

統計数理

第67巻 第2号
(通巻130号)

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

目次

創立75周年記念号

創立75周年記念号発刊にあたって	
田村 義保	153
統計数理研究所創立75周年にあたって	
椿 広計	155
モデリング変革の4半世紀を振り返って [研究資料]	
樋口 知之	157
統計科学の文法と論理的推論 [研究ノート]	
椿 広計	165
時系列解析における状態空間モデルの利用 [総合報告]	
北川 源四郎	181
情報量規準 AIC の統計科学に果たしてきた役割 [総合報告]	
小西 貞則	193
統計地震学の発展と地震活動予測：個人的経験と展望 [研究ノート]	
尾形 良彦	215
チューブ法の理論・応用とその周辺 [総合報告]	
栗木 哲	229
粒子フィルタとデータ同化 [研究詳解]	
上野 玄太	241
統計数理研究所における最適化研究 [総合報告]	
土谷 隆	255
標準コウホート表のコウホート分析モデルのデザイン行列について [研究ノート]	
中村 隆	277
統計数理研究所75周年記念 研究業績紹介 [統計数理研究所研究活動]	299

2019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

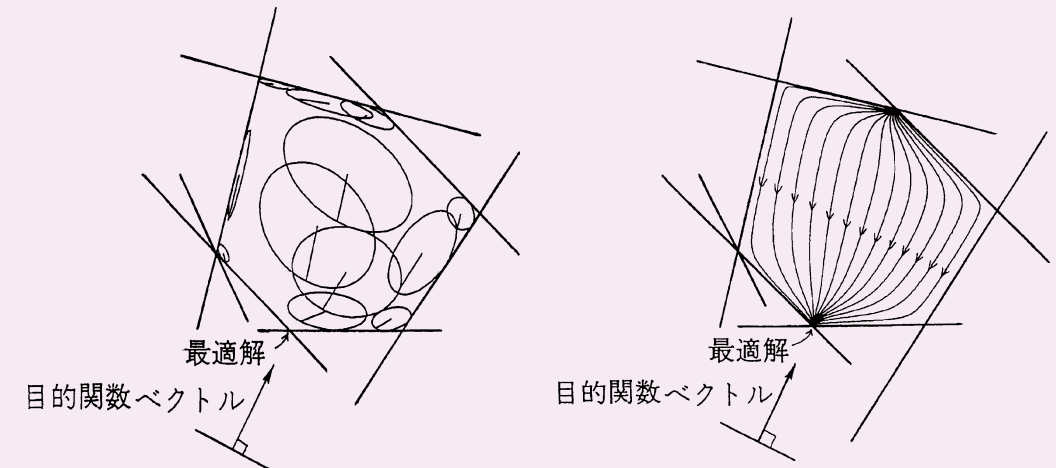
第67巻 第2号

2019

統計数理

Vol. 67, 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 間野 修平
編集委員 伊藤 陽一
志村 隆彰
庄 建倉
中野 慎也
朴 堯星
特集担当編集委員 田村 義保

編集室

池田 広樹 長嶋 昭子 脇地 直子

「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 2019

印刷：笹氣出版印刷株式会社

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

Vol. 67, No. 2

Contents

The 75th Anniversary Issue

Preface	
Yoshiyasu TAMURA	153
The 75th Anniversary of the ISM	
Hiroe TSUBAKI	155
Quarter Century of a Paradigm Shift in a Statistical Modeling Technology	
Tomoyuki HIGUCHI	157
Grammar of Statistical Science and Logical Reasoning	
Hiroe TSUBAKI	165
Use of a State Space Model in Time Series Analysis	
Genshiro KITAGAWA	181
The Role of Information Criterion AIC in Statistical Science	
Sadanori KONISHI	193
The Development of Statistical Seismology: A Personal Experience and View	
Yosihiko OGATA	215
The Tube Method: Theory and Applications	
Satoshi KURIKI	229
Particle Filter and Data Assimilation	
Genta UENO	241
A Historic Review of Research on Optimization at the Institute of Statistical Mathematics	
Takashi TSUCHIYA	255
The Design Matrices of an Age-Period-Cohort Model for a Standard Cohort Table	
Takashi NAKAMURA	277
Abstracts of Research Works	299

December, 2019

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 267 ページを参照

創立 75 周年記念号発刊にあたって

田村 義保[†] (オーガナイザー)

統計数理研究所は 1944 年 6 月に創立され、2019 年 6 月に 75 周年を迎えた。75 周年記念事業の一環として学術雑誌「統計数理」で特集を企画した。特集は計 9 件の総合報告、研究詳解等と椿所長による巻頭言と 2018 年 4 月 1 日時点で在職した所員による活動報告(2019 年 6 月 5 日の 75 周年記念式典時に発行)から構成されている。統計数理研究所の 75 年間の発展については椿所長の巻頭言に詳細がある。また、樋口前所長の論文は、この 25 年間の統計モデリングの変遷についてふれている。この特集が統計学の研究者だけでなくデータサイエンスを実践している方の研究、実践に役立てば幸いである。

竹内啓先生の「歴史と統計学」(日本経済新聞出版社、2018)は統計学の発展を詳述している名著である。21 世紀の統計学の状況にも少しはふれられている。しかし、この本に書かれている以上に統計学の研究者の研究対象・手法は大きく変化しているように考える。それを示すために、「統計数理」の創立 50 周年記念特集号にある論文のタイトルの一部を示す。詳しくは統計数理研究所 Web ページの刊行物「統計数理」にアクセスして欲しい(表 1)。

50 周年記念号の論文テーマ、著者をどのように選んだかについては、発刊当時の所長であった清水良一元所長の「創立 50 周年記念号発刊にあたって」にもない。その時点で重要と考えられたテーマを選んだものとする。数理計画法(内点法)の論文が多いことから、この選択基準であると思われる。調査関係の論文が多いのも 1993 年に国民性の調査が行われ、国際比較のための調査も多く実施されていたことに関係しているためと考える。50 周年特集の著者で、現在も所員である研究者は 5 名(著者合計 22 名)である。また、50 周年記念、75 周年記念ともに執筆した著者は 2 名(2 名とも元所員)である。定年により退官・退職した職員も多いが、現在も、他の組織で現役で研究している者も多い。統計数理研究所の人材交流が盛んであり、統計学界の発展に寄与・貢献していることの証左であるとする。

その他の研究テーマは、執筆した研究者の当時の主要なテーマであるように見える。数理的

表 1. 創立 50 周年記念号論文名.

第 42 卷 No.1	第 42 卷 No.2
1 多次元集中解析法—集中曲線・曲面による統計記述システム—	1 分布の起源—ノンパラメトリックな統計的不確定性関係と統計基礎方程式—
2 多重サンプリングによる非最小位相系の適応制御	2 統計的推測における手法の妥当性
	3 AIC のゆらぎについて
中略	中略
8 真核生物の初期進化	9 データ解析の電子ジャーナル EJDA の実働化—registered ftp の提案と実装
9 非線型可積分系とじゃんけんモデル	

[†] 統計数理研究所 名誉教授：〒190-8562 東京都立川市緑町 10-3

な基礎研究が多いように考える。今回の特集で選んだテーマは、1994年から2017年末までに統計数理研究所の研究者が行った研究の中で、これから先の統計学の発展のために、論文として残しておいた方が良いと考えたものである。最終的な選定は樋口前所長のご意見を参考にして決定した。研究業績報告は、各研究者の最近の主要研究テーマを自選していただいた。これらの中には、現在の統計学界の主要テーマも、これから重要になってくるテーマも含まれている。論文、活動報告の業績とも計算機インテンシブな研究が増えているように考える。統計数理研究所彙報第2巻(1954)や第7巻(1957)には松下嘉米元部長による「十周年にあたりて」、末綱恕一元所長による「十五周年記念日を迎えて」という当時の計算機に関する寄稿がある。統計数理研究所の研究の発展をささえてきたものとして計算機存在は重要である。公的な記録として残すために、ここ25年の主要な計算機を表2にまとめた。速度では250MFlopsから1.49PFlopsと約 5.72×10^6 倍、主記憶では256MBから144TBと約 0.59×10^6 倍となっている。有名なムーアの法則では5年で性能は約10倍になるので、 10^5 倍程度になっていれば良いのであるが、それよりも性能はあがっている。ちなみに、インターネット性能は512kbpsから20Gbpsへと約 3.41×10^4 倍となっている。

AI、ビッグデータ、データサイエンスという言葉が日常、耳にする。統計学の重要性が理解され、研究所には追い風が吹いている。この追い風をうまく活かして、統計数理研究所がさらなる発展をしていくのは、現在の所員の活躍によることが多い。さらに発展した姿で、100周年を迎えて欲しい。

表 2. 1994 年以降にリース契約で導入した主な計算機.

導入年	計算機名と主な仕様
1994	HITAC S3600/120 (主記憶 256MB, 拡張記憶 2GB, 250MFlops)
1996	HITAC S3800/162 (主記憶 1GB, 拡張記憶 4GB, 4GFlops)
1996	IBM SP2 (48 ノード, 主記憶 12GB)
1999	HITAC SR8000 (20 ノード, 主記憶 8GB/ノード 8GFlops/ノード)
2000	Origin2000 (64CPU, 主記憶 48GB, R12000(300MHz))
2004	SGI Altix3700 (主記憶 1920GB, ピーク 5.2GFlops×256)
2006	HP XC4000 (計算ノード, ProLiant DL 145G2 128 ノード (Opteron 2.6GHz×2 主記憶 640GB))
2010	Fujitsu PRIMERGY RX200S5 (360 ノード (2880 コア) 主記憶 48GB×160+24GB×200, ピーク 93.76GFlops×360)
2014	SGI ICE-X を中心としたシステム Intel E5-2697v2 2.7GHz/12 コア 400 ノード (2CPU+128GB 主記憶) 207TFlops 50.0TB(合計)
2015	SGI ICE-X を中心としたシステム増設 Intel Haswell (コード名) 2.4GHz/14 コア 120 ノード (2CPU+256GB 主記憶/ノード) 129.0TFlops 30.0TB (初期との合算 336.3TFlops)
2018	HPE SGI 8600 を中心としたシステム Intel Xeon Gold 6154 3.0GHz/18 コア 376 ノード (2CPU+384GB 主記憶) GPU 計算ノード (8 ノード) 上記+GPU : NVIDIA P100 x 4 システム合計理論性能 : 1.49PFlops 総 Core 数 : 13824core 総主記憶容量 : 144TB

統計数理研究所創立 75 周年にあたって

所長 椿 広計†

統計数理研究所は、令和元年 2019 年 6 月 5 日に設立 75 周年を迎え、その記念事業を 2019 年 度に展開し、「統計数理」誌でも所員ならびに関係者の寄稿による記念号を発刊することとなった。「統計数理」誌もその前身である講究録から数えると発刊 75 周年ということになる。「統計数理」誌が創立記念号を編集するのは、第 42 巻第 1 号の創立 50 周年記念号から数えて 25 年ぶりである。

75 年前の設立を簡単に振り返ると、1943 年 11 月に学術研究会議の「統計数学を中心とする統計科学に関する研究所の設立」の建議が、同会議会員の掛谷宗一東京帝国大学理学部教授からなされた。この建議に基づき、1944 年 6 月 3 日勅令 385 号「統計数理研究所官制」が公布され、同年 6 月 5 日(月)官報 5215 号に掲載された。官制第一条にその設置目的が記されており「統計数理研究所ハ文部大臣ノ管理ニ属シ確率ニ関スル数理及其ノ應用ノ研究ヲ掌リ竝ニ其ノ研究ノ連絡、統一及促進図ル」となっている。同盟時事月報第 8 巻第 6 号(通号 217 号)1944 年 6 月 4 日記事には、「決戦下軍事上、生産或いは国防計画の樹立等の面において統計処理の重要性増大に鑑み、文部省では統計数理研究所を創設することとなり、五日官制を交付、初代研究所長に東京帝大理学部教授専任の所員 6 人、助手 6 人、書記 2 人とされ、東京都下谷區上野公園帝國學士院内に置かれた。」と報じられている。実際の発足時所員は、専任所員としての河田龍夫、坂本平八、松下嘉米男、兼任所員として掛谷宗一(所長、東京帝国大学理学部長)、北川敏男(九州帝国大学教授)、伊藤清(名古屋帝国大学助教授)、増山元三郎(気象台技師)、角谷静夫(大阪大学助教授)、秋月康夫(第三高等学校教授)、佐藤良一郎(東京高等師範学校教授)であった。戦後日本の数学、統計学をけん引した研究者である。

統計数理研究所の歴史を振り返ると、第 2 次世界大戦という非常時における発足期、昭和 20 年代戦後占領政策による社会の計量化を目指した統計調査重視の転換期、高度成長時代の産学連携統計研究の推進期、そして文部省直轄研究所から大学共同利用機関へ、これが創立 50 年までに起きたことである。なお、統計数理研究所及び「統計数理」誌 50 周年の歩みについては、第 42 巻第 1 号の清水良一所長(当時)の巻頭言を参照するのがよからう。

一方、直近 25 年間の統計数理研究所の大きな動きとしては、以下の 4 つを挙げることができよう。第一は、2004 年創立 60 周年に大学共同利用機関法人情報・システム研究機構 統計数理研究所となるとともに、国立大学法人総合研究大学院大学複合科学研究科統計科学専攻が設置され、統計科学専攻を支える基盤研究機関となったことである。その後、統計科学専攻での学位取得者は、現在日本の数理・データサイエンス教育研究のリーダーとなっている者も多い。2019 年 11 月現在学位取得者は 141 名に達している。

第二は、創立 60 周年前後から、統計数理の基礎研究を支える基幹研究系活動に、統計数理の活用を支える戦略研究センター、後の NOE(Network of Excellence)活動につながる研究センター活動が開始されたことである。これを通じて、統計数理によってその活動が発展しうる産官学の様々なコミュニティに対して統計数理研究所が繋がるということが可能になった。

第三は、2009 年に、広尾から立川市への移転が完了し、研究設備環境が充実したことであ

† 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

る。創立 75 周年は、立川移転 10 周年でもある。

第四は、2011 年 4 月に統計思考院を設置し、統計思考力育成事業を開始したことである。これらを通じて、多様な学術支援に必要な、データ基盤、知識基盤、人材基盤という三位一体の基盤構築事業が、情報・システム研究機構の中で推進可能になったのである。

現在、データ中心科学、データサイエンスの活用力が、学術のみならず産業競争力の源泉となるデータ駆動型社会の到来を迎えている。国際的にもこの 20 年間、データサイエンス分野への優秀な人材の投下が加速しており、日本が必ずしも競争優位な状況にはなっていない。一方、次世代競争力の源泉となる人工知能やロボティクスなどの基幹部分を支え、進化させる基礎研究としての統計的学習・推論や統計的モデリング研究の役割は、極めて本質的なものとなった。また、この種の基礎研究を学術・社会に分かりやすくかつ速やかに展開し、議論し、良い方法を共有する場の必要性も日々増している。

日本の統計数理中核研究組織としての統計数理研究所と、そのコミュニケーション機能を担う「統計数理」誌は、これら社会ニーズに応え、統計数理研究所を支え、統計数理研究所が支える多くのコミュニティとともに、実りあふれる 100 周年に向けた歩みを着実に進めなければならない。

モデリング変革の 4 半世紀を振り返って

樋口 知之[†]

(受付 2019 年 5 月 27 日；改訂 7 月 5 日；採択 7 月 16 日)

要 旨

私が研究所に入ったのは平成元年 4 月、また所長を退任すると同時に研究所を退職するのは、平成が終わる直前の平成 31 年 3 月であったため、平成の 30 年間は私にとって研究者人生そのものである。本稿では、データにもとづく予測や判別といった統計的思考の根幹を成す、統計的モデリングにかかわる技術の大きな進展を、実体験した私の視点で概括する。

キーワード：ベイズモデリング，粒子フィルタ，カーネル法，深層学習。

1. はじめに：入り口は線形・ガウスモデリング

私が入所した当時(1989 年)は、その 10 年ほど前から研究が活発化していたベイズモデリング研究の成熟期にあった。線形・ガウス型の事前分布および尤度関数に関する理論は確立し、応用の観点からも、いくつかの課題では研究所の研究成果が標準的方法として認知されるなど、研究進展も著しかった。正確に言えば、統計数理研究所では、滑らかさを表現する事前分布と周辺尤度最大化法を採用した、経験ベイズ法が活発に研究されていた(赤池・北川, 1994, 1995; 石黒 他, 2004; 伊庭 編, 2018)。私は学部生の頃から雑誌「数理科学」の愛読者で、ある時、ベイズモデリングの特集号(赤池, 1981)があり、その柔軟な表現能力に大変な感銘を受けた。その解説原稿のベースとなる国際会議録(Akaike, 1980)のコピーが、地震学を専門とされる松浦充宏東大教授の研究室にあることを知り、研究室に行ってコピーさせてもらい(劣悪な多数回のコピーの連続で、式の細かい部分がかかなり判読しづらくなっていた)すぐさま論文内容の検証をした記憶がある。実はその論文にはタイポ的ミスが複数あり、「世界的権威の赤池先生の論文に間違いを見つけた」と学生の私はかなり昂揚し、当時所長をされておられた赤池先生に不躰にもコンタクトを取ってしまった。もちろん、タイポ的ミスなので結果に間違いは無かったのだが、今と違い取るに足らない些細なタイポミスを知り得ようもなく、研究を効率的にすすめることが難しい、逆に言えば、大変長閑な時代であった。経験ベイズ法を採用した赤池先生の 1980 年の論文は、相当な議論をベイズ統計学の分野で巻き起こし、その当時、ベイズ統計学の第一人者であった英国の Smith 教授らベイズ統計学者とは激しいやりとりがあったと赤池先生から聞いている(赤池 他, 2007)。なお、Smith 教授は現在(2019 年)、2015 年に英国政府の肝入で設立された人工知能とデータサイエンスの国立研究所、アラン・チューリング研究所の所長を務めている。後段にも触れるが、英国の人工知能研究においては統計学の存在感が重厚である。

現象や既存知識のモデリングにおいては、計算上からくる線形・ガウス性の制約は、研究者の自然な思考を妨げる。そのため、その制約を、“人手”でもって解析的に緩める論文が 1980

[†] 中央大学 理工学部：〒112-8551 東京都文京区春日 1-13-27

年代は数多く出版されていた。事実、研究所の先生方は、現象の振る舞いを決して過度に抽象化することなく、計算が容易となる空間への非線形写像を“巧みに見だし”，データに潜む新しい知識を次々とあぶり出していた。「数理〇〇学」という当該専門分野の研究者の持つ、先入観やある種の価値観に大きく依存した数理表現でなく、“データに『因子とメカニズム(機序)』を語らせる”ことにこだわっていた研究所の先生方は、間違いなくデータ駆動科学の先導者である。先生方の玄人芸的非線形モデリングのアプローチと平行して、計算機集約的な方法の開発も活発化していた。高い次元の潜在変数がつくる超高次元の分布から、マルコフ連鎖でもって直接的に実現値を得るマルコフ連鎖モンテカルロ法(伊庭 他, 2005)や、その分布を低次元の分布に分解し(いわゆる逐次ベイズフィルタ)、低次元の分布の逐次更新を数値的に得る「非線形フィルタ」(北川, 2005)などである。私は学位を宇宙プラズマ物理の領域で得たため、統計力学の自由エネルギーや転送積分とほぼ同一の発想にもとづく両手法を自然に理解でき、大いに魅惑された(伊庭, 2003)。

2. 粒子フィルタ：非線形・非ガウスモデルの計算実現

平成にはいってしばらくして北川先生(第十代所長)は、非線形フィルタの適用範囲を大幅に拡張するモンテカルロフィルタを提案された(Kitagawa, 1993, 1996)。この手法は今日において、粒子フィルタの手法群の中で、原始的ではあるがレガシーとして高く評価されている。別分野においても、全く同じ手法がほぼ同時期に英国の Blake らによって提案され、大きな注目を浴びていた(Isard and Blake, 1998)。このアルゴリズムは、Condensation(CONditional DENsity propagATIOn)と呼ばれ、物体追尾分野(特に軍事分野)で盛んに研究開発された。Blake はマイクロソフトの簡易 3D カメラである Kinect の開発にも参加し、後にアラン・チューリング研究所の初代所長に就任している。

当時、北川先生の近くで研究していた自分は、モンテカルロフィルタの大規模並列コンピュータへの実装容易性に衝撃を受け、さまざまな非線形・非ガウス型の時系列モデルへの適用を試みた。また、世界同時的に多くの研究者が、アルゴリズムの高度化と現実問題への適用をすすめる、21 世紀早々にシュプリングから書籍が出版された(Doucet et al., 2001)。私も著者の一人として参加するチャンスを得、当時新進気鋭の海外の若い研究者と交流できたことは、その後の自分の研究者人生にとって大きな糧となった。この友人の多くは現在、後述する深層学習に研究の軸足を移しているが、そのことは偶然でなく、むしろモデリング技術の発展を考えれば極めて自然なことと考える。

粒子フィルタは、時系列モデルが所与であれば、パラメータや状態変数の推定に関して原理的には万能である。もちろん、モンテカルロ誤差(有限サンプルによる表現限界)からくる、分布の表現能力の喪失(いわゆる退化問題)や、尤度値の不安定性など、数値的課題は避けられないが、問題や経験に即して“人”が対象を自由にモデリングできるようになった点は、モデリング技術の発展において大きな飛躍であった。特に、ロボティクス分野への粒子フィルタのインパクトは大きなものがある。前述したシュプリングの本にも寄稿している Thrun らは、ビッグデータから簡便にロボットを制御する方法論を確立した(Thrun et al., 2007)。彼らは、統計学とロボティクスの“真の結婚”を実現した立役者である。さらに彼らはその技術をもって米国国防高等研究計画局(通称 DARPA)のプロジェクト「グランドチャレンジ(2007 年はアーバンチャレンジ)」にスタンフォード大チームとして参加し、自律型無人自動運転車の技術開発に多大なる貢献をした。後に Thrun はグーグル(Google)に入り、Google Street View を開発している。その後、グーグル・グラスを開発した Google-X の社長を務め、現在、グーグルの創設者である Page と一緒に、空飛ぶ自動車 Kitty Hawk の実現に取り組んでいる。

3. カーネル法：非線形モデリング

2000年代にはいって、非線形モデリングの観点からは、別の形で大きな進歩があった（樋口, 2014）。それはカーネル法の登場である（Schölkopf and Smola, 2002; Shawe-Taylor and Cristianini, 2004; 赤穂, 2008; 福水, 2010）。1998年にGoogleが、また2004年にはFacebookが創業されるなど、ビッグデータを新しい情報サービスという価値に転換することに成功した企業が当時、続々と誕生していた。ビッグデータを用いて容易なタスクは、識別関数の構築を通じた判別や分別である。カーネル法は、データ空間で複雑な識別境界面を推定するのではなく、あえて超高次元の特徴ベクトル空間を構成し、そこで豊富に蓄積された線形のモデリング技術を活用する手法である。もちろん、データ間の類似度を規定するカーネル関数は“人”が設定しなければならないが、データ空間での非線形モデリングに求められる巧みの技の多くを、カーネルトリックに押しつけられた点は大きい。パターン認識手法の多くはデータどうしの内積計算を含むため、カーネルトリックによって既存の線形モデリングの諸手法が非線形版に自然に一般化された。カーネル法研究の進展は、グラフ間の類似度を計るさまざまなアイデアを喚起し、その後、分子構造や原子配列のようなグラフ構造による記述が本質的に重要である、バイオインフォマティクスやマテリアルズインフォマティクスといった分野で大きく花開いていく（Schölkopf et al., 2004）。

2000年代はビッグデータの登場により、統計的モデリングの主たる興味である生成モデルの構築から、カーネル法と最適化の活用による、複雑な識別関数（識別モデル）の自動構築に、研究のトレンドがシフトした点は指摘しておきたい。特に、1990年代後半から非常に多くの応用成果を生み出したサポートベクターマシンは、カーネル化技法の恩恵を最も受けたと言える（Boser et al., 1996）。それと同時に、機械学習の言葉で代表される研究者コミュニティが世界的に育っていったのもこの時代である。その理由として、データに基づくさまざまな統計的推測手法やデータマイニングのアルゴリズムが、教師あり学習、教師無し学習、そして強化学習といった、機械学習の3つのタイプで整理され、異分野からの研究者の参入障壁が低くなった点は大きい。

