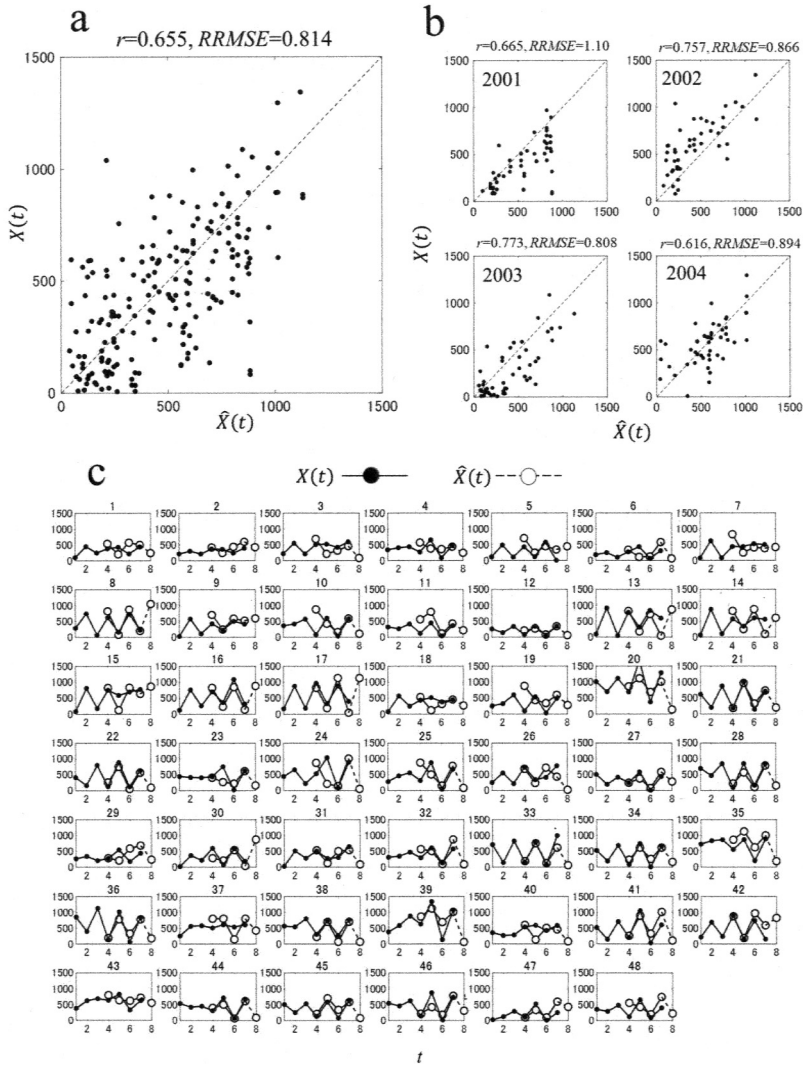


図 5. 48 個体のウンシュウミカン個体群の決定性検出-収量時系列集合に対する Lorenz 類推法の適用において最近傍 1 点のみ使用. a. 短期予測 ($p = 1$) における収量実値対予測値プロット (2001–2004). b. 年毎の短期予測 ($p = 1$) における収量実値対予測値プロット. c. 個体毎の収量実値 $x(t)$ (中実) および 1 年先予測値 $\hat{x}(t)$ (中空) の時系列比較. 横軸の $t = 1$ から 7 はそれぞれ 1998 年から 2004 年に対応する.

2001 から 2004 年 ($t = 4 - 7$) まで通年の予測精度は $r(1) = 0.665$, $RRMSE(1) = 0.814$ である. 図 5b は 2001 年から 2004 年まで年ごとの予測精度を示した. 図 5c に個体毎に実収量 (中実) と 1 年先予測収量 (中空) を比較した. なお, 予測器 $L(t_N, 1, 2)$ が用意している探索可能ベクトル点数は $(t_N - m) \times M$ なので, 2001 年から 2004 年までそれぞれ $M = 48$ から $4M = 192$ 点である. 2001 年のみ $RRMSE(1) = 1.10 > 1$ であり収量の変動予測は失敗している. これは, 探索可能ベクトル点数が $M = 48$ と最も小さいことが要因と考えられる. なお, 2005 年 ($t = 8$) の収量は測定されていないので予測値のみを中空で示してある.

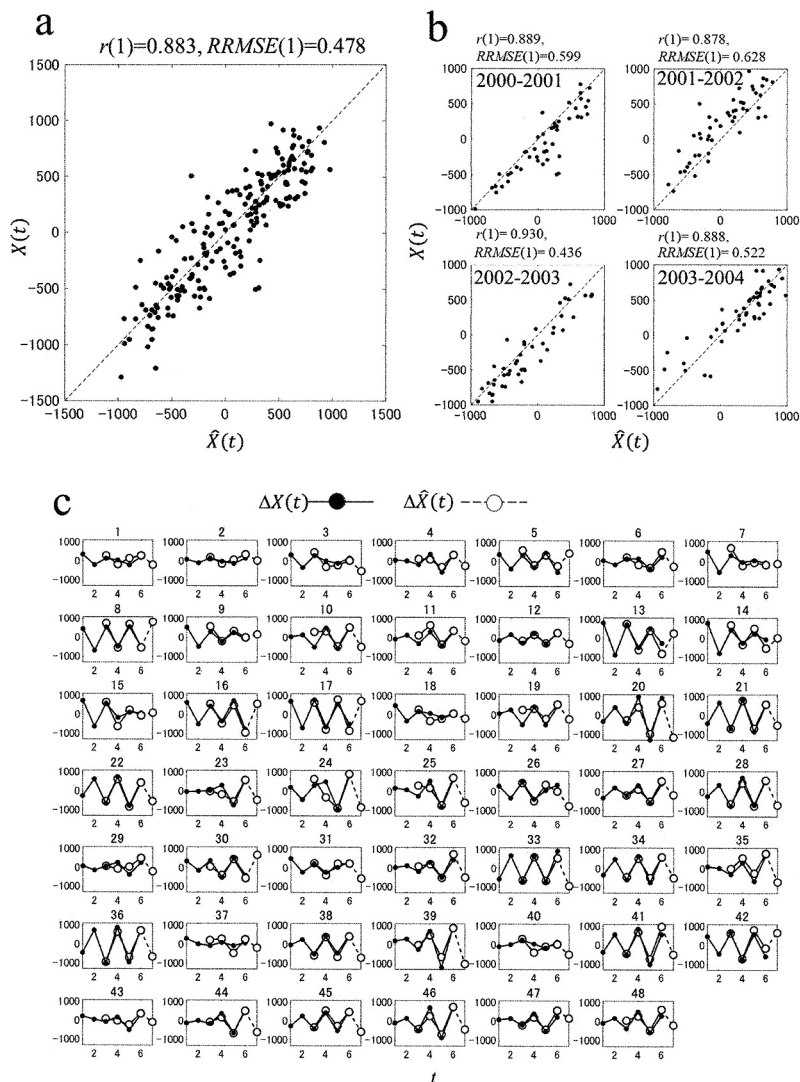


図 6. 48 個体のウンシュウミカン個体群の決定性検出-収量差分時系列集合に対する Lorenz 類推法の適用において最近傍 1 点のみ使用. a. 短期予測 ($p = 1$) における収量差分の実値対予測値プロット (2001–2004). b. 年毎の短期予測 ($p = 1$) における収量差分の実値対予測値プロット. c. 個体毎の収量差分実値 $\Delta x(t)$ (中実) および 1 年先予測値 $\Delta \hat{x}(t)$ (中空) の時系列を比較. 横軸の 1 から 6 はそれぞれ [2003–2001] [2001–2002] [2002–2003] [2003–2004] の期間に対応する.

また、実データの決定論的非線形予測においてカオスの誤判別があった場合、差分値の予測精度が非常に低くなることが指摘されている (池口 他, 2000). 誤判別の有無を確認するために、収量差分の予測精度を図 6 に示した. ここで、収量差分 $\Delta x(t+1) = x(t+1) - x(t)$ に対して収量差分予測値は $\Delta \hat{x}(t+1) = \hat{x}(t+1) - x(t)$ で定義する. 図 6a のように通年で相関係数 $r(1) = 0.883$ および $RRMSE(1) = 0.478$ と高い予測精度が得られた. また [2000–2001],

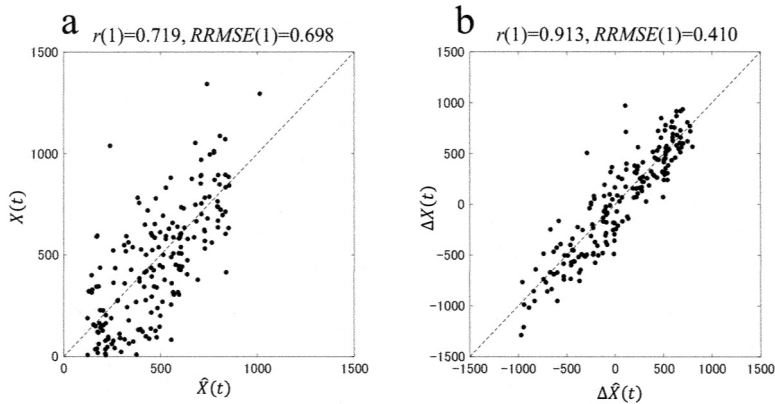


図 7. 48 個体のウンシュウミカン個体群の決定性検出-収量時系列集合に対する Lorenz 類推法の適用において最近傍 10 点を使用し精度向上を図った。a. 収量の実値対予測値プロット, b. 収量差分の実値対予測値プロット。

[2001–2002], [2002–2003] および [2003–2004] においてもそれぞれ予測精度は高い。

図 5 および図 6 で示したように、最近傍 1 点のみでもある程度の予測精度が得られており、図 4 で予め調べておいた RBM 共通ノイズモデルにおける短期予測結果と合致している。このことは、ウンシュウミカンの隔年結果現象は決定論的ダイナミクス(内生要因)が第 1 義的であって、環境要因(外生要因)は 2 次的要因であることを示唆している。

Lorenz 類推法の改良として複数の近傍ベクトルに対して、予測対象ベクトルとの距離に応じた重み平均を用いたり、ヤコビ行列の推定などが試みられている。ここでは 10 個の近傍ベクトルを用いて、単純に算術平均して予測ベクトルを推定し精度向上を試みた。収量と収量差分についての予測結果を図 7a および図 7b に示した。収量については、相関係数 $r(1) = 0.719$ および $RRMSE(1) = 0.698$ と図 5 に比較して予測精度は向上している。また、収量差分についても、 $r(1) = 0.913$ および $RRMSE(1) = 0.410$ と図 6 に比較して精度が向上している。このように、1 年先予測など現場応用を目指す場合には、近傍点数や近傍の設定など試行して経験的に予測精度の向上を図ることになる。

5.2 ピスタチオ

ピスタチオは換金作物としてカリフォルニア州での栽培が盛んである。強い同相同期を呈し、州全体の収穫量からも 2 年周期の隔年結果が顕著に観察できる (Rosenstock et al., 2010, 2011; Noble et al., 2018)。ピスタチオは雌雄異株であり、雌株に十分な花粉が供給されるように雄株が一定間隔で配置されている。当該の果樹園は東西 777 m, 南北 416 m で 9,526 本の個体が $5.2\text{ m} \times 6.4\text{ m}$ で配置されている。雄株は 26 m 間隔で配置され、すべての雌株は雄株から 14 m 以内に存在し十分な花粉が供給されている。そのため、花粉結合による同期ではなく、これまで共通ノイズ同期としてモデル化してきた。ところが、収量の空間分布に強い距離相関の存在が確認された。Ising モデルによるアナロジーから、当該果樹園において近接する個体間に相互作用、例えば、根系ネットワークによる物質交換の存在が示唆された。これを受けて、2.2.3 項で示したように、RBM に(2.7)–(2.9)式を追加することで、RBM の結合写像格子 (Kaneko, 1985) における共通ノイズ同期モデルを提案した (Sakai et al., 2022)。また、(2.10)–(2.13)式によって RBM の制御パラメータの空間勾配 α も含んだ 4 つのパラメータ ($R_C, e_C, \varepsilon, \alpha$) 推定は、同相同期強度 (Prasad et al., 2017; Sakai et al., 2019) を評価関数としたパラメータスタディに

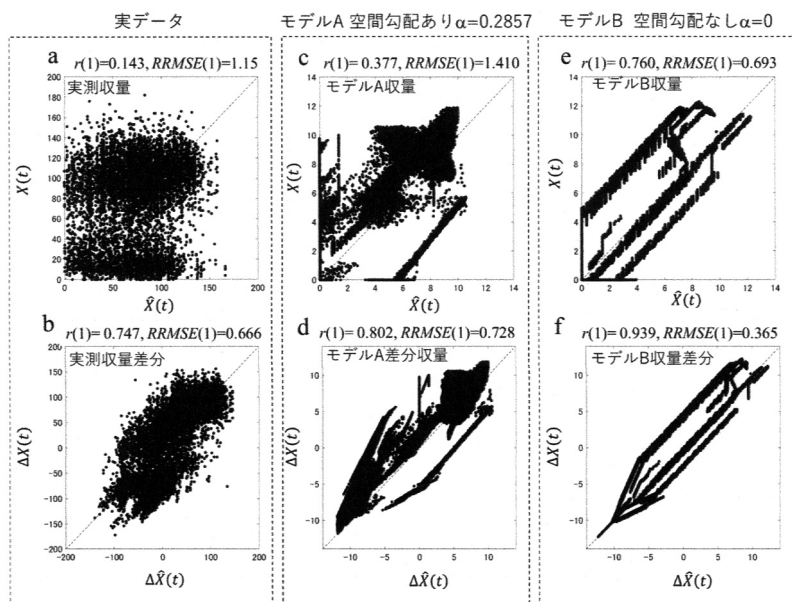


図 8. ピスタチオ実データとモデル A およびモデル B に対する 1 年先予測. 実データ収量の時系列集合とモデル A およびモデル B の時系列集合に対する Lorenz 類推法による 1 年先収量予測を行った. モデル A は (2.7)–(2.11) 式で与えられ, パラメータ $(R_C, e_C, \varepsilon, \alpha) = (1.4, 0.2, 0.05, 0.2857)$ では空間勾配 α を含んでいる. モデル B は (2.7)–(2.9) 式で与えられ, パラメータ $(R_C, e_C, \varepsilon, \alpha) = (1.4, 0.2, 0.05, 0)$ は空間勾配 α を含まない. a. 実データ収量の実値対予測値プロット, $r(1) = 0.143$, $RRMSE(1) = 1.15$. b. 実データ収量差分の実値対予測値プロット, $r(1) = 0.747$, $RRMSE(1) = 0.666$. c. モデル A 収量の実値対予測値プロット, $r(1) = 0.377$, $RRMSE(1) = 1.41$. d. モデル A 収量差分の実値対予測値プロット, $r(1) = 0.802$, $RRMSE(1) = 0.728$. e. モデル B 収量の実値対予測値プロット, $r(1) = 0.760$, $RRMSE(1) = 0.693$. f. モデル B 収量差分の実値対予測値プロット, $r(1) = 0.939$, $RRMSE(1) = 0.365$. 実データは (Sakai et al., 2022) の Appendix で Excel ファイルとして公開しており, 単位は kg/樹木個体である. $M = 9,526$ 本であるが欠損値を含んでいる. モデル A およびモデル B においては, すべての $M = 9,526$ の個体に対して $P_0 = 10$, $L_T = 100$ として数値実験を行い, モデルパラメータ同定における評価関数は収量ではなく, 同期強度の空間分布によって行った.

よった. 空間勾配を考慮したモデル A : $(R_C, e_C, \varepsilon, \alpha) = (1.4, 0.2, 0.05, 0.2857)$ と考慮しないモデル B : $(R_C, e_C, \varepsilon, \alpha) = (1.4, 0.2, 0.05, 0)$ による数値実験によってそれぞれの時系列集合を得た. 実測時系列集合とモデル A およびモデル B の時系列集合に対して, Lorenz 類推法による 1 年先収量予測を実施した (図 8).

当該果樹園においては $M = 9,526$ 個体の収量データ (kg/個体) を 6 年間取得したが, 栽培管理上排除した個体などがあり, 6 年間欠損なく観測できたのは 3,127 個体であったので, 観測時系列集合のサイズは, $N = 6$ (年), $M = 3,127$ (本) である. 一方, モデル A とモデル B の数値実験における時系列集合のサイズは, $N = 10$ (年), $M = 9,526$ (本) である. (4.1) 式で得られた再構成ダイナミクスは, $(N - m + 1) \times M$ 個のベクトル点から成っている.

実測収量の短期予測精度は $r(1) = 0.143$ および $RRMSE(1) = 1.15$ であり予測はできていな

い(図 8a)．一方、実測収量差分の短期予測精度は $r(1) = 0.747$ および $RRMSE(1) = 0.666$ であり(図 8b)、 $RRMSE(1)$ が 1.0 よりもある程度小さいため、年次間の増減については予測が出来ている。

実データ(収量)に対する同期強度を評価関数としたパラメータ推定の結果、空間勾配 α を組み込んだモデル A : $(R_C, e_C, \varepsilon, \alpha) = (1.4, 0.2, 0.05, 0.2857)$ が適当とされたが、モデル A の収量の予測精度は $r(1) = 0.377$ および $RRMSE(1) = 1.41$ となり予測不能であった(図 8c)．一方、収量差分に関しては、 $r(1) = 0.802$ および $RRMSE(1) = 0.728$ とある程度の予測精度を示した(図 8d)．これらの結果は図 8a-b の実データの場合と同様であった。

一般的な空間勾配が無い果樹園を想定しているモデル B : $(R_C, e_C, \varepsilon, \alpha) = (1.4, 0.2, 0.05, 0)$ においては、収量の短期予測精度は $r(1) = 0.760$ および $RRMSE(1) = 0.693$ と、ある程度の予測が出来たといえる(図 8e)．収量差分の短期予測では $r(1) = 0.939$ および $RRMSE(1) = 0.365$ と高い予測精度を示した(図 8f)．これらは、図 7 で示した 48 個体のウンシュウミカンの短期予測と同等の精度である。

以上の結果から、制御パラメータの空間勾配 α の存在は、当該果樹園もしくは当該地域に特異的な属性と考えられる．図 8e-f で示したモデル B の予測結果は、一般的なピスタチオ果樹園では短期予測が可能であることを示唆しており、今後、現場応用を通じての検証が期待される。

ウンシュウミカンやピスタチオ等の樹木作物では、周期 2 年の隔年結果が栽培上問題となる．一方、気象データなど外生的な変数を用いず、過去の収量データのみから 1 年先の差分値を予測できていることは、隔年結果が内生的なダイナミクスに因る変動が外生要因に優越していることを示している．また、栽培現場では、もちろん高い精度での収量予測が望ましいが、個体毎に翌年の増減が収量差分として高精度で予測できる提案手法は、栽培技術・栽培計画・資材調達など営農現場において有益である。

5.3 アンサンブル再構成法による収量予測技術の課題

バタフライ効果で知られる Lorenz はポアンカレと並ぶカオス理論の生みの親と言えよう(Lorenz, 1963)．Lorenz 類推法は最初の決定論的非線形予測法として紹介されているが、彼が気象データの予測に用いたところ期待した予測精度が得られなかったとされる(池口 他, 2000)．しかし、彼の提案した手法は、本稿で示したようにカンキツ類やナッツ類などの豊凶予測には極めて有効である．また、1 ステップ($p = 1$)とは最短時間単位であるが、農作業栽培の観点からは $p = 1$ 年は十分に長期であるので営農の観点からは将来予測が出来ている．果樹収量時系列集合からのダイナミクスのアンサンブル再構成が有効であれば、決定論的非線形予測だけでなく、決定論的非線形時系列解析の諸手法の活用が期待できるが、個体レベルおよびそれよりマクロまたミクロな諸現象においての課題を下記に示す。

5.3.1 個体レベル

多数の集合時系列を収集・蓄積すれば、集合サイズ(個体数)や時系列サイズ(観測年数)の拡大によって予測器の性能向上を図ることができる．気象データなど外生要因をモデルに組み込む工夫をすれば、予測性能を向上させることができる．ダイナミクスが再構成できれば、本稿で紹介した決定論的非線形予測の他に、相関次元、Lyapunov 指数、推移誤差等の諸量を求めることができる、そしてこれらが、果樹園栽培の新たな評価変数になるかもしれない．また、再構成ダイナミクスへのカオス制御(Sakai and Noguchi, 2009)は、熟練農業者による吟味を経て、摘花、摘果、剪定、接ぎ木、肥培管理などの栽培技術に寄与できるかもしれない。

5.3.2 マクロレベル

個体レベルのRBMダイナミクスはカオス振動子でもある。カオス同期 (Pecora and Carroll, 1990, 2015; Rosenblum et al., 1996)によって上位の階層にもダイナミクスが創発しうる。現に、ウンシュウミカンでは、個体→果樹園→集荷組織→市町村→府県→全国という少なくとも6つの過程において隔年結果が存在しているのであるが、個体から集荷組織レベルまではRBMのもつ内生的要因と外生的要因(気象)という自然要素に拠っていると考えてよい。一方、農林業センサスからも明らかのように、全国のウンシュウミカン収穫量は明確な2年周期の変動を示し、年次間の変動量は10%から20%程度で、ほとんど確実に「成り年」と「不成り年」が交互に繰り返される。そして、当該業界ではこれがあたかも自然の法則のように受け止められている。しかし、個体の平均収量から類推して約1億本のウンシュウミカンが栽培されているが、日本列島の西半分全域で共通のノイズ印加が毎年実現しているとは到底考えられない (Sviridova and Sakai, 2016)。見込み予測行動による創発という行動経済学的仮説についての検証を試みているところであるが、そのためには、6つの過程のどこかで機能しているスイッチング行動を実データから探索しなければならない。いずれにせよ、全国収穫量の2年周期の隔年結果は事実であり、このメカニズムの解明は大変興味深い研究課題である。