4. スパースモデリング：非ガウスモデリング

2000年代半ばから2010年代にかけて非ガウスモデリングにおいても大きな発展が、スパースモデリングの普及（伊庭 編, 2017; 科研費新学術領域研究「スパースモデリング」, 2019）によってもたらされた（樋口, 2014, 2016a）。応用ドメインや計測現場における統計的モデリングの成否は、膨大な説明（属性）変数群の中からタスクの解決に有効な特徴ベクトルの適切な構築にあると断言できる。今、興味ある対象を膨大な説明変数の中から少数個で線形回帰表現する問題を考える。何次の回帰モデルとするのか、さらにどの変数を使うかで、膨大な数の回帰モデルが存在する。モデル数は組み合わせ爆発しており、現実にはAICなどの情報量規準による最適モデルの探索（変数選択）は基本貧弱である。一方スパースモデリングでは、回帰係数（変換行列内の要素）にL1正則化（絶対値誤差の最小化）を加えた上でパラメータ推定を行う（川野他, 2018）。最適化関数の形がL1であることから、L1正則化は非ガウスモデリングと言える。この最適化の結果として、重要な説明変数のセットを自動的に浮きあがらせる。

正則化自体は、概念的には古い、汎化能力を高める一つの方策である。前述した1980年代のベイズモデリングの成熟期には、パラメータに対してL2（二乗誤差の最小化）制約を加えることで正則化を実現していた。一方、L1正則化の採用により目的関数の解析的な性質を病的にするかわりに、得られるパラメータの最適解に癖が出ることを意図するアプローチも以前から実はあった。ここでの癖というのは、値が小さいパラメータ値は“大胆にも”ゼロに自動的

にセットされてしまう性質を指す。これにより、属性変数に関するパラメータの値がゼロとなれば、それは事実上その属性変数は不要であることを意味し、変数選択が最適化により実現できるのである。このアプローチは、病的な最適化関数に対して高速に最適解を求めるアルゴリズムが提案されたことを端緒として、1990 年半ば以降この 20 年間、理論的にもまた応用の広がりの意味でも精力的に研究されてきた (Hastie et al., 2008)。スパースモデリング研究の隆盛を後押しした別要因として、最適化法の計算高速化と汎用的ソフトウェアの実用化などの、最適化手法のコモディティ化がある。これは、スパースモデリングはもちろん、前述のカーネル法の進展に大きな役割を果たした (樋口, 2014)。

スパース性のようなデータ空間の低ランク性を陰に陽に活用するアイデアは、1990 年代後半から 2000 年代にかけて、信号処理の成分分離 (複雑なノイズ成分の除去) に有力な手法を多数導き、事実、独立成分分析 (ICA: Independent Component Analysis) (村田, 2004)、Compressed Sensing (CS) や、Non-negative Matrix Factorization (NMF) (亀岡, 2010)、テンソル分解にもそのアイデアを確認することができる。このように、説明変数の選択といった、限定されたモデリング技術に関しては、スパースモデリングにより非ガウスモデリングが実現されたと言えよう。なお、画像、音声、自然言語の処理においては、現在その課題は、End-to-End と呼ばれる深層学習の技術によって大幅に自動化されている。一方、カーネル法は、データ空間から高次元の特徴ベクトルを構成する方法なので、情報圧縮の文脈では逆センスの手法である。

5. 深層学習：非線形モデリング

2010 年代にはいつの最大の衝撃は、特定のタスクでの深層学習の圧倒的パフォーマンスである。その性能の高さは、入出国時の自動顔画像判別、スマートスピーカーでの音声認識、多言語自動翻訳など、私たちの生活に身近な製品として具現化している。これらの技術は人の働き方や社会の構造そのものにも直接的に影響を与えていることから、深層学習はこの四半世紀の情報科学技術における最大のブレイクスルーであると言わざるを得ない。

過去に AI ブームは、1950 年代後半から 1960 年代前半と、1980 年代の 2 回あった (樋口, 2016b)。初回は、一つの脳神経活動を数式でモデル化し、それらを複数組み合わせコンピュータに処理させた、極初歩的なニューラルネットワーク (NN: Neural Network) の提案により巻き起こった。二回目では、「第五世代コンピューティング」がまずブームを誘起し、1980 年代半ばあたりから、ニューロ・ファジー家電製品の開発・販売に代表されるような、第二次ニューロブームがそれに続いた。その後、計算に莫大な時間がかかることや、試行錯誤による構造やパラメータチューニングの面倒さ、また理論構築困難 (精度保証など) により、NN 研究への興味が失われ、AI にとって冬の時代が長く続く。筆者は第二次 AI ブームの時代からベイズ統計学の立場で AI にかかわってきたので、AI 研究とのつきあいも 30 年近くになる。特に、「第五世代コンピューティング」のあとの旧通商産業省にとっては新情報技術の最後となる大型プロジェクト、「リアルワールドコンピューティング (RWC) 計画」 (大津, 1993) の理論・アルゴリズム基盤領域 (前期 5 年) へ参加した。RWC が終結した直後からビッグデータの時代が到来する。

従来の教師あり機械学習は、まずは問題毎に元データから“適切な”特徴ベクトルを構成し、この特徴ベクトルを入力として、諸々の目的関数を最大 (最小) 化するように、学習器内の膨大なパラメータの値を決定・推測する。ここで学習器とは、サポートベクターマシンや、ブースティング、ノンパラメトリックベイズなどを指す。パラメータ学習アルゴリズムは学習器毎に異なるが、一度、データが与えられればあとの学習プロセスはマシンにお任せである。一方、特徴ベクトルの構成法はほぼ人間の知的作業に委ねられており、実はこの構成法が予測・判別性能といったパフォーマンスをほとんど決めていると言っても過言でない。特に一般物体認

識では、画素値ベクトルから特徴ベクトルを構成する、さまざまな手法の研究が長年にわたって行われてきた。深層学習では、この特徴ベクトルの構成作業をあえて省略し、画素値ベクトルを直接 NN の入力とする。

深層学習で使われるニューラルネットワークは、層数が大幅に増えた以外、第二次ニューロブームの時のものと違いは無く、そのパラメータ学習アルゴリズムも、Back propagation を基本とする以前のものと大差はない。深層学習では層数が大幅に増えた結果、パラメータ数も爆発的に増え、学習アルゴリズムもさほど賢くなっているわけでもないので、必然と計算リソースはこれまでとは桁違いに必要となる(樋口, 2018)。ただし、上述したように、それまでの特徴ベクトル構成法は機械学習の『匠の技』と言え、“機械”学習にもかかわらず、人間の判断が最も性能向上に大切という羊頭狗肉の面もあったが、その問題を特定領域では基本的に解決した点は画期的である。

深層学習では、通常、最適化関数は L2 であるため、誤差にガウス性を仮定している。よって、非線形・ガウスモデリング技術は、深層学習により、少なくとも予測・判別性能の観点からはほぼ極みに達したと言える。これまでの深層学習の大きな成功事例は、入力データが画像、音声、テキストなどの、データ構造が簡単なものに集中しているが、その制約を逆手にとって、さまざまな入力データをあえて画像などに変換すれば (Poplin et al., 2018)、特徴ベクトルの選択問題を回避可能とも言える。

非線形モデリングの自動化に革命的インパクトを与えた深層学習にも、課題は数多く残っている(樋口, 2016a, 2016b, 2018)。まず、さまざまな泥臭い工夫に、統一的理論(ガイドライン)がない。第二次ニューロブームの時には、層数やノード数の適切な選択問題や学習係数の客観的決定法が、情報量規準の観点などから真剣に研究された。もはや深層学習においては、これらは完全に(当座?)議論の対象外となっている。よって、精度保証のような理論構築は極めて困難なままである。まさに、深層学習の問題毎の最適化が“黒魔術”とも呼ばれる所以である。もう一つの弱点は、推定結果が(ほぼ)完全にブラックボックス化している点である。この NN の本質的な弱点は、第二次ニューロブームの時も、統計学を中心として研究者から、また NN を実問題に適用した産業界からも、大きな不満としてさんざん指摘されてきた。今の応用開発現場では、予測や判別性能が足りなければ、層数をさらに積み増したり、パラメータ学習に使う計算時間を 1, 2 週間延ばすなど、とても系統的とは言えない職人技が横行している。今後、深層学習がさらに大きな飛躍を遂げるためには、精度保証のための理論研究がすすむことが必須であろう。

6. 今後の統計的モデリング

非ガウス・非線形モデリングを自動化する試みは、この四半世紀に大きな飛躍を遂げた。スマートセンサ、スマホ、バイオメトリクス、インターネット調査、SNS (Social Network Service) からの即時的集計などなど、データ取得の環境はますます“ビッグデータの”になっていく。膨大なデータ量は統計的モデリングの自動化に大きく寄与する一方、精度(信頼度)はサンプル毎にバラバラで、欠測や異常値の混入も頻発、また属性変数の次元も巨大といった、量と質のアンバランスが極端なデータの解析がさらに求められる。このような状況では、まず、目的に直結する属性変数の選択(言い換えれば、特徴ベクトルの構成)や、サンプルの信頼性の評価により、真に必要なデータセットの作成、つまり「スマートデータ」の生成こそが大切になる。統計学は、外部に作成してもらった「スマートデータ」の丁寧で精緻な分析のみを担うのか、それともビッグデータからスマートデータを構成するプロセスへ踏み込むのか、正念場に來ている。

一般社会においては、ビッグデータ × AI(人工知能)により大きな存在感を示す米中のブラッ

トフォーマー(7 Sisters と通常呼ばれる)に対する風当たりも強くなってきている。今後は、モデリング技術の向上を計る上で、モデルの“人”による解釈可能性、帰納法で構築された意思決定システムの説明責任、データのバイアスと社会的偏見の分離など、人に寄り添った視点が極めて重要になってくる(内閣府, 2018)。プライバシーや Fairness を SNS 業界でどう位置づけ、その適切な管理とモニターを実現する技術が今まさに求められている。個人的には、その解決は永久に解決に至らぬ「永遠の難問」と考えるが、しかしながら、下に示すような具体的な技術課題は一般社会に喫緊に対応策を示さねばならない。

- Accountability (結果の説明責任力)
- Interpretability (結果の解釈可能性)
- Reproducibility (結果の再現性)
- Transparency (手続きの透明性)

これらの課題解決には、以下の問題点

- Data imbalance (どんなにビッグデータになろうとも、データそのものがバイアスのある母集団からの情報収集とならざるを得ない現実)
- Sampling strategy (ビッグデータからの統計的判断に利用するデータのサンプリング方策)
- Generative model (データ生成を模擬する確率モデル)

を深く認識することが肝要であり、それらはすべて Foundation of Statistics につながる、統計学の原理の本質的理解そのものと言える。

参 考 文 献

- 赤穂昭太郎 (2008). 『カーネル多変量解析—非線形データ解析の新しい展開(シリーズ確率と情報の科学)』, 岩波書店, 東京.
- 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.
- 赤池弘次 (1981). モデルによってデータを測る, 数理科学, **19**(3), 7–10.
- 赤池弘次, 北川源四郎 編 (1994). 『時系列解析の実際 I』, 朝倉書店, 東京.
- 赤池弘次, 北川源四郎 編 (1995). 『時系列解析の実際 II』, 朝倉書店, 東京.
- 赤池弘次, 樋口知之, 川崎能典, 石黒真木夫 (2007). 『赤池弘次: 統計科学を語る—1993 年. 駆け出し研究者によるインタビュー記録—』, 統計数理研究所, 東京.
- Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1996). A training algorithm for optimal margin classifier, *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, **5**, DOI:10.1145/130385.130401.
- Doucet, A., de Freitas, N. and Gordon, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York.
- 福水健次 (2010). 『カーネル法入門—正定値カーネルによるデータ解析(シリーズ多変量データの統計科学)』, 朝倉書店, 東京.
- Hastie, T., Tibshirani, R. and Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, Springer Series in Statistics, Springer, New York.
- 樋口知之 (2014). 統計数理の誕生とその広がり, 横幹, **8**(1), 14–21.
- 樋口知之 (2016a). スモールデータ, ビッグデータ, そしてスマートデータ—人工知能ブームの中での統計学—, 統計, **67**(1), 9–14.
- 樋口知之 (2016b). 人工知能はみよみまねマシンの究極形, 情報管理, **59**(5), 331–335.
- 樋口知之 (2018). データ関連の数理技術の変遷: 深層学習は春秋戦国時代の秦となるのか?, 人工知能,

33(2), 116–123.

- 伊庭幸人 (2003). 『ベイズ統計と統計物理』, 岩波講座 物理の世界 物理と情報 3, 岩波書店, 東京.
- 伊庭幸人 編 (2017). スパースモデリングと多変量データ解析, 『岩波データサイエンスシリーズ Vol.5』, 岩波書店, 東京.
- 伊庭幸人 編 (2018). 『ベイズモデリングの世界』, 岩波書店, 東京.
- 伊庭幸人, 種村正美, 大森裕浩, 和合肇, 佐藤整尚, 高橋明彦 (2005). 『計算統計 II マルコフ連鎖モンテカルロ法とその周辺』, 統計科学のフロンティア 12, 岩波書店, 東京.
- Isard, M. and Blake, A. (1998). CONDENSATION — Conditional density propagation for visual tracking, *International Journal of Computer Vision*, **29**(1), 5–28.
- 石黒真木夫, 松本隆, 乾敏郎, 田辺国土 (2004). 『階層ベイズモデルとその周辺—時系列・画像・認知への応用』, 統計科学のフロンティア 4, 岩波書店, 東京.
- 科研費新学術領域研究「スパースモデリング」(2019). <http://sparse-modeling.jp/> (2019年7月5日).
- 亀岡弘和 (2010). 非負値行列因子分解入門～音響信号処理を題材として, <http://www.kecl.ntt.co.jp/people/kameoka.hirokazu/publications/Kameoka2010Subspace07.pdf> (2019年7月5日).
- 川野秀一, 松井秀俊, 廣瀬慧 (2018). 『スパース推定法による統計モデリング』, 統計学 One Point 6, 共立出版, 東京.
- Kitagawa, G. (1993). Monte Carlo filtering and smoothing method for non-Gaussian nonlinear state space model, Research Memo., No.462, The Institute of Statistical Mathematics, Tokyo.
- Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of computational and graphical statistics*, **5**(1), 1–25.
- 北川源四郎 (2005). 『時系列解析入門』, 岩波書店, 東京.
- 村田昇 (2004). 『入門 独立成分分析』, 東京電機大学出版局, 東京.
- 内閣府 (2018). 人間中心の AI 社会原則(案), https://www8.cao.go.jp/cstp/ai_gensoku.pdf (2019年7月5日).
- 大津展之 (1993). リアルワールドコンピューティング研究計画 1. 情報処理の新たなパラダイムを目指して, 情報処理, **34**(12), 1423–1428.
- Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y. and DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks, *Nature Biotechnology*, **36**, 983–987.
- Schölkopf, B. and Smola, A. J. (2002). *Learning with Kernels*, The MIT Press, Cambridge.
- Schölkopf, B., Tsuda, K. and Vert, J.-P. (eds.) (2004). *Kernel Methods in Computational Biology*, The MIT Press, Cambridge.
- Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*, Cambridge University Press, Cambridge.
- Thrun, S., Burgard, W. and Fox, D. (2007). 『確率ロボティクス』(上田隆一 訳), 毎日コミュニケーションズ, 東京.

Quarter Century of a Paradigm Shift in a Statistical Modeling Technology

Tomoyuki Higuchi

Faculty of Science and Engineering, Chuo University

I entered the Institute of Statistical Mathematics in April 1989, the beginning year of the Heisei period. Just before the end of the Heisei period, in March, 2019, I stepped down from the director and at the same time retired from the institute. Therefore, the “Heisei 30 years” are the life of a researcher for me. In this article, I summarize the major advances in statistical modeling technologies that form the basis of statistical thinking such as data-based prediction and discrimination from my point of view.

統計科学の文法と論理的推論

椿 広計†

(受付 2019 年 1 月 5 日; 改訂 3 月 24 日; 採択 3 月 26 日)

要 旨

本論文は, Goertz and Mahoney (2012)に従って, 法則を秩序とする科学の主要な研究プロセスが統計的・定量的探求プロセスであるのに対して, プログラムを秩序とする科学の主要な研究プロセスが, 論理的・質的探求プロセスであるとの仮説を提示する. その裏付けのために, 統計科学的立場から, 定量的プロセスと論理的プロセスの差異が顕著となる幾つかの状況を示す. 更に, プログラム科学と法則科学との融合可能性についても議論する.

キーワード: 科学の文法, プログラム, 法則, 定量的研究, 定性的研究.

1. 問題意識の変遷—法則の認識科学と法則科学から設計科学とプログラム科学へ

本論は, 吉田民人が主導した「新しい学術の体系」で示された「プログラム科学」(日本学術会議運営審議会附置新しい学術体系委員会, 2003, 吉田, 2013)の主要な研究プロセスが, Goertz and Mahoney (2012)が示した社会科学における論理的・質的探求プロセスであるとの仮説を提示する. この仮説を補強するために, 統計科学によって実証される法則科学的探究と論理的なプログラム科学の探究プロセスとの方法論的差異についても幾つかの事例を示す. 特に, 集合・数え上げ・分類といった質的・離散数理技法の多くが, 万象を記述する法則の科学的認識ではなく, 人間が生存の便宜のために開発した情報縮約的便法と考えられることも指摘する.

筆者は科学哲学の専門家ではなく統計家であり, 統計的实践の成功や失敗の経験を通じて, 科学的法則の探究プロセスやその支援技法について暗黙知を形成してきた. 従って, 本論では, 筆者の統計家の経験を基に, 法則科学ないしは統計科学の視点から, プログラム科学ないしは論理的推論の特性を指摘しただけである. 決して統計的推論が論理的推論に優越するという主張を意図してはいない.

筆者は, 横断型基幹科学技術連合活動の中で, 「新しい学術の体系」の解釈作業(日本学術会議自動制御研究連絡委員会工学共通基盤研究連絡委員会自動制御学専門委員会, 2005)に参加し, 吉田の科学論を知った. その中で, 価値に依存しない「あるものの追求」を目的とする「認識科学」形成の標準プロセスは, 統計科学創成を興した Pearson (1892)の「科学の文法」が原点であると再認識した. 吉田は, 価値に依存した「あるべきものの追求」を目的とする「設計科学」の標準プロセスを「情報循環」とした. しかし, 当時の筆者には, 設計科学のための情報循環プロセスが, 明確に標準化されているとは考えられなかった.

そこで, Tsubaki et al. (2008)は,

1) 価値(関係者の声)の選択

† 統計数理研究所: 〒190-8562 東京都立川市緑町 10-3

- 2) 関係者の声を技術の声に翻訳(価値実現システムの選択)
- 3) 技術システムの最適化
- 4) システムが実現する価値の社会への注入(実装)

からなる、「技術開発の文法」を提案し、そのプロセスを支援する数理技術の適用例を示した。

次に、筆者は、新技術・新製品開発の標準プロセスモデルと、その各フェイズにおける支援方法群を示す指針を国際標準とする活動を ISO/TC 69 “Applications of statistical methods” で開始した。2009 年に ISO/TC69/SC8 “Application of statistical and related methodology for new technology and product development” が設置され、2015 年には ISO16355 シリーズの第 1 部を発行した(ISO, 2015)。標準化活動と並行して、設計科学の文法を産学共同研究する「価値創生プロセス実践知開発ネットワーク(VCP-Net 研究会)」も組織した。この研究会では、管理技法、数理技法、統計技法の目的と機能から統計的に分類すると共に、技術開発プロセスのあるフェイズで活用される技法が、前のフェイズから引き渡される入力情報、後のフェイズに引き渡す出力情報の要求品質明示化などを研究した(VCP-Net 研究会編, 2014)。このような標準化や研究が、吉田が提起した情報循環を高質化・効率化すると考えたからである。

吉田(2013)の「プログラム科学」を再考するきっかけとなったのは、2016 年から開始した「顧客の期待を超えたエクセレント・サービス」を実現するプロセスの標準化活動である。物質やエネルギーを支配する秩序と、サービスを制御する秩序とでは、それに対処する数理科学的方法が本質的に異なることをどうプロセスに反映させるかが課題なのである(椿, 2017)。実際、1980 年代からソフトウェアやサービスの品質経営分野では、事実に基づく問題解決に資する統計的方法よりも、あるべき姿を求めて、言語情報を論理的に分析する技術が有効と考えられてきた。

椿(2018)では、プログラムを記述する記号列の事象への当てはめでは、連続的回帰モデルより、離散層別モデルの方がプログラム科学のためには有効ではないかと考えた。顧みれば、土橋 他(1985)では、サービス分野での言語情報収集とその定量的分析の必要性、統計的方法に限定しても、層別原理の徹底、特に今日では第 2 世代人工知能と呼ばれる Breiman et al.(1984)の CART による層別の自動化への注目を呼び掛けていた。

こうした中、Goertz and Mahoney(2012)が、数学が定性的(質的)研究、統計学が定量的研究のエンジンであり、因果推論も両者では本質的に異なるという指摘に触れた。それを通じて数理的学術形成プロセスの 2 つのパラダイムの乖離が、人工的・離散的なプログラム科学と、天然的・連続的な法則科学との数理的差異に起因するという仮説を形成するに至った。本論は、そのような問題意識を試論としてまとめたものである。

第 2 章でこれまでの筆者の「科学の文法」に関わる認識を総括する。3 章では、吉田が提唱したプログラム科学における認識が、法則科学における認識とは異なる離散的選択原理によって支えられていることを示す。4 章では、生物の生存にとって必要な、認識の離散化を支援する様々な数理的方法が人間によって生み出されていることを述べる。5 章では、これらの考察を受けてプログラム科学の文法生成の課題と法則科学文法との融合可能性について議論する。

2. 科学の文法

2.1 法則認識科学の文法

Pearson(1892)は、「人間が法則(Law)を自然(Nature)に付与する」、事実に基づく科学形成のプロセスモデルを「科学の文法」として提示した。観察に基づく記述に基づき現象への法則を付与する手続き自体を科学的行為とした。そのプロセスは、単純化すれば、観察に基づく事実の秩序に関する分析、その総合、批判的検証からなる一連の追究プロセスである。Pearson は、

科学的法則記述の文法としての標準プロセスを支援する数理的方法群の開発も開始した。ヒストグラム、標準偏差、相関分析、重回帰分析、主成分分析、モーメント法(ピアソン系分布族への当てはめや混合正規分布)、適合度などである。研究者が、法則記述のプロセスと記述に用いる数理技法体系を同時提供したのである。

20世紀前半には、計量生物学、計量心理学、計量経済学といった計量諸科学が創生され、統計的記述は自然現象のみならず社会現象までその適用範囲を広げると共に、技法も進化を続けた。特に、相関概念や適合度概念は、人間が付与する法則のパフォーマンス尺度であり、より便宜性の高い法則を求める実証科学の営みを支える基幹概念となった。

有限のデータがもたらす有限の情報量に基づいて、統計的方法が人間の法則に関する認識の地平を広げるためには、3つのステップがある。

第一ステップでは、関心のある現象に対する既存知識を表現した統計モデルによる予測値とは乖離する実測値、いわゆる外れ値を検出し、それを問題視するか否かを検討する。この異常検知操作が、統計的方法による問題発見の基本形である。

第二ステップでは、データが外れ値となったことに影響を与えたと考えられる原因候補(要因)を考察し、要因に関わる変数を新たな入力データとして統計モデルに追加する。この際、必要に応じて新たなデータを統計モデル同定にとって最も情報効率が高くなるように採取する。

第三ステップでは、統計モデルをデータに当てはめ、その結果を解釈し、要因から原因を絞り込み(モデル選択)、現象に対する知識を更新する。

なお、既存知識の地平を表現し、現象に当てはめられる統計モデルには、物理学のような演繹的科学を基に、現象の静態や動態を記述する数理モデルを原型としたモデルも含まれる。計量科学は、理論なき実証と批判されることも多いが、「科学の文法」は、法則進化を目指すプロセスモデルであり、統計的方法を数理技術として制御や予測に応用する技術操作とは目的が異なる。

第二、第三のステップが、統計分析ないしは要因分析と呼ばれ、統計教育・研究の中心となった。この統計分析のあるべきプロセスを Cox and Donnelly (2011)は、次のようにまとめ“Ideal Sequence”と呼んでいる。

1. 研究すべき問題、仮説の定式化
2. 関連するデータの探索と適切なデータを採取する研究の計画と実施
3. データ解析
4. 適切な意思決定に繋がる結果の解釈

確かに、研究すべき問題や仮説の発想は、研究者の先見性や多様性受容能力も肝要である。しかし、上記第一ステップで示したように既存統計モデルからの外れ値という事実に着目すれば、新たな「突き止められる原因(Assignable Cause)」が明らかになる可能性が強い。この常識を基本原理として明示化し、社会実装したのは統計的品質管理学の創成者、Shewhart (1939)であり、実証科学の自律的改善のサイクル形成に寄与した。

2.2 設計科学の文法—統計数理的マネジメントの社会実装

吉田に主導された日本学術会議運営審議会附置新しい学術体系委員会(2003)は、価値に依拠した科学を「設計科学」と呼び、従来の価値に依存せず現象の背景原理に関する解釈を追究する科学を「認識科学」と呼んだ。吉田は設計科学の基本的方法を「情報循環」とした。

一方、Shewhart (1939)は、表1に示すように統計科学的マネジメントを大量生産のマネジメント更には一般社会のマネジメントに進化させた。Shewhartの影響を受け Deming, 石川馨らは、PDCA サイクルと「問題解決型 QC ストーリー」という統計的改善活動の標準プロセスを

表 1. Shewhart による統計科学プロセスの社会実装.

フェイズ	適用範囲		
	統計科学	狭義品質管理	マネジメント全般
Plan (計画)	仮説の提示	仕様の提示	達成すべき 目的の提示
Do (実施)	観察・実験	生産	目的を達成しよう とする行為
Check (検証)	仮説の検定	仕様適合性の検査	目的が達成された か否かの検証

社会実装した。これらは、まさに価値を追求するための汎用的プロセスモデルであり、「設計科学の文法」と呼ぶべきものであった(椿, 2019)。

設計科学が認識科学と異なるのは、価値関数の選択, 社会的制約条件の選択, 価値関数の制約付き最適化といった, 個々の人間の価値観や倫理観に依存するプロセスが明確に存在することである。このため, 数理計画や統計的決定理論が重要な位置づけになる。

認識科学の進化のためにも最適化や制約が用いられる。しかし, 認識科学ないしは法則科学では, 物質や情報の法則の背後にある変分原理や尤度原理に基づく天然の最適化や, 既存法則成立を制約とした現象認識といった, 価値には依存しない最適化や制約を用いるのである。

ここで, 価値最適化を実現する条件設定のための統計的方法開発の歴史を振り返る。多くは, 法則を進化させるための統計モデル当てはめという意識よりは, 必要な比較評価技法の開発である。Gosset は, 平均値の比較のための t 統計量を開発し, ギネスビールの品質改善を行った (Student, 1908; Box, 1987)。Fisher (1935) は, 実験計画法を提唱し, 農産物の収量や品質の改善に寄与した。田口 (2000) による機能性評価実験 (動特性 SN 比) は, 利用したい理想の法則 (機能) に現実がどのくらい近いのか, それをどう改善するかの体系的評価技法を示したものであり, 設計科学の発展に横断的に寄与した (椿・河村, 2008)。このようにデータに基づく方法が, 現象の認識だけでなく, 様々な要因に駆動されたシステムを統計的に比較して, 人間が利用するシステムの最適化にも利用されてきた。

実社会における最適化の対象は, 支配秩序としての物質やエネルギーなどの法則を利用する現象には限らない。整数計画法に代表される離散最適化を必要とする秩序体系が厳然と存在する。これを以下では, 吉田に従ってプログラムと呼ぶことにする。

3. プログラムの認識科学

3.1 秩序原理としてのプログラム

吉田 (2013) は, 設計科学と認識科学, 物質科学, 生物科学, 社会科学など様々な科学の分類軸を提案した。ここでは, 「法則科学」と「プログラム科学」という分類に注目する。物質科学は, 「法則」という違反もなければ例外もない秩序に支配されている。法則は, 物理学モデルやその影響をうける統計的定量モデルで近似可能である。

一方、生物学や社会科学は「プログラム」と呼ばれる、違反事象や例外事象を許容する秩序原理にも支配される。生物学は遺伝子のようなシグナル列(シグナル性プログラム)に支配され、社会科学はシンボル列で記述されるプログラム(シンボル性プログラム)に支配されるとした。

Pearson (1892)は、20世紀には物理学は数学との融合が進み、生物学は物理学との融合が進むと述べた。一方、Pearson (1892)はlawという用語が「法」という社会秩序にも用いられていることを意識しており、それが「法則」とは異なることを強調していた。また、Pearsonは、現実に認識される法則には、誤差があることも指摘した。これは様々な未知原因が観察現象を駆動するからであり、それら未知要因に起因する変動を記述する数理ツールとして確率モデル利用を提案したのである。なお、竹内(2010)も人間の「無知」の部分としての偶然概念を取り上げており、この種の認識は確率モデルの利用を推進した近代統計学の立場だと考えられる。

Pearson にとっての科学の進化は、利用可能な法則が、人間科学・社会科学も含む多くの分野に広がり、法則の現象近似パフォーマンスが上がることだった。これに対して、吉田のプログラムは、未知の原因が交絡していなくても、例外や違反が存在することを前提とする。逆説的だが、プログラム秩序を利用するものからすれば、プログラムの有用性は、例外や違反が生じる割合が小さいこととなる。しかし、プログラムの違反や例外は、無知のために確率変動と見なした原因で起きるのではない。異なる論理を共有する少数派と呼ばれる人間の判断には確信的違反や例外も存在する。

3.2 プログラム認識科学成立に必要な統計的方法と論理的方法

プログラム科学を実効化させるには、符号列を原因情報として、結果系パフォーマンスを認識評価する方法が必要となる。プログラム科学が事実に基づく方法を必要とするならば、統計的には2.2節で述べた離散化されたグループ間の事実に基づく評価・比較という方法体系が必要となる。

プログラムが何らかの法則を駆動し結果として量的パフォーマンスが得られる場合には、Fisher (1935)の実験計画法と多元配置分散分析を異なるシンボルやシグナルの作用といった質的制御因子に限定して適用すればよい。プログラム間の定量的結果比較を行えば、パフォーマンスの効率的認識と共に最適化まで行える。時系列的シンボル列の順序構造がパフォーマンスに与える影響評価は、医学統計ではFieler (1940)、農業統計ではCochran et al. (1941)のクロス・オーバー計画と呼ばれる実験計画法技術に埋め込まれた。

一方、離散的入力の質的出力パフォーマンスを評価する離散データ解析も統計学の枠組みで整備されてはいた。符号系列の質的パフォーマンスに関する影響評価についてもクロス・オーバー計画は利用可能であり、符号列投入の時期・順序・直前符号の残存影響など様々な検討が可能である。広津・椿(1984)は、3種の薬剤のうち2種類を様々な順序で一定期間ずつ投与し、前期・後期の有効性の優劣比較を行うクロス・オーバー試験から有効な治療法を定める統計モデルを議論した。表2のように薬剤名をシンボル化すれば、これは最も単純なシンボル性プログラム評価のためのデータと考えられる。空間的二値シンボル列(0-1列)と離散的パフォーマンスとの関係性を認識することは、文字認識など、背景法則が存在しない人工概念に対する統計的パターン認識は、統計的機械学習分野で日常的に用いられている。

しかし、離散現象の背後に連続的な法則を想定して、離散観測量から法則の性能評価することと、その種の法則科学的前提無しにプログラムのパフォーマンスを例外や違反の割合で評価することには、本質的な差異がある。

ある記号列のパフォーマンスに対する影響を評価する最も単純な状況は、ある原因系記号の有無と結果の成否をデータ化した表3のような2×2分割表で表現される。この分割表のある記

表 2. 3 薬剤 2 期間クロス・オーバー臨床試験治療時期優劣比較結果(広津・椿, 1984).

	AB	AC	CB	BC	CA	AC
記号 1 勝	3	7	10	4	8	4
記号 2 勝	9	9	12	13	13	10
引き分け	8	2	4	4	2	5

表 3. シンボル評価の最も単純な例.

		結果の成否		
		結果+	結果-	行合計
原因の有無	シンボル有	N_{++}	N_{+-}	N_{+}
	シンボル無	N_{-+}	N_{--}	N_{-}
	列合計	N_{+}	N_{-}	$N_{..}$

表 4. 例外のないシンボル性プログラム.

		結果の成否		
		結果+	結果-	行合計
原因の有無	シンボル有	N_{+}	0	N_{+}
	シンボル無	0	N_{-}	N_{-}
	列合計	N_{+}	N_{-}	$N_{..}$

号の有無, 結果の成否に任意の数量化を行い決定係数の二乗(相関係数)を求めたのが, 2×2 分割表の関連性尺度である, Pearson の ϕ 係数の二乗となる. 更にその $N_{..}$ 倍が, 分割表の独立性に関する Pearson カイ二乗適合度検定統計量である. Pearson は, 離散変量の世界でも連続量の世界同様, 相関分析で背後の法則を性能評価できると考えていた. 実際, メンデルの法則のような客観科学的離散法則を検討するには, 統計モデルによる法則評価は有効であった.

一方, 表 3 で入力記号のプログラムとしての有効性を評価するならば, プログラムが例外なく意図通り動作した割合を評価することになる. 例外の無い理想状態とは表 4 に示したような状態である. 確かに表 4 は ϕ 係数が 1 になる場合である. しかし, 独立性を統計的に評価するのではなく, 理想状態からのずれを測定するのがプログラム評価である. 論理的には表 4 の第 1 行が成り立つことは記号の存在が, 結果成立の十分条件, 第 2 行が成立するのが必要条件である. その種の論理がどの程度成り立っているかを定性的・論理的研究では重視してきたことは Goertz and Mahoney (2012) が, 強調した通りである. 表 3 で, 十分性に関するプログラムの性能は N_{++}/N_{+} , 必要性に関する性能は N_{--}/N_{-} で評価すればよい. 偽陽性率, 偽陰性率

といった検査性能評価が、医学では行われてきた。パターン認識の閾値評価にも ROC 曲線を用いることも多いが、これがパフォーマンスの背後にある法則を評価したものではなく、プログラムのパフォーマンスの論理性能評価と考えられる。

実際、Goertz and Mahoney (2012)は、科学的接近を統計的な量的方法と数学的・論理的な質的方法に峻別した。彼らが指摘したように、社会科学でも統計的接近が徐々に重視されてきている。しかし、筆者は、離散的シンボルやシグナルが本質的な秩序であるプログラム科学では質的方法が、法則科学では量的方法が適しているという仮説を持っている。4章では、筆者の限られた統計的体験の中で、統計的方法周辺に存在する量と質との問題の差異が、鮮明に現れた状況を紹介する。

4. 人間の生存に必要なだった離散化とそれに対峙する統計的方法

4.1 離散的認識における量的情報の捨象—離散的認識の人為性

人間の認識する事実は、時間的・空間的に何らかのバンドパスフィルタリングが施されている。実際、人間は極めて微小な領域の事実は認識できない。椿他 (1982)は、視覚も帯域上のバンドパス特性を持ち、それが錯視現象の原因と考えた。生物全般が生存に適した進化の中で、適切な情報の強調や不要な情報を捨象するための帯域特性を保有するようになったと推察する。このことが所定の主要特性を共有する集合の同一視と、その要素の数え上げという人間あるいは生物特有の操作を導いている。プログラム科学の背景にある離散性は、人工物としての数学に基づく認識と強く関連する。

数学が最初に扱うのは、自然数を用いた数え上げである。統計教育でもヒストグラムなどの度数分布の表示は第一歩である。更に理論度数分布としてのポアソン分布は、小数の法則として事故件数の従う分布として多用されている。

しかし、数え上げが可能なのは、集合としての要素が数学的に定義されているからである。リンゴが3個、ミカンが2個あったとき、果物が何個あるかという問いと、ミカンが何個あるかという問いとでは答えは異なる。約100gのミカンは何個あるかと問われれば、人間は $100 \pm \varepsilon$ gのミカンを同一視して数え上げを行う。事故であるか否かの判断も人間が事故の範囲を定めていることによって可能となる。数え上げ操作は、包含関係を人間が定義し、有限集合論的認識を支援するための重要な人為的操作である。

しかし、法則科学的観点からすれば、離散確率分布も、連続確率分布の極限分布として表現可能なことに留意すべきである。統計学で人間が量的現象を捨象し、数え上げの対象にしていることを端的に示すのが、ポアソン分布には「散らばり母数 (Dispersion Parameter)」が存在しないとの誤解である。Morris (1982)が、自然指数型分布族で分散関数が期待値 μ の1次式となるのは、離散分布であるポアソン分布に限ることを証明したことも誤解を助長させている。もちろん、これは非負整数値分布を前提にした証明である。

実際に、ポアソン変量の σ 倍の確率変数は、 $0, \sigma, 2\sigma, \dots$ といった離散的値をとり、分散関数 $V(\mu) = \sigma^2 \mu$ となる指数型分布族である。そのような状況を量的観測ではなく、量を捨象して数え上げとするのが人間の判断としては経済的なことは否定しない。

しかし法則科学にとって重要なのは、この散らばり母数を有するポアソン分布は、散らばり母数を持つ指数型連続分布族の極限として理想化された分布だという認識を持つことである。すなわち、分散関数 $V(\mu) = \sigma^2 \mu^p$ で特徴づけられるガンマ複合ポアソン分布($1 \leq p \leq 2$)で非負連続値分布である Tweedie 分布の極限である (Jørgensen, 1987)。

Tweedie 分布は、同一分布に従うガンマ変量を散らばり母数の無いポアソン変数個を加える、量の加法操作から得られる。この量的操作結果の観測が、より法則科学に近い認識行為であ

る。ポアソン変量とは、分散 0 のガンマ変量を確率的個数で加え合わせているから離散変量となっているだけである。

非負整数値をとる 2 変量ポアソン分布に相関係数概念を導入することは、通常、擬似尤度あるいは一般化推定方程式でしかできない。同一のポアソン変数個、別々のガンマ変量を加えた 2 変量 Tweedie 分布は、実在する指数型分布族であり、母相関係数がパラメータとして存在し、量的関連性を記述することができる確率モデルとなっている (Iwasaki and Tsubaki, 2005)。

本来、統計モデルによる法則科学的認識は、主として人間の周辺に内在する量的性質に関わる方法として開発されてきたことを意識すべきである。これは、人間の認識の特性・捨象・価値観・論理に依存しない法則を追究する営みであった。

4.2 論理的測定モデルと統計的測定モデル—実在する概念とそうでない概念

Bollen (1989) は、統計的測定モデルの第 1 段階としての概念の理論的定義を挙げた。その好例として、CIA (1982) の「テロリズム」の定義が示されている。当時の CIA の定義が、「暴力と関連する行為」かつ「政治的意図を持つ行為」かつ「直接の被害者を越えた範囲に影響を与えることを意図した行為」という 3 特性を有することが、テロリズム事象を定義する必要十分条件であることが明確になっているからである。この種の考察は用語学 (Terminology) から導かれる離散的かつ集合論的なものである。ISO 1087-1:2000 は、国際標準化活動における用語の定義に関わる理論と方法を標準化したものである (ISO, 2000)。実際、ここでは、概念は特性群の特定の結合から構成される知識、内包 (Intension) は、概念を形成する特性の集合として定義されている。従って、CIA (1982) の理論的定義から Bollen (1989) が特性を列挙した方法で、概念の成立は複数特性を共有する必要十分条件として定義され、定義された概念に対象が属するか否かは論理的に定まる。概念に対応する英語なり日本語等の用語が存在しても、概念の定義は数理的・論理的に確定する。しかし、唯名論の立場をとらずとも、テロリズムといった概念は、物質科学の世界には実在せず、人間が認識の便宜のために対象を分類した数学的概念にすぎない。

それにも関わらず、Bollen (1989) は、用語学的概念定義の次の段階として、統計的測定モデル、例えば因子モデルに関連する事実には当てはめる手続きを与えている。このとき特性と概念との関係は量的法則モデルで記述される。例えば、テロリズムという論理的概念の背景に、テロらしさといった潜在標準正規確率変数 X を想定し、暴力行為らしさ Y_1 、政治的行為らしさ Y_2 、範囲の広範性らしさ Y_3 を目的とした行為らしさという特性をアナログスケールなどで近似的に連続変数と考えられる尺度で観測することで、連立方程式モデルである因子モデル

$$(4.1) \quad Y_j = \alpha_j + \beta_j X + \varepsilon_j, j = 1, \dots, 3, \varepsilon_j \sim N(0, \sigma_j^2)$$

を当てはめ、潜在因子 X の測定方法が導かれるのである。

確かに、この種の考え方は、背後にある潜在因子が実在する概念の場合には有効である。Spearman (1904) は、一般知能因子が存在し、それが所定の試験項目に正の影響を与えると考えた。実在する潜在変数を客観測定するために、適切な観測項目が設計されることは、計量心理学分野だけでなく物理測定分野ですら行われてきた法則科学的営みである (例えば、牧 他, 2004)。

Bollen (1989) の理論的用語定義は、論理学・用語学的概念測定にとって重要だが、この唯名論的概念の測定に、法則科学的統計モデル (4.1) を適用するには、次の問題がある。統計的測定モデル (4.1) の下では、全くテロらしさが無い行為、すなわち連続変数として X がマイナス無限大になったとき、暴力行為・政治的行為・行為が影響する範囲もマイナス無限大になる。すなわち、テロらしさが全くなければ、暴力行為らしさも全くなくなることを意味する。しか

し、全くテロらしくない行為であっても政治家の演説のように政治的行為らしさをもつ事象は存在する。

ここでも、Goertz and Mahoney (2012)が指摘したように、連続的な法則に基づく統計的推論と、必要条件や十分条件といった論理的推論には乖離が生じている。

以下では、具体的に論理的推論モデルと統計的離散測定モデルとの乖離を例示したい。そのために、ある母集団に属する人が、ある事象を q 個の特性から定義される概念に適合すると判断する確率を P_0 、その事象が特性 j ($j = 1, \dots, q$) を有すると判断する確率を P_j とし、どのような確率間の構造を考えるのが妥当かを調べる。このとき、次の3条件を要求するのが論理的条件と整合的である。

条件1: $P_0 = f(P_1, P_2, \dots, P_q)$ は、 P_j について単調増加関数である。

条件2: ある特性 j について $P_j = 0$ ならば、 $P_0 = 0$ となる。

条件3: $P_0 = 1$ ならば、全ての j について $P_j = 1$ となる。

条件2は、ある特性の完全な欠落は概念の完全に不成立となることを意味する。条件3は、完全な概念の成立は、全特性が完全に成立することを意味する。注意が必要なのはある特性 j について $P_j = 1$ となっても、必ずしも P_0 は1とはならないことである。

この種の条件を満たす関数形 f は、一意に定まるわけではない。例えば条件3は、

$$(4.2) \quad \log(1 - P_0) = \alpha_j + \beta_j \log(1 - P_j), \quad \beta_j > 0$$

が満たされれば成立する。(4.2)は、 $\log(1 - P_j) = \alpha_j^* + \beta_j^* \log(1 - P_0)$ と書き換えれば、あまり自然ではないが、 P_0 の統計的測定モデルと見なせる。一方、条件2は、例えば

$$(4.3) \quad \log P_0 = \gamma + \delta \sum_{j=1}^q \log P_j, \quad \delta > 0$$

が満たされれば成立する。そこで、離散統計モデルとしては全く異形だが

$$(4.4) \quad \log(1 - P_0) = \alpha_j + \beta_j \log(1 - P_j) + \delta \sum_{k=1}^q \log P_k \quad \beta_j > 0, \quad \delta > 0 \quad j = 1, \dots, q$$

というモデルを考えると、条件1から3は満たされる。これを概念が原因で特性が結果とする測定モデル(4.1)に類似させて書き直せば、

$$(4.5) \quad \log(1 - P_j) = \alpha_j^* + \beta_j^* \log(1 - P_0) + \delta^* \sum_{k=1}^q \log P_k$$

となる。(4.5)は右辺第3項の存在のため、統計的測定モデルとはならない。(4.5)式左辺並びに右辺2項をロジットに置き換えることが統計的には自然だが、これを入れると $P_0 = 0$ ならば、全ての特性 j について $P_j = 0$ を要求することになり、論理的な条件を不必要に追加したことになる。

このように論理的測定を記述する確率モデルの定式化は一意ではないが、いずれにしても統計的測定モデルとは本質的に異なる。従って、論理的概念測定と統計的概念測定は異なる行為と考えなければならない。

4.3 論理的分類と統計的分配

ある特性の有無に関する認識を統計的に扱うために、4.2節ではその特性が有るか否かの回答に曖昧さが無く、各個人については True か False が定まることを前提にした。このとき、有

限混合分布といった頻度論的確率分布が用いられる。しかし、各個人が True なり False と事象を判断することは、人為的論理操作と考えられる場合が多い。

公的統計分野では、収集したデータを集計するために、家計統計では家計簿に記述された情報から商品分類、事業所統計では調査票情報から事業所の産業分類を行わなければならない。しかし、家計統計では、被調査者が 200 円支払った商品が、発泡酒なのかビールなのかは、商品の定義が明確かつ網羅性があるのならば、統計的教師付き分類問題として明確に定式化できる。一方、調査対象事業所の売り上げが 1500 万円として、その事業所は製造業としての売り上げもあれば、サービス業としての売り上げもあるという状況がある。このとき事業所の産業分類は、売り上げを各業種に按分して按分率の高い業種に形式的に分類を行うこととなる。この按分率は、論理的分類を前提とする有限混合分布の混合確率ではない。背後にあるのは連続量の分配である。

このように人間が介在する分類問題の背後には様々な曖昧さがあるため、Goertz and Mahoney (2012) が、示唆するようにテロらしさ、暴力行為らしさといった曖昧性をファジイ集合論で表現しようという立場も生じる。

しかし、事象自体が頻度論的かつ論理的に確定するのではなく、法則科学的実態を持つ量が加法的に観測されていることを統計モデル化することは可能である。椿 (1999), Tomosada and Tsubaki (2011) は、画像処理の Mixed Pixel 問題が、分類問題ではなく按分問題(粒子混合)であるという立場の統計モデルを提案した。以下にその概要を示す。

p 次元観測変量 \mathbf{Y} は、 K 個の細分された変量の和であるとする。按分対象となるカテゴリは Q 個あり、細分された変量は、あるカテゴリ q に属し、 K_q 個存在すると仮定する。このとき、 \mathbf{Y} は、次のように量の加法操作で表現される。

$$\mathbf{Y} = \sum_{q=1}^Q \sum_{k=1}^{K_q} \mathbf{X}_{qk}$$

但し、 \mathbf{X}_{qk} は、カテゴリ q に属する細分変量を意味する。さらに、 \mathbf{X}_{qk} が互いに独立に、モーメント母関数 $\phi_q(\mathbf{t}) = E[\exp(\mathbf{t}^T \mathbf{X}_{qk})]$ を持つ同一分布に従うと仮定する。このとき、カテゴリ q のデータだけからなる観測変量 \mathbf{Y}_q のモーメント母関数は $\psi_q(\mathbf{t}) = \phi_q(\mathbf{t})^{1/K}$ となる。

従って、 $p_q = K_q/K$ と置けば、観測変量 \mathbf{Y} の従う分布のモーメント母関数は(4.6)のようになる。

$$(4.6) \quad \psi_{\mathbf{Y}}(\mathbf{t}) = \prod_{q=1}^Q \phi_q(\mathbf{t})^{p_q}$$

従って、そのキュムラント母関数は、純粹観測ベクトルの従う分布のキュムラント母関数が、比率 p_q で荷重和されたものになる。通常の頻度論的混合では、密度関数あるいはモーメント母関数が混合確率で荷重和されるのと、期待値構造は一致するが、2 次のモーメント構造は異なる。

このように按分的混合は、量の加法操作を背景機序として前提にする、論理的分類ではなく量の分配率の推定問題である。ファジイ理論が扱うような人間の分類行為には、論理学と法則科学との中間的状況も存在すると考えられる。そのような場合への一つの対処法として、確率変数 X_i のモーメント母関数 $\varphi_i(\theta) = E[\exp(\theta X_i)]$, $i = 1, \dots, I$ に対して、一般化キュムラント母関数を

$$(4.7) \quad \eta_i(\theta, t) = \{\varphi_i(\theta)^t - 1\}/t$$

と定義することが考えられる。このとき、

$$\begin{aligned}\eta_i(\theta, t) &= \left[\left\{ 1 + \theta E[X_i] + \frac{\theta^2}{2} E[X_i^2] + \dots \right\}^t - 1 \right] / t \\ &= \theta E[X_i] + \frac{\theta^2}{2} E[X_i^2] + \dots + \theta^2 (t-1) / 2 \left\{ E[X_i] + \frac{\theta}{2} E[X_i^2] + \dots \right\}^2 + \dots \\ &= \theta E[X_i] + \frac{\theta^2}{2} \{ E[X_i^2] + (t-1) E[X_i]^2 \} + o(\theta^2)\end{aligned}$$

と展開されるので、一次一般化キュムラントは、 $E[X_i]$ 、二次一般化キュムラントは、 $E[X_i^2] + (t-1)E[X_i]^2$ となる。

この一般化キュムラント母関数の混合を $\eta(\theta, t) = \sum_{i=1}^I p_i \eta_i(\theta, t)$ を確率モデルとして採用すれば、頻度論的(論理的)分類と連続的分配とを一つのパラメータで繋ぐことができる。なお、混合後の一般化キュムラントは、一次一般化キュムラントが $c_1 = \sum_{i=1}^I p_i E[X_i]$ 、二次一般化キュムラントが $c_2 = \sum_{i=1}^I p_i \{ t E[X_i]^2 + \text{Var}[X_i] \}$ となる。

量の加法的操作は法則科学を記述する方法論であり、4.1節で論じたように、ポアソン分布をガンマ分布も含まれる Tweedie 連続分布の極限として捉えれば、量の按分率推定問題がベータ分布やディリクレ分布と同等の意味を持つ、二項分布や多項分布を極限として導いていることが分かる。

統計的分類問題、パターン認識問題は、論理的推論を背景に持つ、プログラム科学的方法論である。しかし、逆に法則科学的の立場から、分類問題と分配問題との差異に留意する必要がある。

5. おわりに：プログラム科学の文法を目指して

5.1 法則科学的検討を進めるべきプログラム科学的問題

一見、プログラム秩序、離散的シンボルで表記されている現象であっても、背後に連続量が存在し、閾値を設定した人為的切断でシンボルが形成されている場合が多い。このような量的情報を論理的に扱う必然性はない。第一段階で法則科学的洞察を行い、第2段階でそこに閾値を入れたときのパフォーマンス評価を行えば良い。

一方で、この種の人為的切断がない、純粋な記号と量的パフォーマンスの関係は法則科学的には、どのように考えれば良いのだろうか。これは質と量との混合問題である。これが、端的に現れるのが遺伝子情報といった純粋な記号列が、物質科学的法則に対してシグナルとして作用する生物学である。

先ず量的科学として、関心のある量に影響を与える量を網羅し、記号が関心のある量に直接影響を与える可能性だけでなく、関心のある量に直接影響する量に影響を与えている可能性をも探求することは最低限必要である。ただし、これは法則科学的研究である。

一方、 M 個の純粋記号からなるシグナル記号列が、量にどのような影響を与えるかは、論理的・演繹的考察も重要である。実際、シグナル記号列の量への影響は論理のパズル、鍵をあけるような組み合わせ最適化を必要とする。しかし、統計的方法が開錠を効率化する可能性は大きい。それが、3章で紹介した M 水準の質的要因からなる実験計画法で問題を定式化することである。ただし、連続要因の場合と異なり、高次の交互作用効果が、鍵を解くのに重要となることは容易に予想される。従って、高次交互作用項(論理型入力変数の積の項)も考慮したモデルをデータに当てはめることが必要となる。

この種のモデル同定を限られた標本から効率的に行う方法は、少なくとも2つ現存すると考えている。一つはカーネル法の適用により高次交互作用をモデルに投入することである。もう

一つは尺度化と非線形性に関する射影追跡回帰 (Friedman et al., 1981)により, 尺度間交互作用の消去を試みることである. 前者は, サポートベクターマシン, 後者はニューラルネットワークなどの統計的機械学習技術として, 今や日常的に用いられるようになった. この種のセミパラメトリックモデルを効率的に同定するデータ収集技法としての逐次実験計画法も強化学習と呼ばれるようになった. 実験計画法の配置理論にとって, 重要な研究分野であるが, その最適計画については筆者には専門性はない.