5.3.3 ミクロレベル

個体レベルより下位の階層においてRBMを引き起こすメカニズムが当然に存在しているはずである。開花遺伝子の発現 (Shalom et al., 2012)、花芽形成ホルモン等の寄与が無ければRBMは振動子として作動できない。よって、この機能を担う物質とそのメカニズムの解明は大変興味深いテーマである (Satake and Kelly, 2021)、それとともに、2.2.3項の拡散結合モデルとして紹介したように、根系ネットワークによる物質交換が明らかに樹木豊凶に影響を与えている (Sakai et al., 2022)。また、牧草地は1年生の牧草からなっているが、収穫後に耕起し翌春に播種というのではなく、収穫後に残された地下茎ネットワーク(ルートマット)によって再生される。そのため、ダイナミクスの存在を仮定することは的外れではない (Tilman and Wedin, 1991)。

6. カオスと農業

農業機械の研究者にとっては、播種・耕うん・肥培管理また病害虫・野生動物防除など通じて密度依存性は身近な概念である。密度依存性をモデル化したLogistic写像 (May, 1976)の周期倍分岐のカオスも農業現場での経験と平仄があう。トラクタ非線形動力学の要素過程であるBouncing Ball DynamicsはLogistic写像と全く等価な分岐構造(周期倍分岐)を持っている (Watanabe and Sakai, 2019)。本稿で紹介した隔年結果の数理は昇圧型DC-DC変換器の数理と完全に等価で (Banerjee et al., 2000; 酒井, 2023)、周期増加分岐のカオスルートを示す。これらは単なる偶然の一致なのであろうか。低次元の非線形ダイナミクスは農業のいたるところに転がっている (酒井, 1993, 1995, 1997, 2023; Sakai, 2001)。農業は必然的に人間が見て触れる量の変化を扱う。そこでは、再現性と予測および類推可能性を備えたものでなければ農業技術としては成立していない。科学革命後の産業技術基盤となった線形理論建設の舞台裏で、様々な非線形現象への関心も着実に育まれていたのであろうし、ようやくポアンカレが3体問題として可視化してくれたともいえる (高橋 他, 1998)。その後さらに1世紀を要したが、カオス、非線形力学、同期理論、グラフ、ネットワーク、粒子系のモデリングと解析、そしてリモートセンシングなどによるオブザーバの収集能力、さらに手元のパソコンの計算パワーによって実証可能な技術基盤を農業技術として構築できる道具立てを近代科学はようやく揃えてくれたように思う。本稿で提案したアンサンブル再構成法は果樹栽培に限らず、カオス集団力学の決定

論的時系列解析法としての一般性を有するように定式化できている。農業とカオスは親和性が高いのである。

謝 辞

文科省科学研究費 20K21347 の補助を受けた。

参 考 文 献

- Abe, T., Tachiki, Y., Kon, H., Nagasaka, A., Onodera, K., Minamino, K., Han, Q. and Satake, A. (2016). Parameterisation and validation of a resource budget model for masting using spatiotemporal flowering data of individual trees, *Ecology Letters*, **19**, 1129–1139.
- Banerjee, S., Karthik, M. S., Yuan, G. and Yorke, J. A. (2000). Bifurcations in one-dimensional piecewise smooth maps-theory and applications in switching circuits, *IEEE Transactions on Circuits and Systems I*, **47**, 389–394.
- Crone, E. E., Polansky, L. and Lesica, P. (2005). Empirical model of pollen limitation, resource acquisition, and mast seeding by a bee-pollinated wildflower, *American Naturalist*, **166**, 396–408.
- Esmaili, S., Hastings, A., Abbott, K., Machta, J. and Nareddy, V. R. (2021). Density dependent resource budget model for alternate bearing, *Journal of Theoretical Biology*, **509**, 110498.
- Esmaili, S., Hastings, A., Abbott, K. C., Machta, J. and Nareddy, V. R. (2022). Noise-induced versus intrinsic oscillation in ecological systems, *Ecology Letters*, **25**, 814–827.
- Gaion, L. A. and Carvalho, R. F. (2018). Long-distance signaling: What grafting has revealed?, *Journal of Plant Growth Regulation*, **37**, 694–704.
- 池口徹, 合原一幸 (1997). 力学系の埋め込み定理と時系列データからのアトラクタ再構成(〈特集〉力学系理論-応用数理における新しい展開), *応用数理*, **7**(4), 260–270.
- 池口徹, 小室元政, 山田泰司 (2000). 『カオス時系列解析の基礎と応用』(合原一幸 編), 産業図書, 東京.
- Isagi, Y., Sugimura, K., Sumida, A. and Ito, H. (1997). How does masting happen and synchronize?, *Journal of Theoretical Biology*, **187**, 231–239.
- Ishibashi, A. and Sakai, K. (2019). Dispersal of allergenic pollen from *Cryptomeria japonica* and *Chamaecyparis obtusa*: Characteristic annual fluctuation patterns caused by intermittent phase synchronisations, *Scientific Reports*, **9**, 11479.
- Kaneko, K. (1985). Spatiotemporal intermittency in coupled map lattices, *Progress of Theoretical Physics*, **74**(5), 1033–1044.
- Kaneko, K. (1990). Globally coupled chaos violates the law of large numbers but not the central-limit theorem, *Physical Review Letters*, **65**, 1391–1394.
- Klein, T., Siegwolf, R. T. W. and Körner, C. (2016). Belowground carbon trade among tall trees in a temperate forest, *Science*, **352**(6283), 342–344.
- Koenig, W. D. and Knopes, J. M. H. (2005). The mystery of masting in trees, *American Scientist*, **93**, 340–347.
- Lev-Yadun, S. (2011). Why should trees have natural root grafts?, *Tree Physiology*, **31**, 575–578.
- Lorenz, E. N. (1963). Deterministic nonperiodic flow, *Journal of Atmospheric Sciences*, **20**, 130–141.
- Lyles, D., Rosenstock, T. S., Hastings, A. and Brown, P. H. (2009). The role of large environmental noise in masting: General model and example from pistachio trees, *Journal of Theoretical Biology*, **259**(4), 701–713.
- May, R. (1976). Simple mathematical models with very complicated dynamics, *Nature*, **261**, 459–467.
- Noble, A. E., Machta, J. and Hastings, A. (2015). Emergent long-range synchronization of oscillating

- ecological populations without external forcing described by Ising universality, *Nature Communications*, **6**, 6664.
- Noble, A. E., Rosenstock, T. S., Brown, P. B., Machta, J. and Hastings, A. (2018). Spatial patterns of tree yield explained by endogenous forces through a correspondence between the Ising model and ecology, *The Proceedings of the National Academy of Sciences*, **115**(8), 1825–1830.
- Pecora, L. M. and Carroll, T. L. (1990). Synchronisation in chaotic systems, *Physical Review Letters*, **64**, 821–824.
- Pecora, L. M. and Carroll, T. L. (2015). Synchronisation of chaotic systems, *Chaos*, **25**, 097611.
- Prasad, A., Sakai, K. and Hoshino, Y. (2017). Direct coupling: a possible strategy to control fruit production in alternate bearing, *Scientific Reports*, **7**, 39890.
- Rees, M., Kelly, D. and Bjørnstad, O. N. (2002). Snow tussocks, chaos, and the evolution of mast seeding, *American Naturalist*, **160**, 44–59.
- Rosenblum, M. G., Pikovsky, A. S. and Kurths, J. (1996). Phase synchronization of chaotic oscillators, *Physical Review Letters*, **76**(11), 1804–1807.
- Rosenstock, T. S., Rosa, U. A., Plant, R. E. and Brown, P. H. (2010). A reevaluation of alternate bearing in pistachio, *Scientia Horticulturae*, **124**, 149–152.
- Rosenstock, T. S., Hastings, A., Koenig, W. D., Lyles, D. J. and Brown, P. H. (2011). Testing Moran's theorem in an agroecosystem, *Oikos*, **120**, 1434–1440.
- 酒井憲司 (1993). 非線形農学の萌芽, 数理科学, **31**(9), 43–48.
- 酒井憲司 (1995). 耕地のカオス, 数理科学, **33**(3), 67–72.
- 酒井憲司 (1997). 『カオス農学入門』, 朝倉書店, 東京.
- Sakai, K. (2001). *Nonlinear Dynamics and Chaos in Agriculture Systems*, Elsevier Science, Amsterdam.
- 酒井憲司 (2023). カオス同期に基づく自然への理解と農業応用, 電気学会誌, **143**(1), 29–32.
- Sakai, K. and Noguchi, Y. (2009). Controlling chaos (OGY) implemented on a reconstructed ecological two-dimensional map, *Chaos, Solitons and Fractals*, **41**(1), 630–641.
- Sakai, K., Noguchi, Y. and Asada, S. (2008). Detecting chaos in a citrus orchard: Reconstruction of nonlinear dynamics from very short ecological time series, *Chaos, Solitons and Fractals*, **38**(5), 1274–1282.
- Sakai, K., Hoshino, Y., Prasad, A., Fukamachi, A. and Ishibashi, A. (2019). Period-3 dominant phase synchronisation of *Zelkova serrata*: Border-collision bifurcation observed in a plant population, *Scientific Reports*, **9**, 15568.
- Sakai, K., Brown, P., Rosenstock, T., Upadhyaya, S. and Hastings, A. (2022). Spatial phase synchronisation of pistachio alternate bearing: Common-noise-induced synchronisation of coupled chaotic oscillators, *Chaos, Solitons and Fractals*, **165**(2), 112764.
- 佐竹暁子 (2007). 理論と実証分析の相互フィードバック: 植物の繁殖同調モデルを例に, 日本生態学会誌, **57**, 200–207.
- Satake, A. and Iwasa, Y. (2000). Pollen coupling of forest trees: Forming synchronized and periodic reproduction out of chaos, *Journal of Theoretical Biology*, **203**, 63–84.
- Satake, A. and Kelly, D. (2021). Studying the genetic basis of masting, *Philosophical Transactions of the Royal Society B*, **376**(1839), 20210116.
- Shalom, L., Samuels, S., Zur, N., Shlizerman, L., Zemach, H., Weissberg, M., Ophir, R., Blumwald, E. and Sadka, A. (2012). Alternate bearing in citrus: Changes in the expression of flowering control genes and in global gene expression in ON- versus OFF-crop trees, *PLoS One*, **7**(10), e46930.
- Simard, S. W., Pery, D. A., Jones, M. D., Myrold, D. D., Durall, D. M. and Molina, R. (1997). Net transfer of carbon between ectomycorrhizal tree species in the field, *Nature*, **388**, 579–582.

- Sviridova, N. and Sakai, K. (2016). Common noise induced synchronization on collective dynamics of citrus production, *Journal of the Japanese Society of Agricultural Machinery and Food Engineers*, **78**(3), 221–226.
- 高橋陽一郎, 山口昌哉, 辻下徹 (1998). 『数学—複雑系の科学と現代思想』, 青土社, 東京.
- Tilman, D. and Wedin, D. (1991). Oscillations and chaos in the dynamics of a perennial grass, *Nature*, **353**, 653–655.
- Watanabe, M. and Sakai, K. (2019). Impact dynamics model for a nonlinear bouncing tractor during inclined passage, *Biosystems Engineering*, **182**, 84–94.
- Ye, X. and Sakai, K. (2016). A new modified resource budget model for nonlinear dynamics in citrus production, *Chaos Solitons and Fractals*, **87**, 51–60.

Deterministic Nonlinear Prediction of Tree Crops
—Ensemble Reconstruction of Discrete Dynamics and a Lorenz Method for
One-year Forward Prediction of Individual Yield—

Kenshi Sakai

Institute of Agricultural Science, Tokyo University of Agriculture and Technology

Deterministic chaos is generated from low-dimensional dynamics with orbital instability. Discriminating chaotic behavior from observed time series is the main objective of nonlinear time series analysis. Deterministic nonlinear prediction is a powerful tool for this purpose but requires large time series sizes. Because crop yields can only be measured once per year, the observed yield time series size is very small. However, the number of crop individuals in a field or orchard is large; thus, the size of the time series sets is large. In addition, the reproductive patterns of many fruit trees can be assumed to have low-dimensional nonlinear dynamics. In this paper, we apply deterministic nonlinear prediction to yield time series sets obtained from orchards and present practical examples of one-year-ahead forecasting of yields for *Citrus unshiu* and Pistachio (*Pistacia vera*).

力学系を組み込んだ変分自己符号化器による 顕微鏡動画モデリング：ゼニゴケ精子の 運動を例として

近藤 洋平^{1,2}・南野 尚紀³・上田 貴志^{2,3}

(受付 2023 年 4 月 27 日；改訂 7 月 16 日；採択 7 月 21 日)

要 旨

顕微鏡による観察は、フックによって細胞が発見された17世紀から現代に至るまで、生物学において最も重要な研究方法の一つである。また近年では、光学技術と蛍光タンパク質を用いたツールの発展によって、多くの現象がこれまでになく高い時空間分解能で観測可能になっている。しかし、生み出される顕微鏡動画データは複雑かつ高次元であり、そこから生物学的に興味のあるパラメータを抽出することは必ずしも容易ではない。この問題に対し我々は、教師なし特徴抽出のための変分自己符号化器に力学系を組み込んで顕微鏡動画をモデリングすることで、力学系の状態変数およびパラメータの形で観察対象の生物を特徴づける情報を得る枠組みを提案する。本稿では、植物学において重要な地位を獲得しつつあるモデル生物であるゼニゴケ (*Marchantia polymorpha*) の精子運動様式を対象に、提案する枠組みの有用性を検証した。力学系として位相振動子を用いることで、精子の回転運動の位相とその速度を、暗視野顕微鏡による高速イメージングデータから抽出することができた。

キーワード：力学系、変分自己符号化器、画像解析、植物学、ゼニゴケ。

1. はじめに

生物の形態や行動といった表現型と、遺伝子およびそれから翻訳されるタンパク質の作用とを関連づけることは、現代の生物学において基盤的なアプローチの一つである。例えば、陸上植物の起源の解明において重要な位置を占めているモデル種であるゼニゴケ (*Marchantia polymorpha*) について、その生殖に重要な精子の運動の様式とその分子メカニズムが植物学者の関心を集めている。そこで、精子の運動に変調をきたすよう遺伝子改変された様々な変異体が作出されている。しかし変異体の精子運動を分析する上で課題となるのは、データが顕微鏡動画という複雑かつ高次元の形式をとっている点である。そのような生の高次元データから運動を特徴づけるパラメータ、例えば泳動中の精子回転周期など、を抽出することは容易ではない。この難しさは、最もよく研究されているヒトの精子運動の様式についてさえ顕微鏡動画の分析に関する新たな論文が発表され続けていることからわかる(例えば Saggiorato et al., 2017)。

¹ 自然科学研究機構 生命創成探究センター：〒444-8787 愛知県岡崎市明大寺町東山5-1

² 総合研究大学院大学 先端学術院：〒444-8787 愛知県岡崎市明大寺町東山5-1

³ 基礎生物学研究所：〒444-8585 愛知県岡崎市明大寺町西郷中38

そこで本論文では、精子運動の力学系モデルと変分自己符号化器 (Variational auto-encoder, VAE) (Kingma and Welling, 2014; Rezende et al., 2014) による教師なし特徴抽出の組み合わせによって、顕微鏡動画からの直接的な精子運動周期の抽出を試みる。すなわち、顕微鏡動画を自己符号化器のエンコーダーによって低次元化し、その特徴空間における時間発展を力学系によって制約し、自己符号化器のデコーダーによって未来の時点での顕微鏡画像を予測するという設定の上で、データに基づき力学系と自己符号化を誤差逆伝播法によって同時に学習する。これによって、学習後の力学系の状態変数とパラメータという形で精子回転運動の特徴が抽出できると期待される。

精子回転運動の力学系モデルとして、精子と周囲の流体力学まで考慮した詳細なものも選択肢にあるが、ここでは回転運動を抽象化した単純な振動子を考える。振動子モデル、特に振幅の自由度を無視した位相振動子は、単細胞の緑藻であるクラミドモナスの鞭毛が示す回転同期現象のモデル化などに既に応用がある (Klindt et al., 2017)。また、振動子を採用した状態空間モデルによって時系列から振動成分を抽出する振動子分解と呼ばれる手法も脳イメージングデータを対象に成功を納めており (Matsuda et al., 2022)、生物学分野でのデータ同化の文脈における振動子の有効性を示している。

周辺分野における関連する研究として、ロボティクス等への応用を念頭に、力学系モデルと教師なし学習による特徴抽出器の組み合わせによって動画から直接に力学系モデルのパラメータを推定する試みがなされている。ただし、問題の難しさから、撮影されている物体のセグメンテーションやトラッキングといった追加の情報を利用する手法が多い。そういった追加情報を用いず動画データのみから学習を行う試みとしては Le Guen らの PhyDNet (Le Guen and Thome, 2020) や Jaques らの手法 (Jaques et al., 2020) が挙げられるが、成功しているのは少数の質点の運動といった極めて単純な動態や人工データを扱うケースに限られ、応用までには距離がある。本研究も単純な動画および力学系モデルにおけるものではあるが、実際のデータを用いた生物学的に興味深いパラメータの定量化として、より広い文脈においても意義深い例であると言える。

2. モデルと学習

顕微鏡動画に対して、画像を低次元化するためのエンコーダー・デコーダーと、低次元の中間層における時間発展を記述する位相振動子を考える。モデルの概要を図1に示す。このモデルに入力される長さ T の動画を $X = [x_0, x_1, \dots, x_{T-1}]$, $x_i \in \mathbb{R}^{H \times W}$ とすると、エンコーダーを介して中間層の変数は次のようにサンプリングされる。

$$(2.1) \quad z = \text{Enc}(X)$$

$$(2.2) \quad \phi_0 \sim \text{WrappedNormal}(\phi; \mu_\phi(z), V_\phi)$$

$$(2.3) \quad p \sim \text{Normal}(p; \mu(z), V(z)),$$

ここで $p \in \mathbb{R}^P$ は、例えば画像に写り込む精子以外の夾雑物の有無といった、位相の進行に依らない画像間の変動を記述するための非位相変数である。本研究では p の次元として常に $P = 10$ を用いた。位相変数の初期値 $\phi_0 \in [0, 2\pi)$ は巻き込み正規分布、非位相変数 p は正規分布に従い、それらの分布の母数 $\mu_\phi(z), \mu(z), V(z)$ は潜在変数 z から

$$(2.4) \quad (\mu_\phi(z), \mu(z), V(z))^\top = Bz + C$$

と計算される。 $B \in \mathbb{R}^{3 \times P}$, $C \in \mathbb{R}^3$ はエンコーダーのパラメータと同時に学習によって決まる。一方 V_ϕ は目的関数の評価を簡単にするために定数として扱う。本研究のモデルでは簡単なた

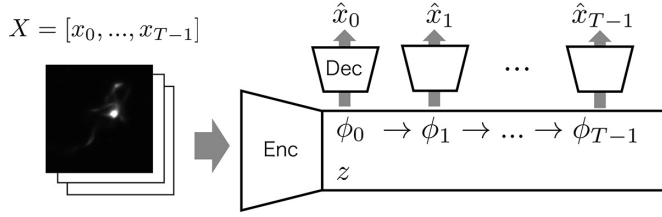


図 1. モデルの概要. 変分自己符号化器の中間層に位相振動子が埋め込まれ, 顕微鏡動画を自己符号化する.

表 1. エンコーダーおよびデコーダーの構造. In: 入力チャンネルサイズ, Out: 出力チャンネルサイズ, Conv: 畳み込みニューラルネットワーク, Trans Conv: 転置畳み込みニューラルネットワーク, SiLU: シグモイド線形ユニット関数 $\text{SiLU}(x) = x/(1 + e^{-x})$. 全ての畳み込み層においてフィルター適用間隔は 2 とした.

	Layer type	In	Out	Filter size	Nonlinear Transformation
Encoder	Conv	T	32	(4, 4)	SiLU
	Conv	32	32	(4, 4)	SiLU
	Conv	32	64	(4, 4)	SiLU
	Conv	64	64	(4, 4)	SiLU
Decoder	Linear	2 + P	$64 \times 6 \times 6$		SiLU
	Trans Conv	64	64	(4, 4)	SiLU
	Trans Conv	64	32	(4, 4)	SiLU
	Trans Conv	32	32	(4, 4)	SiLU
	Trans Conv	32	1	(4, 4)	Sigmoid

めに振幅自由度を無視した位相振動子を採用し, 位相の時間発展を次のように書く.

$$(2.5) \quad \phi_{t+1} = \phi_t + \omega \Delta t.$$

位相速度 ω は初期値を $[0, 0.01]$ の一様分布からサンプルし, ニューラルネットワークのパラメータと同時に学習する.

これら位相変数と非位相変数は, デコーダーによって対応する時点の顕微鏡画像へと復号される.

$$(2.6) \quad \hat{x}_t = \text{Dec}(\cos(\phi_t), \sin(\phi_t), p).$$

そしてこのモデルは元の動画を再構成した $\hat{X} = [\hat{x}_0, \dots, \hat{x}_{T-1}]$ を出力する.

本研究で採用したエンコーダーとデコーダーの構造については表 1 にまとめた. 以上の設定によって, このモデルのパラメータの総数は 318,807 となる.