5.2 純粹プログラム科学とその文法

人間社会を動かす, 法や政策は吉田のシンボル性プログラムであり, その成果も経済的効果のような量に変換されない, 社会構成員の感性的反応となることも多い. この種の純粹プログラム科学に近い問題については, ロジックツリーと MECE (Mutually Exclusive and Collectively Exhaustive) と呼ばれる本来純粹数学的接近を適用することが学術のみならず, 社会全般に広がっている. 日本の EBPM (Evidence Based Policy Making) 推進においても, 統計的方法と共に大きな位置づけを占めるようになってきている (行政改革推進本部, 2017). この種の方法論が統計科学の文法とは一線を画するものであることが, 本論の主張だった. しかし, 品質経営分野で統計的問題解決の標準シナリオ以外に, 論理的方法が主眼となる「課題達成型 QC ストーリー」と呼ばれる標準シナリオが 1980 年代から普及していることを再度強調しておきたい (狩野, 1993).

- 1) テーマの選定
- 2) 課題の明確化と目標設定
- 3) 方策の立案・選定
- 4) 成功シナリオの追究と実施
- 5) 効果の確認
- 6) 歯止め

これを支える論理的技法群 (ロジックツリーをも含む新 QC 七つ道具など) も整備されていた. 仮に統計的方法がその有効性を発揮するとすれば, プログラムの質的パフォーマンスを評価するシンボル列に関する If-Then 論理の発見であり, これは Breiman et al. (1984) の CART, すなわち自動層別技術の適用である. 今日的には, 層別原理をアンサンブル学習で高性能化する Breiman (2001) の Random Forest も有力な記号要因解析技法となろう.

一方で, プログラム科学が科学として成立する論理条件は, 科学的プロセスとしての Pearson の科学の文法, あるいは Shewhart, Deming, 石川の PDCA サイクルとの整合性である. なぜならば, この種の実証的プロセスが, 論理的方法論ですら自律的進化と再現性の確保のために重要だからである. 例えば,

- 1) あるべきプログラム・パフォーマンスの明確な定義
- 2) プログラム・パフォーマンスにおける看過できない違反や例外の発見
- 3) パフォーマンスに影響をあたえる可能性のある記号列の論理的網羅
- 4) 記号列の必要十分性の観点からの比較・評価
- 5) 最適な記号列プログラムの実装
- 6) プログラムの実施とそのパフォーマンスのモニタリング

といった課題達成型 QC ストーリーと PDCA サイクルを結合したプロセスモデルが, プログラム科学の文法として有効な可能性はある. しかし, それも論理科学的方法の実践の中で実証されるべき仮説にすぎない. 統計数理科学が将来 2 つの科学的方法のスコープマネジメントを

適切に実現できる日がくることを期待して、本論を閉じたい。

査読者から示された論点と謝辞

本ノートに対しては、査読者より本質的な3つの論点を提示いただいている。第一は、プログラム科学と法則科学とが、科学の特徴づけというよりは、外界との関係性の違いであるとの指摘である。自身の論理ないしはモデルを変更しつづける、すなわち、外界からのフィードバックが不可欠か否かということではないかという見解である。第2は、プログラム科学においても、現実ないしは観察から独立ではなく、出発点は帰納的観察に基づいているのではないかという指摘である。この2つの指摘は、行動モデルの観点からは明らかに正しく、筆者も同意する。生物の進化ですら、外界とのフィードバックがトリガーで起きていることに間違いはない。また、プログラム科学が支配的とされる生物科学、社会科学においても、法則はその基盤的秩序の一部を担っていることは、吉田の学術の体系でも前提になっている。筆者自体の疑問は、法則を認識し、それに基づいて何らかの設計に反映する統計科学的方法論と、プログラムを設計する方法論とに、科学的探究プロセスが少し異なるのではないかということである。もちろん、後者に対しても帰納的接近が有用であることについては、何ら疑いを持ってはいない。

第3の論点は、一般化キュムラント関数のように、頻度論的分類と連続的分配とを接続すること自体が、プログラム科学と法則科学との区別の必要性がないことを意味しないかという疑問である。この指摘は、筆者の視点が統計科学に近すぎる、すなわち法則科学的立場で頻度論的分類を眺めていることに気づかされた。科学的法則と論理的秩序を混合させること自体に実学的意義が有るかについては、査読者から提起された課題として強く意識したい。以上、本ノートに貴重な論点を示唆すると共に、様々な誤りを指摘していただいた査読者に深甚の謝意と敬意を捧げたい。

参 考 文 献

- Bollen, K. A. (1989). *Structural Equations with Latent Variables*, Wiley, Hoboken.
- Box, J. F. (1987). Guinness, Gosset, Fisher, and small samples, *Statistical Science*, **2**(1), 45-52.
- Breiman, L. (2001). Random forests, *Machine Learning*, **45**(1), 5-32.
- Breiman, L., Friedman, J., Stone, C. J. and Olshen, T. A. (1984). *Classification and Regression Trees*, Wadsworth, Belmont.
- Central Intelligence Agency (1982). Patterns of international terrorism: 1981, <https://www.cia.gov/library/readingroom/document/cia-rdp85m00363r000901960025-4>.
- Cochran, W. G., Autrey, K. M. and Cannon, C. Y. (1941). A double change-over design for daily cattle feeding experiments, *Journal of Dairy Science*, **24**(11), 937-951.
- Cox, D. R. and Donnelly, C. A. (2011). *Principles of Applied Statistics*, Cambridge University Press, Cambridge.
- Fieler, E. C. (1940). The biological standardization of insulin, Supplement to *the Journal of the Royal Statistical Society*, **7**(1), 1-64.
- Fisher, R. A. (1935). *The Design of Experiments*, Oliver and Boyd, Edinburgh.
- Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression, *Journal of American Statistical Association*, **76**, 817-823.
- Goertz, G. and Mahoney, J. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*, Princeton University Press, Princeton. (西川賢, 今井真士 訳 (2015). 『社会学のパラダイム論争：2つの文化の物語』, 勁草書房, 東京.)

- 行政改革推進本部 (2017). 「EBPM の試行的検証」EBPM 推進に向けた取組み・ロジックモデルなどについて, https://www.cas.go.jp/jp/seisaku/gyoukaku/H27_review/H29_fall_open_review/z7.pdf.
- 広津千尋, 椿広計 (1984). クロス・オーバー計画でとられた臨床データの統計モデル, 「実験データの解析の理論的背景」, 京都大学数理解析研究所講究録, **526**, 46-70.
- ISO (2000). *Terminology Work — Vocabulary — Part 1: Theory and Application*, ISO 1087-1:2000, International Organization for Standardization, Geneva.
- ISO (2015). *Application of Statistical and Related Methods to New Technology and Product Development Process — Part 1: General Principles and Perspectives of Quality Function Deployment (QFD)*, ISO 16355-1: 2015, International Organization for Standardization, Geneva.
- Iwasaki, M. and Tsubaki, H. (2005). A new bivariate distribution in natural exponential family, *Metrika*, **62**, 323-336.
- Jørgensen, B. (1987). Exponential dispersion models, *the Journal of the Royal Statistical Society, Series B*, **49**(2), 127-162.
- 狩野紀昭 監修 (1993). 『QC サークルのための課題達成型 QC ストーリー』, 日科技連, 東京.
- 牧雅康, 沖一雄, 椿広計, 安岡善文 (2004). 潜在要因の影響を考慮した水質計測の因果モデリング, 日本リモートセンシング学会誌, **24**(2), 153-162.
- Morris, C. (1982). Natural exponential families with quadratic variance functions, *Annals of Statistics*, **10**(1), 65-80.
- 日本学術会議自動制御研究連絡委員会工学共通基盤研究連絡委員会自動制御学専門委員会 (2005). 横断型基幹科学技術としての制御学の役割—「知の統合」を目指す研究・教育の促進に向けて, 自動制御研究連絡委員会工学共通基盤研究連絡委員会自動制御学専門委員会報告, <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-19-t1031-14.pdf>.
- 日本学術会議運営審議会附置新しい学術体系委員会 (2003). 新しい学術の体系—社会のための学術と文理の融合—, 日本学術会議, <http://www.scj.go.jp/ja/info/kohyo/18pdf/1829.pdf>.
- Pearson, K. (1892). *The Grammar of Science*, Walter Scott, London.
- Shewhart, W.A. (1939). *Statistical Method from the Viewpoint of Quality Control* (eds. W. E. Deming), The Graduate School, the Department of Agriculture, Washington D.C. (坂元平八 訳 (1960). 『品質管理の基礎概念—品質管理の観点からみた統計的方法』, 岩波書店, 東京.)
- Spearman, C. (1904). 'General intelligence,' objectively determined and measured, *The American Journal of Psychology*, **15**(2), 201-292.
- Student (1908). The probable error of a mean, *Biometrika*, **6**(1), 1-25.
- 田口玄一 (2000). 『ロバスト設計のための機能性評価—効率的開発の方法』, 日本規格協会, 東京.
- 竹内啓 (2010). 『偶然とは何か—その積極的意味』, 岩波新書 1269, 岩波書店, 東京.
- Tomosada, M. and Tsubaki, H. (2011). Novel mixture model for mixed pixel classification of multi-spectral image data, *International Journal of Mathematical Models and Methods in Applied Science*, **5**(1), 175-183.
- 椿広計 (1999). リモートセンシングにおける分類 (2) ミクセルの識別に関する統計的考察, 日本リモートセンシング学会誌, **19**(1), 75-80.
- 椿広計 (2017). サービスと品質管理, 品質管理学会等の活動を振り返る, 標準化と品質管理, **70**(10), 2-8.
- 椿広計 (2018). データ駆動型社会の人と品と質とのマネジメント, 応用統計学, **47**(2&3), 89-98.
- 椿広計 (2019). 第 1 部日本社会における統計科学の展開, 第 2 章日本の品質管理活動と統計科学 (国友直人, 山本拓 編), 『統計と日本社会 データサイエンス時代の展開』, 東京大学出版会, 東京.
- 椿広計, 河村敏彦 (2008). 『設計科学におけるタグチメソッド—パラメータ設計の体系化と新たな SN 比解析』, 日科技連, 東京.
- 椿広計, 鈴木守, 花泉弘, 藤村貞夫 (1982). 視覚の帯域通過特性に基づく錯視現象の解析, 計測自動制御学会, 日本機械学会, 日本自動制御協会第 25 回自走制御講演会要旨集, 509-510.

- Tsubaki, H., Nishina, K. and Yamada, S. (eds.) (2008). *The Grammar of Technology Development*, Springer, Tokyo.
- 土橋俊人, 高須久, 椿広計 (1985). どう解析するかこのデータ, どう収集するかこの言語情報, 品質, **15**(3), 56-66.
- VCP-Net 研究会 編著 (2014). 『知の巡りをよくする手法の連携活用—サービス・製品の価値を高める価値創生プロセスのデザイン』(大藤正, 黒河英俊 監修), 日本規格協会, 東京.
- 吉田民人 (2013). 『近代科学の情報論的転回 プログラム科学論』(吉田民人論集編集委員会 編), 勁草書房, 東京.

Grammar of Statistical Science and Logical Reasoning

Hiroe Tsubaki

The Institute of Statistical Mathematics

Although the standardized research process of science should be a statistical and quantitative exploration process in order to give appropriate laws to the Nature or societies, the main research process of science with program as an artificial and weak order may be logical and qualitative, as suggested by Goertz and Mahoney (2012). From a statistical scientific point of view, I illustrate some situations in which the difference between quantitative and logical processes is noticeable. I also discuss the possibility of unifying the two scientific approaches along the Grammar of Science by Pearson (1892).

時系列解析における状態空間モデルの利用

北川 源四郎†

(受付 2019 年 1 月 7 日; 改訂 3 月 18 日; 採択 3 月 26 日)

要 旨

時系列解析において状態空間モデルは、統計的制御における最適制御則の導出や ARMA モデルの最尤推定など陽な形で解が得られない問題に対する解を求めるための逐次計算を実現するための計算手段として導入されたが、1980 年頃からは非定常時系列モデリング、非線形モデリング、信号分離、異常値処理、自己組織型モデリング、データ同化など、様々な時系列モデリングを統一的に取り扱うためのプラットフォームとしての役割りを果たしてきた。その一方で、状態空間モデルに関連した状態推定のために、カルマンフィルタ、ガウス和フィルタ、非ガウス型フィルタ、粒子フィルタなどの様々なアルゴリズムが開発されてきた。

本項では、統計数理研究所の研究を中心に、状態空間モデルと関連する計算法およびその応用について概観する。

キーワード：状態推定，非定常モデル，非線形モデル，カルマンフィルタ，非ガウス型フィルタ，粒子フィルタ。

1. 状態空間モデルとカルマンフィルタによる状態の推定

線形・ガウス型状態空間モデルは、 d_m 次元時系列 y_n の変動を下記のように 2 つのモデルで表現する。

$$(1.1) \quad x_n = F_n x_{n-1} + G_n v_n \quad (\text{システムモデル})$$

$$(1.2) \quad y_n = H_n x_n + w_n \quad (\text{観測モデル})$$

ただし、 x_n は d_k 次元状態ベクトル、 F_n, G_n, H_n はそれぞれ $d_k \times d_k, d_k \times d_\ell, d_m \times d_k$ の行列である。 $v_n \sim N(0, Q_n)$ と $w_n \sim N(0, R_n)$ はそれぞれ d_ℓ 次元および d_m 次元の正規分布に従う白色雑音でシステムノイズ、観測ノイズと呼ばれる。通常はこの線形・ガウス型状態空間モデルのことを単に状態空間モデルと呼ぶことが多い。物理的なシステムでは、過去の情報は現在を經由して将来に伝えられるが、状態空間モデルにおいてはこの状態 x_n だけが分かれば、時系列の将来の動きは予測できることになる。

状態空間モデルに関して、時系列 $Y_j = \{y_1, \dots, y_j\}$ が与えられたとき状態 x_n を推定する問題が状態推定である。観測値の最終時点 j と推定する状態の時刻 n の大小関係により、予測 ($n > j$ の場合)、フィルタ ($n = j$ の場合)、平滑化 ($n < j$ の場合) と区別される。時系列の予測、制御、補間、パラメータ推定、成分分解などの問題のほとんどがこの状態推定を利用することによって統一的に解決できる。

† 東京大学 数理・情報教育研究センター：〒113-8556 東京都文京区弥生 2-11-16

初期状態 x_0 の分布が正規分布 $N(x_0, V_0)$ で与えられるとき、状態 x_n の条件付き分布 $p(x_n|Y_j)$ は正規分布となる。したがって、線形・ガウス型状態空間モデルに対しては、カルマンフィルタによって効率的に条件付き平均 $x_{n|j}$ と分散共分散行列 $V_{n|j}$ を求めることによって状態推定が実現できる (Anderson and Moore, 1979; 片山, 2005; 北川, 2005)。

このように、状態空間モデルを利用すると、時系列解析に用いられる様々なモデルを統一的に取り扱うことができ、しかも予測、平滑化、成分分解、パラメータ推定などに必要となる状態推定をデータ数に比例する計算量で効率的に計算できる。当初は、統計的制御における最適制御則の導出や ARMA モデルの最尤推定など陽な形で最適解が得られない問題に対して、逐次計算によって解を求めるための計算手段として導入されたが、1980 年頃からは非定常時系列モデリング、非線形モデリング、信号分離、異常値処理、自己組織型モデリング、データ同化など、様々な時系列モデリングの問題を統一的に取り扱うためのプラットフォームとしての役割りを果たしてきた。その一方で、状態空間モデルに関連した状態推定のために、カルマンフィルタ、ガウス和フィルタ、非ガウス型フィルタ、粒子フィルタなどの様々なアルゴリズムが開発されてきた。

本項では、統計数理研究所の研究を中心に、状態空間モデリングの発展と関連する計算法およびその応用について概観する。

2. 線形定常時系列解析における状態空間モデルの利用

2.1 統計的制御

化学プラント、火力発電所や大型船舶のように巨大で強い外乱を受けるシステムの制御では、制御対象となるシステムのモデルを求めることが困難なために、最適制御理論の適用が困難であった。Akaike (1970) は多変量 AR モデルを制御入力 r_n を持つ状態空間モデル

$$(2.1) \quad \begin{aligned} x_n &= Fx_{n-1} + Gr_n + v_n \\ y_n &= Hx_n \end{aligned}$$

で表現し、2 次評価基準の期待値

$$(2.2) \quad J(x_0, r) = E \left[\sum_{n=1}^L \{x_n^T Q x_n + r_{n-1}^T R r_{n-1}\} \right]$$

を最小化する制御入力を決定する方式を提案した。この方法は、セメントの焼成炉や火力発電所ボイラーの制御や船舶のオートパイロットの設計に適用され実用化されたが⁸ (赤池・北川, 1994, 1995; Ohtsu et al., 2015)、これが時系列解析における、状態空間モデル利用のさきがけとなった。

2.2 時系列モデルの尤度計算とパラメータ推定

状態空間モデルに含まれる未知のパラメータを θ とし、長さ N の時系列 $Y_N = \{y_1, \dots, y_N\}$ が与えられるとき、時系列モデルの尤度は同時確率密度関数を用いて $L(\theta) = f_N(Y_N|\theta)$ と定義される。ここで $f_n(Y_n|\theta) = f_{n-1}(Y_{n-1}|\theta)p(y_n|Y_{n-1}, \theta)$ という分解を繰り返し適用して、 d_m 次の密度関数に分解していくと最終的には

$$(2.3) \quad L(\theta) = \prod_{n=1}^N p(y_n|Y_{n-1}, \theta)$$

と表現できる。ただし $\varepsilon_n = y_n - H_n x_{n|n-1}$, $\Sigma_n = H_n^T V_{n|n-1} H_n$ とするとき、 $p(y_n|Y_{n-1}, \theta) =$

$(2\pi)^{-m/2} |\Sigma_n|^{-1/2} \exp\{-\varepsilon_n^T \Sigma_n^{-1} \varepsilon_n / 2\}$ である。したがって、一般に時系列モデルの対数尤度は

$$(2.4) \quad \log L(\theta) = \sum_{n=1}^N \log p_n(y_n | Y_{n-1}, \theta) = -\frac{d_m N}{2} \log 2\pi - \frac{1}{2} \sum_{n=1}^N \log |\Sigma_n| - \frac{1}{2} \sum_{n=1}^N \varepsilon_n^T \Sigma_n^{-1} \varepsilon_n$$

によって与えられる。

ここで、 $x_{n|n-1}$ と $V_{n|n-1}$ はカルマンフィルタで得られるので、状態空間モデルで表現できる時系列モデルに対してはカルマンフィルタの副産物として、対数尤度が統一的かつ自動的に計算できることになる。時系列モデルのパラメータの最尤推定値を求めるためには、数値的最適化によりこの対数尤度を最大とするパラメータを求めればよい。

2.2.1 ARMA モデルの最尤推定

AR(自己回帰)モデルに対しては Yule-Walker 法のほか、Parcor 法 (Burg, 1967), Householder-最小二乗法 (Akaike et al., 1975, 1979; Kitagawa and Akaike, 1978, 1981) などの多くの実用的な推定法が提案されている (北川, 2005)。これらは近似的な最尤推定量を与えるものと解釈できるが、ARMA モデルに対しては残差のフィードバックがあるために精度の良い同様の近似解を得ることができなかった。Box and Jenkins (1970) では、backcasting によってイノベーション系列を生成し、近似尤度を計算する方法が利用されている。

Akaike (1974, 1978) は ARMA(m, ℓ) モデル

$$(2.5) \quad y_n - \sum_{j=1}^m a_j y_{n-j} = v_n - \sum_{j=1}^{\ell} b_j v_{n-j},$$

に対して $H = [100 \cdots 0]$, $Q = \sigma^2$, $R = 0$,

$$(2.6) \quad F = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ a_k & a_{k-1} & \cdots & a_1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ r_1 \\ \vdots \\ r_{k-1} \end{bmatrix}$$

と定義することによって線形・ガウス型状態空間モデルで表現できることを利用して最尤推定を実現した。ここで、 r_j は ARMA モデルから定まるインパルス応答関数、 $k = \max(m, \ell + 1)$ である。ただし、Akaike (1974) ではカルマンフィルタは用いられておらず、また Akaike (1978) および Akaike et al. (1979) では、より高速な Morf et al. (1974) のアルゴリズムが用いられている。

2.3 欠測値の処理

欠測値の処理は統計学において重要な課題であり、EM アルゴリズムが利用されることが多い。しかしながら、状態空間表現とカルマンフィルタを用いると欠測値の処理は極めて明快かつ簡単である (Jones, 1980)。観測値 y_n が実際には観測されなかった場合には、 $Y_n = Y_{n-1}$ となるので、 $p(x_n | Y_n) = p(x_n | Y_{n-1})$ が成り立つ。これは予測分布とフィルタ分布が等しいことを意味するので、ある観測値が欠測となった場合には、対応するフィルタのステップを省略するだけでよいことになる。

特に、欠測値を推定したい場合には、フィルタリングを行った後に、平滑化を行えば、その平滑化分布が欠測値の推定値を与えることになる。ただし、パラメータ推定などでは、欠測値の補間をする必要はなく、実際に観測されたデータだけを用いて尤度を計算することができ

る。Kitagawa and Akaike (1981), 北川 (2005)には、顕著なスペクトルピークを持つ時系列の場合には、数 10 パーセントの欠測値があってもよいパラメータ推定や欠測値の補間ができる事が示されている。

2.4 不等間隔時系列

時系列が不等間隔に観測されている場合には、推定すべき信号が確率微分方程式で表される連続時間のモデル

$$(2.7) \quad f_t^{(k)} = a_1 f_t^{(k-1)} + \cdots + a_k f_t + v_t$$

によって生成されているものと仮定する。ただし、 $f_t^{(k)}$ は f_t の k 階微分、 v_t はウィーナー過程 W_t の独立増分で分散 σ^2 とする。このとき、この不等間隔時系列は係数 a_1, \dots, a_k から求められる遷移行列 F_t を用いて

$$(2.8) \quad \begin{aligned} x_t &= F_{t-s} x_s + G v_{t,s} \\ y_t &= H x_t + w_t \end{aligned}$$

と状態空間表現が得られる。ただし、 $G = I_k$, $H = [10 \cdots 0]$, $x_t = [f_t, f_t^{(1)}, \dots, f_t^{(k-1)}]^T$

$$(2.9) \quad v_{t,s} = \int_s^t F_{t-u} B dW_u, \quad \text{Cov}(v_{t,s}) = \int_s^t F_{t-u} B B^T F_{t-u}^T du$$