上で定義したモデルのような確率的自己符号化器を学習するための目的関数として, 変分サイズ法の処方箋に従い, 次のような対数周辺尤度の変分下限が用いられる.

$$(2.7) \quad \log p(X) \geq \mathbf{E}_{q(z|X)}[\log p(X|z)] - D_{\text{KL}}(q(z|X)||p(z)),$$

ここで D_{KL} はカルバック・ライブラー情報量を表す. 提案分布 $q(z|X)$ はエンコーダーによってモデル化される. このように隠れ変数の事後分布を入力から推定するニューラルネットワークを学習するという手法は Amortized inference と呼ばれ深層学習の文脈で多くの例がある. 物理モデルを変分自己符号化器に組み込んだ定式化を扱っている Takeishi and Kalousis, 2021 を

参考にすると、本研究のモデルでは目的関数は

$$(2.8) \quad L(\theta, X) = \mathbf{E}_{q(\phi_0, p|X, \theta)} \log p(X|\phi_0, p, \theta) - D_{\text{KL}}(q(\phi_0, p|X)||p(\phi_0, p))$$

となる。ここでは学習されるモデルパラメータをまとめて θ と書いており、 θ はエンコーダー・デコーダーのパラメータに加えて式(2.4)の B, C と位相速度 ω を含んでいる。右辺第一項は、決定論的なデコーダーについては観測ノイズとして各次元で独立な標準正規分布を考慮して二乗誤差とすることが多く、本研究もそれに従う。右辺第二項については、隠れ変数 (ϕ_0, p) の事前分布がそれぞれ一様分布と標準正規分布とし、位相変数に対する提案分布である巻き込み正規分布に関して分散は学習しないため、学習に影響するのは次のように容易に計算できる部分のみである。

$$(2.9) \quad D_{\text{KL}}(q(\phi_0, p|X)||p(\phi_0, p)) \rightarrow D_{\text{KL}}(\text{Normal}(\mu(X, \theta), V(X, \theta))||\text{Normal}(0, I)).$$

この $\mu(X, \theta), V(X, \theta)$ は式(2.3)に現れる非位相変数の平均 $\mu(X)$ と分散 $V(X)$ と同じものであるが、ここではパラメータ依存性を明示した。

変分下限を目的関数とする確率的自己符号化器すなわち変分自己符号化器の最適化には、計算コストの軽減と局所解の回避を兼ねて、勾配法によるパラメータの更新毎にデータの一部のみを用いる確率的勾配降下法が使用される。本研究でも、全体の長さ $N(> T)$ の動画データを

$$(2.10) \quad [x_0, \dots, x_{N-1}] \rightarrow [x_{0:T-1}, x_{1:T}, \dots, x_{N-T:N-1}]$$

のように分割した上で確率勾配降下法によってモデルを適用した。具体的には、深層学習ライブラリの PyTorch (version 2.0.0) (Paszke et al., 2019) を用いて、本研究で扱う顕微鏡動画それぞれについて学習率 $\eta = 0.0003$ とした Adam オプティマイザーによって 1000 エポック最適化した。本研究では、予測でなくデータ圧縮が主眼であること、目的関数が対数周辺尤度の変分下限であることから、バリデーションデータを利用した Early stopping は実施しなかった。ただし、変分自己符号化器においてもエンコーダーの過学習がおこりうることを示されており (Zhang et al., 2022)、それに対処するための学習アルゴリズムも提案されているが、その実効性は検証途上であると考えられるため、本研究では採用しなかった。

3. データ

先行研究の手順 (Minamino et al., in preparation) に従い、我々はゼニゴケの野生株 Takaragaike-1 から蒸留水によって精子懸濁液を採取し、暗視野顕微鏡によって観察した。そこで得られた 30 ms 間隔の経時観察動画をもとに画像解析ソフトウェア ilastik (Berg et al., 2019) によって個々の精子をトラッキングし、96 px \times 96 px のサイズの動画データを八本作成した。八本の動画のフレーム数は 100, 114, 170, 136, 100, 100, 135, 76 である。

4. 結果と考察

予測時点数を $T = 20$ 、非位相変数の次元を $P = 10$ 、初期位相の分散を $V_\phi = (0.02)^2$ とした学習の結果、位相振動子を埋め込んだ変分自己符号化器は精子の回転運動を再構成するようになった。目的関数値およびその一部である KL 罰則値、すなわち式(2.8)の右辺第二項の値、のエポック履歴を図 2 に示す。本研究で解析した動画間では目的関数値の振る舞いに顕著な違いがみられない。より詳しく調べるために、典型的な再構成例を図 3 に示す。エンコーダー・デコーダーの学習には動画全長を用いているために必ずしも驚くべきことではないが、隠れ変数の推定に用いた入力動画の一部 ($x_{20:19+T}$) よりも先の時点についても、安定した画像再構成が

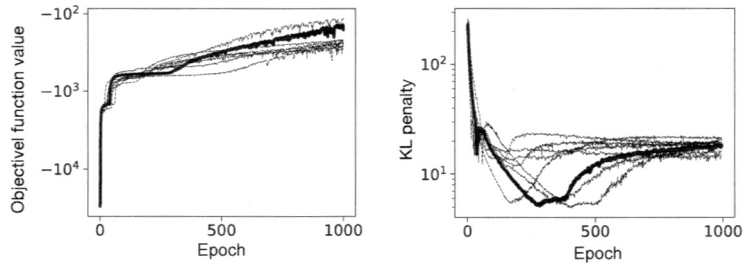


図 2. 八本の精子運動動画データそれぞれについての目的関数値(左)および KL 罰則値(右)のエポック履歴. いずれも確率勾配法におけるミニバッチあたりの値であり, 各ミニバッチは 160 枚の画像を含んでいる. 図 3-5 で用いた動画に関する結果を実線で示した.

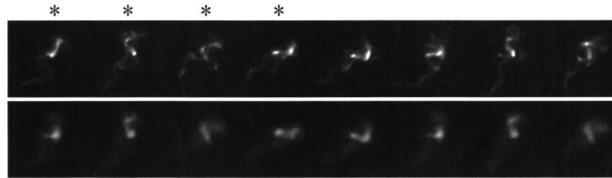


図 3. 一つの精子動画データによって学習したモデルについて, 元の動画の一部(上)と再構成された動画(下). *印は動画の再構成にあたって位相変数と非位相変数の推定に用いられた入力動画の範囲を示す.

なされていることがわかる. ただし再構成画像では鞭毛のような微細な構造は失われ, 明瞭に確認できるのは細胞体のみである.

同じ学習器と動画データを用いて, スライディング・ウィンドウ方式によって動画全長に渡って位相変数 ϕ_0 と非位相変数 p の推定値を計算した結果が図 4 である. 図に示されている値はエンコーダーによって計算された (ϕ, p) の平均値を使い分散を無視している, すなわち式(2.2)および(2.3)における $\mu_\phi(z)$ および $\mu(z)$ を用いている. 言い換えると, 図 4 は隠れ変数の MAP 推定値を示している. また非位相変数 p については煩雑になるのを避けるために (p_0, p_1, p_2) のみ表示している. ここから精子の細胞体の回転運動の周期が安定していることがわかる. 一方で, 位相の進行に依らない画像の変動を吸収するために導入した p については, 期待通り位相変数との明確な相関がない.

そこで我々は, p に符号化されている情報を具体的に調べるために, p の値をエンコーダーによって決められた値から変更し, 再構成画像がいかに変わるかを調べた. 図 5 に各 p_i について $[p_i - 5, p_i - 2, p_i, p_i + 2, p_i + 5]$ と変更した際の再構成画像を示した. p は学習時に標準正規分布に近づくようにカルバック・ライブラー情報量によって制御されるため, 図 4 右からもわかるようにその推定値は概ね $[-2, +2]$ の範囲にとどまる. この範囲の p の変化では再構成画像への影響は視覚的には顕著なものではない. またこの範囲を超えて p を増減させても, 再構成画像にわずかな明暗パターンの変化を引き起こすにとどまることが多く, 容易に解釈が可能な結果ではなかった. 変分自己符号化器の隠れ変数に符号化される画像の特徴が必ずしも解釈可能でない問題は広く認識されており, これを改善するために隠れ変数の分布に様々な制約をかけるなどの手法が提案されているが, 複雑な実データに対する成功例は乏しく, さらなる技術の発展が待たれる.

最後に, 同じモデルと学習方法によって, 八本の動画全てを個別に解析した結果を図 6 に示

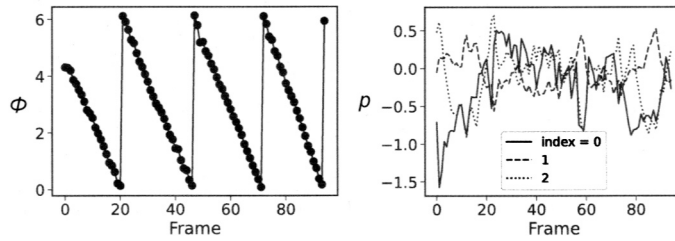


図 4. スライディング・ウィンドウ方式で推定された位相変数(左)と非位相変数(右)の時系列. モデルおよびデータは図 3 で使用したものと同じである.

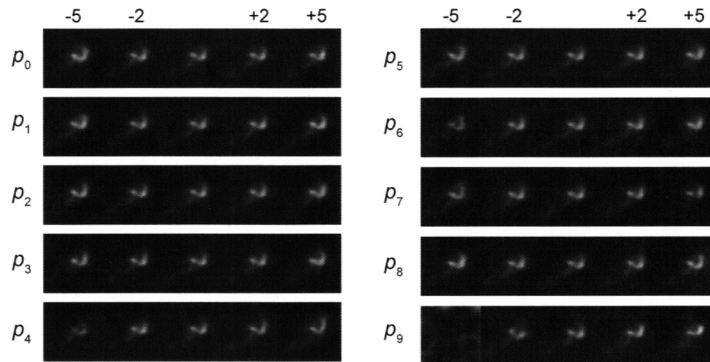


図 5. 各非位相変数 $\{\rho_i\}$ を元の値から変更して再構成した画像. モデルおよびデータは図 3 および図 4 で使用したものと同じである.

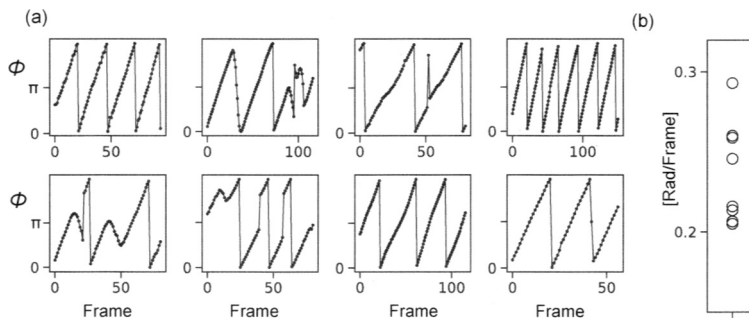


図 6. (a) 八本の精子動画それぞれについて同様の解析を実施して推定された位相時系列. (b) 学習によって得られた位相速度 ω .

す. どの精子も観察時間内では安定した位相速度を持つが, 精子間の差は大きく, 個体差とも呼ぶべきものがあると思われる. また, 一部の精子では位相の跳びが発生している. これはトラッキングされているものではない精子の写り込みなどによって説明できる場合もあるが, 詳細はさらなる研究が必要である. 抽出された回転速度は実時間の単位では 1.3 ± 0.16 Hz であり, これは肉眼による半定量的な評価と一致している. しかし一方で, 精子運動分析専用の観察装置および画像解析ソフトウェアを用いた研究 (Furuichi and Matsuura, 2016) では精子回転を反映していると想定される “lateral head displacement” の頻度について 17–19 Hz というかなり大きな値が報告されている. ここで留意すべきは, ゼニゴケ精子の頭部からは二本の鞭毛が

生えており、さらにそれぞれ異なる運動パターンを示すとの報告もある(宮村 他, 2002)ことから精子頭部の運動パターンは時空間的にかなり複雑になっている可能性があることである。我々の結果と先行研究との差が、注目している運動成分の違いに由来するのか、または精子の状態や周囲の流体の性質の違いに由来するのか、今後明らかにしていくことは精子運動のさらなる理解につながるため重要である。

5. おわりに

本研究では、ゼニゴケ精子の運動様式を特徴づけるパラメータを顕微鏡動画から直接得るという目的に対し、位相振動子を組み込んだ変分自己符号化器を提案し、その有効性を示した。我々の提案モデルは最小限のものであり、その力学系モデル・自己符号化器・目的関数について様々な拡張が考えうるため、以下で簡単に議論したい。

まず力学系モデルの拡張として興味深いのは複数の振動子、さらには振動子間の相互作用を考慮することである。このようなモデルは結合振動子と呼ばれ、理論的な研究によってその挙動が詳細に明らかになっている。単細胞生物の遊泳に関する研究が示すように(Klindt et al., 2017)、結合振動子モデリングによってゼニゴケ精子が持つ二本の鞭毛がいかにして協働し細胞を駆動しているかを明らかにできると期待できる。ただし図3からわかるように、そもそも本研究のモデルでは鞭毛の運動の詳細を捉えられていない。これを解決するためには、より時空間分解能の高いデータを用いることと、より高解像度の画像生成に向けたデコーダーを用いることの両方が必要になるだろう。また本研究で無視した振幅自由度について、ゼニゴケ精子における知見はまだ乏しいものの、遊泳細胞の状態や環境の変化によって鞭毛が振幅変動を示す例は多い。そのため複素ギンツブルグ-ランダウ方程式等のモデルによって振幅自由度を考慮したモデリングも検討に値する。

自己符号化器について、単純により高性能の深層学習器を採用するというを超えて興味深い拡張として、物体中心表現学習(例えばLocatello et al., 2020)のような手法によって、複数の物体が写る画像から単一の精子のみを抽出するというものが考えられる。それによって、精子以外の夾雑物の写り込みに対する頑強性の獲得や、精子間の相互作用の学習が可能になるかもしれない。

変分自己符号化器の目的関数に対しては先行研究で様々な正則化項の追加が提案されており、隠れ変数間の統計的独立性による解釈可能性の改善(Chen et al., 2018)や、入力画像への敵対的攻撃に対する頑強性の向上(Cemgil et al., 2020)が報告されている。ただし、正則化項の追加によって変分下限の近似が緩まり、それが隠れ変数の事後分布の推定におよぼす影響は必ずしも明らかでない。一方で我々が有用と考えているのは、よく使われる変分下限とは異なる周辺対数尤度下限である因子化情報量規準(Hayashi et al., 2015)であり、これはモデルの学習と変数選択を同時に行うことを可能にする(Kondo et al., 2016)。例えば、上で述べたように複数の振動子を考慮する場合には振動子の個数の選択が問題になり、実際松田らは脳活動時系列の振動子による状態空間モデリングにあたって赤池情報量規準を採用しモデル選択している(Matsuda et al., 2022)。我々のモデルは深層学習器を含むためモデル選択を目的として学習を繰り返すことは計算コストが大きいという問題があるが、因子化情報量規準はその解決策になりうる。また、本研究を含め多くのケースで式(2.7)の右辺第一項が表す画像再構成誤差を単純に二乗誤差によって評価しているが、代わりに観察機器の光学的性質を考慮した観測ノイズの分布を用いるなどの工夫の余地はまだまだあるように思われる。

本稿ではゼニゴケ野生株に関する結果のみを示したが、変異体の解析についても現在進行中である。本研究で得られた知見が、未だ謎に包まれている植物精子の運動様式の解明につなが

ることを期待したい。

謝 辞

本研究は MEXT 科研費新学術領域研究 No. 19H05675 の助成を受けたものです。計算機資源に関して基礎生物学研究所・超階層生物学センターの支援に厚く感謝いたします。原稿について富沢瑤子研究員に貴重な助言をいただきました。

参 考 文 献

- Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J. I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F. A. and Kreshuk, A. (2019). ilastik: interactive machine learning for (bio)image analysis, *Nature Methods*, **16**(12), 1226–1232.
- Cemgil, A. T., Ghaisas, S., Dvijotham, K., Gowal, S. and Kohli, P. (2020). The autoencoding variational autoencoder, *Proceedings of the 34th Conference on Neural Information Processing Systems*, 15077–15087.
- Chen, R. T. Q., Li, X., Grosse, R. and Duvenaud, D. (2018). Isolating sources of disentanglement in VAEs, *Proceedings of the 32nd Conference on Neural Information Processing Systems*, 2615–2625.
- Furuichi, T. and Matsuura, K. (2016). Kinetic analysis on the motility of liverwort sperms using a microscopic computer-assisted sperm analyzing system, *Environmental Control in Biology*, **54**(1), 45–49.
- Hayashi, K., Maeda, S. and Fujimaki, R. (2015). Rebuilding factorized information criterion: Asymptotically accurate marginal likelihood, *Proceedings of the 32nd International Conference on Machine Learning*, 1358–1366.
- Jaques, M., Burke, M. and Hospedales, T. M. (2020). Physics-as-inverse-graphics: Unsupervised physical parameter estimation from video, *Proceedings of the 8th International Conference on Learning Representations*, 1–16.
- Kingma, D. and Welling, M. (2014). Auto-encoding variational bases, *2nd International Conference on Learning Representations, Conference Track Proceedings*, <https://arxiv.org/abs/1312.6114>.
- Klindt, G. S., Ruloff, C., Wagner, C. and Friedrich, B. M. (2017). In-phase and anti-phase flagellar synchronization by waveform compliance and basal coupling, *New Journal of Physics*, **19**(11), 113052.
- Kondo, Y., Maeda, S. and Hayashi, K. (2016). Bayesian masking: Sparse Bayesian estimation with weaker shrinkage bias, *Asian Conference on Machine Learning*, 49–64.
- Le Guen, V. and Thome, N. (2020). Disentangling physical dynamics from unknown factors for unsupervised video prediction, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11471–11481.
- Locatello, F., Weissenborn, D., Unterthiner, T., Mahendran, A., Heigold, G., Uszkoreit, J., Dosovitskiy, A. and Kipf, T. (2020). Object-centric learning with slot attention, *Proceedings of the 34th Conference on Neural Information Processing Systems*, 11525–11538.
- Matsuda, T., Homae, F., Watanabe, H., Taga, G. and Komaki, F. (2022). Oscillator decomposition of infant fNIRS data, *PLoS Computational Biology*, **18**(3), e1009985.
- 宮村新一, 松永 茂, 堀 輝三 (2002). ゼニゴケ精子鞭毛運動の高速度ビデオによる解析, *蘚苔類研究*, **8**(3), 79–83.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A.,

- Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. and Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library, *Proceedings of the 33rd Conference on Neural Information Processing Systems*, 8024–8035.
- Rezende, D. J., Mohamed, S. and Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models, *Proceedings of the 31st International Conference on Machine Learning*, 1278–1286.
- Saggiorato, G., Alvarez, L., Jikeli, J. F., Kaupp, U. B., Gompper, G. and Elgeti, J. (2017). Human sperm steer with second harmonics of the flagellar beat, *Nature Communications*, **8**(1), 1415.
- Takeishi, N. and Kalousis, A. (2021). Physics-integrated variational autoencoders for robust and interpretable generative modeling, *Proceedings of the 35th Conference on Neural Information Processing Systems*, 14809–14821.
- Zhang, M., Hayes, P. and Barber, D. (2022). Generalization gap in amortized inference, *Proceedings of the 36th Conference on Neural Information Processing Systems*, <https://arxiv.org/abs/2205.11640>.

Dynamical-system-integrated Variational Autoencoder
for Modeling of Time-lapse Microscopy Data:
Application to Liverwort Spermatozoid Motion

Yohei Kondo^{1,2}, Naoki Minamino³ and Takashi Ueda^{2,3}

¹ExCELLS, National Institutes of Natural Sciences

²Graduate Institute for Advanced Studies, SOKENDAI

³National Institute for Basic Biology

Microscopy has been one of the most important research methods in biology from the 17th century, when Hooke discovered the cell, to the present day. In recent years, the development of optical technology and fluorescent protein-based tools has revealed various biological phenomena with unprecedented spatio-temporal resolution. However, such microscopic movie data are complex and high-dimensional, and thus it is not always easy to obtain biologically-relevant parameters from them. To address the issue, we propose a framework for modeling of microscopic movie by integrating a dynamical system into a variational autoencoder, a deep-learning model for unsupervised feature extraction. This allows us to extract biological information as state variables and parameters of the integrated dynamical system. In this article, we have validated our framework by analyzing spermatozoid motion of the liverwort *Marchantia polymorpha*, an emerging model organism for plant biology. Using a phase oscillator, we have succeeded to estimate the phase and speed of spermatozoid rotation from our high-speed dark-field video microscopy.

位相的データ解析による銀河分布の定量化と バリオン音響振動抽出

竹内 努^{1,2}・河野 海¹・クレ スチエータ^{1,3,4}・西澤 淳^{5,6,7}・
村上 広耶¹・馬 海霞¹・本武 陽一⁸

(受付 2022年9月21日; 改訂 2023年3月16日; 採択 3月20日)

要 旨

宇宙空間において、銀河は銀河団、銀河群、フィラメント(1次元的な銀河の連なり)、ヴォイド(銀河の存在しない領域)といった構造を形作り不均一に分布している。これを宇宙の大規模構造と呼ぶ。宇宙論では通常の物質をバリオンと呼び、我々が直接観測するのはこのバリオンが形作る大規模構造である。

大規模構造を形成する主要因は重力不安定だが、これとは別種の構造の源となる機構が存在する。これがバリオン-光子脱結合の時期に生じるバリオンの音波である。これはバリオン音響振動(BAO)として知られ、銀河の空間分布に刻印されている。本研究では、銀河分布のBAO信号を検証するために、位相的データ解析(topological data analysis: TDA)の方法で銀河の空間分布を解析した。TDAでは、点群データを「穴」の集合として解釈し、その幾何学的構造からパーシステントホモロジー群を構成することで、1つ1つの穴構造の特徴を抽出する手法である。従って、それぞれの穴の情報を保持するパーシステントホモロジー群と穴の位置情報などを組み合わせることで、TDAによって点群データから定義される穴集合のそれぞれの穴のサイズ、位置などの特徴が定量評価できる。

我々はまず、バリオンの物理あり、なしの場合のシミュレーションデータセットを解析し、TDAのパフォーマンスを調べた。この結果、TDAは、宇宙の大規模構造の中からBAO信号を実際に検出できることがわかった。バリオンありのシミュレーションはBAOからの有意な信号を示したのに対し、バリオンなしデータはこの信号が検出されなかった。この結果をもとに、次にスローンデジタルスカイサーベイデータリリース14(SDSS DR14)のeBOSS(extended Baryon Oscillation Spectroscopic Survey)から $z < 1.0$ のクエーサー(活動銀河核の一種)を抽出してサンプルを構成し、TDAを適用した。この解析で、我々は $r \sim 150$ [Mpc]のスケールで特徴的な穴(中空のシェル)を発見した。これは、銀河/クエーサー分布に刻印されたBAOの信

¹名古屋大学 素粒子宇宙物理学専攻: 〒464-8602 愛知県名古屋市千種区不老町

²統計数理研究所 客員: 〒190-8562 東京都立川市緑町10-3

³国立天文台 科学研究部: 〒181-8588 東京都三鷹市大沢2-21-1

⁴日本学術振興会特別研究員(PD)

⁵岐阜聖徳学園大学 DX推進センター: 〒501-6194 岐阜県岐阜市柳津高桑西1-1

⁶名古屋大学 高等研究院: 〒464-8601 愛知県名古屋市千種区不老町

⁷名古屋大学 素粒子宇宙起源研究所: 〒464-8602 愛知県名古屋市千種区不老町

⁸統計数理研究所(現 一橋大学 ソーシャル・データサイエンス研究科): 〒186-8601 東京都国立市中2-1)

号に正確に対応している。この解析は2000個のクエーサーからなるサブサンプルで実行した。これはTDAが、たとえ非常に疎なサンプリングであっても中空の穴のような位相的構造を検出するのに有効であることを示している。

キーワード：位相的データ解析，パーシステントホモロジー，銀河分布，宇宙の大規模構造，バリオン音響振動。

1. はじめに

1.1 宇宙の物質分布の特徴づけと記述方法

各宇宙年齢における物質および銀河の分布には、宇宙の構造の形成と進化に関する基本的な情報が含まれている (e.g., Peebles, 1980; Efstathiou and Silk, 1983; Bernardeau et al., 2002). まず強調しておかなくてはならない事実として、宇宙論が理論的に予言するのは物質分布あるいは銀河分布の統計的性質であり、時空間のどの位置にどの銀河が誕生するかなど具体的な天体についての予言力はない。よって、宇宙論の理論と観測の比較は統計的記述を通じて行うしかない。この理由から、物質分布のゆらぎの統計的特性を特徴づけるために、膨大な数の洗練された方法が提案されてきた (e.g., Peebles, 1980; Martinez and Saar, 2001; Bernardeau et al., 2002). その中でも、 n 点相関関数は、観測される銀河分布 (e.g., Peebles, 1980; Martinez and Saar, 2001; Bernardeau et al., 2002)の解析で最も一般的でよく研究されている方法である。宇宙空間のある位置 \vec{x} における銀河の数密度 $n(\vec{x})$ を測定し、その統計平均 $\langle n \rangle$ と比較する。宇宙は我々の住んでいるもの1つしか観測できないが、仮想的に多数の宇宙の集合(アンサンブル: ensemble)を想定し、ある量 Q の多数の宇宙での平均を考えることを統計平均あるいはアンサンブル平均と呼び、記号 $\langle Q \rangle$ で表す。これに対し、空間内で平均した値を体積平均とよび、記号 \bar{Q} で表す。宇宙論では、この2種類の平均が一致するという仮定(宇宙論的エルゴード仮説)に基づいて議論を展開する。

この準備のもと、 d^3x_1, d^3x_2 という2つの微小体積を考える。双方の微小体積中に銀河が見つかる同時確率を $\mathcal{P}(\vec{x}_1, \vec{x}_2)d^3x_1d^3x_2$ とする。銀河分布が一様、即ち無相関(クラスタリングがない)場合、微小体積内に銀河が見つかる確率は

$$(1.1) \quad \mathcal{P}(\vec{x})d^3x = \langle n \rangle d^3x,$$

なので、体積 d^3x_1 内の銀河と体積 d^3x_2 内に銀河が見つかる同時確率は

$$(1.2) \quad \mathcal{P}(\vec{x}_1, \vec{x}_2)d^3x_1d^3x_2 = \langle n \rangle^2 d^3x_1d^3x_2$$

となる。銀河がクラスタリングしている場合、同時確率は

$$(1.3) \quad \mathcal{P}(\vec{x}_1, \vec{x}_2)d^3x_1d^3x_2 = \langle n(\vec{x}_1)n(\vec{x}_2) \rangle d^3x_1d^3x_2 \equiv \langle n \rangle^2 [1 + \xi(\vec{x}_1, \vec{x}_2)] d^3x_1d^3x_2$$

となり、一様な場合に比べて $\langle n \rangle^2 \xi(\vec{x}_1, \vec{x}_2)d^3x_1d^3x_2$ だけ高くなる。この $\xi(\vec{x}_1, \vec{x}_2)$ を銀河分布の2点相関関数と呼ぶ。同様に、一般に n 点相関関数は n 個の微小体積内に銀河を見出す同時確率を用いて定義される。

ここまでは銀河分布を点群として定式化した。銀河分布を議論する際には数密度の分布を平滑化し、密度場として扱うことも多い。点群と密度場の数学的扱いは異なるが、宇宙物理ではこの違いは重視しないで同様に扱うことが多く、本稿でもそれに倣う。宇宙論的定式化のため、銀河分布の密度場 $\rho(\vec{x})$ の密度ゆらぎ

$$(1.4) \quad \delta(\vec{x}) \equiv \frac{\rho(\vec{x}) - \bar{\rho}}{\bar{\rho}}$$

を考えるのが便利である。ここで $\bar{\rho}(\bar{x})$ は平均密度を表す。密度 $\rho(\bar{x})$ は点群としての銀河の数密度、あるいは平滑化した銀河分布の質量密度のどちらを表現する場合もありうる。密度ゆらぎ $\delta(\bar{x})$ を用いると、相関関数 $\xi(\bar{x}_1, \bar{x}_2)$ は

$$(1.5) \quad \xi(\bar{x}_1, \bar{x}_2) = \frac{\langle [\rho(\bar{x}_1) - \bar{\rho}] [\rho(\bar{x}_2) - \bar{\rho}] \rangle}{\bar{\rho}^2} = \langle \delta(\bar{x}_1) \delta(\bar{x}_2) \rangle$$

と表せる。なお、宇宙論では宇宙は一様等方であると仮定するので、対称性より

$$(1.6) \quad \xi(\bar{x}_1, \bar{x}_2) = \xi(|\bar{x}_1 - \bar{x}_2|) \equiv \xi(r)$$

として扱うことが一般的である。式(1.5)の形で表すと、相関関数が密度場の2次のキュミュラントになっていることが分かる。一般的に、ある空間上のベクトル $\bar{x} = (x_1, \dots, x_n)^\top$ を添字とする確率変数 $X(\bar{x})$ を確率場 (random field) とよぶ。数学的には、ある確率場は無次元までのモーメントあるいはキュミュラントで一意に定まることから、キュミュラントと等価な n 点相関関数によって確率場が決まる。これは確率場の特性関数のキュミュラント展開定理を用いて示せる (e.g., Daley and Vere-Jones, 2003)。即ち、相関関数を測定することによって原理的には宇宙の密度場の統計的性質を特定できる。

しかし実際には、高次の n 点相関の推定は極めて難しく非現実的であるため、高次相関の情報を扱う別のアプローチが検討されている。銀河分布の3次元構造の発見 (e.g., de Lapparent et al., 1986) により、平滑化した密度場の幾何学 (トポロジー) を用いて大規模構造を定量化し、背後にある物理過程を特定しようというアイデアが提示された (e.g., Shandarin and Zeldovich, 1983; Gott et al., 1986, 1987, 1989; Boerner and Mo, 1989)。早くから用いられてきたのがジーナス (種数, genus: G) によるトポロジーの特徴付けである (Gott et al., 1986)。銀河分布の密度ゆらぎ $\delta(\bar{x})$ の標準偏差 σ の ν 倍 (ν は定数) を閾値とすれば、 $\nu\sigma$ に対する上側レベル集合が各 ν について定義できる。これをエクスカージョン集合 (ここでは E_ν と表記) とよび、 $\rho(\bar{x}) = \nu\sigma$ となる等密度面がその表面となる。密度場を考えているので、 E_ν は3次元の集合である。銀河分布のジーナスは、平滑化した銀河密度場のオイラー標数の $-1/2$ 、即ち

$$(1.7) \quad G = -\frac{1}{2} \int K dA,$$

(K : 等密度面のガウス曲率, dA : 等密度面上の面積要素) と定義される。なお、ここでの定義は宇宙論で用いられるもので、数学における通常定義と1だけずれていることに注意されたい (松原, 1996)。エクスカージョン集合 E_ν の単位体積あたりのジーナスを ν の関数として表した量 $G(\nu)$ をジーナスカーブと呼ぶ。ジーナスカーブを銀河分布の集約統計量として用い、宇宙論的研究を行う方法がジーナス統計である。このアプローチは、大規模構造の重力による成長を密度分布の初期値からの微小なずれ (摂動) として扱うことで、大規模構造進化を記述する理論体系、いわゆる構造形成の摂動論的アプローチに関連して開発されてきた (e.g., Matsubara, 1994, 1996, 2003; Matsubara and Suto, 1996; Matsubara et al., 2022; Matsubara and Yokoyama, 1996; Pogosyan et al., 2009; Gay et al., 2012)。摂動論を用いると、ジーナスカーブは大規模構造のスケールではモーメントの関数として解析的に書けるので (e.g., Matsubara, 1994, 1996; Matsubara and Suto, 1996; Matsubara and Yokoyama, 1996)、これによって位相幾何学的情報を定量的に表現できる。特にジーナスカーブはガウス場と他の確率場を区別できるため、大規模構造を解析するための一般的な道具として定着している。

ジーナス統計の方法を含む、より一般的な体系がいわゆる“ミンコフスキー汎関数 (Minkowski functionals)” である (e.g., Hikage et al., 2003; Matsubara, 2003; Sullivan et al., 2019)。ミンコフスキー汎関数の大規模構造への応用でも式(1.4)で定義した平滑化した銀河密度場の密度ゆ

らぎを考える。この場合ミンコフスキー汎関数 M_k は $k = 0$ については

$$(1.8) \quad M_0 = \frac{1}{V} \int_{E_\nu} dV,$$

$k = 1, 2, 3$ については閾値 ν の等密度面上の曲率の積分によって定義されるので、原理的には銀河密度場はミンコフスキー汎関数によって完全に特定できる (Schmalzing and Buchert, 1997; Schmalzing and Gorski, 1998) (これ以降、本論文を通して $k = 0, 1, 2, 3$ は次元の意味でのみ用いる)。実際、この方法は宇宙論的解析に広く応用されている (密度場の解析: e.g., Mecke et al., 1994; Kerscher et al., 1997; Beisbart et al., 2002; Hikage et al., 2003; Kerscher and Tikhonov, 2010; Wiegand et al., 2014; Wiegand and Eisenstein, 2017; 赤方偏移歪み: e.g., Appleby et al., 2018; 弱重力レンズ宇宙論: e.g., Spurio Mancini et al., 2018; Vicinanza et al., 2019; Parroni et al., 2020; 宇宙再電離: e.g., Yoshiura et al., 2017; Bag et al., 2018, 2019; Chen et al., 2019; 宇宙論/重力理論の検証: e.g., Ling et al., 2015; Junaid and Pogosyan, 2015; Shiraishi et al., 2016; Fang et al., 2017; および宇宙マイクロ波背景放射 (CMB): e.g., Matsubara, 2010; Planck Collaboration et al., 2016; Santos et al., 2016; Chingangbam et al., 2017; Ganesan and Chingangbam, 2017; Joby et al., 2019)。

1.2 ミンコフスキー汎関数から位相的データ解析へ

このように、銀河密度場は密度場のトポロジーによって定量的評価ができる。上で述べたジーナス統計は、点群データの性質を特徴付ける。ジーナスのように、点群データ全体をその穴構造についての一つの特徴量で表現するのではなく、点群データに含まれる個別の穴構造についてのスケールのような特徴をそれぞれ保持できる表現がパーシステントホモロジー群である (TDA: e.g., Edelsbrunner et al., 2002; Zomorodian and Carlsson, 2004; Wasserman, 2018)。以下ではパーシステントホモロジー群に基づくデータ解析法である TDA についての簡易な説明を行う。

データが持つ一つ一つの穴構造を評価できる TDA を用いることで、宇宙論におけるデータ解析をより発展させることができる。本研究で対象となる点群データの穴構造は、銀河や任意の天体などで構成される N 体系の各点に配置された半径 r の n 次元球の集合によって形成される。宇宙論分野黎明期のこのアプローチによる研究は、基本的にはランダムパーコレーションの応用であった (Gott et al., 1986)。すなわち、各銀河の周りに球を置いて半径 r を増加させ、半径 r の関数として考えている体積内の連結成分の幾何学的構造を調べる方法である。これは TDA で用いられる考え方そのものである。

萌芽的な意味での TDA の概念の誕生は 90 年代に遡るが、実質的な誕生は 2000 年代に入ってからである (see e.g., Edelsbrunner et al., 2002; Edelsbrunner and Harer, 2010)。TDA では、よくデータセットをパーシステントホモロジー群として表現することが行われる。ランダムパーコレーションと同様、パーシステントホモロジー群では点群データの各点の周りに半径 r の (中空でない) 球を考え、さまざまな次元の穴構造の生成と消滅を扱う。この過程を図 1 で概略的に説明する。2 次元平面内に 3 つの点からなる集合を考える。半径 r が小さい場合、これは点の周りに 3 つの円盤が置かれた構造である (図 1: 左)。次に r がある値より大きくなると、円盤が連結してサイクル構造が形成される (図 1: 中央)、これが穴の誕生とみなされる。さらに r が大きくなると、サイクル構造の内部が円盤で埋め尽くされて、穴は消滅する (図 1: 右)。ある穴が発生 (誕生) する r を x 座標、消滅 (死亡) する r を y 座標として、データに存在する全ての穴構造の生成と消滅をそれぞれプロットして構成されるのがパーシステンス図 (persistence diagram) と呼ばれるパーシステントホモロジー群の中心的な表現方法である。 n 次元空間中の

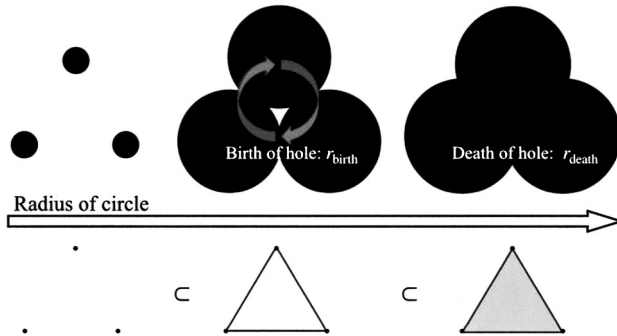


図 1. パーシステントホモロジー群の概念図 (Kono et al., 2020). 位相的情報は、離散的なデータ点からの半径 r の n 次元の球で構成された「穴」によって特徴づけられる。

点群データに対してパーシステンス図は、0次元の穴(連結成分)、1次元の穴(リング構造)、2次元の穴(シェル構造)…といった複数の次元の穴に対して個別に定義できる。より厳密な定義と方法論は2章で示す。

天体物理学や宇宙論における近代的な TDA の応用例はここ数年で徐々に増加している。Pranav et al. (2017, 2019a), Feldbrugge et al. (2019), Wilding et al. (2021), そして Biagetti et al. (2021) は一般の確率場への応用を念頭に置き、ガウスおよび非ガウス確率場の確率的ホモロジーに関する詳細な定式化を示した。この方法により、(Pranav et al., 2019b) は CMB に特異な信号を検出した。Cole and Shiu (2018) は、TDA を適用して CMB の非ガウス性を評価する方法を開発した。Xu et al. (2019) は、 N 体シミュレーション (Libeskind et al., 2018) から生成された大きな銀河データ集合に TDA を適用し、シミュレーションデータで 23 個のヴォイドを検出した。これは、TDA のヴォイド検出における有用性を示しており、従来の方法 (e.g., CLASSIC: Kitaura and Angulo, 2012) の結果ともよく一致している。さらに彼らは、MassiveNu シミュレーション (Liu et al., 2018) と比較して宇宙論的パラメータを制限し、パラメータ間の縮退を解く方法について検証した。Elbers and van de Weygaert (2019) と Giri and Mellema (2021) は TDA による宇宙再電離期の電離バブルの構造と進化、およびサイズ分布とトポロジーの定量的推定の有効性について議論した。

1.3 バリオン音響振動(BAO)

宇宙の構造形成は質量において支配的なダークマターが駆動する。しかし、電磁波で観測できる宇宙、即ち目に見える宇宙の姿を形作るのは電磁波と相互作用する通常の物質のなす構造である。宇宙論では通常の物質をバリオンとよぶ慣習があるので、本論文でも以降通常物質をバリオンと表現する。バリオンは、電磁相互作用(たとえば、輻射加熱/冷却、ガス圧、流体力学的過程)を介して複雑な進化をたどる。このような非自明な現象の典型例が、初期の宇宙のバリオン-輻射 2 流体で生成されたバリオン音響振動(BAO)である (Peebles and Yu, 1970; Sunyaev and Zeldovich, 1970)。ここでは、BAO と関連する観測量について簡単に説明する。

初期宇宙のバリオン-光子プラズマにおける点源による初期摂動を考える。プラズマでは、バリオンと光子がトムソン散乱(Thomson scattering)を解して結合し、単一の流体としてふるまう。光子は光速で動き、また非常に多数存在するため、バリオン-輻射流体は密度に対してとてつもなく大きな圧力を持つ。圧力は周囲と均一になろうとし、その結果外側に向かって伝播する球形の音波を生じる。この時期の音速は、次式によって評価される。

$$(1.9) \quad c_s = \frac{c}{\sqrt{3 \left(1 + \frac{3\rho_B}{4\rho_R}\right)}}.$$

ここで c は光速であり, ρ はエネルギー密度を, 下付き文字 B と R はそれぞれバリオンと輻射の成分を表す. 式(1.9)は, バリオンと輻射の脱結合前の音速が, 初期の光速の $\sim 57\%$ に及ぶことを示している. 音響的地平線(sound horizon)の最終半径 r_s は, 次の式で与えられる.

$$(1.10) \quad \begin{aligned} r_s &= \int_0^{t_{\text{dec}}} (1+z)c_s dt \\ &= \int_{z_{\text{dec}}}^{\infty} \frac{c_s}{H_0 \sqrt{\Omega_{R,0}(1+z)^4 + \Omega_{M,0}(1+z)^3 + \Omega_{\Lambda,0}}} dz. \end{aligned}$$

ここで z は宇宙論的赤方偏移, t は宇宙年齢, H_0 はハッブルパラメータ, $\Omega_{R,0}$ は輻射のエネルギー密度パラメータ, $\Omega_{M,0}$ は物質のエネルギー密度パラメータ, $\Omega_{\Lambda,0}$ は宇宙項の密度パラメータ, 下付き添字 dec は光子と物質の脱結合の時刻での値を, 0 は現在時刻での値を意味する. それぞれの詳細は付録 A.3 節を参照のこと. 我々が観測するのは, 大規模構造に刻印された, 原始揺らぎから生じた多くの音波の重ね合わせである. BAO は定義上バリオンに生じる現象だが, バリオンと暗黒物質の密度ゆらぎの構造は重力相互作用を通して作用し合うため, 最終状態では一致することが知られている (バリオンの追いつき (catch-up): e.g., Ma and Bertschinger, 1995; Eisenstein et al., 2007).

BAO のスケールは, 共動座標 (宇宙膨張による距離の拡大の影響を除いた座標系. 付録 A を参照) で一定である. このことから, BAO シグナルは基本的には銀河の 2 点相関関数で検出することができる. しかし, BAO スケールは銀河の典型的なスケールと比較して非常に大きいため, 2 点相関に現れる BAO シグナルはかなり弱い. したがって, BAO シグナルを測定するには, 広大な観測体積において高密度でサンプリングされた大規模銀河探査が必要とされる.

Sloan Digital Sky Survey (SDSS) は, 全天の 3 分の 1 をカバーする最も広い可視測光・分光探査である. SDSS により, 2 点相関関数で約 150 Mpc の BAO シグナルが検出された (Eisenstein et al., 2005). ここで単位 [pc] はパーセクを意味し, $1 \text{ pc} = 3.08 \times 10^{18} \text{ cm}$ である. Mpc, Gpc はそれぞれメガパーセク, ギガパーセクと読み, 10^6 pc , 10^9 pc を意味する. とはいえ, 伝統的方法による BAO 解析は SDSS のような大規模広域探査データでのみようやく可能になる.