である。これによって、確率微分方程式モデルを不等間隔データにあてはめることができる (Kitagawa, 1984)。特に、 $f_t^{(k)} = dW_t$ というモデルを想定すると、不等間隔データの平滑化を行うことができる。

3. 非定常時系列解析への利用

赤池は Akaike (1980a, 1980b)においてベイズモデリングに基づく経済時系列の新しい季節調整法を提案した。季節調整法では時系列を

$$(3.1) \quad y_n = T_n + S_n + w_n$$

とトレンド成分 T_n , 季節成分 S_n , 不規則成分 w_n の 3 つ以上の成分に分解するが、未知数がデータ数より多いために、通常の最小二乗法や最尤法では意味のある解が得られない。赤池は対数尤度関数にペナルティ項

$$(3.2) \quad d^2(T_n - 2T_{n-1} + T_{n-2})^2 + r^2(S_n - S_{n-12})^2 + z^2(S_n + \cdots + S_{n-11})^2$$

を付加した最小化問題を考え、更にベイズ型情報量規準 ABIC を導入することによって、新しい季節調整法を提案した (Akaike, 1980b)。

この方法は、単に新しい季節調整法を開発したということに止まらず、データ数を超える大規模パラメトリックモデリングに適用可能な汎用的な方法を提供するものであり、現代的に言えば L_2 正則化パラメータを客観的に決定する方法を提案したという意味で画期的であった。

この方法を季節成分がない最も簡単な場合について、モデルの観点から見直してみるとトレンド成分に関して

$$(3.3) \quad \begin{aligned} T_n &= 2T_{n-1} - T_{n-2} + v_n, & v_n &\sim N(0, \tau^2) \\ y_n &= T_n + w_n, & w_n &\sim N(0, \sigma^2) \end{aligned}$$

という 2 つのモデルを想定することと同等であることがわかる。この二つのモデルは 2 次元の

状態空間モデルとして表現できることから、トレンド推定や季節調整の問題が状態空間モデルによって解決できること、および平滑化パラメータ d^2 が分散比 σ^2/τ^2 に対応することを示唆している。

3.1 非定常時系列モデリングによる成分分解

一般に、状態空間モデルによる成分分解は次のように定式化できる。時系列 y_n を次のように J 個の成分 $x_n^{(j)}$, $j = 1, \dots, J$ と観測ノイズ w_n に分解するものとする。

$$(3.4) \quad y_n = x_n^{(1)} + \dots + x_n^{(J)} + w_n$$

さらに、それぞれの成分 $x_n^{(j)}$ は状態空間モデルで

$$(3.5) \quad x_n^{(j)} = F^{(j)}x_{n-1}^{(j)} + G^{(j)}v_n \quad (\text{システムモデル})$$

$$(3.6) \quad y_n = H^{(j)}x_n^{(j)} + w_n \quad (\text{観測モデル})$$

と表現できるものとする。このとき、

$$(3.7) \quad F = \begin{bmatrix} F^{(1)} & & \\ & \ddots & \\ & & F^{(J)} \end{bmatrix}, \quad G = \begin{bmatrix} G^{(1)} & & \\ & \ddots & \\ & & G^{(J)} \end{bmatrix}, \quad x_n = \begin{bmatrix} x_n^{(1)} \\ \vdots \\ x_n^{(J)} \end{bmatrix}$$

$$H = [H^{(1)}, \dots, H^{(J)}]$$

とおくと、時系列の変動を J 個の成分の和として表現する状態空間モデルが得られる。それぞれの成分モデルを具体的に定めることによって、成分分解のための具体的な状態空間モデルが定まる。このモデルを用いて、時系列を J 個の成分 $x_n^{(1)}, \dots, x_n^{(J)}$ と観測ノイズ w_n に分解することができる。以下では、この方法にもとづく、季節調整法、時変分散の推定法、時変係数モデルの推定法について順次紹介する。

3.2 状態空間モデルによる季節調整

観測した時系列をトレンドと呼ばれる傾向成分と季節成分および偶然変動に成分分解する方法が季節調整法である。

トレンドの代表的モデルとしては、2次のトレンドモデル $t_n = 2t_{n-1} - t_{n-2} + v_n$ が用いられることが多い。一方、時系列の変動のうち毎年繰り返して現れる変動パターンが季節成分である。1年周期の季節成分モデルとしては、 $s_n = s_{n-12} + u_n$ が最も自然であるが、トレンドのモデルと共通の固有根を持つため、識別性がないことが知られている。そのため、このトレンドモデルから共通の固有根を除去することによって得られるモデル、すなわち1周期の合計がほぼ0になるとしたモデル

$$(3.8) \quad s_n + \dots + s_{n-p+1} = v_n, \quad v_n \sim N(0, \tau^2)$$

が利用される (北川, 2005)。この季節成分モデルの状態空間表現は

$$(3.9) \quad x_n^{(2)} = \begin{bmatrix} s_n \\ s_{n-1} \\ \vdots \\ s_{n-p+1} \end{bmatrix}, \quad F^{(2)} = \begin{bmatrix} -1 & -1 & \dots & -1 \\ 1 & & & \\ & \ddots & & \\ & & & 1 \end{bmatrix}, \quad G^{(2)} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$H^{(2)} = [1 \ 0 \ \dots \ 0]$$

で与えられる。 p は周期の長さで、月次データの 1 年周期をモデル化する場合は $p = 12$ である。

この標準的季節調整モデルは $F = \begin{bmatrix} F^{(1)} & O \\ O & F^{(2)} \end{bmatrix}$, $G = \begin{bmatrix} G^{(1)} & O \\ O & G^{(2)} \end{bmatrix}$, $H = [H^{(1)}, H^{(2)}]$ とすることにより状態空間モデルで表現できる。ただし、 O は適当な次元の 0 行列である。状態空間モデルを利用する方法の特長は、この標準的な季節調整法を簡単に拡張や改良できることであり、曜日効果項や定常 AR 成分を含むモデルに容易に拡張することができる (Kitagawa, 1981; Gersch and Kitagawa, 1983; Kitagawa and Gersch, 1984, 1996; 北川, 2005)。

3.3 時変係数 AR モデル

自己回帰モデル

$$(3.10) \quad y_n = \sum_{j=1}^m a_j y_{n-j} + w_n, \quad w_n \sim N(0, \sigma^2)$$

は定常時系列解析の標準的モデルとして利用されている (赤池・中川, 1972; 北川, 2005)。この AR モデルにおいて自己回帰係数 a_j を時間とともに変化する係数 $a_{n,j}$ で置き換えたものが時変係数 AR モデルである。ここで係数変化のモデルとしてランダムウォークモデル

$$(3.11) \quad a_{n,j} = a_{n-1,j} + v_{n,j}, \quad v_{n,j} \sim N(0, \tau^2)$$

を導入し、 m 個の時変 AR 係数を縦に並べて状態ベクトルを $x_n = [a_{n,1}, \dots, a_{n,m}]^T$ と定義すると、自己回帰モデルと係数の変化を表すランダムウォークモデルは状態空間モデルによって表現することができる。ただし、 $v_n \sim N(0, \tau^2 I_m)$, (I_m は $m \times m$ の単位行列)

$$(3.12) \quad F = G = I_m, \quad H_n = (y_{n-1}, \dots, y_{n-m})$$

である。したがって、時間変化する AR 係数をデータから推定することができる。また、推定された時変係数 $a_{j,n}$ から時間とともに変化するスペクトルを推定することができる (Kitagawa, 1983)。

3.4 時変分散モデル

時系列 y_n は正規分布に従うが、その分散は時々刻々と変化する、 $y_n = \sigma_n \varepsilon_n$, $\varepsilon_n \sim N(0, 1)$ となる場合を考える。このとき

$$(3.13) \quad \log y_n^2 = \log \sigma_n^2 + \log \varepsilon_n^2$$

となるので、時変分散とノイズが分離される。ここで分散 σ_n^2 の時間変化に、例えば、ランダムウォークモデル

$$(3.14) \quad \log \sigma_n^2 = \log \sigma_{n-1}^2 + v_n$$

を仮定すると、この変換したデータのトレンドから時変分散を推定することができる。ただし、観測ノイズ $\log \varepsilon_n^2$ は二重指数分布に従うが、カルマンフィルタによる推定のためにはこれを正規分布で近似する必要がある (北川, 2005)。この方法は、地震波の分散の推定や、金融時系列の確率的ボラティリティの推定に用いられる。

4. 非線形・非ガウス型状態空間モデリング

状態空間モデルは様々なモデルを統一的に取り扱うことができ、しかも効率的な計算法が利用できる便利なモデリング・プラットフォームであったが、1980 年代の後半になると線形・ガ

ウス型モデルでは対応できない問題が多くなり、一般化した非線形・非ガウス型状態空間モデル

$$(4.1) \quad x_n = F_n(x_{n-1}, v_n) \quad (\text{システムモデル})$$

$$(4.2) \quad y_n = H_n(x_n) + w_n \quad (\text{観測モデル})$$

が利用されるようになった。ここで、 v_n と w_n はそれぞれ、確率密度関数 $q(v)$, $r(w)$ に従う d_v 次元と d_w 次元の白色雑音、 F_n は d_k 次元、 H_n は d_m 次元の非線形関数である。

ただし、モデルが非線形あるいはノイズの分布が非ガウス型の場合には、状態の分布は非ガウス型の分布となるので、直接カルマンフィルタを適用することができず、フィルタリングのためにはこれらの分布を何らかの形で近似することが必要である。非ガウス型の分布の近似法の違いによって様々なフィルタのアルゴリズムが得られている。

(1) 正規近似による方法：非ガウス型の分布をひとつの正規分布で近似する方法で、非線形時系列モデルに適用できる拡張カルマンフィルタ (Anderson and Moore, 1979; 片山, 2011) が知られている。ただし、真の分布が2峰以上の場合には、よい近似は得られない。

(2) 混合正規近似による方法：非ガウス型の分布を $p(x_n) = \sum_{i=1}^m \alpha_{in} \varphi_i(x_n)$ と複数の正規分布 φ_i の加重和で近似する方法である。この場合、モデルが線形であれば予測分布およびフィルタ分布もガウス分布の和で表現でき、しかもそれらのパラメータを m 個のカルマンフィルタで計算できるという大きな利点がある (Kitagawa, 1994; Kitagawa and Gersch, 1996)。ただし、実際には、正規分布の項数が時間の進行とともに爆発的に増大するので、この方法の実装のためには各時間ステップで、項数の削減を行うことが不可欠である。

(3) 数値的近似による方法：非ガウス型分布を多数の分点を持つ階段関数で近似するもので、分点数を増やせば任意の分布を高い精度で近似できる。非ガウス型フィルタ (Kitagawa, 1987; 北川, 2005) では、予測分布、フィルタ分布および平滑化分布を階段関数近似などを用いて数値的に表現し、状態推定の公式を数値積分などの数値計算によって実現している。計算機の高速度化によって、この方法では、低次元の状態空間モデルに対しては、極めて精緻な結果を得ることができる。ただし、数値積分が必要なので4次元以上の連続状態の状態空間モデルに対しては適用が困難である。

(4) 粒子近似による方法：粒子近似では、他の密度関数を近似する方法とは異なり、独立に生成されたとみなせる多数の粒子を用いて状態の分布を表現する。この近似に基づくフィルタとして粒子フィルタ (Gordon et al., 1993; Kitagawa, 1993, 1996) が提案され、複雑なモデルの場合でも実装が容易なために様々な問題に適用されている (Doucet et al., 2001)。ただし、必然的にサンプリング誤差を伴うので、少数の粒子を用いる場合には注意が必要である。粒子数の変化に伴う、粒子フィルタの精度の違いに関しては Kitagawa (2014) に詳しい。

4.1 非ガウス型分布モデル

状態空間モデルのノイズ項を適切な非ガウス型分布に変更することによって、正規分布モデルとは異なる面白い結果が得られる。

4.1.1 構造変化の検出

システムノイズをコーシー分布や、ピアソン分布族などの裾の重い分布に変更することによって、トレンドや信号のジャンプが自動的に検出できる。混合ガウス分布や正規分布と一様分布の和なども利用される (Kitagawa, 1987, 1996; Kitagawa and Gersch, 1996)。

4.1.2 異常値の検出

観測ノイズの分布に、コーシー分布や混合正規分布などを用いると異常値の影響を軽減でき、ロバスト化が自動的に実現できる (Kitagawa and Matsumoto, 1996). Kitagawa (1994) にはガウス和フィルタによって、季節調整などの高次元状態空間モデルに対しても混合正規分布を用いた異常値処理が可能であることが示されている.

4.1.3 非対称分布の利用：変動分散の推定

3.4 節で示した分散変動の推定では、二重指数分布 $r(w) = \exp(x - e^w)$ をこれと同じ平均、分散をもつ正規分布で近似していたが、非ガウス型フィルタあるいは粒子フィルタを用いると、二重指数分布をそのまま用いて、より正確に変動分散の推定値を求めることができる (北川, 2005).

さらに分散 σ_n^2 の時間変化が

$$(4.3) \quad \log \sigma_n^2 = \log \sigma_{n-1}^2 + v_n$$

と表されるとき、状態を $x_n = \log \sigma_n^2$ と定義すると、時系列の状態空間モデルは

$$(4.4) \quad \begin{aligned} x_n &= x_{n-1} + v_n \\ y_n &= \exp(x_n/2)w_n \end{aligned}$$

と表される. このように、時変分散モデルは一般型の非線形状態空間モデルにおいて $F(x, v) = x + v$, $H(x, w) = \exp(x/2)w$ とおけばよい.

4.2 非線形平滑化

システムモデルが非線形の場合、従来は拡張カルマンフィルタによって状態を推定することが多かった. しかしながら、真の予測分布が多峰になることがある場合には、拡張カルマンフィルタは全く見当違いの結果を与える場合があることが知られている. しかしこのような場合でも、非線形モデルをそのまま利用して、非ガウス型フィルタあるいは粒子フィルタにより平滑化分布を求めると、真の状態の変化をよく再現できることが知られている (Kitagawa, 1991; 北川, 2005).

4.3 自己組織型の状態空間モデル：状態とパラメータの同時推定

状態空間モデル (1.1), (1.2) が未知パラメータ θ を含む場合、拡大した状態を $z_n = [x_n^T, \theta^T]^T$ と定義し、 θ に対する時間変化のモデルを例えば $\theta_n = \theta_{n-1} + u_n$ と仮定すると、状態 z_n に対する状態空間モデルが得られる. このモデルを用いた非線形フィルタ・平滑化によって、状態とパラメータの同時推定が実現できる (Kitagawa, 1998). 上記のモデルのノイズ項 u_n の分散を 0 とすると固定パラメータの推定、分散に正值を与えると時間的に変化するパラメータが推定できる.

4.4 離散過程

非ガウス型フィルタの導出過程を検討すると、状態 x_{n+1} および時系列 y_n の条件付き分布が状態 x_n だけで定まることが本質的であることがわかる. したがって、非線形・非ガウス型状態空間モデルを更に一般化した一般化状態空間モデル

$$(4.5) \quad x_n \sim q_n(\cdot | x_{n-1}) \quad (\text{システムモデル})$$

$$(4.6) \quad y_n \sim r_n(\cdot | x_n) \quad (\text{観測モデル})$$

に対しても全く同様にフィルタを導出することができる。

この拡張したモデルを用いると、二項系列やポアソン系列などに従う離散系列も通常の時系列と全く同じように取り扱うことができる。

4.5 データ同化

データ同化は観測データとの統合によりシミュレーションモデルを現実に合わせて方法であり、海洋学、気象学に止まらず多くの分野で使われるようになってきている。逐次型のデータ同化は、状態空間モデル観点からは極めて自然な方法であり、アンサンブル・カルマンフィルタあるいは粒子フィルタが利用されている(樋口 他, 2011)。

5. まとめ

時系列解析は1960年代まではスペクトル解析などの周波数領域解析が中心であったが、1970年代になるとフィードバックを持つ多変量システムの解析や制御のためには、時間順序を明示的に表現した時系列モデルの利用が必要であることが認識されるようになった。当初は、ARモデルやARMAモデルなどの定常時系列モデルの同定や推定されたモデルを用いた分析や予測・制御の方法が個別に研究されていたが、状態空間モデルの導入により殆どの線形・定常時系列モデルが統一かつ効率的に取り扱うことができるようになった。このような、モデリングの過程において情報量規準AICが指導原理として大きな役割を果たしたことは言うまでもない。

1980年代に入ると、状態空間モデルの自由にモデルを合成できるというメリットを活かして様々なタイプの非定常時系列モデルが開発されたが、合成モデルは一種のベイズモデルであり、非定常時系列解析法の発展はベイズモデリングの発展と軌を一にしていたといえる。更に1990年代になると、非線形性や非ガウス性を持った時系列の分析が重要になり、非線形・非ガウス型状態空間モデルが利用されるようになった。ただし、この場合には、従来のカルマンフィルタに替わって状態空間モデルが非線形の場合やノイズ分布が非ガウス型の場合でも適用できる新しいタイプの逐次フィルタおよび平滑化の方法が必要となった。その後、粒子フィルタを始めとする統計計算法の開発によって、複雑なモデルを駆使した時系列解析が可能になり、状態空間モデルは広範な応用分野で利用されるようになってきている。

ビッグデータが出現した現在では、データ同化などにおいて巨大な次元の状態空間モデルが必要になっている。今後は、そのような場合にも適用可能な計算法の開発だけでなく、データサイエンスの課題に即した新しいモデル形式自体の開発も期待される。

参 考 文 献

- Akaike, H. (1970). Autoregressive model fitting for control, *Annals of the Institute of Statistical Mathematics*, **23**(2), 163-180.
- Akaike, H. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes, *Annals of the Institute of Statistical Mathematics*, **26**(3), 363-387.
- Akaike, H. (1978). Covariance matrix computation of the state variable of stationary Gaussian process, *Annals of the Institute of Statistical Mathematics*, **30**, 499-504.
- Akaike, H. (1980a). Likelihood and the Bayes procedure, *Bayesian Statistics* (eds. J. M. Bernardo et al.), 1-13, University Press, Valencia.

- Akaike, H. (1980b). Seasonal adjustment by a Bayesian modeling, *Journal of Time Series Analysis*, **1**(1), 1-13.
- 赤池弘次, 北川源四郎 編 (1994). 『時系列解析の実際 I』, 朝倉書店, 東京.
- 赤池弘次, 北川源四郎 編 (1995). 『時系列解析の実際 II』, 朝倉書店, 東京.
- 赤池弘次, 中川東一郎 (1972). 『ダイナミックシステムの統計的解析と制御』, サイエンス社, 東京.
- Akaike, H., Arahata, E. and Ozaki, T. (1975). TIMSAC-74, A time series analysis and control program package-(1), *Computer Science Monographs*, No.5, The Institute of Statistical Mathematics, Tokyo.
- Akaike, H., Kitagawa, G., Arahata, E. and Tada, F. (1979). TIMSAC-78, *Computer Science Monographs*, No.11, The Institute of Statistical Mathematics, Tokyo.
- Anderson, B. D. and Moore, J. B. (1979). *Optimal Filtering*, Prentice-Hall, Englewood Cliffs.
- Box, G. E. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*, John Wiley and Sons, New Jersey.
- Burg, J. P. (1967). Maximum entropy spectral analysis, *Proceedings of 37th Annual International Meeting*, Society of Exploration Geophysics, Oklahoma City, Oklahoma, Oct. 31, 1967.
- Doucet, A., De Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*, Springer, New York.
- Gersch, W. and Kitagawa, G. (1983). The prediction of time series with trends and seasonalities, *Journal of Business and Economic Statistics*, **1**(3), 253-264.
- Gordon, N. J., Salmond, D. J. and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings F (Radar and Signal Processing)*, **140**(2), 107-113.
- 樋口知之, 上野玄太, 中野慎也, 中村和幸, 吉田 亮 (2011). 『データ同化入門=次世代のシミュレーション技術』, 朝倉書店, 東京.
- Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics*, **22**(3), 389-395.
- 片山 徹 (2011). 『非線形カルマンフィルタ』, 朝倉書店, 東京.
- Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive filter, *Journal of Time Series Analysis*, **2**(2), 103-116.
- Kitagawa, G. (1983). Changing spectrum estimation, *Journal of Sound and Vibration*, **89**(3), 433-445.
- Kitagawa, G. (1984). State space modeling of nonstationary time series and smoothing of unequally spaced data, *Time Series Analysis of Irregularly Observed Data*, 189-210, Springer, New York.
- Kitagawa, G. (1987). Non-Gaussian state space modeling of nonstationary time series, *Journal of the American Statistical Association*, **82**(400), 1032-1063.
- Kitagawa, G. (1991). A nonlinear smoothing method for time series analysis, *Statistica Sinica*, **1**(2), 371-388.
- Kitagawa, G. (1993). Monte Carlo filtering and smoothing method for non-Gaussian nonlinear state space model, Research Memo., No.462, The Institute of Statistical Mathematics, Tokyo.
- Kitagawa, G. (1994). The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother, *Annals of the Institute of Statistical Mathematics*, **46**(4), 605-623.
- Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics*, **5**(1), 1-25.
- Kitagawa, G. (1998). A self-organizing state-space model, *Journal of the American Statistical Association*, **93**(443), 1203-1215.
- 北川源四郎 (2005). 『時系列解析入門』, 岩波書店, 東京.
- Kitagawa, G. (2014). Computational aspects of sequential Monte Carlo filter and smoother, *Annals of the Institute of Statistical Mathematics*, **66**(3), 443-471.
- Kitagawa, G. and Akaike, H. (1978). A procedure for the modeling of non-stationary time series, *Annals*

- of the Institute of Statistical Mathematics*, **30**(2), 351-363.
- Kitagawa, G. and Akaike, H. (1981). *On TIMSAC-78, Applied Time Series Analysis II* (ed. D. Findley), 499-547, Academic Press, Massachusetts.
- Kitagawa, G. and Gersch, W (1984). A smoothness prior state space modeling of time series with trend and seasonality, *Journal of American Statistical Association*, **79**(386), 378-389.
- Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statistics, No.116, Springer, New York.
- Kitagawa, G. and Matsumoto, N. (1996). Detection of coseismic changes of underground water level, *Journal of the American Statistical Association*, **91**(434), 521-528.
- Morf, M., Sidhu, G.S. and Kailath, T. (1974). Some new algorithms for recursive estimation in constant, linear, discrete-time systems, *IEEE Transactions on Automatic Control*, **AC-19**, 315-323.
- Ohtsu, K., Peng, H. and Kitagawa, G. (2015). *Time Series Modeling for Analysis and Control: Advanced Autopilot and Monitoring Systems*, Springer, Tokyo.

Use of a State Space Model in Time Series Analysis

Genshiro Kitagawa

Mathematics and Informatics Center, University of Tokyo

In time series analysis, the state space model was initially used as means for realizing the maximum likelihood estimation of the ARMA model and for obtaining an optimum gain in statistical control. Since 1980, it has played a role as a platform for handling various kinds of time series modeling, such as nonstationary time series modeling, nonlinear modeling, signal extraction, missing-value processing, self-organizing state-space modeling and data assimilation. In parallel to the development of the use of the state-space model, various algorithms extending the conventional Kalman filter, such as Gaussian-sum filter, non-Gaussian filter, particle filter, etc. have been developed for state estimation of the generalized state-space model. In this article, we focus on the research of the Institute of Statistical Mathematics, and outline the development of state-space models and related calculational methods and their applications.

情報量規準 AIC の統計科学に果たしてきた役割

小西 貞則[†]

(受付 2019 年 1 月 21 日; 改訂 4 月 12 日; 採択 4 月 25 日)

要 旨

情報量規準 AIC は、導出の基本概念を尤度原理と Kullback-Leibler 情報量に置き、これを予測という視点から理論を展開したことが本質的であった。モデリングの過程におけるモデルの評価と選択は、多様なモデルとその推定法が提唱される度に問題が提起され、AIC の基本的考え方を理論的・実際の側面から研究することによって、新たなモデル評価基準の提唱へと繋がっていった。本論文では、AIC の果たしてきた役割を概観し、一般に情報量規準と呼ばれるモデル評価基準がどのように提唱されてきたかを述べる。また、ベイズアプローチに基づく予測分布モデル等の評価を目的として提唱された、AIC に基礎を置く情報量規準についてもふれる。

キーワード：AIC, ABIC, BIC, DIC, GIC, PIC, TIC, WAIC.

1. はじめに

データの背後にある現象の解明と予測・制御、そして新たな知識発見のための基礎的な役割を担う現象のモデリングに、本質的な役割を果たしてきたのが情報量規準 AIC (Akaike Information Criterion) である。Akaike (1973, 1974) の提唱した AIC は、最尤法によって推定したモデルを確率分布で表現し、その良さを Kullback-Leibler 情報量 (Kullback and Leibler, 1951) によって予測の視点から評価したことで、極めて適用範囲の広い柔軟な手法となり、諸分野の現象解明に大きく寄与してきた。赤池・北川 編 (1994, 1995) や Bozdogan ed. (1994), Parzen, Tanabe and Kitagawa (1998) には、自然科学はもとより社会科学の様々な分野で AIC が情報抽出や予測・制御にどのように寄与したかを紹介している。また、赤池弘次博士の第 22 回京都賞受賞を記念して 2007 年に出版された「赤池情報量規準 AIC」(室田・土谷 編, 2007) には、自らの言葉で情報量規準 AIC 導入に至る経緯とその効果について述べている。

情報量規準 AIC は、候補として挙げたモデル集合の中で、近似モデルの良さを相対比較することを目的とし、導出の基本概念を尤度原理と Kullback-Leibler 情報量に置き、これを予測という視点から理論を展開したことが本質的であった。これは、統計科学の尤度原理と情報科学の情報理論を融合することによって、モデルの評価と選択に新たな方向性を提起したといえる。

蓄積されたデータに内包される有用な情報を抽出、活用するため、これまでに様々なモデルとモデルの推定法が提唱されてきた。モデルの推定法という観点からみると、確率分布で表現されたモデルを、最尤法、正則化法、 L_1 ノルム型正則化法、ベイズアプローチなど、それぞれの手法の特徴を考慮して推定する。さらに、モデリングの過程において重要な役割を果たすの

[†] 中央大学 理工学部: 〒112-8551 東京都文京区春日 1-13-27 (現 九州大学大学院 数理学研究院: 〒819-0395 福岡市西区元岡 744)

が、推定したモデルの評価と選択である。この問題に多くの研究者が取り組み、設定したモデルとその推定法に対応して AIC の基本理念を理論的・実際の側面から研究し、新たなモデル評価基準の提唱へと繋がって行った。

本稿では、AIC 導出の理論をもう一度振り返ってみることから始め、一連のモデリングのプロセスの中で、AIC の果たしてきた役割を概観し、一般に情報量規準と呼ばれるモデル評価基準がどのように提唱されてきたかを述べる。2 節で AIC 導出の過程を整理し、情報量規準と呼ばれるモデル評価基準を定式化する。3 節で、多種多様なモデルと推定法に対応して、AIC 導出の基本理念を展開して新たに提唱されたモデル評価基準について述べる。4 節では、ベイズアプローチによって構築されたモデルの評価を目的として、AIC の基本的な考え方に基づいて導出されたいくつかのモデル評価基準について述べる。5 節では、Akaike (1980b) の提唱した ABIC (Akaike Information Criterion) を紹介すると共に、AIC としばしば比較の対象として取り上げられる BIC (Schwarz, 1978) との相違点等についてふれる。6 節では、モデル選択の不確定性とそれに対処する一つの方法である Akaike ウェイト (Akaike, 1978b, 1979; Burnham and Anderson, 2002) について述べる。

2. 情報量規準

現象解明のためのモデリングは、当該分野の知識とデータをもとにモデル集合を想定し、この中から現象発生の確率的メカニズムを最もよく近似するモデルを評価し選択する。本節では、このモデルの評価・選択という問題に対して、情報量規準がどのように定式化されてきたかを、Akaike (1973, 1974) の基本的な考え方を踏襲して整理する。

2.1 AIC 導出の基本的考え方

いま、データ $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ は、未知の密度関数 $g(\mathbf{y})$ (確率分布関数 $G(\mathbf{y})$) に従って生成されたとする。データを発生した $g(\mathbf{y})$ は、真の分布、あるいは真のモデルと考える。観測された有限個のデータ \mathbf{y} に内在する情報を抽出するために、確率分布によって表現されたモデル集合 $\{f(\mathbf{y}|\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$ を想定し、モデルに含まれる p 次元パラメータベクトル $\boldsymbol{\theta}$ を、推定量 $\hat{\boldsymbol{\theta}}$ で置き換えた $f(\mathbf{y}|\hat{\boldsymbol{\theta}})$ で真のモデル $g(\mathbf{y})$ を近似する。推定したモデル $f(\mathbf{y}|\hat{\boldsymbol{\theta}})$ は、データを発生した真のモデル $g(\mathbf{y})$ との近さを測ることによってその良さを評価する。Akaike (1973, 1974) は、分布間の距離を測る基準として Kullback-Leibler 情報量 (K-L 情報量) を採用し、モデルの評価を予測の視点から捉えることによって AIC 導出に繋がった。これは、以下のように述べることができる。

推定したモデル $f(\mathbf{y}|\hat{\boldsymbol{\theta}})$ とデータを発生した真のモデル $g(\mathbf{y})$ との距離は、予測の視点を入れて K-L 情報量で測るとき、次の式で与えられる。

$$(2.1) \quad I\{g(z), f(z|\hat{\boldsymbol{\theta}})\} = E_G \left[\log \frac{g(Z)}{f(Z|\hat{\boldsymbol{\theta}})} \right] = E_G[\log g(Z)] - E_G[\log f(Z|\hat{\boldsymbol{\theta}})].$$

ここで、期待値は $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{y})$ を固定して真の分布 G に関してとる。予測の視点とは、データ \mathbf{y} とは独立に、真の分布からランダムに採られた将来のデータ $Z = z$ の従う分布 $g(z)$ を、モデル $f(z|\hat{\boldsymbol{\theta}})$ で近似したときの平均的な良さを測ることを意味する。

(2.1) 式の K-L 情報量の右辺第一項 $E_G[\log g(Z)]$ は、個々のモデルに依存せず一定であることから、第 2 項の期待対数尤度と呼ばれる $E_G[\log f(Z|\hat{\boldsymbol{\theta}})]$ の値が大きいモデルほど真のモデルに近いといえる。期待対数尤度は、真のモデルに依存する未知の量である。そこで、 n 個のデータを発生した未知の真の分布 G を、各データ点 y_i に確率 $1/n$ を付与した経験分布関数 \hat{G} で推定する。これは、離散型確率分布の期待値をとることから

$$E_G[\log f(Z|\hat{\theta})] = \frac{1}{n} \sum_{i=1}^n \log f(y_i|\hat{\theta}) = \frac{1}{n} \log f(\mathbf{y}|\hat{\theta})$$

となり、モデル $f(\mathbf{y}|\hat{\theta})$ の対数尤度 $\ell_n(\hat{\theta}) = \log f(\mathbf{y}|\hat{\theta})$ が求まる。したがって、モデルの対数尤度は期待対数尤度 ($\times n$) の一つの推定量である。

しかし、対数尤度は期待対数尤度 ($\times n$) の推定量ではあるが、データ \mathbf{y} とは独立に真のモデル g から発生した将来のデータ \mathbf{z} に基づく対数尤度 $f(\mathbf{z}|\hat{\theta}(\mathbf{y}))$ ではなく、モデルの推定に用いたデータ \mathbf{y} を再び利用した $f(\mathbf{y}|\hat{\theta}(\mathbf{y}))$ で推定していることから、推定のバイアス

$$(2.2) \quad \log f(\mathbf{y}|\hat{\theta}(\mathbf{y})) - nE_G[\log f(Z|\hat{\theta}(\mathbf{y}))]$$

を生じる原因となっている。これは、一般に $\log f(\mathbf{z}|\hat{\theta}(\mathbf{y})) < \log f(\mathbf{y}|\hat{\theta}(\mathbf{y}))$ となることから分かる。(2.2)式は、ある特定のデータ \mathbf{y} に対するバイアスであるが、大きさ n のデータを g から繰り返し抽出したときの平均的なバイアスは

$$(2.3) \quad b(G) = E_{G(\mathbf{y})}[\log f(\mathbf{Y}|\hat{\theta}(\mathbf{Y})) - nE_{G(\mathbf{z})}[\log f(Z|\hat{\theta}(\mathbf{Y}))]]$$

で与えられる。ここで、期待値は \mathbf{Y} の同時分布 $\prod_{i=1}^n g(y_i)$ に関してとる。したがって、このバイアスを何らかの方法で求めて、もし、バイアスがデータを生成した真の確率分布 G に依存していれば、 $b(G)$ の一致推定量 $\hat{b}(G)$ で対数尤度のバイアスを補正した $\ell_n(\hat{\theta}) - \hat{b}(G)$ が期待対数尤度 ($\times n$) の推定量として求まる。一般に、 -2 を掛けた

$$(2.4) \quad \text{IC} = -2 \log f(\mathbf{y}|\hat{\theta}) + 2\hat{b}(G)$$

を、K-L 情報量の推定量として導かれたモデル評価基準であることから情報量規準という。IC 値が小さいモデルほど K-L 情報量の値も小さく、真のモデルに近いといえる。

情報量規準 AIC は、最尤法によって推定したモデル $f(\mathbf{y}|\hat{\theta}_{ML})$ を評価するための基準で、期待対数尤度 ($\times (-2n)$) の近似推定量として導かれ、次の式で与えられた。

$$(2.5) \quad \text{AIC} = -2 \log f(\mathbf{y}|\hat{\theta}_{ML}) + 2 \quad (\text{モデルの自由パラメータ数})$$

ただし、 $\hat{\theta}_{ML}$ は θ の最尤推定量とし、 $\log f(\mathbf{y}|\hat{\theta}_{ML})$ は n 次元データベクトル \mathbf{y} に基づくモデルの最大対数尤度である。最大対数尤度で期待対数尤度を推定したとき、平均的にどの程度過大に推定しているかを表す(2.3)式のバイアスが、結果としてモデルの自由パラメータ数と一致することを示している。AIC の値を最小とするモデルを選択する方法は、AIC 最小化法と呼ばれる。

多数のパラメータで特徴付けられたモデルほど、観測したデータへのモデルの当てはまりはよい。しかし、複雑すぎるとモデルは将来の現象予測に有効に働かない。AIC は予測の観点から最適なモデルを選択するための評価基準で、モデルのデータへの適合度を最大対数尤度 $\log f(\mathbf{y}|\hat{\theta}_{ML})$ で捉え、モデルの自由パラメータ数をモデルの複雑さに対するペナルティとして組み込んでいるといえる。

2.2 情報量規準の定式化

(2.3)式のバイアス補正項 $b(G)$ は、モデルを最尤法で推定するか、あるいは正則化法などで推定するかによって、また真のモデルと想定したモデルの関係をどう捉えるかによって異なる形をとる。いま、最尤法で推定したモデルを $f(\mathbf{y}|\hat{\theta}_{ML})$ とする。このとき、(2.3)式のバイアス $b(G)$ は、最尤推定量の漸近的性質(例えば、小西・北川, 2004, p.42)を用いると、データ数 n に対して漸近的に $b(G) = \text{tr}\{J^{-1}(G)I(G)\}$ となる。ただし、 $J(G)$ 、 $I(G)$ は次式で定義される

$p \times p$ 行列とし, 式中 $\partial/\partial\boldsymbol{\theta} = (\partial/\partial\theta_1, \dots, \partial/\partial\theta_p)^T$ は転置ベクトルを表す.

$$(2.6) \quad J(G) = -E_G \left[\frac{\partial^2 \log f(Z|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} \right], \quad I(G) = E_G \left[\frac{\partial \log f(Z|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \frac{\partial \log f(Z|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T} \right].$$

いま, $\hat{J}(G), \hat{I}(G)$ をそれぞれ $J(G), I(G)$ の一致推定量とすると

$$(2.7) \quad \text{TIC} = -2 \sum_{i=1}^n \log f(y_i|\hat{\boldsymbol{\theta}}_{ML}) + 2\text{tr}\{\hat{J}^{-1}(G)\hat{I}(G)\}$$

が求まる. これは, 竹内 (1976) によって与えられ, 情報量規準 TIC と呼ばれている.

ここで, 想定したパラメトリックモデル $\{f(y|\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\}$ の中に真のモデル $g(y)$ が含まれる, すなわち, ある $\boldsymbol{\theta}_0 \in \Theta$ に対して $g(y) = f(y|\boldsymbol{\theta}_0) (F(y|\boldsymbol{\theta}_0))$ となるものが存在すると仮定する. このとき, (2.6) 式の期待値を $G = F$ でとると $J(F) = I(F)$ が成立し $\text{tr}\{J^{-1}(F)I(F)\} = p$ となり, 情報量規準 AIC が導かれる. $I(F)$ はフィッシャー情報行列である. AIC は, TIC の漸近バイアスをモデルの自由パラメータ数で近似した評価規準であるといえる. 導出の詳細は, 小西・北川 (2004, 3 章), Konishi and Kitagawa (2008, Chapter 3) を参照されたい.

Akaike (1974) の論文では, 想定したモデル集合の中に真のモデルは含まれていないという仮定のもとで AIC 最小化法を議論し, データを発生した真のモデルの近傍に適切にパラメトリックモデルを想定すれば, 最尤法に基づくモデルの対数尤度のバイアスはモデルの自由パラメータ数で近似できると述べている. これによって, 情報量規準 AIC は, 個々のモデルに対して漸近バイアス $\text{tr}\{\hat{J}^{-1}(G)\hat{I}(G)\}$ を解析的に導出する必要がなくなり, また, パラメータ数 p は当然未知の確率分布 G にも依存しないことから, バイアスの推定による変動も取り除かれ, 適用上極めて柔軟な手法となったといえる.

情報量規準の構成においては, モデル $f(z|\hat{\boldsymbol{\theta}})$ の期待対数尤度 $E_G[\log f(Z|\hat{\boldsymbol{\theta}})]$ を予測の観点から推定することが本質的であった. これは, 観測データ \mathbf{y} に基づいて構築したモデルを, 真のモデルからランダムに抽出した将来のデータ z でモデルを評価するという考え方を定式化したことで実現した. 同様に, 予測の観点から種々の予測誤差を捉えることを可能とした極めて汎用性の高い手法が, Stone (1974) によるクロス・バリデーション (Cross-Validation; 交差検証法) である.

クロス・バリデーションは, 観測データ \mathbf{y} のみに基づいて予測の観点からモデルを評価する方法で, モデルの推定に用いるデータとモデルの評価に用いるデータを分離して行う. クロス・バリデーションによると期待対数尤度は,

$$(2.8) \quad \sum_{i=1}^n \log f(y_i|\hat{\boldsymbol{\theta}}_{ML}^{(-i)})$$

と推定される. ただし, $\hat{\boldsymbol{\theta}}_{ML}^{(-i)}$ は, n 個の観測データの中から i 番目のデータ y_i を取り除いた残りの $(n-1)$ 個のデータに基づく最尤推定値とする. Stone (1977) は, クロス・バリデーションによるモデル評価基準と AIC は漸近的に同等であることを示した. その証明は, 候補モデル集合の中に真のモデルは含まれていないという仮定のもとで行っており, 導出の過程で (2.7) 式の TIC を与えている.

一般に, K-L 情報量に基づいて予測の視点から導かれる AIC タイプの情報量規準は, 期待対数尤度の推定量を求める問題に帰着される. 同様に, 期待対数尤度を予測の視点からクロス・バリデーションによって推定したのが (2.8) 式であった. このことから, Konishi and Kitagawa (2008, p.245) は, 次節で述べる汎関数理論を用いることによって, AIC タイプの情報量規準はクロス・バリデーションと漸近的に同等であることを示した.

Sugiura (1978)は、ガウス型線形回帰モデルに対して、真のモデルが想定したモデルに含まれる場合に、(2.3)式の期待対数尤度のバイアスを精密に求め、修正情報量規準 AIC_c と呼ばれる評価基準を提唱した。Hurvich and Tsai (1989), Fujikoshi and Satoh (1997), 藤越・杉山 (2012), McQuarrie and Tsai (1998)等は、時系列モデル、多変量回帰モデルに対して正規性の仮定のもとでバイアス補正項を求めて、AIC を修正した形の情報量規準を提案している。修正情報量規準は、正規性等の条件下で求められているが、パラメータ数 p に比べてデータ数 n がそれほど多くないときは、実際上有効であることが数値的に検証されている(例えば、Burnham and Anderson, 2002 を参照されたい)。

3. 最尤法の枠組みを外した情報量規準

では、ロバスト推定、正則化最尤法など、最尤法を含むより広いクラスの推定法によって構築されたモデルの評価を可能とする情報量規準は、どのように構成すればよいであろうか。本節では、K-L 情報量の推定量として導かれたいくつかの情報量規準について述べる。

3.1 一般化情報量規準 GIC

最尤法を含むより広いクラスの推定法で構築したモデルの評価を、統計的汎関数に基づくアプローチによって可能にしたのが、一般化情報量規準 GIC (generalized information criterion: Konishi and Kitagawa, 1996)である。

データを発生した真のモデル $g(y)$ は、候補モデル集合 $\{f(y|\theta); \theta \in \Theta \subset R^p\}$ に含まれないとする。このとき、モデルのパラメータは $g(y)(G(y))$ に従って発生したデータによって推定される。そこで、一般にパラメータ θ_i の推定量 $\hat{\theta}_i$ は、確率分布 G の実数値関数、すなわちある統計的汎関数 $T_i(G)$ が存在して、 n 個のデータそれぞれに等確率 $1/n$ をもつ経験分布関数 \hat{G} に対して $\hat{\theta}_i = T_i(\hat{G}) (i = 1, 2, \dots, p)$ で与えられるとする。この $T_i(G)$ を第 i 要素とする p 次元汎関数ベクトルを $\mathbf{T}(G) = (T_1(G), \dots, T_p(G))^T$ とすると、 p 次元推定量は $\hat{\theta} = \mathbf{T}(\hat{G})$ で与えられる。例えば、標本平均 $\bar{y}_n = n^{-1} \sum_{i=1}^n y_i$ を定義する汎関数は $T_\mu(G) = \int y dG(y)$ であり、この汎関数 T_μ によって $\bar{y}_n = T_\mu(\hat{G}) = \int y d\hat{G}(y)$ で与えられることが分かる。標本数 n を無限大とすると、経験分布関数 \hat{G} は真の分布 G に法則収束することから、 $\hat{\theta} = \mathbf{T}(\hat{G})$ は $\theta = \mathbf{T}(G)$ に対して一致性をもつ推定量である。

一般化情報量規準 GIC は、 $f(y|\theta)$ のパラメータを汎関数で定義される推定量 $\hat{\theta} = \mathbf{T}(\hat{G})$ で置き換えたモデル $f(y|\hat{\theta})$ の評価基準で、次の式で与えられた。

$$(3.1) \quad \text{GIC} = -2 \sum_{i=1}^n \log f(y_i|\hat{\theta}) + \frac{2}{n} \sum_{i=1}^n \text{tr} \left\{ \mathbf{T}^{(1)}(y_i; \hat{G}) \frac{\partial \log f(y_i|\theta)}{\partial \theta^T} \Bigg|_{\theta=\hat{\theta}} \right\}.$$

ただし、 $\mathbf{T}^{(1)}(y; \hat{G})$ は、その第 i 要素 $T^{(1)}(y; \hat{G})$ が次の式で与えられる点 \hat{G} での汎関数微分で、 p 次元経験影響関数ベクトルと呼ばれる。

$$T_i^{(1)}(y; \hat{G}) = \lim_{\epsilon \rightarrow 0} \frac{T_i((1-\epsilon)\hat{G} + \epsilon\delta_y) - T_i(\hat{G})}{\epsilon}.$$

ここで、 δ_y は点 y 上に確率 1 をもつ分布とする。影響関数は、ロバスト推定において、分布のわずかな変化に対して推定値がどれだけ変化するかを調べるために用いられた (Huber, 1981; Hampel et al., 1986)。

一般化情報量規準 GIC は、最尤法をはじめとしてロバスト推定法、様々な L_2 ノルム正則化項をもつ正則化最尤法などによって推定されたモデルの評価を可能とするモデル評価基準である。これらの推定量は、一般に標本空間とパラメータ空間の直積空間上で定義された実数値関

数 $\psi_i(y, \theta)$ に対して、次の同時方程式の解 $\hat{\theta}$ として与えられる。

$$(3.2) \quad \sum_{i=1}^n \psi_j(y_i, \hat{\theta}) = 0, \quad j = 1, 2, \dots, p.$$

ここで、 $\psi = (\psi_1, \psi_2, \dots, \psi_p)^T$ とベクトル表示して、これを ψ -関数と呼ぶ。最尤推定量 $\hat{\theta}_{ML}$ 、正則化最尤推定量 $\hat{\theta}_R$ は、それぞれ

$$(3.3) \quad \psi_{ML}(y, \theta) = \frac{\partial \log f(y|\theta)}{\partial \theta}, \quad \psi_R(y, \theta) = \frac{\partial \{\log f(y|\theta) - \lambda R(\theta)\}}{\partial \theta}$$

としたときの解である。ただし、 $R(\theta)$ は正則化項、 $\lambda > 0$ は正則化パラメータと呼ばれ、モデルのデータへの適合度と当てはめたモデルの滑らかさを連続的に調整する役割を果たす。

この ψ -関数に対して、(3.1)式の GIC の影響関数は

$$(3.4) \quad \mathbf{T}^{(1)}(y, G) = J(\psi, G)^{-1} \psi(y, G)$$

で与えられる。ただし、 $J(\psi, G)$ は、次式で与えられる $p \times p$ 行列で、(2.6)式の行列 $J(G)$ に相当する。

$$J(\psi, G) = -E_G \left[\frac{\partial \psi(Z, \theta)^T}{\partial \theta} \right].$$

ここで、(3.4)式の影響関数を(3.1)式の GIC へ代入すると、(3.2)式の同時方程式の解として与えられる推定量 $\hat{\theta}$ に基づくモデル $f(y|\hat{\theta})$ の評価基準

$$(3.5) \quad \text{GIC}_R = -2 \sum_{i=1}^n \log f(y_i|\hat{\theta}) + 2\text{tr}\{J(\psi, \hat{G})^{-1} I(\psi, \hat{G})\}$$

が求まる。ただし、 $I(\psi, G)$ は

$$I(\psi, G) = E_G \left[\psi(Z, G) \frac{\partial \log f(Z|\theta)}{\partial \theta^T} \right]$$

で与えられる $p \times p$ 行列で、これは(2.6)式の行列 $I(G)$ に対応する。

情報量規準 GIC_R のバイアス補正項の推定値は、一般に実数値関数 $h(z|\theta)$ の期待値 $E_G[h(Z|\theta)]$ ($\theta = \mathbf{T}(G)$) を、経験分布関数 \hat{G} に関する期待値 $E_{\hat{G}}[h(Z|\hat{\theta})] = n^{-1} \sum_{i=1}^n h(y_i|\hat{\theta})$ ($\hat{\theta} = \mathbf{T}(\hat{G})$) で推定した結果を用いている。

特別な場合として、(3.3)式の ψ_{ML} を(3.5)式へ代入すると最尤法に基づく TIC が求まる。さらに、Fisher 一致性の概念 ($\mathbf{T}(F_\theta) = \theta; F_\theta = F(y|\theta)$) を適用することによって、 M 推定などのロバスト推定に対しても AIC のバイアス補正項であるモデルの自由パラメータ数に対応する結果が求まり、AIC は M 推定量に基づくモデルの評価基準へと自然に拡張される (小西・北川, 2004, p.77; Konishi and Kitagawa, 2008, p.131)。GIC の導出とその応用および精密化については、Konishi and Kitagawa (1996), Konishi (1999, 2002), Konishi and Kitagawa (2003), 小西・北川 (2004, 4 章), Konishi and Kitagawa (2008, Chapter 5) を、統計的汎関数については、von Mises (1947), Fernholz (1983) などを参照されたい。

確率過程に対する情報量規準は、Uchida and Yoshida (2001, 2004) によって与えられた。Lv and Liu (2014) は、モデル集合を一般化線形モデル (McCullagh and Nelder, 1989) として、候補モデル集合の中には真のモデルは含まれないという仮定のもとで、AIC タイプのモデル評価基準を求めた。結果は、(2.6)式の行列 J, I に対応するものを一般化線形モデルのもとで求めているが、GIC の特別な場合と考えられる。Shen and Ye (2002), Shen, Huang and Ye (2004)

は、それぞれガウス分布と指数型分布族に対して、期待対数尤度の近似的に不偏な推定量として導いた適応型モデル評価基準を提唱した。これらは、AIC 導出の基本概念から導かれたものであるが、汎用性という点では問題が残る。

3.2 正則化法と平滑化パラメータの選択

非線形回帰モデルの関数推定に対しては、最尤法は有効に機能しない場合が多く、このため対数尤度に曲線(曲面)の局所変動の程度を考慮に入れた正則化最尤法(罰則付き最尤法)が用いられる。その際、平滑化パラメータ(正則化パラメータ)がモデルの複雑さの程度を調整し、データへの過適合による汎化能力の低下を抑制する働きをする。本節では、非線形回帰モデリングの過程で本質的な平滑化パラメータの選択に用いられてきたモデル評価基準について述べる。

いま、目的変数 y と p 次元説明変数 x に関して観測された n 組のデータ集合に、回帰モデル $y = u(x; \beta) + \varepsilon$ を当てはめるとする。現象の平均構造を捉える回帰関数 $u(x; \beta)$ に対して、スプライン、 B -スプライン、動径関数などを仮定してモデル化する。これらのモデルを統一的に表すと、回帰関数を非線形関数 $b_j(x)$ の線形結合とした

$$(3.6) \quad y = \sum_{j=1}^m \beta_j b_j(x) + \varepsilon, \quad \varepsilon \sim F(\varepsilon)$$

で与えられ、基底展開法に基づく非線形回帰モデルと呼ばれる(例えば、Hastie, Tibshirani and Friedman, 2009, 5 章; 小西, 2010, 3 章)。

基底展開に基づく非線形回帰モデルは、対数尤度関数にペナルティ項(正則化項)を課した正則化最尤法、すなわち $\log f(y|\beta) - \lambda R_n(\beta)$ の最大化によって推定する。正則化項 $R_n(\beta)$ としては、関数の曲率を考慮した 2 階微分の積分の離散近似、パラメータ β の差分や 2 乗和等が説明変数の次元と分析目的に応じて用いられる(小西・北川, 2004, p.92)。正則化法は、Good and Gaskins (1971)によって密度推定の枠組みで提唱され、その後、縮小推定量や本稿 5 節で述べるように、ベイズモデルとの関係が明らかにされた(Akaike, 1980b; Kitagawa and Gersch, 1984, 1996; Shibata, 1989)。

正則化最尤法によって推定したモデルの複雑さの程度は、平滑化パラメータ λ に加えて基底関数の個数 m にも依存する。そのため、平滑化の程度を調整するこれらのパラメータの値を決める問題をモデル選択として捉え、AIC に基づく様々なモデル評価基準が提唱された。Hastie and Tibshirani (1990)は、AIC のバイアス補正項である自由パラメータ数を、基底関数の個数と平滑化パラメータを含む有効自由度 (effective degrees of freedom) で置き換えたモデル評価基準を提唱した。その後、ガウス型線形回帰モデルの枠組みで求められた修正情報量規準 AIC_c (Sugiura, 1978)に含まれる変数の個数を有効自由度で置き換えた評価基準も提唱された(Hurvich, Simonoff and Tsai, 1998, 等)。しかし、限られた設定のもとでの数値比較の有効性は認められるが、理論的整合性には課題が残る。

これに対して、汎関数の枠組みで導出した GIC の特別な場合として与えられた(3.5)式の GIC_R へ、正則化最尤推定量を与える(3.3)式の ψ_R を代入すると、平滑化パラメータ λ をもつ正則化最尤法に基づくモデルの評価基準が求まる。この結果を用いて、基底展開法に基づく非線形回帰モデルを正則化最尤法によって推定したときの平滑化パラメータの選択、基底関数の個数を決める評価基準を導出してモデリングに組み込んだ解析手法が提案された(Imoto and Konishi, 2003; Ando, Konishi and Imoto, 2008; Kawano and Konishi, 2011; Tateishi and Konishi, 2011; Kawano, Misumi and Konishi, 2012; Park and Konishi, 2017 等)。また、 GIC_R は、関数データ解析 (Ramsay and Silverman, 2005)において、経時的に観測・測定されたデータの関数

化にも適用された (Araki et al., 2009a, 2009b; Kayano, Dozono and Konishi, 2010; Matsui and Konishi, 2011 等).