よって, データへの要求がより少なくてすむ BAO シグナルの検出法および定量化の開拓は, 宇宙論モデルの制限において極めて重要な課題である. 本研究ではまず TDA を導入し, 銀河の空間分布データへの応用を紹介する. そして有意性検定の方法についても議論し, TDA が BAO スケールの評価に有用であることを示す. さらに, TDA の長所である逆解析についても議論する. これは, 銀河分布上に刻まれたバリオン音響振動に対応するトポロジー構造の視覚化を可能にする初めての方法である.

本論文は次のように構成されている. 2 章では, TDA の概念と方法, 逆解析を紹介する. 次に, 3 章ではバリオン物理ありとなしの場合のシミュレーションデータを用い, BAO 研究の道具としての TDA の有効性を検証する. 4 章では SDSS DR14 (Pâris et al., 2018) の extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al., 2016) Large Scale Structure (LSS) catalog (Ata et al., 2018) に TDA を適用した結果を示す. 5 章では, 本論文で導入した方法論の結果と将来展望についてのまとめと議論を述べる. 本稿の物理的内容は Kono et al. (2020) の内容と議論を基礎にしている.

本論文では *Planck* 衛星による CMB 観測で推定された宇宙論パラメータ $h = H_0/100 \text{ [km s}^{-1} \text{ Mpc}^{-1}] = 0.6766$, $\Omega_{\Lambda,0} = 0.6889$, $\Omega_{M,0} = 0.3111$, $\ln(10^{10} A_s) = 3.047$ を採

用する。この物理的意味については付録 A.3 節を参照していただきたい。

2. 位相的データ解析 (TDA)

TDA では、 N 次元ユークリッド空間 \mathbb{R}^N 中の有限個の点群 P の位相幾何学的性質を検証する (この論文全体を通して、 \mathbb{R} , \mathbb{N} と \mathbb{N}_0 は、それぞれ実数、自然数、および非負自然数の集合を表す)。点群 P が有限集合であるため、伝統的なトポロジーの議論を単純に適用しても P の特徴に関する重要な情報を抽出することはできない。TDA は、スケール (e.g., Edelsbrunner et al., 2002; Boissonnat et al., 2018) を変えてもある程度保持される (persistent) 性質を調べることにより、点群データが持つ意味のある位相幾何学的特徴を抽出できる。このため、TDA で基本的な役割を果たすこのような幾何学的モデルを導入する。この解析の主要な道具がパーシステントホモロジー群である。

2.1 パーシステントホモロジー群

2.1.1 点データの「形」

複雑な構造を持つ大規模データの分析では、データの性質を代表する少数の特徴量によってデータを縮約して表現することが重要である。このため、次元縮小 (dimensionality reduction) が中心的役割を果たす。例えば、以下のようなユークリッド空間に設定された点群データを考える。

$$(2.1) \quad \mathcal{P} \equiv \{x_i \in \mathbb{R}^N \mid i = 1, \dots, m\}.$$

例えばデータ \mathcal{P} に周期的な構造がある場合はフーリエ変換が非常にうまく機能し、データ点がクラスタリングしている場合は相関関数による解析が有効である (e.g., Peebles, 1980)。

しかし、このようなアプローチは、非周期的、あるいは明確なクラスター構造を持たない複雑なデータでは有効ではない。フーリエ変換や相関関数を用いる従来の解析方法では、このような点群の形状をうまく特徴付けることは難しい。代わりに、各データ点を中心とする球を置き、その球の結合関係としてデータ点の接続関係を導入することでデータの特徴付ける方法が有効である。例えば接続関係を穴構造の数によって特徴づけると、それはドーナツとコーヒーカップを同一視するような同相変換に対して不変なホモロジー群として、点群データを大胆に縮約して表現したことに対応する。一方これでは、穴の大きさのようなスケールの情報が失われてしまう問題がある。そこで前述の、各データ点に置いた球の半径を大きくしていく際の穴構造の生成・消滅を抽出することが有用となる。これによって、穴の形状やサイズといった情報を保持した点群データの縮約表現が得られる。これがホモロジー群を拡張した“パーシステントホモロジー群”である。

2.1.2 点群データにおける「穴」の代数表現

前述の通り、点群データの分析ではランダムパーコレーションのように、各点の周りに球を置いて半径を増加させ、半径の関数として考えている体積内の連結成分の穴構造を調べることが有用である。このような点群を元に構成された球の集合は、球という凸閉集合の集合と見なすことができる。凸閉集合の有限個の集まりで被覆された構造があった場合、その構造はトポロジカルな特徴を保存 (ホモトピー同値) しながら、代数操作が可能な単体複体による表現に変換できる。単体複体は図形の構造を表現したものであり、 k 次元の三角形に対応する概念である単体の組み合わせとして構成される。

一般には単体複体は位相空間の構造を持たないが、ここではまず具体的なイメージを持ちやすいユークリッド空間を導入した単体複体について説明する。ユークリッド空間内において k

単体は以下で定義される.

Definition 1. (ユークリッド空間内での単体(simplex)) \mathbb{R}^n 内の $k+1$ 個の点 p_0, p_1, \dots, p_k で構成される k 個のベクトル $p_0\vec{p}_1, p_0\vec{p}_2, \dots, p_0\vec{p}_k$ が一次独立な関係にある場合, その $k+1$ 個の点によって k 次単体 $|p_0, p_1, \dots, p_k|$ は以下のように定義される.

$$|p_0, p_1, \dots, p_k| := \left\{ \sum_{i=0}^k \lambda_i p_i \mid \lambda_i \geq 0 (i=0, 1, \dots, k), \sum_{i=0}^k \lambda_i = 1 \right\}$$

具体的には, 0 次元の単体は点, 1 次元の単体は三角形, 2 次元の単体は四面体に対応する (図 2). この単体をずれなく綺麗に貼り合わせたものが単体複体である. ずれなく貼り合わせるという操作に基づき, ユークリッド空間内における単体複体は以下のように定義される.

Definition 2. (ユークリッド空間内での単体複体(simplicial complex)) \mathbb{R}^n 内の有限個の単体の集まり K が以下の条件を満たすとき, K を単体複体と呼ぶ.

- (1) K に属する単体 τ の面 σ が K に含まれる.
- (2) 2 個の単体 $\tau, \sigma \in K$ の共通部分 $\tau \cap \sigma$ は空集合でない限り, τ の面かつ σ の面である.

ここで単体の面とは, 直観的には 4 面体なら境界(表面)の 4 枚の 3 角形という 1 次元低い単体の集合, 線分なら両端の 2 点という 0 次元単体の集合を指す. 図 3 にユークリッド空間内での単体複体の例を示す.

前述したように, 単体複体そのものは位相空間によらず, 以下のような組み合わせ論的構造として定義される.

Definition 3. (単体複体(simplicial complex)) 有限集合 V 上の単体複体 X とは, 次の条件

- (1) $\{a\} \in X$ ($\forall a \in V$),
- (2) $\sigma \in X$ のとき, $\tau \subset \sigma$ ならば $\tau \in X$.

を満たす V の部分集合の族である.

X における単体の最大次元は複体 X の次元と呼ばれ, $\dim X$ と書く. また, X の k 単体の集合を X_k と表記する.

次に, 点群データから単体複体を構成する手順を説明する. より具体的には, 本研究における「穴」の抽出では, 単体複体の一種である Čech 複体を構成する.

\mathcal{P} 上の Čech 複体は次のように定義される.

Definition 4. (Čech 複体(Čech complex)) 点群 $\mathcal{P} \in \mathbb{R}^N$ を考える. それぞれの x_i の周りに半径 r の球

$$(2.2) \quad B_r(x_i) = \{x \in \mathbb{R}^N; \|x - x_i\| \leq r\},$$

を置く. ここで $\|x\|$ はユークリッドノルムである. Čech 複体 $\mathcal{C}(\mathcal{P}, r)$ とは, \mathcal{P} 上の単体複体で, それを構成する単体が

$$(2.3) \quad \{x_{i_0}, \dots, x_{i_k}\} \in \mathcal{C}(\mathcal{P}, r) \Leftrightarrow \bigcap_{j=0}^k B_r(x_{i_j}) \neq \emptyset$$

で定義されるものである.

すなわち点群 \mathcal{P} の周りに置いた半径 r の球の関係(重なり)から単体を構成する. 半径 r を変

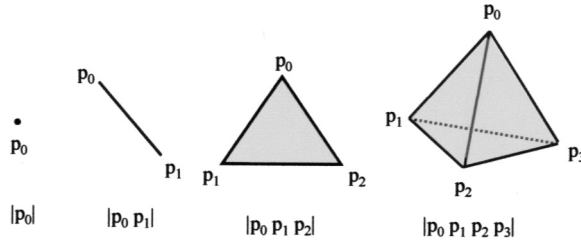


図 2. 単体 (simplex) の例. 図中の影をつけた部分は中身が詰まっていることを示している.

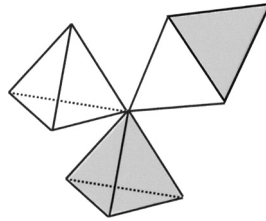


図 3. 単体複体 (simplicial complex) の例. 図 2 と同様に, 影をつけた部分は中空ではなく中身の詰まった図形であることを示す.

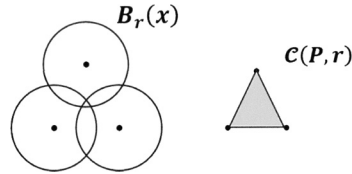


図 4. Čech 複体の例. 左パネル: x_i の周りの半径 r の球の和集合 B , 右パネル: 対応する Čech 複体 C .

化させたときの Čech 複体の振舞を図 1 に示す. 図 1 からわかる通り, 球の和集合と対応する Čech 複体は連続変形によって互いに移りあえる. これは, トポロジーの基本的定理の一つである脈体定理 (Nerve Theorem) から導かれる (e.g., Kozlov, 2008). つまり, 点群の連結性をモデル化した半径 r の球の和集合 $\bigcap_{j=0}^k B_r(x_{ij})$ と単体複体 $C(P, r)$ は, 連続変形によって互いに変形できることが保証される (ホモトピー同値). したがって, 単体複体には穴に関する情報が保存される. このようにして, 点群データから代数操作が可能な「穴」の表現が実現される.

Čech 複体はデータ \mathcal{P} と半径 r によって決まる. 半径 r が小さいと Čech 複体は元のデータ点群 \mathcal{P} と同じ離散構造を持つ. 一方 r を増大させるとデータ点は球によって連結し始め, r を十分大きく取ると最終的に全体が連結する (図 4). つまり前述のように, r はデータ点間の連結性を扱うための「解像度」とみなすことができる. 半径 $r_1 < r_2 < \dots < r_n$ に対応して Čech 複体の増大列 $C(\mathcal{P}, r_1) \subset C(\mathcal{P}, r_2) \subset \dots \subset C(\mathcal{P}, r_n)$ が得られる. 一般にこのような単体複体の増大列をフィルトレーションと呼ぶ.

実際の応用では, Čech 複体と共通の性質を持ちつつ計算機上での扱いに適した alpha 複体を使用されることが多い. Alpha 複体は, ヴォロノイ分割 (Voronoi tessellation) を用いた Čech 複体の改良版である. ここでは alpha 複体の厳密な導入はせず, 図 5 で概念のみ示すにとどめる. 以下の解析では alpha 複体を採用する.

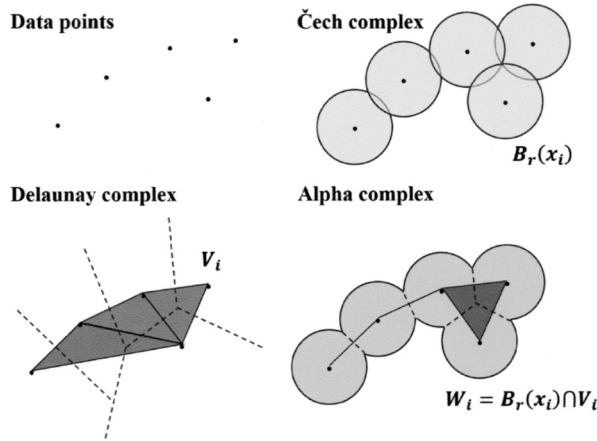


図 5. 点データ集合の例(左上), その Čech 複体(右上), Delauney 複体(左下), alpha 複体(右下) (Kono et al., 2020).

2.1.3 ホモロジー

次に, 得られた単体複体から「穴」構造を代数的に算出する方法を説明する. 穴構造は, ホモロジー群として表現される. ホモロジー (homology) 群は, 穴に対応する幾何学的意味を持つ代数的な対象である. ホモロジー群の代数的計算方法は複数あるが, その中の 1 つについて, 以下に概説する.

単体複体 X に含まれる全ての頂点の集合を $V = \{1, \dots, m\}$ とする. X に含まれる全ての k 単体の集合を X_k とする. X_k に含まれるある k -単体 $\sigma_k = |v_0, \dots, v_k|$ の頂点に順序を設定する. つまり, 頂点集合に順序 $v_0 < \dots < v_k$ を設定する. このようにして向きを指定した k -単体を $\langle \sigma_k \rangle \equiv \langle v_0, \dots, v_k \rangle$ とする. この向きを指定した k -単体 $\langle \sigma_k \rangle$ を基底として整数 \mathbb{Z} を係数とするベクトル空間を導入する.

$$(2.4) \quad C_k(X) \equiv \left\{ \sum_{\sigma_k \in X_k} a_{\sigma_k} \langle \sigma_k \rangle \mid a_{\sigma_k} \in \mathbb{Z} \right\},$$

ここで, $k > \dim X$ あるいは $k < 0$ ならば $C_k(X) = 0$ である.

単体複体の穴を定義するため, サイクル構造を導入する. サイクル構造とは, 結合した単体を辿って 1 周回って元に戻ってくる構造である. k 次元の穴は, X_k に含まれる k 次元の単体を組み合わせてできるサイクル構造で, X_{k+1} に含まれる $k+1$ 単体の境界となっていない構造のことを指す. 従って, サイクル構造を抽出するには, X_k から境界のない k 単体の集合を見つければよい. このように, 穴構造の抽出には単体の境界を算出することが必要となる. 向きが指定された k -単体 $\langle \sigma_k \rangle = \langle v_0, \dots, v_k \rangle$ の境界は, 以下で定義される境界作用素 ∂_k を適用することで得られる.

Definition 5. (単体の境界作用素) 境界作用素 ∂_k は

$$(2.5) \quad \partial_k \langle v_0, \dots, v_k \rangle \equiv \sum_{i=0}^k (-1)^i \langle v_0, \dots, \check{v}_i, \dots, v_k \rangle,$$

$$(2.6) \quad \partial_k \sum_{\sigma_k \in X_k} a_{\sigma_k} \langle \sigma_k \rangle = \sum_{\sigma_k \in X_k} a_{\sigma_k} \partial_k \langle \sigma_k \rangle$$

と定義される(ここで \tilde{v}_i は i 番目の成分を取り除いたことを表す).

前述したように, 単体複体 X に含まれる k 次元の穴は, X に含まれる境界のない k -単体の集まりで, X に含まれる $k+1$ -単体の集まりの境界になっていないもののことをいう. サイクルに対応する X に含まれる境界のない k -単体複体 c の集合は,

$$(2.7) \quad \text{Ker } \partial_k := \{c \in C_k(X) \mid \partial_k(c) = 0\}$$

与えられる. また, X に含まれる $k+1$ -単体の境界の集合は,

$$(2.8) \quad \text{Im } \partial_{k+1} := \{c \in C_k(X) \mid c = \partial_{k+1}(c'), c' \in C_{k+1}(X)\}$$

与えられる. 従って, k 次元のサイクルで $k+1$ 次元の単体複体の境界でないものは, これらの商 $\text{Ker } \partial_k / \text{Im } \partial_{k+1}$ をとることで与えられる.

このように, 単体複体 X に存在する k 次元の穴構造に対応する数学的意味を持つ, k 次ホモロジー群は以下のように定義される.

Definition 6. (ホモロジー (homology)) 商ベクトル空間

$$(2.9) \quad H_k(X) = \text{Ker } \partial_k / \text{Im } \partial_{k+1}$$

は単体複体 X の k 次ホモロジー群である. なお $k=0$ のホモロジー群 H_0 と Hubble パラメーター H_0 とを混同しないように注意されたい.

このように, データ \mathcal{D} に内在する穴の情報は, その Čech 複体 $\mathcal{C}(\mathcal{D}, r)$ のホモロジー群を計算することで抽出できる. しかし, この方法では穴のスケール情報を抽出することはできず, またある半径 r_2 で穴が出現し, 半径 r_3 で消滅するといった穴のパーシステンス (persistence) についての検証をすることもできない. こういった情報を抽出するためには, パーシステントホモロジー群を導入する必要がある.

2.1.4 パーシステントホモロジー群

穴の生成消滅の情報を持つパーシステントホモロジー群も, ホモロジー群と同様に代数的演算によって求められる. そのためまず, 半径 r を増大させ, 新たな単体に加わった半径を記録する. また, ちょうど t 番目 ($0 \leq t \leq n$) の単体が追加された半径 r_t を用いた場合の単体複体を $X^t := \mathcal{C}(\mathcal{D}, r_t)$ として記録する. このようにして, 単体の発生回数 n の数の半径 r_t と単体複体 X^t のセットが得られる. 次に, この r_t と X^t を用いて, 半径 r の増大によって生成・消滅する穴構造を代数的に算出することを考える. $X^t \subset X^{t+1}$ であることから, 穴の消滅は必ずその穴を境界とする高次の単体複体の誕生により生じる. 従って, 図形構造の変化を捉える半径としては, 単体の発生時刻を保持すれば十分である. そこで, 単体が発生した時刻に, その単体を要素とする基底空間を設定する. 例えば, r_t で k 単体 $\langle \sigma_k \rangle$ が誕生したとすると, それは以下のように表現される.

$$(2.10) \quad e_{\langle \sigma_k \rangle} = \begin{pmatrix} 0, & 0, & \dots, & \langle \sigma_k \rangle, & \dots, & 0, & 0 \\ 1 & 2 & & t & & n-1 & n \end{pmatrix}$$

単体複体 X^t は, このような基底空間で表現された単体の和として代数的に表現することができる.

ホモロジー群の算出と同様に, 上で与えた基底空間で表現された単体複体からサイクルや境界の発生・消滅を算出することで, 穴構造の発生・消滅を抽出することを考える. その際に重要な点は, 境界に対応する単体・単体複体の発生とサイクルのような図形構造に対応する単体

複体の発生する半径が異なるという点である．単体の発生半径の情報しか持たない基底空間の元で，サイクル構造の存在を算出するために，基底に作用して時刻を進展(t を増大)させる作用素 x を導入する．つまり，さきほどの単体 $e_{\langle\sigma_k\rangle}$ に x を作用させると，以下のような変化を与える作用素である．

$$xe_{\langle\sigma_k\rangle} = x(0, 0, \dots, \langle\sigma_k\rangle, \dots, 0, 0) = (0, 0, \dots, 0, \langle\sigma_k\rangle, \dots, 0, 0)$$

$$\begin{matrix} & 1 & 2 & & t & & n-1 & n & & 1 & 2 & & t & t+1 & & n-1 & n \end{matrix}$$

この作用素を用いた上で，境界の生成・消滅の情報を算出する境界作用素が以下で与えられる．

$$(2.11) \quad \partial_k(e_{\langle\sigma_k\rangle}) = \sum_{i=0}^k (x^{b(\sigma_k)-b(\sigma_{k-1}^i)})e_{\langle\sigma_{k-1}^i\rangle}, \quad \sigma_k \in X_k,$$

ここで， $b(\cdot)$ は，その単体が生じた時刻 t を返す関数， $\langle\sigma_{k-1}^i\rangle$ は， k 単体 $\sigma_k = \langle v_0, \dots, v_k \rangle$ の面 $\langle v_0, \dots, \check{v}_i, \dots, v_k \rangle$ を表すとした．パーシステントホモロジー群の境界作用素 ∂_k に対して，ホモロジー群のときと同様に，サイクルに対応する $\text{Ker } \partial_k$ と $k+1$ 次元の単体複体の境界に対応する $\text{Im } \partial_{k+1}$ の商を取ることで，生成時刻と生成から消滅までの区間情報を保持した穴構造が抽出される．このようにして，穴の生成・消滅の情報を保存するパーシステントホモロジー群が算出される．

一般的には，パーシステントホモロジー群 $PH_k(\mathcal{X})$ は X^i のホモロジー群 $H_k(X^i)$ の間の写像列として定義される．

Definition 7. (パーシステントホモロジー (persistent homology) 群) k 次のパーシステントホモロジー群 $PH_k(\mathcal{X})$ は

$$(2.12) \quad PH_k(\mathcal{X}) : H_k(X^1) \longrightarrow \dots \longrightarrow H_k(X^i) \longrightarrow \dots \longrightarrow H_k(X^n)$$

として定義される．

ここで，ある単体 $\sigma \in X^n$ が誕生する「時刻」を $t_\sigma \equiv \min \{t | \sigma \in X^t\}$ とした．

2.1.5 パーシステンス図

パーシステントホモロジー群 $PH_k(\mathcal{X})$ について， $b_i, d_i \in \{1, \dots, n\}$ ($b_i \leq d_i$) および非負整数 ℓ_k が存在し， $PH_k(\mathcal{X})$ がホモロジー群の構造を保存しながら以下のように一意に分解できることが知られている．

$$(2.13) \quad PH_k(\mathcal{X}) \simeq \bigoplus_{i=1}^{\ell_k} I[b_i, d_i].$$

ここで，区間 $I[b, d]$ は k 次元の穴が X_b で誕生し， X_d で死滅することを表現し， \simeq はホモトピー同値を表すものとした．

式(2.13)で表される直和分解より，パーシステントホモロジー群は誕生と死滅の対 (b_i, d_i) (パーシステント対)によって一意に定まるとわかる．パーシステント対の集合をプロットしたパーシステンス図 (persistence diagram) が定義できる．

Definition 8. (パーシステンス図 (persistence diagram))

$$(2.14) \quad D_k(\mathcal{X}) = \{(b_i, d_i) | i = 1, \dots, \ell_k\}$$

を \mathcal{X} の k 次パーシステンス図と呼ぶ．

パーシステント対 (誕生-死滅対) においては $b_i < d_i$ であるため，パーシステンス図のすべての

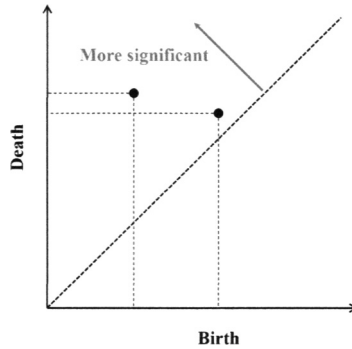


図 6. パーシステンス図の模式図. 対角線近くに位置するデータ点は位相的にはノイズと見なされる一方, 対角線から遠い点はデータ点群の持つ有意でロバストな構造を表す.