AIC に基づくモデル評価基準は, 一般化加法モデル GAM (generalized additive model; Hastie and Tibshirani, 1990) における Wood, Pya and Säfken (2016) や混合効果モデルに対する Liang, Wu and Zou (2008), Yu and Yau (2012), Misumi and Konishi (2016) など, 様々な手法のモデリングの過程で用いられて, モデルの評価と選択に貢献してきた.

Shibata (1989) は正則化法によるモデルとその評価について議論し, Regularized Information Criterion (RIC) を提唱した. Murata, Yoshizawa and Amari (1994) は, ニューラルネットワークモデルの最適なパラメータ数, あるいは隠れ層の個数の決定を目的とした Network Information Criterion (NIC) を提唱した. さらに, 正則化項を考慮した損失関数に基づくモデルの推定と評価を議論している. これらは, それぞれのモデリングの目的に合わせて, AIC 導出の基本的な考え方を踏襲して提唱されたモデル評価基準である.

3.3 スパースモデリング

データ数に比してモデルのパラメータ数が大幅に上回る大規模モデリングでは, モデルの推定とモデルの評価を分離して行うことの限界が指摘された. 一つは, 候補となるモデルが多数に上ることによる計算量の限界, 一つはモデル選択の信頼性 (Brieman, 1996) などが挙げられる. このような状況の中で回帰モデリング, 特に, 線形回帰モデルの推定と変数選択に新たな方向性を示したのが, lasso (least absolute shrinkage and selection operator; Tibshirani, 1996) であった. これは, 損失関数に回帰係数の絶対値 (L_1 ノルム) の和を正則化項として付与した推定法で, その特徴はモデルの推定と変数選択を同時に実行できる点にあった. このため, 高次元線形回帰モデルに対する有効なモデリングとして注目を集め, 様々な L_1 型正則化線形回帰モデリング (スパースモデリング) の研究が急速に進展した (川野 他, 2010; Konishi, 2014, Section 2.3; Hastie, Tibshirani and Wainwright, 2015; 廣瀬, 2016; 川野・松井・廣瀬, 2018 等).

スパースモデリングでは, 調整パラメータ λ の値の増加に伴って, 回帰係数の推定値は 0 へと縮退する. 基本的には, 調整パラメータの値を与えたもとでモデルをスパース推定し, その結果 0 でない回帰係数の推定値に対応する説明変数の個数をモデルの自由パラメータ数として AIC や 5 節 (5.2) 式の BIC を用いて評価するプロセスを繰り返すことは可能である. この方法に対して, 様々なスパース推定法の特徴, データ数とパラメータ数との関係やモデル選択の一致性等を考慮した理論研究が進展し, 新たなモデル評価基準が提唱された.

Efron et al. (2004), Zou, Hastie and Tibshirani (2007) は, Stein のリスク不偏推定の枠組みで lasso に対してモデルの自由度を与え, AIC, BIC, Mallows (1973) の C_p に基づいた評価基準を検討した. Kato (2009) は微分幾何学的アプローチによって, より広い lasso タイプの自由度の不偏推定について議論した. モデルの自由度については, Ye (1998), Efron (2004) を併せて参照されたい. Zhang, Li and Tsai (2010), Fan and Tang (2013) は, AIC のバイアス項の 2 と対応する BIC の $\log n$ を, データ数 n に依存する正の実数列で置き換えてモデルの複雑さを制御することで, 調整パラメータの選択を議論している. Hirose, Tateishi and Konishi (2013) は, 様々なスパース回帰モデリングに対する自由度を数値的に計算するアルゴリズムを提唱し, AIC, 修正情報量規準 AIC_c , BIC, Mallows' C_p などに基づくモデル評価基準による調整パラメータの選択法を与えた. Ninomiya and Kawano (2016) と Umezu et al. (2019) は, それぞれ lasso と bridge (Frank and Friedman, 1993), SCAD (smoothly clipped absolute deviation; Fan and Li, 2001) などの非凸正則化法に対して, 一般化線形モデルの枠組みで AIC 導出の基本概念に基づいてモデル評価基準を提唱した.

BIC は, 候補モデル集合に真のモデルは含まれているとしたとき一致性をもち, しかも AIC

よりはより単純なモデルを選択する傾向にある。このような理由により、スパースモデリングの調整パラメータの選択に、BIC をもとにしたモデル評価基準が提唱されている。Wang, Li and Tsai (2007) は、SCAD の調整パラメータの選択に対して (5.2) 式の BIC の自由パラメータ数をモデルの自由度で置き換えた評価基準を提唱した。Wang, Li and Leng (2009) は、lasso, SCAD を含む L_1 正則化法に対して、BIC を基準とした調整パラメータ選択法に対して理論的整合性を議論している。

3.4 ブートストラップ情報量規準

前節までに述べた情報量規準は、データ発生の確率構造とモデル推定に関して、それぞれ異なる条件下で漸近理論に基づいて導出された。それに対して、ブートストラップ情報量規準は、個々のモデルの対数尤度のバイアスをブートストラップ法 (Efron, 1979) を適用して数値的に近似したものである (Ishiguro, Sakamoto and Kitagawa, 1997; Konishi and Kitagawa, 1996)。なお、本節ではデータ \mathbf{y} とブートストラップ標本 \mathbf{y}^* の違いをモデルの中で示すため、推定量 $\hat{\theta}$ を $\hat{\theta}(\mathbf{y})$ と表す。

情報量規準構成においては、推定したモデル $f(\mathbf{y}|\hat{\theta}(\mathbf{y}))$ の期待対数尤度 $nE_G[\log f(Z|\hat{\theta}(\mathbf{y}))]$ を対数尤度 $\log f(\mathbf{y}|\hat{\theta}(\mathbf{y}))$ で推定したときのバイアスの補正が本質的であった。ブートストラップ法の基本的な考え方は、未知の確率分布 G からの標本 $\mathbf{y} = \{y_1, \dots, y_n\}$ に基づく推測過程を、データから推定した既知の確率分布である経験分布関数 \hat{G} からの標本であるブートストラップ標本 $\mathbf{y}^* = \{y_1^*, \dots, y_n^*\}$ に置き換えて実行する点にある。このため、ブートストラップ標本 \mathbf{y}^* に基づいて推定したモデルを $f(\mathbf{y}|\hat{\theta}(\mathbf{y}^*))$ とする。

次に経験分布関数を真の分布としたときの $f(\mathbf{y}|\hat{\theta}(\mathbf{y}^*))$ の期待対数尤度は、 \hat{G} が n 個の各データに等確率 $1/n$ をもつ離散型確率分布の確率分布関数であることから

$$E_G[\log f(Z|\hat{\theta}(\mathbf{y}^*))] = \int \log f(z|\hat{\theta}(\mathbf{y}^*)) d\hat{G}(z) = \frac{1}{n} \sum_{i=1}^n \log f(y_i|\hat{\theta}(\mathbf{y}^*)) = \frac{1}{n} \log f(\mathbf{y}|\hat{\theta}(\mathbf{y}^*))$$

となる。一方、期待対数尤度の一つの推定量である対数尤度は、モデルをブートストラップ標本によって推定し、推定したモデル $f(\mathbf{y}|\hat{\theta}(\mathbf{y}^*))$ の評価を再びブートストラップ標本を用いて行うことから、 $\log f(\mathbf{y}^*|\hat{\theta}(\mathbf{y}^*))$ で与えられる。従って、ブートストラップ法によって期待対数尤度を対数尤度で推定したときのバイアスは、

$$E_G[\log f(\mathbf{y}^*|\hat{\theta}(\mathbf{y}^*)) - \log f(\mathbf{y}|\hat{\theta}(\mathbf{y}^*))]$$

と推定される。

この期待値は、 \hat{G} が既知の確率分布 (経験分布関数) であることを利用して、モンテカルロ法によって数値的に近似できるところにブートストラップ法の最大の特徴がある。すなわち、経験分布関数からの大きさ n の標本の反復抽出とは、観測データからの大きさ n の標本の復元抽出の反復と同値 (小西・越智・大森, 2008, p.9) であることを利用して

$$b(\hat{G}) \approx \frac{1}{B} \sum_{i=1}^B \{\log f(\mathbf{y}^*(i)|\hat{\theta}^*(i)) - \log f(\mathbf{y}|\hat{\theta}^*(i))\} := b_B(\hat{G})$$

と近似する。ただし、 $\mathbf{y}^*(i)$ は i 番目のブートストラップ標本、 $\hat{\theta}^*(i)$ は i 番目のブートストラップ標本に基づく推定値とする。このとき、対数尤度のバイアスを補正した情報量規準 EIC (extended information criterion) は、

$$(3.7) \quad \text{EIC} = -2 \sum_{i=1}^n \log f(y_i | \hat{\theta}) + 2b_B(\hat{G})$$

で与えられる。

ブートストラップ法は、実行プロセスの中で解析的アプローチを、観測データ自身を反復抽出(リサンプリング)するというモンテカルロ計算法で置き換えたことにより、極めて緩やかな仮定のもとで、より複雑な問題に適用できる柔軟な統計手法となった。しかし、バイアス推定の標本変動に加えて、ブートストラップリサンプリングによる変動が生じることから、バイアス項の差異でモデルの違いを見るときには十分注意を払う必要がある。このブートストラップバイアス推定の確率変動を減少させるための方法が、Konishi and Kitagawa (1996), Kitagawa and Konishi (2010)によって提案された。また、Konishi and Kitagawa (1996)は、ブートストラップバイアス推定および変動減少法の理論的整合性を汎関数の枠組みで証明した。

4. ベイズモデルの評価基準

本節では、ベイズアプローチによって構築されたモデルの評価を目的として、AIC 導出の基本的な考え方を踏襲して提唱されたいくつかのモデル評価基準について述べる。

データ \mathbf{y} を発生した真のモデル $g(y)$ に対して、想定したモデル集合を $\{f(y|\theta); \theta \in \Theta \subset R^p\}$ とし、パラメータベクトル θ の事前分布を $\pi(\theta)$ とする。このとき、データ \mathbf{y} に対する θ の事後分布は、

$$(4.1) \quad \pi(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\int f(\mathbf{y}|\theta)\pi(\theta)d\theta}$$

である。さらに、真のモデル g からランダムに抽出された将来のデータ z に対して、データ \mathbf{y} を与えたもとで、モデル $f(z|\theta)$ の事後分布に関する期待値

$$(4.2) \quad h(z|\mathbf{y}) = E_{\pi(\theta|\mathbf{y})}[f(z|\theta)] = \int f(z|\theta)\pi(\theta|\mathbf{y})d\theta$$

として与えられるのが予測分布である。

4.1 ベイズ予測分布の情報量規準

予測分布モデルの評価を K-L 情報量に基づいて行うとき、期待対数尤度 $E_G[\log h(Z|\mathbf{y})]$ の推定が本質的となる。AIC の場合と同様に未知の確率分布 G を経験分布関数 \hat{G} で置き換えると、 $E_{\hat{G}}[\log h(Z|\mathbf{y})] = n^{-1} \sum_{i=1}^n \log h(y_i|\mathbf{y}) = n^{-1} \log h(\mathbf{y}|\mathbf{y})$ が求まる。従って、予測分布モデルの期待対数尤度を対数尤度で推定したときのバイアスは

$$b_{\text{pred}}(G) = E_{G(\mathbf{y})}[\log h(\mathbf{Y}|\mathbf{Y}) - nE_{G(z)}[\log h(Z|\mathbf{Y})]]$$

で与えられ、バイアスを補正した

$$(4.3) \quad \text{IC}_{\text{pred}} = -2 \log h(\mathbf{y}|\mathbf{y}) + 2b_{\text{pred}}(\hat{G})$$

が予測分布に対する情報量規準である (Akaike, 1980a)。

Konishi and Kitagawa (1996, p.878)は、汎関数の枠組みでこのバイアスを求めて、予測分布モデルに対する情報量規準を導出した。さらに、積分のラプラス近似 (Tierney and Kadane, 1986; Davison, 1986)を用いて、最尤法によって推定したモデル $f(z|\hat{\theta}_{ML})$ に対して、予測分布を $h(z|\mathbf{y}) = f(z|\hat{\theta}_{ML}) + O_p(n^{-1})$ と近似して、TIC, AIC と同様の情報量規準が求まることを示した。Kitagawa (1997)は、モデルと事前分布がともに多変量正規分布と仮定した線形ガウス

型ベイズモデルの情報量規準を導出し、これを PIC (predictive information criterion) と呼んだ。

4.2 逸脱度情報量規準 DIC

Spiegelhalter et al. (2002) は、ベイズの観点から AIC と同タイプのモデル評価基準を提唱し、これを DIC (deviance information criterion; 逸脱度情報量規準) と呼んだ。AIC の評価の対象とするモデルは、最尤法によって推定したモデル $f(z|\hat{\theta}_{ML})$ であった。これに対して、DIC は最尤推定量をパラメータのベイズ推定量である事後平均 $\hat{\theta}_B = E_{\pi(\theta|\mathbf{y})}[\theta|\mathbf{y}]$ で置き換えたモデル $f(z|\hat{\theta}_B)$ を評価の対象とした。モデル $f(z|\hat{\theta}_B)$ と真のモデルとの近さを、K-L 情報量で予測の観点から測ったとき、AIC のモデルの自由パラメータ数に対応する有効自由度を次の式で与えた。

$$b_{\text{DIC}} = 2\{\log f(\mathbf{y}|\hat{\theta}_B) - E_{\pi(\theta|\mathbf{y})}[\log f(\mathbf{y}|\theta)]\}$$

従って、バイアスを補正した DIC は

$$(4.4) \quad \text{DIC} = -2\log f(\mathbf{y}|\hat{\theta}_B) + 2b_{\text{DIC}}$$

で与えられる。

一般に、最大対数尤度 $\log f(\mathbf{y}|\hat{\theta})$ がモデルのデータへの当てはまりの良さを表すのに対して、逆に当てはまりの悪さを表す $-2\log f(\mathbf{y}|\hat{\theta})$ を逸脱度という。DIC の $-2\log f(\mathbf{y}|\hat{\theta}_B)$ は事後平均に対する逸脱度に基づいており、この意味で逸脱度情報量規準と呼ばれている。

DIC の有効自由度 b_{DIC} は、ベイズモデルの設定によっては負の値を採ることもあり、このため有効自由度を事後分布に関する $\log f(\mathbf{y}|\theta)$ の分散 $b_{\text{DICa}} = 2\text{Var}_{\pi(\theta|\mathbf{y})}\{\log f(\mathbf{y}|\theta)\}$ とした DIC も提案されている (Gelman et al., 2013)。

4.3 情報量規準 WAIC

Spiegelhalter et al. (2014) では、DIC の果たしてきた役割を再考するとともに、いくつかのデメリットも議論されている。これに対して、Watanabe (2009, 2010) の提唱した WAIC (widely applicable information criterion) は、ベイズモデリングの過程の中にマルコフ連鎖モンテカルロ法による予測分布の積分計算を組み込んだ汎用性の高い情報量規準として用いられている。

WAIC は、(4.2) 式の予測分布に対して期待対数尤度 $\sum_{i=1}^n E_G[\log h(z_i|\mathbf{y})]$ を

$$\sum_{i=1}^n \log h(y_i|\mathbf{y}) = \sum_{i=1}^n \log \int f(y_i|\theta)\pi(\theta|\mathbf{y})d\theta$$

で推定したときのバイアスを

$$b_{\text{WAIC}} = 2 \left\{ \sum_{i=1}^n \log h(y_i|\mathbf{y}) - \sum_{i=1}^n E_{\pi(\theta|\mathbf{y})}[\log f(y_i|\theta)] \right\}$$

で与えた。この結果、ベイズ型予測分布モデルの評価を可能とする WAIC は、

$$(4.5) \quad \text{WAIC} = -2 \sum_{i=1}^n \log h(y_i|\mathbf{y}) + 2b_{\text{WAIC}}$$

で与えられた。その特徴は、事後分布 $\pi(\theta|\mathbf{y})$ から反復発生させた θ_s によって

$$b_{\text{WAIC}} = 2 \sum_{i=1}^n \left\{ \log \left(\frac{1}{S} \sum_{s=1}^S f(y_i|\theta_s) \right) - \frac{1}{S} \sum_{s=1}^S \log f(y_i|\theta_s) \right\}$$

とシミュレーションによって求めることができ、また特異モデルにも適用可能であることにあ
る。さらに、DIC の事後分布に関する分散に基づく有効自由度に対応する

$$b_{\text{WAICa}} = \sum_{i=1}^n \text{Var}_{\pi(\theta|\mathbf{y})} \{\log f(y_i|\theta)\} = \sum_{i=1}^n [E_{\pi(\theta|\mathbf{y})} \{\{\log f(y_i|\theta)\}^2\} - \{E_{\pi(\theta|\mathbf{y})} [\log f(y_i|\theta)]\}^2]$$

で補正した WAIC を提唱した。

5. 赤池ベイズ情報量規準 ABIC とベイズ型モデル評価基準 BIC

AIC としばしば比較の対象として取り上げられるモデル評価基準が、Schwarz (1978)によっ
て提唱された BIC である。BIC は、候補モデル集合の中でベイズ理論による事後確率が最も
高くなるようなモデルを最適なモデルとして選択するもので、K-L 情報量に基づいて予測の
視点から導出された AIC とは考え方を異にしている。従って、BIC は、Bayesian information
criterion というよりも、ベイズ型モデル評価基準と呼ぶ方が適切である。

一方、Akaike (1980b)の提唱した ABIC(Akaike Bayesian information criterion)は、AIC の考
え方に基づいてベイズモデルの評価を目的としたもので、この点でベイズ情報量規準といえる
評価基準である。本節では、AIC、ABIC と BIC の違いを見るために BIC 導出の概略を示し
た後、ABIC について述べる。また、モデル選択の一致性についてふれる。

5.1 モデルの事後確率と BIC

いま、 r 個のモデルの候補を M_1, M_2, \dots, M_r とし、各モデル M_i は確率分布モデル $f_i(y|\theta_i)$
($\theta_i \in \Theta_i \subset R^{p_i}$)とパラメータベクトル θ_i の事前分布 $\pi_i(\theta_i)$ によって特徴付けられているとす
る。ベイズアプローチでは、データ \mathbf{y} が観測されたとき、パラメータベクトル θ_i の事前分布
 $\pi_i(\theta_i)$ で積分した

$$p_i(\mathbf{y}) = \int f_i(\mathbf{y}|\theta_i)\pi_i(\theta_i)d\theta_i$$

を評価の対象とする。この $p_i(\mathbf{y})$ は、データ \mathbf{y} がモデル M_i から観測される尤もらしさを表し
ており、周辺尤度あるいは周辺分布と呼ばれる。

次に、 i 番目のモデルが生起する事前確率を $P(M_i)$ とすると、 i 番目のモデルの事後確率は、
ベイズの定理より

$$(5.1) \quad P(M_i|\mathbf{y}) = \frac{p_i(\mathbf{y})P(M_i)}{\sum_{j=1}^r p_j(\mathbf{y})P(M_j)}, \quad i = 1, 2, \dots, r$$

で与えられる。この事後確率は、データ \mathbf{y} が観測されたとき、そのデータが i 番目のモデルか
ら生起する確率であり、従って、 r 個のモデルの中から一つのモデルを選択するとき、事後確
率最大のモデルを採用するのが自然である。これは、(5.1)式の分母がすべてのモデルに共通で
あることから、分子の $p_i(\mathbf{y})P(M_i)$ を最大にするモデルを選択することと同等である。さらに、
事前確率 $P(M_i)$ はすべて等しいとした場合には、データの周辺尤度 $p_i(\mathbf{y})$ を最大にするモデル
を選択することになる。

Schwarz (1978)の提唱した BIC は、積分で表された周辺尤度を、積分のラプラス近似
(Barndorff-Nielsen and Cox, 1989, p.169; 宮田, 2018)を用いて近似した結果得られたもので、一
般に、周辺尤度 $p(\mathbf{y})$ に対して次の式で与えられた。

$$(5.2) \quad -2 \log p(\mathbf{y}) = -2 \log \left\{ \int f(\mathbf{y}|\theta)\pi(\theta)d\theta \right\} \approx -2 \log f(\mathbf{y}|\hat{\theta}_{ML}) + p \log n.$$

ただし、 $\hat{\theta}_{ML}$ はモデル $f(y|\theta)$ の p 次元パラメータベクトル θ の最尤推定値である。BIC の値を最小とするモデルを最適なモデルとして選択する。導出の詳細は、小西・北川 (2004, p.155), Konishi and Kitagawa (2008, p.215) を参照されたい。

導出方法からも分かるように、BIC は想定したモデル集合の中に真のモデルが含まれるという仮定を陽に置いているわけではない。事前確率は、想定した各モデルが真のモデルである確率を与えていると考えることができる。しかし、導出の過程ですべて等しいとしており、またモデルの複雑さを表すパラメータ数の項には、パラメータの事前分布の情報も現れていない。この問題に対して、積分のラプラス近似を精密化して導いたのが、一般化ベイズ型モデル評価基準 GBIC (Konishi, Ando and Imoto, 2004) である。

5.2 赤池ベイズ情報量規準 ABIC

ABIC は、季節調整法の大規模モデリングの過程で用いられた正則化法をベイズの観点から捉えて、ベイズモデリングに本質的な評価基準として提唱されたもので、以下のように述べることができる。

一般に正則化最尤法は、対数尤度関数に正則化項あるいはペナルティ項と呼ばれる制約 $R(\theta, \lambda)$ を課した正則化対数尤度関数

$$\log f(\mathbf{y}|\theta) - \frac{1}{2}R(\theta, \lambda)$$

の最大化によってパラメータを推定する。ここで、 $\lambda (\in \Lambda \subset R^q; q < p)$ は q 次元パラメータとし、正則化法の枠組みでは平滑化パラメータに相当する。この式は、

$$\log f(\mathbf{y}|\theta) + \log \left\{ \exp \left(-\frac{1}{2}R(\theta, \lambda) \right) \right\} = \log \left\{ f(\mathbf{y}|\theta) \exp \left(-\frac{1}{2}R(\theta, \lambda) \right) \right\}.$$