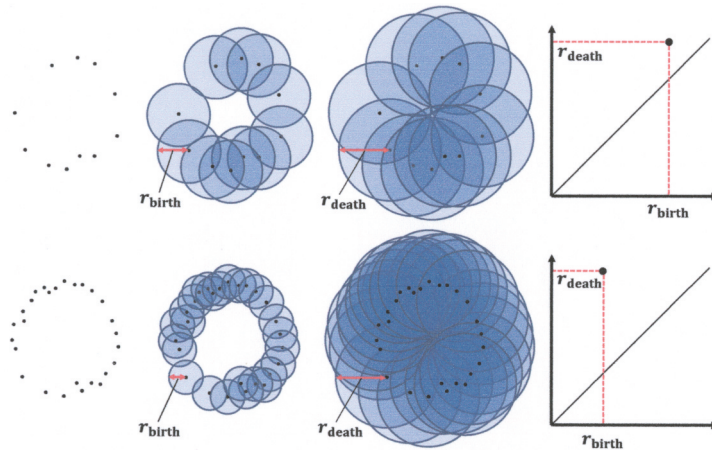


図 7. 点データの密度がパーシステンス図上の構造の生成(birth)および消滅(death)に与える影響を表した概念図 (Kono et al., 2020). 左から右の順に, 点データ, データ上に半径 r_{birth} の球(ball)を描いた図, 半径 r_{death} の球を描いた図, および対応する H_1 のパーシステンス図を示す. 上段の 4 つの図が低密度データ点の場合, 下段が高密度の場合である.

点是对角線よりも上に存在する. 定義により, 対角線の近くにある点の「寿命」は短く, 誕生後すぐに死滅することが示されている. 対照的に, 対角線から遠く離れた点は長く生存できる穴(生成元)を表す. 図 6 はこれを端的に示している. 本研究は銀河分布の空間構造を議論するため, 以下では誕生と死滅に対応する半径を b と d ではなく r_{birth} および r_{death} と表記する. 直感的には, r_{birth} は H_k を構成する点の平均間隔を表し, r_{death} は構造のサイズを表している.

パーシステンス図の重要な特性の 1 つについて言及する. リング状の構造を持つ点群データを考える. 図 7 ではこのような点群データにおける 2 つの場合のパーシステンス図が示されている. 上のパネルはデータが低密度の場合のパーシステントホモロジー群と対応するパーシステンス図, 下のパネルは高密度の場合の同じ図を示している. 低密度のデータは, 実際には図 7 の高密度データの部分集合となっている. つまり, 2 つのデータはリング構造に関して基本的に同じ情報を共有している. ただし, 高密度データの r_{birth} and r_{death} の間隔は低密度

データの間隔よりも小さい。これは、2つのデータが同じ構造を持つ場合、データの密度が高い(サンプリングレートが高い)ほど検出された特徴が有意に長い寿命を持つ可能性が高くなることを示唆する。図 7 では低密度データの r_{birth} が大きく r_{death} に近いことが見て取れる。これは r_{birth} が上記のデータ点間の平均距離の指標であることの直接的な帰結である。対照的に、 r_{death} は注目する構造のサイズの指標であるためほとんど変化しない。この事実は、次章以降で示すシミュレーションおよび観測データの解析で検出された特徴の有意性についての議論と関連する。

2.1.6 ボトルネック距離

様々な応用において、異なる 2つのパーシステンス図 D_1 と D_2 の差異を比較する方法が必要になる。パーシステンス図はベクトルとならないため、この比較には注意が必要である。これを可能にする方法の 1つがボトルネック距離(bottleneck distance)である (e.g., Edelsbrunner and Harer, 2010; Kerber et al., 2017; Cohen-Steiner et al., 2007; Chazal et al., 2008)。

Definition 9. (ボトルネック距離(bottleneck distance)) ボトルネック距離は

$$(2.15) \quad d_b(D_1, D_2) \equiv \inf_m \sup_{x_1 \in |D_1|} \|x_1 - m(x_1)\|_\infty$$

と定義される。ここで $\|x\|_\infty$ はチェビシェフ距離(Chebyshev distance) $L^\infty \equiv \max\{|x_1|, |x_2|\}$ で、 $|D|$ はパーシステンス図 D に対角線 (x, x) 上を重なりを許して埋め尽くす無限個の生成元の集合 Δ を加えた集合 $|D| \equiv D \cup \Delta$, m は $|D_1|$ と $|D_2|$ の間の全単射である。対角線の多重集合 Δ は、対応する相手のない生成元の対応相手として用意された、寿命が 0 の生成元の集合である。これによって、2つのパーシステンス図の生成元の個数が異なるような場合でも直接比較が可能となる。

ボトルネック距離の概念を図 8 に示す。2つのパーシステンス図が完全に同一である場合、ボトルネック距離は最小値となる。ここでパーシステンス図の距離の安定性を定義する。

Definition 10. (パーシステンス図の距離の安定性) ハウスドルフ距離 D_H を

$$(2.16) \quad D_H = \max \left\{ \sup_{x \in X} (x, Y), \sup_{y \in Y} (y, X) \right\}$$

とする。パーシステンス図の距離 d が安定とは、ある定数 L が存在し、任意の点集合 $X = \{x_i\}$, $Y = \{y_j\}$ に対し

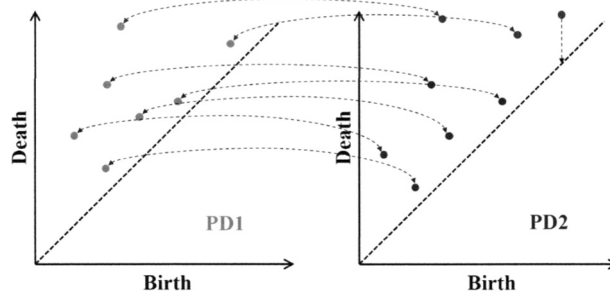


図 8. 2つのパーシステンス図間のボトルネック距離を計算する方法の概念。他方のパーシステンス図に対応する点がない点は、対角線に存在するとみなして計算する。

$$(2.17) \quad d(D_1, D_2) \leq LD_H(X, Y)$$

が成り立つことをいう (Lipschitz 連続性)。

ポトルネック距離はこの意味での安定性が保証されており (Cohen-Steiner et al., 2007), パーシステンス図の検証のために用いることができる。

2.2 パーシステンス図の逆解析

2点相関関数による従来の BAO 解析では, データにおける宇宙論的な構造の位置や形状を特定することはできない。これは, この種の解析が相関の統計的な特徴のみを表現するため, データの持つ平均的な特徴しか抽出できないからである。TDA では, 得られたパーシステンス図の各生成元と対応する図形構造を特定することができる。このような逆解析は, 解析結果の物理的議論を進める上で有用である。半径 r についてのフィルトレーションを通じて, 対応する単体複体 $\sigma(r)$ の位相的情報は保持される。この場合, 逆解析の目的は特異的に長寿命な穴 (生成元) $(r_{\text{birth},i}, r_{\text{death},i})$ (つまりパーシステンス図に現れた有意な点に対応するデータ点群の特徴) に対応する単体複体 $(\sigma(r_{\text{birth},i}), \sigma(r_{\text{death},i}))$ を見つけることである。例えば, 穴が消滅する際の穴を構成する単体複体やそのときの半径 $r_{\text{death},i}$ から, その穴の特徴的な構造や位置, スケールといった情報が推定できる。本研究では R TDA パッケージを用いて逆解析を行った。R TDA では, 逆解析だけでなく, 異なるパーシステンス図の間の距離, パーシステンス図にある穴構造の存在についての検定などを実施することができる。R TDA はまた C++ で書かれている位相的データ解析アルゴリズム GUDHI (The GUDHI Project, 2015), Dionysus (Morozov, 2007), PHAT (Bauer et al., 2012) のインターフェースも与える。

分析対象である 3次元点群データから検出された 3次元の穴 (シェル) 構造の逆解析のために, R パッケージ SCHU (<https://github.com/xinxuyale/SCHU>; Xu et al., 2019) を用いた。SCHU は Dionysus での位相的データ解析に最適化されている。これにより, パーシステンス図を元に検出された生成元に対応する特徴的な図形構造の位置と形状を特定することが可能になる。

逆解析法の研究は急速に発展しているため, 他のさまざまな洗練された方法が提案されている (たとえば Obayashi et al., 2018 を参照されたい)。様々な方法の比較は本研究の範囲外であり, 今後の研究で行う予定である。

3. 宇宙論的シミュレーションを用いた解析法の性能評価

3.1 データ

本研究で開発した方法の性能を検証するため, まず一連の暗黒物質のみの宇宙論的 N 体シミュレーションを実行し, データを生成した。この計算では公開 N 体シミュレーションコード Gadget-2 (Springel, 2005) を用い, 初期条件は Crocce et al. (2006) の方法を用いて生成した。初期条件は赤方偏移 $z = 20$ に設定し, シミュレーションボックスは通常の BAO スケール ~ 150 Mpc よりも十分に大きい一辺 2 Gpc に設定した。BAO 検出のためには質量分解能は重要でないため, シミュレーションボックスには 256^3 個の粒子を配置した。これは, 粒子あたりの質量 $6.4 \times 10^{13} M_{\odot}$ (M_{\odot} は太陽質量) に対応する。このシミュレーションには冷たい暗黒物質 (cold dark matter: CDM) 粒子しか含まれていないが, 本研究で注目する時代は光子-バリオン脱結合のかなり後であり, また見る必要があるのは重力だけなので, BAO の特徴は初期条件にのみ刻印しておけば一般的に十分である。初期ゆらぎのパワースペクトルは, 線型ボルツマン方程式を数値的に解くコード CLASS (Lesgourgues, 2011) を用いて計算している。これを元に $z = 0$ でのシミュレーションのスナップショットを取り, パワースペクトル解析を行うこ

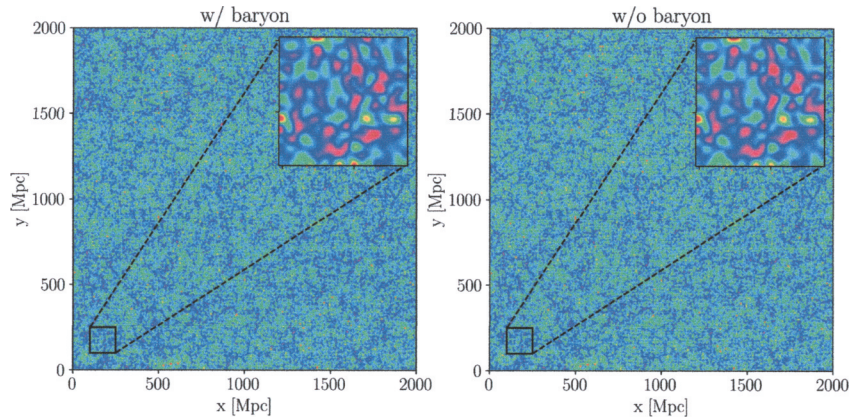


図 9. 赤方偏移 $z = 0$ でのバリオンあり(右)およびなし(左)の場合のシミュレーションデータ。インサートは 100–250 Mpc のボックス領域の拡大図を表す。密度揺らぎは z 軸方向に 100 Mpc の厚さの領域を射影しており、見やすくするために arcsinh をとった値を示している。

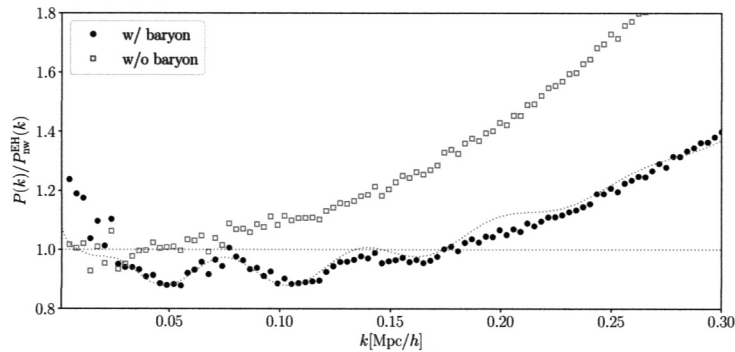


図 10. シミュレーションによる $z = 0$ でのパワースペクトル (Kono et al., 2020). Eisenstein and Hu (1998) のバリオン振動のない線型パワースペクトルで規格化されている。黒丸はバリオン振動ありのシミュレーションデータ、点線は同じ宇宙論的パラメーターのセットを用い、ハローフィットモデル (Takahashi et al., 2012) によって計算した解析解である。比較のため、バリオン振動なしのシミュレーションデータを四角で示す。バリオンありと比べ、 $k < 0.2 \text{ Mpc}/h$ において振動がみられない。

とによってデータに BAO が確かに存在することが確認できる。

コントロールサンプルとして、BAO なしで初期パワースペクトルを置き換えたまったく同じシミュレーションセットを生成した。これは初期パワースペクトルを計算する際、物質の総質量(CDM+ バリオン)を変更せずにバリオンの量を極限まで減らすことで実現している。すなわち、通常モデルでは $\Omega_B = 0.049$ であるが、コントロールサンプルは $\Omega_B = 0.002204$ とした。以降便宜上両者を BAO あり/なし、やバリオンあり/なしと表現し区別することにする。図 9 はバリオンがある場合とない場合のシミュレーションのスナップショットを示している。どちらのシミュレーションセットにも同じランダムシードを使用しているため、目視によって違いを確認することは非常に難しい。一方で、図 10 は両者のシミュレーションセットのバ

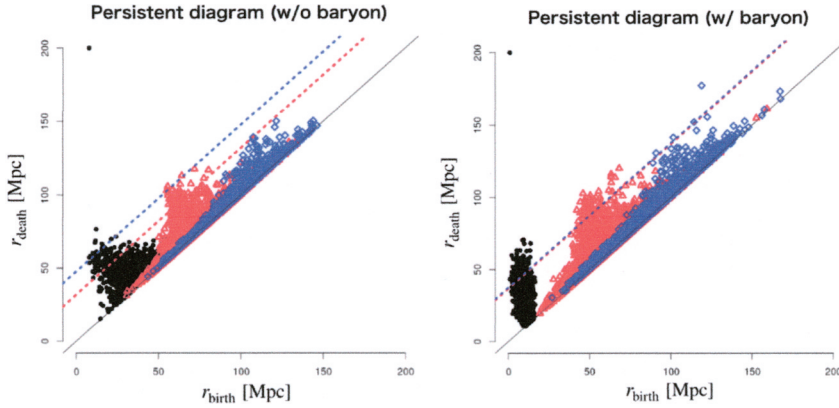


図 11. シミュレーションデータのパーシステンス図. 黒い点, 赤い三角形, 青いひし形はそれぞれ H_0 , H_1 , H_2 を表す. 黒い対角線は $r_{\text{birth}} = r_{\text{death}}$ である. 左パネルはバリオンなし, 右パネルはバリオンありシミュレーションの結果を示している.

ワースペクトルを示している. それぞれのワースペクトルは見やすさのために, BAO の振動的な特徴を平滑化したワースペクトル (Eisenstein and Hu, 1998) によって規格化されている. 予想されるとおり, 得られたコントロールサンプルのワースペクトルは BAO の信号を示さない. 本論文では, 計算機システム上の制約のため 256^3 粒子の親サンプルから 2000 個の粒子をランダムに抽出したサブサンプルを用いた. 今後はコードを並列化すると同時に計算機をアップグレードすることで, はるかに大きなデータセットが処理可能になる予定である.

3.2 シミュレーションデータの解析結果

シミュレーションデータのパーシステンス図を検証しよう. 検出された信号の統計的有意性を評価するには, 信頼区間を計算する方法が必要になる. しかし, パーシステンス図の信頼区間推定は自明な議論ではなく, いくつかの方法が提案されている (e.g., Fasy et al., 2014). そのひとつとして, ここではボトルネック距離を使用する方法を採用する. 本研究ではまず, ブートストラップリサンプリングでサンプルのセットを生成する. ここから対を抽出し, その 2 つのパーシステンス図の間の距離をボトルネック距離で測ることによって信頼区間を推定する.

シミュレーションで得られたパーシステンス図を図 11 に示す. 左のパネルはバリオンなし, 右のパネルはバリオンありの結果を表す. 黒い点, 赤い三角形, 青いひし形はそれぞれ 0 次 (H_0), 1 次 (H_1), 2 次 (H_2) のパーシステントホモロジー群に対応する. 赤と青の点線は, それぞれ H_1 と H_2 の 90-% 信頼区間である. これらパーシステンス図の信頼区間はリサンプリング回数 $N_{\text{boot}} = 30$ のブートストラップリサンプリングによる方法で計算したものである. この研究では, 対応する信頼区間の上にある点を統計的に有意であるとみなす.

図 11 の右パネルで示されているように, バリオンありデータから得られた H_2 には 4 個の有意な生成元, すなわち 3 次元の穴 (青いひし形) が検出された. 一方バリオンなしデータでは対応するスケールの有意な穴は検出されなかった.

バリオンありのサンプルで検出された有意な穴の r_{birth} と r_{death} を表 1 に示す. 検出された中空の穴の平均半径は r_{death} は 150.16 ± 8.46 [Mpc] であった. 検出個数は 4 つだが, 取得された \bar{r}_{death} は, BAO で期待される半径と一致する. 図 12 に検出された生成元に対応するデータ点群の構造の位置と形状を示す. この 3 次元構造は, 中空の穴となっている.

表 1. シミュレーションデータで検出された $p < 0.2$ の穴の r_{birth} と r_{death} .

No.	r_{birth} [Mpc]	r_{death} [Mpc]
1	116.34	130.65
2	127.42	141.80
3	140.17	151.70
4	144.60	176.47

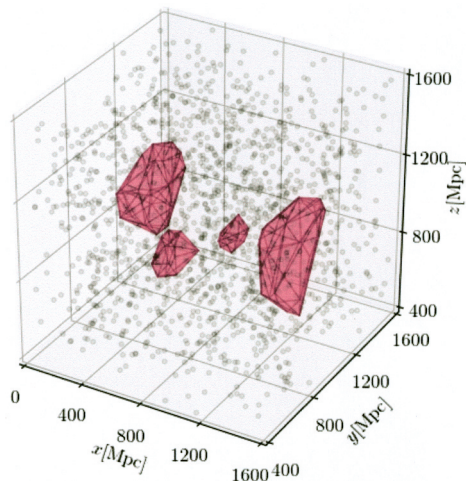


図 12. 逆解析の例. シミュレーションデータにおいて最も優位性の高い 4 個の H_2 に対応する実空間での構造を示す. 網掛け部分は H_2 に対応する構造のすべての構成粒子の凸包であり, 中空の構造となっている.

一方 H_1 に属する生成元(赤い三角形)については, 双方のデータで重要な特徴が得られた. この生成元は 2 次元のループ構造に対応する. バリオンありのデータでは $p < 0.2$ の 17 個の有意な H_1 が検出された. これら有意な生成元の平均 r_{death} は 99.00 ± 2.26 [Mpc] である. バリオンなしデータの H_1 では 34 個の生成元が検出され, $\bar{r}_{\text{death}} = 100.49 \pm 3.24$ [Mpc] となった. H_1 で検出された生成元に対応する図形構造の特徴的スケールの一致は, バリオンの存在によって影響されない 2 次元のループのような構造があることを示唆する. 大規模構造のフィラメントはこのような構造の候補の可能性がある. これとは対照的に, \bar{r}_{birth} はバリオンの有無で違いがあり, バリオンありで 46.13 ± 2.24 [Mpc], バリオンなしで $\bar{r}_{\text{birth}} = 62.34 \pm 2.84$ [Mpc] となった. この解析から, ループを構成する銀河の間の距離はバリオンの存在によって減少することが分かる.

そして, H_0 に属する生成元(黒い点)についてはバリオンの有無でパーシステンス図上の分布に明らかな違いが見られた H_0 は連結成分を現し, 銀河団やフィラメントなどの 1 次元的な連続構造に対応する. バリオンなしのデータでは r_{birth} は対角線に沿って幅広く分布するのに対し, バリオンありでは r_{birth} は低い値に集中している. バリオンありの場合のような r_{birth} が局在したような分布は, 連結構造がある特定の半径で誕生したことを意味する. このことは, 銀河団やフィラメント構造を構成する銀河の近某銀河間距離が均等であり, 従って特定の r_{birth} で一気に結合して連結構造を形成することを意味する. 一方で, r_{death} の分布はバリオンあり・なしに関わらず広く分布していた. このことは, 銀河団やフィラメント間の距離はバリ

オンの影響を受けないことを示唆する．このように H_0 のパーシステンス図の特徴は，バリオンの存在が銀河団やフィラメント構造内の近傍銀河間距離に何らかの制約を与える効果を持つことを示唆する．更なる解析を行うことで， H_0 と H_1 のパーシステンス図上の特徴を物理的に解釈できることが期待される．

4. Sloan Digital Sky Survey Data Release 14 による解析

4.1 SDSS 銀河探査データ

本研究では，Sloan Digital Sky Survey (SDSS) Data Release 14 (DR14) の extended Baryon Oscillation Spectroscopic Survey (eBOSS) (Dawson et al., 2016) 大規模構造カタログ (Ata et al., 2018) のクエーサーデータを用いた．これは SDSS-IV (Blanton et al., 2017) の一部である．観測されたバンド(波長帯)は u, g, r, i , および z である．天球上の探査領域は 2044 deg^2 ，赤方偏移は $0.8 < z < 2.2$ の範囲である．有効面積は北銀河半球(NGC)で 1288 deg^2 ，南銀河半球で 995 deg^2 となっている．クエーサーサンプルは Myers et al. (2015) によって選択されたデータを使用した．このサンプルにおける BAO の存在は Ata et al. (2018) によって検証され，すでに確認されている．

サンプル全体には 147,000 個のクエーサーが含まれているが，本研究ではシミュレーションデータと同様 2000 個の銀河をランダム抽出して用いた．

4.2 SDSS データ解析結果

SDSS DR14 データのパーシステンス図を図 13 に示す．図 13 より，SDSS DR14 データでは H_2 に属する 4 つの有意な生成元すなわち「シェル」が検出された．平均の r_{death} は $146.6 \pm 2.0 \text{ [Mpc]}$ である(表 2)．有意な生成元の数 は 4 個だが， \bar{r}_{birth} は BAO から予想される半径の期待値と一致している． H_2 の逆解析の結果を図 14 に示す．

この解析では， H_1 に属する 19 個の生成元が有意な穴(ループ)として検出されている．これらの生成元の平均 r_{death} は $101.82 \pm 3.54 \text{ [Mpc]}$ となり，シミュレーションデータで得られた \bar{r}_{birth} と一致している．また観測データの r_{birth} は $57.92 \pm 3.05 \text{ [Mpc]}$ となった．そして H_0 ホモロジーの分布は，興味深いことにバリオンなしシミュレーションの分布と類似している．これら低次元ホモロジーの解釈には今後のより詳細な研究が待たれる．

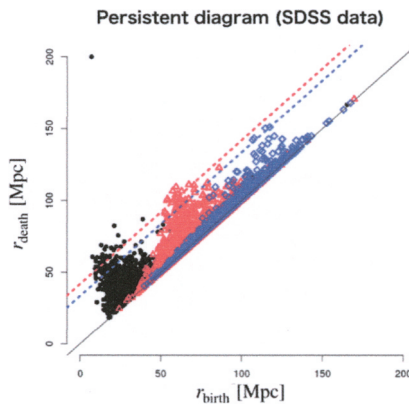
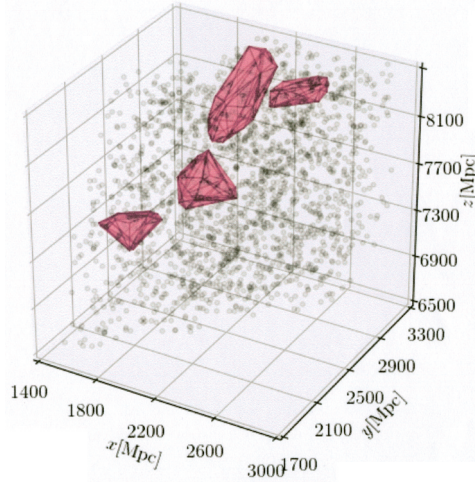


図 13. SDSS データのパーシステンス図．シンボルは図 11 と同様である．

表 2. 観測データ (SDSS DR14) で検出された $p < 0.2$ の r_{birth} および r_{death} .

No.	r_{birth} [Mpc]	r_{death} [Mpc]
1	107.00	141.98
2	107.83	143.30
3	114.86	150.20
4	117.56	150.98

図 14. 実際の銀河探索 (SDSS) データで検出された H_2 に属する 4 つの有意な生成元に対応する実空間構造. 網掛け部分の意味は図 12 と同様.

5. まとめと議論

本研究では、銀河分布の BAO 信号検出における位相的データ解析の有用性を検証した。本研究の位相的データ解析で対象としたパーシステントホモロジー群は、様々なスケールに対応した単体複体の集合列によって定義されたフィルトレーションを構成する。この性質はデータに内在する穴のスケールと形状を分析するのに特に有用である。

まず $N = 256^3$ の N 体シミュレーションを用い、パーシステントホモロジー群に対するバリオンの影響を検証した。バリオンの物理を入れた場合とない場合のシミュレーションで、パーシステンス図は大幅に異なることが示された。バリオンありのデータからは 2 次のパーシステンス図から 4 つの有意な生成元 (穴構造) が検出されたが、バリオンなしのデータからは検出されなかった。また H_1 と H_0 に属する生成元のパーシステンス図上での分布にも明らかな違いが見られた。検出された 1 次元の穴の \bar{r}_{death} は同程度の値を取るが、 \bar{r}_{birth} は大きく異なる。同様に 0 次元の穴の \bar{r}_{birth} は、バリオンの有無で大きく異なる傾向を示した。

次に、実際の観測データとして SDSS のクエーサー探索から抽出した高々 2000 サブサンプルで BAO 信号を検出することに成功した。得られた \bar{r}_{death} は H_2 に対して 146.6 ± 2.0 [Mpc] であり、BAO スケールと一致する。このことは、銀河分布の大規模構造検出においてパーシステントホモロジー群の計算コストが相関関数よりもはるかに安価であることを意味し、BAO の宇宙論的分析の新たな可能性を与える。

本研究の方法による BAO 検出は、伝統的な 2 点相関関数による解析と好対照をなす。2 点

相関関数の推定には、高密度にサンプリングされた銀河データが必要とされるが、位相的データ解析は～2000個の銀河のスパースサンプリングされたデータでBAOのような弱い信号を検出できる。これは、位相的データ解析を活用することで、スパースにサンプリングされた銀河探査も宇宙論的研究に使用できることを意味する。このことは次世代の宇宙論的銀河調査のデザインと戦略に新しい方向性を与えると期待される。

さらに、従来の解析と比較した位相的データ解析の著しい利点は、逆解析によりBAO構造を可視化できることである。パーシステンス図の逆分析が持つ機能は、従来の解析とは一線を画す明確な特長であり、強調に値する。たとえば本研究の直接的発展として、BAO信号に寄与する実空間での銀河分布構造を簡単に特定できるため、伝統的方法に比べさまざまな宇宙論的議論がずっと簡単に可能になる。ただし、最も単純な位相的データ解析はノイズに対して頑健でないという弱点があり、今後のより詳細な応用ではこの点を克服する必要がある。ロバストな位相的データ解析の方法は現在も研究されており（たとえば Vishwanath et al., 2022）。将来の研究ではその成果を取り入れられると期待できる。

謝 辞

本稿について大変貴重な改善点の指摘および助言を頂いた査読者のお2人、および担当編集委員の島谷健一郎氏には心からの感謝を申し上げる。池田思朗、福水健次、栗木哲の各氏からは本研究に関する貴重な議論および有益な示唆を頂いた。ここに深く感謝する。本研究はJSPS科学研究費助成事業(19H05076, 20H01932, 21H01128, 21J23611)、新学術領域研究「加速宇宙」(15H05890 および 16H01096)、住友財団2018年度基礎科学研究助成金(180923)、および統計数学研究所の共同利用「データサイエンス的方法による銀河進化研究の新展開」の補助を受けて行った。

付 録

A. 宇宙論の基礎

A.1 Friedmann-Lemaître-Robertson-Walker 計量とスケール因子

一般相対性理論は、時空がその幾何学的構造によって特徴付けられることを示し、微分幾何学が使用するべき基本的な枠組みであることを示した。微分幾何において、多様体上の局所的性質を記述する基本的な概念が線素(line element)あるいは計量(metric)である。Friedmann-Lemaître-Robertson-Walker (FLRW) 計量は、均一で等方的な時空の(局所的)幾何構造を表すために提唱された計量で、宇宙論のモデルとして広く用いられている。

$$(A.1) \quad ds^2 = g_{\mu\nu} dx_\mu dx_\nu = -c^2 dt^2 + a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right]$$

ここで $g_{\mu\nu}$ が計量テンソル、 t は時刻、 r, θ, ϕ は極座標の動径距離および角度方向の座標を表す。また K はガウス曲率、 $a(t)$ はスケール因子である。スケール因子は宇宙膨張による空間のスケールの変化を表現しており、宇宙論の慣習上 $a_0 \equiv a(t_0) = 1$ (t_0 : 現在の宇宙年齢)と規格化して用いられる。ここでは K は(長さ) $^{-2}$ の次元を持つように定義を採用する。この場合 r は長さの次元を持ち、 $a(t)$ は無次元になる。

スケール因子を導入すると、宇宙の進化による天体間の距離の変化から宇宙膨張による「自明な」距離の変化を除いた正味の変化分を表現することができる。即ち、距離をあらわすベクトルを \vec{r} とすると

$$(A.2) \quad \vec{r} = a(t)\vec{x}$$

と書ける。この \vec{x} は宇宙膨張の影響を除いた距離の指標であり、共動座標 (comoving coordinate) と呼ばれる。

A.2 赤方偏移と宇宙論パラメータ

過去のある時刻 $t = t_{\text{em}}$ で共動座標 $(r, \phi, \theta) = (r_e, 0, 0)$ で放射された光が、時刻 $t = t_0$ において原点 $(r, \phi, \theta) = (0, 0, 0)$ に位置する観測者に到達する状況を考える。相対性理論では、光はヌル測地線 (null geodesic) に沿って移動する。ヌル測地線とは $ds^2 = 0$ を満たす経路のことで、最小作用の原理で実現する軌道と考えてよい。式 (A.1) を用いれば、距離と時間を

$$(A.3) \quad \frac{cdt}{a(t)} = \frac{dr}{\sqrt{1 - Kr^2}}$$

のように関係付けられる。宇宙膨張の光への影響を議論するため、変数 r を

$$(A.4) \quad \chi \equiv \int_0^{r_{\text{em}}} \frac{dr}{\sqrt{1 - Kr^2}}$$

によって座標距離 χ に、時間 t を

$$(A.5) \quad \eta \equiv \int^t \frac{dt'}{a(t')}$$

によって共形時間 η に変換すると、光の満たすべき方程式は

$$(A.6) \quad c(\eta_{\text{obs}} - \eta_{\text{em}}) = \chi$$

となる。ここで η_{obs} は光が観測者に到達した時刻を表す。

時刻 $\eta = \eta_0$ および $\eta = \eta_0 + \delta\eta_0$ で放出された光が、それぞれ η_1 および $\eta_1 + \delta\eta_1$ で観測者に到達したとする。式 (A.6) の右辺は η によらないことから

$$(A.7) \quad \delta\eta_0 = \delta\eta_1$$

すなわち

$$(A.8) \quad \frac{\delta t_0}{a(t_0)} = \frac{\delta t_1}{a(t_1)}$$

となる。時間間隔 δt_0 および δt_1 において、光の位相が保存することから、

$$(A.9) \quad a(t_0)\nu_0 = a(t_1)\nu_1 \iff \frac{\lambda_0}{a(t_0)} = \frac{\lambda_1}{a(t_1)}$$

が得られる。ここで ν は光の振動数で、波長とは $\lambda\nu = c$ の関係がある。

膨張する宇宙では $a(t_0) > a(t_1)$ なので、観測される時刻での波長は放射のときよりも長くなる、即ち $\lambda_0 > \lambda_1$ となる。この波長の伸びが宇宙膨張による赤方偏移と呼ばれている。赤方偏移 z は

$$(A.10) \quad z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{a(t_0)}{a(t_1)} - 1$$

と定義される。

ここで、上記のようにスケール因子 $a(t)$ を $a(t_0) = 1$ となるよう規格化する。このとき $a(t)$ と z との関係は

$$(A.11) \quad z = \frac{1}{a} - 1,$$

あるいは

$$(A.12) \quad a(t) = \frac{1}{1+z}$$

となる。

A.3 フリードマン方程式と宇宙論パラメータ

ここまでの準備, およびアインシュタイン方程式 (Einstein's equation) を用いれば, 膨張する宇宙を記述するための「運動方程式」を導くことができる。アインシュタイン方程式とは, 時空の歪みとエネルギー密度を関連付けた一般相対性理論の基本方程式であり,

$$(A.13) \quad R^{\mu\nu} - \frac{1}{2}Rg^{\mu\nu} + \Lambda g^{\mu\nu} = \frac{8\pi G}{c^4}T^{\mu\nu}$$

と書かれる。ここで $R^{\mu\nu}$ はリッチテンソル (Ricci tensor), $R \equiv R^\alpha_\alpha$ はそのトレースでリッチスカラー (Ricci scalar), G はニュートンの重力定数, Λ は宇宙定数である。詳細は相対性理論の文献に譲るが, リッチテンソルは計量 $g_{\mu\nu}$ とその微分量の関数であり, 時空の歪みを表現する 2 階の共変テンソルである。計量が FLRW メトリックの場合, 時空の運動方程式は

$$(A.14) \quad \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{c^2K}{a^2} + \frac{c^2\Lambda}{3}$$

$$(A.15) \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{c^2\Lambda}{3},$$

となる。これはフリードマン方程式 (Friedmann equations) と呼ばれている。ここで ρ は宇宙の密度, p は圧力を現す。式 (A.14) はアインシュタイン方程式 (式 (A.13)) の 00 成分, 式 (A.15) はトレースから求められる。

さらに, 宇宙論パラメータと呼ばれる以下の量を定義すると, 宇宙膨張の物理量依存性がより具体的に記述できる。

- ハッブルパラメータ:

$$(A.16) \quad H(t) \equiv \frac{\dot{a}(t)}{a(t)}.$$

これは時間依存する量であるが, 観測的には現在値 H_0 が用いられることが多い。また無次元化した $h = H_0/100$ もよく用いる。

- 密度パラメータ:

$$(A.17) \quad \Omega_M(t) \equiv \frac{\rho(t)}{\rho_c(t)} \equiv \frac{8\pi G\rho(t)}{3H^2(t)},$$

ここで ρ_c は臨界密度パラメータと呼ばれる量で, 現在の宇宙では $\rho_c = 1.88 \times 10^{-29} h^2$ [g cm^{-3}] と測定されている。

- 無次元宇宙項パラメータ:

$$(A.18) \quad \Omega_\Lambda(t) \equiv \frac{c^2\Lambda}{3H^2(t)}.$$

- 曲率パラメータ:

$$(A.19) \quad \Omega_K(t) \equiv -\frac{c^2 K}{a(t)^2 H^2(t)}.$$

式(A.16)–(A.19)を用いると、式(A.14)は

$$(A.20) \quad \Omega_M(t) + \Omega_\Lambda(t) + \Omega_K(t) = 1$$

と表せる。これは一般の宇宙年齢 t で成り立つので、現在の宇宙年齢 t_0 では

$$(A.21) \quad \Omega_{M0} + \Omega_{\Lambda0} + \Omega_{K0} = 1$$

となる。

導入で述べているように、本論文では *Planck* 衛星による CMB 観測で推定された宇宙論パラメータ $h = H_0/100$ [$\text{km s}^{-1} \text{Mpc}^{-1}$] = 0.6766, $\Omega_{\Lambda,0} = 0.6889$, $\Omega_{M,0} = 0.3111$, $\ln(10^{10} A_s) = 3.047$ を採用している。曲率パラメータ $\Omega_{K0} = 0$ である。これらの値は、現在の宇宙は空間的に平坦で、宇宙定数(または暗黒エネルギー)により加速膨張している宇宙モデルを強く示唆する。パラメータ A_s は宇宙初期のホライズンスケールの揺らぎの振幅を決めるもので、シミュレーションデータを生成する際にこの値を採用している。

参 考 文 献

- Appleby, S., Chingangbam, P., Park, C., Yogendran, K. P. and Joby, P. K. (2018). Minkowski tensors in three dimensions: Probing the anisotropy generated by redshift space distortion, *Astrophysical Journal*, **863**(2), 200.
- Ata, M., et al. (2018). The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: First measurement of baryon acoustic oscillations between redshift 0.8 and 2.2, *Monthly Notices of the Royal Astronomical Society*, **473**(4), 4773–4794.
- Bag, S., Mondal, R., Sarkar, P., Bharadwaj, S. and Sahni, V. (2018). The shape and size distribution of H II regions near the percolation transition, *Monthly Notices of the Royal Astronomical Society*, **477**(2), 1984–1992.
- Bag, S., Mondal, R., Sarkar, P., Bharadwaj, S., Choudhury, T. R. and Sahni, V. (2019). Studying the morphology of H I isodensity surfaces during reionization using Shapefinders and percolation analysis, *Monthly Notices of the Royal Astronomical Society*, **485**(2), 2235–2251.
- Bauer, U., Kerber, M. and Reininghaus, J. (2012). PHAT, a software library for persistent homology, <https://bitbucket.org/phat-code/phat/src/master/>.
- Beisbart, C., Dahlke, R., Mecke, K. and Wagner, H. (2002). Vector- and tensor-valued descriptors for spatial patterns, *Morphology of Condensed Matter*, Lecture Notes in Physics, Volume 600, 238–260, Springer, New York.
- Bernardeau, F., Colombi, S., Gaztañaga, E. and Scoccimarro, R. (2002). Large-scale structure of the universe and cosmological perturbation theory, *Physics Reports*, **367**(1-3), 1–248.
- Biagetti, M., Cole, A. and Shiu, G. (2021). The persistence of large scale structures I: Primordial non-Gaussianity, *Journal of Cosmology and Astroparticle Physics*, **2021**(4), 061.
- Blanton, M. R., et al. (2017). Sloan Digital Sky Survey IV: Mapping the Milky Way, nearby galaxies, and the distant universe, *Astronomical Journal*, **154**(1), 28.
- Boerner, G. and Mo, H. (1989). A percolation analysis of cluster superclustering, *Astronomy and Astrophysics*, **224**(1-2), 1–9.
- Boissonnat, J.-D., Chazal, F. and Yvinec, M. (2018). *Geometric and Topological Inference*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge.
- Chazal, F., Cohen-Steiner, D., Glisse, M., Guibas, L. J. and Oudot, S. (2008). Proximity of persistence modules and their diagrams, Research Report, RR-6568, INRIA, 29.

- Chen, Z., Xu, Y., Wang, Y. and Chen, X. (2019). Stages of reionization as revealed by the Minkowski functionals, *Astrophysical Journal*, **885**(1), 23.
- Chingangbam, P., Ganesan, V., Yogendran, K. P. and Park, C. (2017). On Minkowski functionals of CMB polarization, *Physics Letters B*, **771**, 67–73.
- Cohen-Steiner, D., Edelsbrunner, H. and Harer, J. (2007). Stability of persistence diagrams, *Discrete & Computational Geometry*, **37**, 103–120.
- Cole, A. and Shiu, G. (2018). Persistent homology and non-Gaussianity, *Journal of Cosmology and Astroparticle Physics*, **2018**(3), 025.
- Croce, M., Pueblas, S. and Scoccimarro, R. (2006). Transients from initial conditions in cosmological simulations, *Monthly Notices of the Royal Astronomical Society*, **373**(1), 369–381.
- Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I, Probability and Its Applications*, 2nd ed., Springer-Verlag, Berlin.
- Dawson, K. S., et al. (2016). The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and early data, *Astronomical Journal*, **151**(2), 44.
- de Lapparent, V., Geller, M. J. and Huchra, J. P. (1986). A slice of the universe, *Astrophysical Journal*, **302**, L1–L4.
- Edelsbrunner, H. and Harer, J. (2010). *Computational Topology — An Introduction*, American Mathematical Society, Providence, Rhode Island, <http://www.ams.org/bookstore-getitem/item=MBK-69>.
- Edelsbrunner, H., Letscher, D. and Zomorodian, A. (2002). Topological persistence and simplification, *Discrete & Computational Geometry*, **28**(4), 511–533.
- Efstathiou, G. and Silk, J. (1983). The formation of galaxies, *Fundamentals of Cosmic Physics*, **9**, 1–138.
- Eisenstein, D. J. and Hu, W. (1998). Baryonic features in the matter transfer function, *Astrophysical Journal*, **496**(2), 605–614.
- Eisenstein, D. J., et al. (2005). Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies, *Astrophysical Journal*, **633**(2), 560–574.
- Eisenstein, D. J., Seo, H.-J., Sirko, E. and Spergel, D. N. (2007). Improving cosmological distance measurements by reconstruction of the baryon acoustic peak, *Astrophysical Journal*, **664**(2), 675–679.
- Elbers, W. and van de Weygaert, R. (2019). Persistent topology of the reionization bubble network — I. Formalism and phenomenology, *Monthly Notices of the Royal Astronomical Society*, **486**(2), 1523–1538.
- Fang, W., Li, B. and Zhao, G.-B. (2017). New probe of departures from general relativity using Minkowski functionals, *Physical Review Letters*, **118**(18), 181301.
- Fasy, B. T., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S. and Singh, A. (2014). Confidence sets for persistence diagrams, *The Annals of Statistics*, **42**(6), 2301–2339.
- Feldbrugge, J., van Engelen, M., van de Weygaert, R., Pranav, P. and Vegter, G. (2019). Stochastic homology of Gaussian vs. non-Gaussian random fields: Graphs towards Betti numbers and persistence diagrams, *Journal of Cosmology and Astroparticle Physics*, **2019**(9), 052.
- Ganesan, V. and Chingangbam, P. (2017). Tensor Minkowski functionals: First application to the CMB, *Journal of Cosmology and Astroparticle Physics*, **2017**(6), 023.
- Gay, C., Pichon, C. and Pogosyan, D. (2012). Non-Gaussian statistics of critical sets in 2D and 3D: Peaks, voids, saddles, genus, and skeleton, *Physical Review D*, **85**(2), 023011.
- Giri, S. K. and Mellema, G. (2021). Measuring the topology of reionization with Betti numbers, *Monthly Notices of the Royal Astronomical Society*, **505**(2), 1863–1877.
- Gott, I., Gott, J. R., III, Melott, A. L. and Dickinson, M. (1986). The sponge-like topology of large-scale structure in the universe, *Astrophysical Journal*, **306**, 341–357.
- Gott, I., Gott, J. R., III, Weinberg, D. H. and Melott, A. L. (1987). A quantitative approach to the topology of large-scale structure, *Astrophysical Journal*, **319**, 1–8.

- Gott, I., Gott, J. R., III, et al. (1989). The topology of large-scale structure. III. Analysis of observations, *Astrophysical Journal*, **340**, 625–646.
- Hikage, C., et al. (2003). Minkowski functionals of SDSS galaxies I: Analysis of excursion sets, *Publications of the Astronomical Society of Japan*, **55**, 911–931.
- Joby, P. K., Chingangbam, P., Ghosh, T., Ganesan, V. and Ravikumar, C. D. (2019). Search for anomalous alignments of structures in Planck data using Minkowski tensors, *Journal of Cosmology and Astroparticle Physics*, **2019**(1), 009.
- Junaid, M. and Pogosyan, D. (2015). Geometrical measures of non-Gaussianity generated from single field inflationary models, *Physical Review D*, **92**(4), 043505.
- Kerber, M., Morozov, D. and Nigmatov, A. (2017). Geometry helps to compare persistence diagrams, *Journal of Experimental Algorithmics*, **22**, 1.4.
- Kerscher, M. and Tikhonov, A. (2010). Morphology of the local volume, *Astronomy and Astrophysics*, **509**, A57.
- Kerscher, M., et al. (1997). Minkowski functionals of Abell/ACO clusters, *Monthly Notices of the Royal Astronomical Society*, **284**(1), 73–84.
- Kitaura, F.-S. and Angulo, R. E. (2012). Linearization with cosmological perturbation theory, *Monthly Notices of the Royal Astronomical Society*, **425**(4), 2443–2454.
- Kono, K. T., Takeuchi, T. T., Cooray, S., Nishizawa, A. J. and Murakami, K. (2020). A study on the baryon acoustic oscillation with topological data analysis, arXiv:2006.02905.
- Kozlov, D. (2008). *Combinatorial Algebraic Topology*, Springer, Berlin.
- Lesgourgues, J. (2011). The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview, arXiv:1104.2932
- Libeskind, N. I., et al. (2018). Tracing the cosmic web, *Monthly Notices of the Royal Astronomical Society*, **473**(1), 1195–1217.
- Ling, C., Wang, Q., Li, R., Li, B., Wang, J. and Gao, L. (2015). Distinguishing general relativity and $f(R)$ gravity with the gravitational lensing Minkowski functionals, *Physical Review D*, **92**(6), 064024.
- Liu, J., Bird, S., Zorrilla Matilla, J. M., Hill, J. C., Haiman, Z., Madhavacheril, M. S., Petri, A. and Spergel, D. N. (2018). MassiveNuS: Cosmological massive neutrino simulations, *Journal of Cosmology and Astroparticle Physics*, **2018**(3), 049.
- Ma, C.-P. and Bertschinger, E. (1995). Cosmological perturbation theory in the synchronous and conformal Newtonian gauges, *Astrophysical Journal*, **455**, 7–52.
- Martinez, V. and Saar, E. (2001). *Statistics of the Galaxy Distribution*, CRC Press, London.
- Matsubara, T. (1994). Analytic expression of the genus in a weakly non-Gaussian field induced by gravity, *Astrophysical Journal*, **434**, L43–L46.
- Matsubara, T. (1996). Statistics of isodensity contours in redshift space, *Astrophysical Journal*, **457**, 13–17.
- 松原 隆彦 (1996). ジーナス統計を用いた宇宙の大構造の定量化, 日本物理學會誌, **51**(4), 266–273.
- Matsubara, T. (2003). Statistics of smoothed cosmic fields in perturbation theory. I. Formulation and useful formulae in second-order perturbation theory, *Astrophysical Journal*, **584**(1), 1–33.
- Matsubara, T. (2010). Analytic Minkowski functionals of the cosmic microwave background: Second-order non-Gaussianity with bispectrum and trispectrum, *Physical Review D*, **81**(8), 083505.
- Matsubara, T. and Suto, Y. (1996). Nonlinear evolution of genus in a primordial random Gaussian density field, *Astrophysical Journal*, **460**, 51–58.
- Matsubara, T. and Yokoyama, J. (1996). Genus statistics of the large-scale structure with non-Gaussian density fields, *Astrophysical Journal*, **463**, 409–419.
- Matsubara, T., Hikage, C. and Kuriki, S. (2022). Minkowski functionals and the nonlinear perturbation theory in the large-scale structure: Second-order effects, *Physical Review D*, **105**(2), 023527.
- Mecke, K. R., Buchert, T. and Wagner, H. (1994). Robust morphological measures for large-scale

- structure in the universe, *Astronomy and Astrophysics*, **288**, 697–704.
- Morozov, D. (2007). Dionysus, a C++ library for computing persistent homology, <https://www.mrzv.org/software/dionysus/>.
- Myers, A. D., et al. (2015). The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Quasar target selection, *Astrophysical Journal Supplement Series*, **221**(2), 27.
- Obayashi, I., Hiraoka, Y. and Kimura, M. (2018). Persistence diagrams with linear machine learning models, *Journal of Applied and Computational Topology*, **1**, 421–449.
- Pâris, I., et al. (2018). The Sloan digital sky survey quasar catalog: Fourteenth data release, *Astronomy and Astrophysics*, **613**, A51.
- Parroni, C., Cardone, V. F., Maoli, R. and Scaramella, R. (2020). Going deep with Minkowski functionals of convergence maps, *Astronomy and Astrophysics*, **633**, A71.
- Peebles, P. J. E. (1980). *The Large-scale Structure of the Universe*, Princeton University Press, New Jersey.
- Peebles, P. J. E. and Yu, J. T. (1970). Primeval adiabatic perturbation in an expanding universe, *Astrophysical Journal*, **162**, 815–836.
- Planck Collaboration, Ade, P. A. R., et al. (2016). Planck 2015 results. XVII. Constraints on primordial non-Gaussianity, *Astronomy and Astrophysics*, **594**, A17.
- Pogosyan, D., Gay, C. and Pichon, C. (2009). Invariant joint distribution of a stationary random field and its derivatives: Euler characteristic and critical point counts in 2 and 3D, *Physical Review D*, **80**(8), 081301.
- Pranav, P., et al. (2017). The topology of the cosmic web in terms of persistent Betti numbers, *Monthly Notices of the Royal Astronomical Society*, **465**(4), 4281–4310.
- Pranav, P., et al. (2019a). Topology and geometry of Gaussian random fields I: on Betti numbers, Euler characteristic, and Minkowski functionals, *Monthly Notices of the Royal Astronomical Society*, **485**(3), 4167–4208.
- Pranav, P., et al. (2019b). Unexpected topology of the temperature fluctuations in the cosmic microwave background, *Astronomy and Astrophysics*, **627**, A163.
- Santos, L., Wang, K. and Zhao, W. (2016). Probing the statistical properties of CMB B-mode polarization through Minkowski functionals, *Journal of Cosmology and Astroparticle Physics*, **2016**(7), 029.
- Schmalzing, J. and Buchert, T. (1997). Beyond genus statistics: A unifying approach to the morphology of cosmic structure, *Astrophysical Journal*, **482**(1), L1–L4.
- Schmalzing, J. and Gorski, K. M. (1998). Minkowski functionals used in the morphological analysis of cosmic microwave background anisotropy maps, *Monthly Notices of the Royal Astronomical Society*, **297**(2), 355–365.
- Shandarin, S. F. and Zeldovich, I. B. (1983). Topology of the large-scale structure of the universe, *Comments on Astrophysics*, **10**, 33–45.
- Shiraishi, M., Hikage, C., Namba, R., Namikawa, T. and Hazumi, M. (2016). Testing statistics of the CMB B-mode polarization toward unambiguously establishing quantum fluctuation of the vacuum, *Physical Review D*, **94**(4), 043506.
- Springel, V. (2005). The cosmological simulation code GADGET-2, *Monthly Notices of the Royal Astronomical Society*, **364**(4), 1105–1134.
- Spurio Mancini, A., et al. (2018). 3D cosmic shear: Numerical challenges, 3D lensing random fields generation, and Minkowski functionals for cosmological inference, *Physical Review D*, **98**(10), 103507.
- Sullivan, J. M., Wiegand, A. and Eisenstein, D. J. (2019). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Evolution of higher-order correlations demonstrated with Minkowski functionals, *Monthly Notices of the Royal Astronomical Society*, **485**(2), 1708–1719.

- Sunyaev, R. A. and Zeldovich, Y. B. (1970). Small-scale fluctuations of relic radiation, *Astrophysics and Space Science*, **7**(1), 3–19.
- Takahashi, R., Sato, M., Nishimichi, T., Taruya, A. and Oguri, M. (2012). Revising the halofit model for the nonlinear matter power spectrum, *Astrophysical Journal*, **761**(2), 152.
- The GUDHI Project (2015). GUDHI User and Reference Manual, <http://gudhi.gforge.inria.fr/doc/latest/>.
- Vicinanza, M., Cardone, V. F., Maoli, R., Scaramella, R., Er, X. and Tereno, I. (2019). Minkowski functionals of convergence maps and the lensing figure of merit, *Physical Review D*, **99**(4), 043534.
- Vishwanath, S., Sriperumbudur, B. K., Fukumizu, K. and Kuriki, S. (2022). Robust topological inference in the presence of outliers, arXiv:2206.01795.
- Wasserman, L. (2018). Topological data analysis, *Annual Review of Statistics and Its Application*, **5**(1), 501–532.
- Wiegand, A. and Eisenstein, D. J. (2017). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Higher order correlations revealed by germ-grain Minkowski functionals, *Monthly Notices of the Royal Astronomical Society*, **467**(3), 3361–3378.
- Wiegand, A., Buchert, T. and Ostermann, M. (2014). Direct Minkowski functional analysis of large redshift surveys: A new high-speed code tested on the luminous red galaxy Sloan Digital Sky Survey-DR7 catalogue, *Monthly Notices of the Royal Astronomical Society*, **443**(1), 241–259.
- Wilding, G., et al. (2021). Persistent homology of the cosmic web — I. Hierarchical topology in Λ CDM cosmologies, *Monthly Notices of the Royal Astronomical Society*, **507**(2), 2968–2990.
- Xu, X., Cisewski-Kehe, J., Green, S. B. and Nagai, D. (2019). Finding cosmic voids and filament loops using topological data analysis, *Astronomy and Computing*, **27**, 34–57.
- Yoshiura, S., Shimabukuro, H., Takahashi, K. and Matsubara, T. (2017). Studying topological structure of 21-cm line fluctuations with 3D Minkowski functionals before reionization, *Monthly Notices of the Royal Astronomical Society*, **465**(1), 394–402.
- Zomorodian, A. and Carlsson, G. (2004). Computing persistent homology, *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, SCG'04, 347–356, Association for Computing Machinery, New York.

Quantification of Galaxy Distribution with Topological Data Analysis and Detection of the Baryon Acoustic Oscillation

Tsutomu T. Takeuchi^{1,2}, Kai T. Kono¹, Suchetha Cooray^{1,3,4}, Atsushi J. Nishizawa^{5,6,7}, Koya Murakami¹, Hai-Xia Ma¹ and Yoh-Ichi Mototake⁸

¹Division of Particle and Astrophysical Science, Nagoya University

²The Institute of Statistical Mathematics

³Division of Science, National Astronomical Observatory of Japan

⁴Research Fellow of the Japan Society of the Promotion of Science (PD)

⁵DX Center, Gifu Shotoku Gakuen University

⁶Institute for Advanced Research, Nagoya University

⁷Kobayashi Maskawa Institute, Nagoya University

⁸The Institute of Statistical Mathematics; Now at Graduate School of Data Science, Hitotsubashi University

Galaxies are distributed inhomogeneously in space, with clusters, groups, filaments and voids. This is called the large-scale structure in the Universe. In astronomy, it is customary to refer to the normal matter as baryons. What we directly observe is the large-scale structure made of baryons. The large-scale structure has formed mainly through the gravitational instability, but there is another origin of the structure exists, which is the acoustic oscillation of baryons at the baryon-photon decoupling. This is imprinted on the spatial distribution of galaxies in the Universe, known as the baryon acoustic oscillation (BAO). In this work, we analyzed the spatial distribution of galaxies with a method from the topological data analysis (TDA), in order to examine the BAO signal in the galaxy distribution. The TDA provides a method to treat various types of “holes” in point set data, by constructing the persistent homology group from the geometric structure of data points and handling the topological information of the dataset. We can obtain the information on the size, position, and statistical significance of the holes in the data. A particularly strong point of the persistent homology is that it can classify the holes by their spatial dimension. We first analyzed the simulation datasets with and without the baryon physics to examine the performance of the PH method. We found that the persistent homology is indeed able to detect the BAO signal among the large-scale structures in the Universe: simulation data with baryon physics present a prominent signal from the BAO, while data without baryon physics does not show this signal. Then, we applied the persistent homology to a quasar (a kind of active galaxies) sample at $z < 1.0$ from extended Baryon Oscillation Spectroscopic Survey in Sloan Digital Sky Survey Data Release 14 (SDSS DR14). We discovered a characteristic hole (a hollow shell) at a scale $r \sim 150$ [Mpc]. This corresponds exactly to the BAO signature imprinted in the galaxy/quasar distribution. We performed this analysis on a small subsample of 2000 quasars. This clearly demonstrates that the PH analysis is very efficient in finding this type of topological structures even if the sampling is very sparse.

Key words: Topological data analysis (TDA), persistent homology, galaxy distribution, the large-scale structure in the Universe, baryon acoustic oscillation.

「統計数理」投稿規程

1. 「統計数理」は、統計科学の深化と発展、そして統計科学を通じた社会への貢献を目指すものである。投稿原稿は、統計科学に関連した内容を持つもので、和文の原稿に限る。
2. 投稿原稿は次の6種とする。
 - a. 原著論文 (Paper)
統計科学の発展に貢献すると考えられる研究結果。
 - b. 総合報告 (Review Article)
特定の主題に関する一連の研究およびその周辺領域の発展を著者の見解に従って総括的、かつ体系的に報告したもの。
 - c. 研究ノート (Letter)
研究速報、新しい発想、提言、問題提起、事例報告など研究上、記録にとどめておく価値があると認められるものや、既発表の論文等に対するコメントで、研究上、記録にとどめておく価値があると認められるもの。
 - d. 研究詳解 (Research Review)
特定の研究領域における理論的あるいは応用的成果を、最近の結果や知見を加えてわかりやすく説明したもの。
 - e. 統計ソフトウェア (Statistical Software)
有用な計算法や解析法に関する短いプログラムおよびサブルーチンのリスト、利用手引き、実行例など。
 - f. 研究資料 (Research Archives)
歴史的なデータ、入手困難なデータや統計的手法の比較検討のために有用なデータ、あるいは、歴史的文献の翻訳や解説など。
いずれも原則として、未発表のものに限る。
3. 投稿された原稿は、編集委員会が選定・依頼した査読者の審査を経て、掲載の可否を決定する。
4. 投稿原稿は電子投稿査読システム <https://www.editorialmanager.com/toukei/> より投稿するものとする。原稿は pdf ファイルとし、必要なフォントはすべて埋め込み、原稿全体を一つのファイルにまとめることとする。論文が採択になった場合、著者は最終稿のソースファイルとハードコピーを提出するものとする。
5. 著作権
 - (1) 掲載される論文等の著作権はその採択をもって統計数理研究所に帰属するものとする。統計数理研究所は、紙媒体の「統計数理」のほか電子媒体などを通じて論文等を公表することができる。特別な事情がある場合は、著者と本編集委員会との間で協議の上措置する。
 - (2) 投稿原稿の中で引用する文章や図表の著作権に関する問題は、著者の責任において処理する。
 - (3) 著者が自分の論文等を複製、転載、翻訳、翻案等の形で利用するのは自由である。この場合、著者は掲載先に出典を明記する。
6. 原稿は次の執筆要項に従って作成する。

「統計数理」執筆要項

1. 原稿は A4 用紙に 1 行 36 字から 40 字で 1 行おき、1 頁あたり 22 行程度とする。原稿の長さは原則として表・図を含めて 30 頁相当以内とし、各ページにページ番号を付す。図表は別紙にまとめ、本文中には挿入箇所のみを指定する。L^AT_EX で原稿を作成する場合は、「統計数理」スタイルファイルの使用を推奨する。
<https://www.ism.ac.jp/editsec/toukei/>
2. 文体は「である体」とし、句読点は「,」「.」を用いる。
3. 原稿は以下の順に書くものとする。

[第 1 頁] 標題, 著者名, 所属名, 和文要旨 (500 字程度, 文献の引用および数式は原則として避ける), 和文キーワード (6 語以内).

[第 2 頁] 英語による標題, 著者名, 所属名, Abstract (450 ワード程度), Key words (6 words and phrases 以内). Abstract は、問題の所在と得られた結果等がそれだけで理解できるようなものとする。

[第3頁以降]

- ① 本文：章、節の番号は、第1章にあたるものは、“1.”、第1章第1節にあたるものは、“1.1” というようにつける。また、式の番号は、章ごとに (2.1), (2.2) のようにし、式の左側に配置する。
 - ② 数式：数式は簡明さを心がけ、添字にさらに添字をつけるのはなるべく避ける。
 - ③ 参考文献：書き方は本要項 第4項を参照。
 - ④ 表：一枚の用紙に一つの表を書く。表の番号は論文中に現れる順に従って、表1、表2、… または、Table 1, Table 2, … のようにする。
 - ⑤ 図：一枚の用紙に一つの図を描く。図はそのまま写真製版できる鮮明なものを用意する。大きさは印刷出来上りの1~2倍とし、トレースが必要な場合は原則として著者が行うものとする。図の番号は論文中に現れる順に従って、図1、図2、… または、Fig. 1, Fig. 2, … のようにする。図は原則としてモノクロ印刷とするが、カラー印刷を必要とする場合は編集委員会に相談すること。
 - ⑥ 注：本文中の注釈は極力避ける。やむを得ず注釈をつける場合は脚注とせず、論文末尾に後注とする。後注は、順番に“1, 2, …”の番号を付け、本文中では上付きで示す。
4. 本文中での参考文献の引用は、著者名(出版年)とする。たとえば、Efron (1982), 清水・湯浅 (1984), Cox and Snell (1981), 坂元 他 (2004), Nakano et al. (2000)。
 5. 参考文献の書き方

[雑誌の場合] いずれの場合も雑誌名は省略しないものとする。

① ページ番号がある文献

著者名(出版年). 標題, 雑誌名, 巻, ページ [始-終].

【例】 Chernoff, H. (1973). The use of faces to represent points in k -dimensional space graphically, *Journal of the American Statistical Association*, **68**, 361-368.

【例】 Bligh, E. G. and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification, *Canadian Journal of Biochemistry and Physiology*, **37**, 911-917, <https://doi.org/10.1139/o59-099>.

② ページ番号がない文献

著者名(出版年). 標題, 雑誌名, 巻 に続けて, DOIがあるものはDOIを入れ, DOIがない場合はURLと最終アクセス日を記載する。

【例】 Lien, G. Y., Kalnay, E. and Miyoshi, T. (2013). Effective assimilation of global precipitation: Simulation experiments, *Tellus A*, **65**, <https://doi.org/10.1175/WAF-D-13-00032.1>.

【例】 Kiraly, F. J. and Oberhauser, H. (2019). Kernels for sequentially ordered data, *Journal of Machine Learning Research*, **20**(31), <http://jmlr.org/papers/v20/16-314.html> (最終アクセス日 2023年10月1日)。

[叢書の中の一巻の場合]

著者名(出版年). 書名 (編集者名), 叢書名, 発行所名, 発行地名。

【例】 Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1983). *Akaike Information Criterion Statistics*, Mathematics and Its Applications, Reidel, Dordrecht.

[単行本等の場合]

著者名(出版年). 書名, 発行所名, 発行地名。

【例】 Cressie, Noel (1993). *Statistics for Spatial Data*, Wiley, New York.

[編集書の中の一部の場合]

著者名(出版年). 標題, 編集書名 (編集者名), 巻, ページ, 発行所名, 発行地名。

【例】 Akaike, H. (1980). Likelihood and the Bayes procedure, *Bayesian Statistics* (eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), 143-166, University Press, Valencia, Spain.

なお、同じ著者によるものが同一年に複数個現れる場合には、(1980a), (1980b) などとして区別する。文献は、日本人も含め、著者名のアルファベット順に並べる。

6. 著者校正は原則として一回とする。その際、印刷上の誤り以外の字句や図版の訂正、挿入、削除等は原則として認めない。