と書き表すことができる。さらに、指数関数の項を規格化して確率分布 $\pi(\theta|\lambda)$ とすると、これはハイパーパラメータ λ によって規定される θ の事前分布である。このとき、データ \mathbf{y} の周辺分布あるいは周辺尤度は

$$p(\mathbf{y}|\lambda) = \int f(\mathbf{y}|\theta)\pi(\theta|\lambda)d\theta$$

で与えられる。

ベイズモデルの周辺分布 $p(\mathbf{y}|\lambda)$ を、ハイパーパラメータ λ を持つパラメトリックモデルと考えると、モデルの評価は次の AIC の枠組みで捉えることができ、その評価基準は

$$(5.3) \quad \text{ABIC} = -2 \left\{ \max_{\lambda} \log p(\mathbf{y}|\lambda) \right\} + 2q = -2 \max_{\lambda} \log \left\{ \int f(\mathbf{y}|\theta)\pi(\theta|\lambda)d\theta \right\} + 2q$$

で与えられる。この評価基準は、Akaike (1980b) によって提唱され、赤池ベイズ情報量規準 (Akaike Bayesian information criterion) と呼ばれている。この方法によると、ベイズモデルのハイパーパラメータ λ は、対数周辺尤度 $\log p(\mathbf{y}|\lambda)$ の最大化、すなわち、最尤法によって推定しており、また、ハイパーパラメータによって特徴付けられた複数のベイズモデルの相対的な良さを比較するには、ABIC 最小化法によってモデルを選択すればよい。

このようにして推定したハイパーパラメータを $\hat{\lambda}$ とすると、パラメータ θ の事前分布 $\pi(\theta|\hat{\lambda})$ に対して、 θ の事後分布

$$\pi(\theta|\mathbf{y}; \hat{\lambda}) = \frac{f(\mathbf{y}|\theta)\pi(\theta|\hat{\lambda})}{\int f(\mathbf{y}|\theta)\pi(\theta|\hat{\lambda})d\theta}$$

が求まる。パラメータ θ の推定値としては通常、事後分布のモード、すなわち $\pi(\theta|y; \hat{\lambda}) \propto f(y|\theta)\pi(\theta|\hat{\lambda})$ を最大とする $\hat{\theta}$ が用いられる。ABIC 最小化法によるモデリングは、周辺分布 $p(y|\lambda)$ に対して最尤法によるハイパーパラメータの推定とモデル選択を行ったあと、パラメータ θ の事後分布 $\pi(\theta|y; \hat{\lambda})$ の最大化によって θ の推定値を求めていることが分かる。

ABIC 最小化法は、当初経済データの季節調整法の開発に用いられた (Akaike, 1980b, 1980c; Akaike and Ishiguro, 1980)。その後、コホート分析 (中村, 1982)、2 値回帰モデル (Ishiguro and Sakamoto, 1983)、地震学 (Ogata, Katsura and Tanemura, 2003; Ogata, 2004; 尾形, 2015)、状態空間モデリング (Kitagawa, 1987, 1996, 1998) など様々な分野のモデリングに利用されている。

5.3 モデル選択の一致性

AIC と BIC は、それぞれ K-L 情報量とベイズ事後確率という異なる基準に基づいて導出されたものであり、適用上異なる性質をもつことは明らかである。一般に、理論的側面からは、BIC はモデル選択に対して一致性を有していることが強調され、これに対して、AIC はミニマックス最適性を有することを両者の違いとして見ることが多い。

Shibata (1976) は、AIC 最小化法によって次数選択を行う場合、標本数 n を無限に大きくするとき、真の次数を選択する確率は 1 とならない、すなわち次数選択の一致性を有しないことを示している。また、Shibata (1983) は、想定したモデルの中に真のモデルが含まれない場合、AIC によって選択されたモデルは、平均 2 乗誤差の意味で漸近的に最適であることを示した。Nishii (1984) は、ガウス型線形回帰モデルに対して AIC, BIC を含むいくつかの評価基準の一致性をもつ条件等について検討した。Akaike (1978a) は、AIC 最小化法のベイズの観点からの捉え方を述べ、多変量ガウス分布モデルに対して、等事前確率の仮定の下で AIC は一致性は有しないが、ミニマックス解を与えることを示した。AIC タイプのモデル選択基準に対するミニマックス最適性と関連論文については、Yang (2005) を参照されたい。

AIC と BIC を比べるとき、モデル選択の一致性という観点からしばしば議論されてきた。一致性は、真のモデルが想定したモデル集合に含まれているという仮定のもとでの証明であることもあり、これに対して様々な指摘がなされた。Burnham and Anderson (2002)、小西・北川 (2004, p.66)、甘利 (2007, 第 II 編 1 章, p.66)、北川 (2007, 第 II 編 2 章, p.83)、Konishi and Kitagawa (2008, p.73) は、一致性をめぐる議論に否定的で、実際問題に対峙し現象解明に向けてモデリングを行うという観点から、それぞれの見解を明確に述べている。

実際、モンテカルロ・シミュレーションによって、AIC, BIC 等のモデル評価基準を比較検証するとき、データを発生させるモデルを真のモデルとすることが多い。さらに、真のモデルは想定したモデルの中に含まれ、しかも単純なモデルからデータを発生させることから、BIC との比較において一致性を有しない AIC が劣るという検証結果がしばしば見受けられる。このような設定のもとでのモデル評価基準の比較は、現象をモデル化するという立場からみると必ずしも実際的ではないことが指摘されている。赤池 (1995, p.199)、Burnham and Anderson (2002, p.20) は、真のモデルとは、本来、無限次元であり、従ってそれを観測された有限個のデータから完全に再現することは現実的でない。そこでモデル集合を想定して、その中から最良の近似モデルを選択するという考えに基づいており、少なくとも想定したモデル集合の中には、データを発生した真のモデルは含まれていないとしている。このような点を踏まえて、Burnham and Anderson (2002) は実データの解析、シミュレーションを通して、AIC の有用性を様々な角度から検証している。

6. モデル選択の不確定性と Akaike ウェイト

情報量規準によるモデル選択では、その値によってデータから構築した複数の候補モデルを順位付けして、データ発生の確率構造を最も良く近似するモデルを一つ選択する。しかし、評価基準はデータに依存しており、それ自身確率変数であることからモデル選択に起因する不確定性を生じる。したがって、評価基準値の近い他の候補モデルとの差をどのように解釈するかという問題に繋がる。

Akaike (1978a, 1978b, 1979, 1983) は、(2.5) 式の AIC に対して $(-1/2)$ AIC が期待対数尤度 ($\times n$) の漸近的な不偏推定量であることから

$$\exp\left(-\frac{1}{2}\text{AIC}\right) = \exp\{\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}_{ML}) - p\} = f(\mathbf{y}|\hat{\boldsymbol{\theta}}_{ML})e^{-p}$$

は最尤法によって推定したモデルの尤度とみなせ、また近似的に事後分布を表しているとした。したがって、この式を利用して AIC 値にそれほど大きな差のない候補モデル集合に対しては、平均化によって融合したモデルの構築を示唆している (Akaike, 1978b, 1979; Bozdogan, 1987)。この考え方を発展させて Akaike (1978b), Burnham and Anderson (2002) は、最小 AIC 値と各候補モデルの AIC 値との差に基づくモデルの相対的な確からしさを、尤度と関連づけて、以下のように定式化している。

いま、データ \mathbf{y} に基づいて推定したモデル集合を $\{f_i(\mathbf{y}|\hat{\boldsymbol{\theta}}_i); i = 1, \dots, r\}$ とする。各モデルに対する AIC 値を AIC_i とし、最小 AIC 値を AIC_{\min} とおく。このとき、両者の差

$$\Delta\text{AIC}_i = \text{AIC}_i - \text{AIC}_{\min}, \quad i = 1, \dots, r$$

は、最小 AIC 値を基にしたモデルの相対的な比較を表している。さらに、

$$\exp\left(-\frac{1}{2}\Delta\text{AIC}_i\right)$$

は、データが与えられたとき、モデル i の近似的な尤度を表すことから、この AIC 値の差を基準化した次の式は Akaike ウェイトと呼ばれる (Burnham and Anderson, 2002, p.75)。

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta\text{AIC}_i\right)}{\sum_{k=1}^r \exp\left(-\frac{1}{2}\Delta\text{AIC}_k\right)}$$

ウェイトの総和は $\sum_{i=1}^r w_i = 1$ であることから、 w_i はモデル集合に属する各モデルの相対的な確からしさの指標となっている。Burnham and Anderson (2002) の提唱したマルチモデル推測とは、Akaike ウェイトを複数のモデルの相対的な良さとして用いて、モデル集合に基づく推論を実行する方法であるといえる。AIC 最小化法によって選択された一つのモデルが、明らかに他のモデルより現象予測に優れたモデルであればよいが、そうでない場合には順序付けられた複数のモデルを融合してモデルの推測を行う方が有効であり、Akaike ウェイトはそのための基準指標を示すと考えることができる。

マルチモデル推測は、モデル選択の不確定性に対処するための一つのアプローチであるといえる。同様に、2つのモデルの AIC の差の有意性の検定や信頼集合を構成して不確定性を検証する方法 (Linhart, 1988; Shimodaira, 1997)、さらに観測データからブートストラップ標本を反復抽出してモデル選択を繰り返して実行し、各モデルが選択される頻度を推定するブートストラップ選択確率に基づく方法などが提唱されている。モデル選択の不確定性については、Kishino and Hasegawa (1989), Shimodaira and Hasegawa (1999), Burnham and Anderson (2002, Chapter 6), 下平 (2004, 2007) を参照されたい。

7. おわりに

本稿では、AIC 導出の理論を振り返ってみることから始め、一連のモデリングのプロセスの中で AIC の果たしてきた役割を再考し、一般に情報量規準と呼ばれるモデル評価基準がどのように提唱されてきたかを述べた。

情報量規準 AIC 導出の基本的考え方は、最尤法によって推定したモデルを確率分布で表現し、それを Kullback-Leibler 情報量によって予測の視点から評価したことであった。この基本概念は、データからの情報を確率分布モデルで捉え、当該分野で蓄積された知識を事前分布としてベイズ理論によってモデルに同化させた予測分布モデルの評価・選択へと繋がった。1980 年に提唱された赤池情報量規準 ABIC は、5.2 節で述べたようにモデルのパラメータ数がデータ数を超えるような大規模モデリングの過程で用いられた正則化法をベイズの観点から捉えて、ベイズモデリングに本質的な評価基準として提唱されたものであった。本稿は、理論的側面から情報量規準について述べてきたが、AIC, ABIC は、自然科学はもとより社会科学の様々な分野で、現象解明のためのモデリングに重要な役割を果たしてきた。

近年、諸科学・産業界では、計測・測定技術、計算機関連技術の高度な発展によって、大規模・高次元データが獲得、蓄積され、次々とデータベース化されつつある。特に、少数かつ高次元データ、大量かつ超高次元データからの効率的な情報抽出技術の開発研究が強く希求され、国際的に研究が推進されている。このような状況の中で、正則化法、 L_1 ノルム正則化法、ベイズモデルは、現象分析のための重要な解析手法として用いられ、適用上の問題点の克服と汎化能力の向上を目指して、AIC タイプのモデル評価基準が提唱されてきたといえる。

情報量規準 AIC の理論は、これからますます多様な形式で獲得される複雑な大規模データの背後に潜む有益な情報やパターンを高効率に抽出・処理するための新しいモデリングの中でも活かされるものと思われる。同時に、過去から学び未来を予測する技術を獲得するためには、新たな発想でモデリングに取り組む必要があることも確かである。

謝 辞

統計数理「創立 75 周年記念号」への発表の機会を与えて下さいました統計数理研究所特任教授の田村義保先生に感謝致します。東京大学特任教授の北川源四郎先生には、数々の貴重なご意見、ご指摘を賜りました。また、査読者には、貴重なご指摘をいただきました。ここに記して厚く御礼申し上げます。

参 考 文 献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), 267-281, Akademiai Kiado, Budapest. (Reproduced in *Breakthroughs in Statistics, Vol.1, Foundations and Basic Theory* (eds. S. Kotz and N.L. Johnson), Springer-Verlag, New York, (1992).)
- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **AC-19**(6), 716-723.
- Akaike, H. (1978a). A Bayesian analysis of the minimum AIC procedure, *Annals of the Institute of Statistical Mathematics*, **30**, 9-14.
- Akaike, H. (1978b). On the likelihood of a time series model, *The Statistician*, **27**, 217-235.
- Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting, *Biometrika*, **66**, 237-242.

- Akaike, H. (1980a). On the use of predictive likelihood of a Gaussian model, *Annals of the Institute of Statistical Mathematics*, **32**, 311-324.
- Akaike, H. (1980b). Likelihood and the Bayes procedure, *Bayesian Statistics* (eds. J. M. Bernardo, M. H. De Groot, D. V. Lindley and A. F. M. Smith), 143-166 (discussion 185-203), University Press, Valencia, Spain.
- Akaike, H. (1980c). Seasonal adjustment by a Bayesian modeling, *Journal of Time Series Analysis*, **1**, 1-13.
- Akaike, H. (1983). Information measures and model selection, *International Statistical Institute*, **44**, 277-291.
- 赤池弘次 (1995). 『時系列解析の心構え, 時系列解析の実際 II』(赤池弘次, 北川源四郎 編), 第 12 章, 朝倉書店, 東京.
- Akaike, H. and Ishiguro, M. (1980). A Bayesian approach to the trading-day adjustment of monthly data, *Time Series Analysis* (eds. O. D. Anderson and M. R. Perryman), 213-226, North-Holland, Amsterdam.
- 赤池弘次, 北川源四郎 編 (1994). 『時系列解析の実際 I』, 朝倉書店, 東京.
- 赤池弘次, 北川源四郎 編 (1995). 『時系列解析の実際 II』, 朝倉書店, 東京.
- 甘利俊一 (2007). 『赤池情報量規準—その思想と新展開, 赤池情報量規準 AIC—モデリング・予測・知識発見—(室田一雄, 土谷隆 編)』, 第 II 編, 1 章, 52-78, 共立出版, 東京.
- Ando, T., Konishi, S. and Imoto, S. (2008). Nonlinear regression modeling via regularized radial basis function networks, *Journal of Statistical Planning and Inference*, **138**, 3616-3633.
- Araki, Y., Konishi, S., Kawano, S. and Matsui, H. (2009a). Functional regression modeling via regularized Gaussian basis expansions, *Annals of the Institute of Statistical Mathematics*, **61**, 811-833.
- Araki, Y., Konishi, S., Kawano, S. and Matsui, H. (2009b). Functional logistic discrimination via regularized basis expansions, *Communications in Statistics — Theory & Methods*, 2944-2957.
- Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, New York.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, **52**, 345-370.
- Bozdogan, H. (ed.) (1994). *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach*, Kluwer Academic Publishers, the Netherlands.
- Breiman, L. (1996). Heuristics of instability and stabilization in model selection, *Annals of Statistics*, **24**, 2350-2383.
- Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information — Theoretical Approach*, Springer-Verlag, New York.
- Davison, A. C. (1986). Approximate predictive likelihood, *Biometrika*, **73**, 323-332.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *Annals of Statistics*, **7**, 1-26.
- Efron, B. (2004). The estimation of prediction error: Covariance penalties and cross-validation, *Journal of the American Statistical Association*, **99**, 619-632.
- Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression, *Annals of Statistics*, **32**, 407-499.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalised likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348-1360.
- Fan, Y. and Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood, *Journal of the Royal Statistical Society*, **B75**, 531-552.
- Fernholz, L. T. (1983). *von Mises Calculus for Statistical Functionals*, Lecture Notes in Statistics, **19**, Springer-Verlag, New York.

- Frank, I. and Friedman, J. (1993). A statistical view of some chemometrics regression tools, *Technometrics*, **35**, 109-148.
- Fujikoshi, Y. and Satoh, K. (1997). Modified AIC and C_p in multivariate linear regression, *Biometrika*, **84**, 707-716.
- 藤越康祝, 杉山高一 (2012). 『多変量モデルの選択』, 朝倉書店, 東京.
- Gelman, A., Carlin, J. C., Stern, H. and Dunson, D. B. (2013). *Bayesian Data Analysis*, 3rd ed., Chapman and Hall/CRC, New York.
- Good, I. J. and Gaskins, R. A. (1971). Nonparametric roughness penalties for probability densities, *Biometrika*, **58**, 255-277.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics, The Approach Based on Influence Functions*, Wiley, New York.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman and Hall, London.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*, 2nd ed., Springer, New York.
- Hastie, T., Tibshirani, R. and Wainwright, M. (2015). *Statistical Learning with Sparsity*, Chapman & Hall/CRC, New York.
- 廣瀬 慧 (2016). スパースモデリングとモデル選択, 電子情報通信学会誌, **99**, 392-399.
- Hirose, K., Tateishi, S. and Konishi, S. (2013). Tuning parameter selection in sparse regression modeling, *Computational Statistics & Data Analysis*, **59**, 28-40.
- Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.
- Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples, *Biometrika*, **76**, 297-307.
- Hurvich, C. M., Simonoff, J. S. and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion, *Journal of the Royal Statistical Society*, **B60**, 271-293.
- Imoto, S. and Konishi, S. (2003). Selection of smoothing parameters in B -spline nonparametric regression models using information criteria, *Annals of the Institute of Statistical Mathematics*, **55**, 671-687.
- Ishiguro, M. and Sakamoto, Y. (1983). A Bayesian approach to binary response curve estimation, *Annals of the Institute of Statistical Mathematics*, **35**, 115-137.
- Ishiguro, M., Sakamoto, Y. and Kitagawa, G. (1997). Bootstrapping log likelihood and EIC, an extension of AIC, *Annals of the Institute of Statistical Mathematics*, **49**, 411-434.
- Kato, K. (2009). On the degrees of freedom in shrinkage estimation, *Journal of Multivariate Analysis*, **100**, 1338-1352.
- Kawano, S. and Konishi, S. (2011). Semi-supervised logistic discrimination via regularized Gaussian basis expansions, *Communications in Statistics — Theory and Methods*, **40**, 2412-2423.
- 川野 秀一, 廣瀬 慧, 立石正平, 小西貞則 (2010). 回帰モデリングと L_1 型正則化法の最近の展開, 日本統計学会誌, **39**, 211-242.
- Kawano, S., Misumi, T. and Konishi, S. (2012). Semi-supervised logistic discrimination via graph-based regularization, *Neural Processing Letters*, **36**, 203-216.
- 川野 秀一, 松井秀俊, 廣瀬 慧 (2018). 『スパース推定法による統計モデリング』, 共立出版, 東京.
- Kayano, M., Dozono, K. and Konishi, S. (2010). Functional cluster analysis via orthonormalized Gaussian basis expansions and its application, *Journal of Classification*, **27**, 211-230.
- Kishino, H. and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, *Journal of Molecular Evolution* **29**, 170-179.
- Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series (with discussion), *Journal of the American Statistical Association*, **82**, 1032-1063.

- Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics*, **5**, 1-25.
- Kitagawa, G. (1997). Information criteria for the predictive evaluation of Bayesian models, *Communications in Statistics — Theory and Methods*, **26**, 2223-2246.
- Kitagawa, G. (1998). Self-organizing state space model, *Journal of the American Statistical Association*, **93**, 1203-1215.
- 北川源四郎 (2007). 『情報量規準と統計的モデリング, 赤池情報量規準 AIC—モデリング・予測・知識発見—(室田一雄, 土谷 隆 編)』, 第 II 編, 2 章, 79-109, 共立出版, 東京.
- Kitagawa, G. and Gersch, W. (1984). A smoothness priors-state space modeling of time series with trend and seasonality, *Journal of the American Statistical Association*, **79**, 378-389.
- Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statistics, **116**, Springer-Verlag, New York.
- Kitagawa, G. and Konishi, S. (2010). Bias and variance reduction techniques for bootstrap information criteria, *Annals of the Institute of Statistical Mathematics*, **62**, 209-234.
- Konishi, S. (1999). Statistical model evaluation and information criteria, *Multivariate Analysis, Design of Experiments and Survey Sampling* (ed. S. Ghosh), 369-399, Marcel Dekker, New York.
- Konishi, S. (2002). Theory for statistical modeling and information criteria—functional approach, *Sugaku Expositions*, **15**(1), 89-106, American Mathematical Society.
- 小西貞則 (2010). 『多変量解析入門—線形から非線形へ—』, 岩波書店, 東京.
- Konishi, S. (2014). *Introduction to Multivariate Analysis: Linear and Nonlinear Modeling*, Chapman & Hall/CRC, New York.
- Konishi, S. and Kitagawa, G. (1996). Generalized information criteria in model selection, *Biometrika*, **83**, 875-890.
- Konishi, S. and Kitagawa, G. (2003). Asymptotic theory for information criteria in model selection—functional approach, *Journal of Statistical Planning and Inference*, **114**, 45-61.
- 小西貞則, 北川源四郎 (2004). 『情報量規準』, 朝倉書店, 東京.
- Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*, Springer, New York.
- Konishi, S., Ando, T. and Imoto, S. (2004). Bayesian information criterion and smoothing parameter selection in radial basis function network, *Biometrika*, **91**, 27-43.
- 小西貞則, 越智義道, 大森裕浩 (2008). 『計算統計学の方法—ブートストラップ, EM アルゴリズム, MCMC—』, 朝倉書店, 東京.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, *Annals of Mathematical Statistics*, **22**, 79-86.
- Liang, H., Wu, H. and Zou, G. (2008). A note on conditional AIC for linear mixed-effects models, *Biometrika*, **95**, 773-778.
- Linhart, H. (1988). A test whether two AICs differ significantly, *South African Statistical Journal*, **22**, 153-161.
- Lv, J. and Liu, J-S. (2014). Model selection principles in misspecified models, *Journal of the Royal Statistical Society*, **B76**, 141-167.
- Mallows, C. L. (1973). Some comments on C_p , *Technometrics*, **15**, 661-675.
- Matsui, H. and Konishi, S. (2011). Variable selection for functional regression models via the L_1 regularization, *Computational Statistics and Data Analysis*, **55**, 3304-3310.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman & Hall/CRC, London.
- McQuarrie, A. D. R. and Tsai, C.-L. (1998). *Regression and Time Series Model Selection*, World Scientific, Singapore.

- Misumi, T. and Konishi, S. (2016). Mixed effects historical varying coefficient model for evaluating dose-response in flexible-dose trials, *Journal of the Royal Statistical Society (Applied Statistics)*, **C65**, 331-344.
- 宮田庸一 (2018). ラプラス近似のベイズ統計学への応用とその周辺, *数学*, **70**, 275-295.
- Murata, N., Yoshizawa, S. and Amari, S. (1994). Network information criterion-determining the number of hidden units for an artificial neural network model, *IEEE Transactions on Neural Networks*, **6**, 865-872.
- 室田一雄, 土谷 隆 編 (2007). 『赤池情報量規準 AIC—モデリング・予測・知識発見—』, 共立出版, 東京.
- 中村 隆 (1982). ベイズ型コウホート・モデル—標準コウホート表への適用—, *統計数理研究所彙報*, **29**, 77-97.
- Ninomiya, Y. and Kawano, S. (2016). AIC for the Lasso in generalized linear models, *Electronic Journal of Statistics*, **10**, 2537-2560.
- Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression, *Annals of Statistics*, **12**, 758-765.
- Ogata, Y. (2004). Space-time model for regional seismicity and detection of crustal stress changes, *Journal of Geophysical Research*, **109**(B3), B03308, doi:10.1029/2003JB002621.
- 尾形良彦 (2015). 地震の確率予測の研究—その展望, *統計数理*, **63**(1), 3-27.
- Ogata, Y., Katsura, K. and Tanemura, M. (2003). Modelling heterogeneous space-time occurrences of earthquakes and its residual analysis, *Applied Statistics*, **52**, 499-509.
- Park, H. and Konishi, S. (2017). Principal component selection via adaptive regularization method and generalized information criterion, *Statistical Papers*, **58**, 147-160.
- Parzen, E., Tanabe, K. and Kitagawa, G. (1998). *Selected Papers of Hirotugu Akaike*, Springer, New York.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed., Springer, New York.
- Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 461-464.
- Shen, X. and Ye, J. (2002). Adaptive model selection, *Journal of the American Statistical Association*, **97**, 210-221.
- Shen, X., Huang, H-C. and Ye, J. (2004). Adaptive model selection and assessment for exponential family distributions, *Technometrics*, **46**, 306-317.
- Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika*, **63**, 117-126.
- Shibata, R. (1983). Asymptotic mean efficiency of a selection of regression variables, *Annals of the Institute of Statistical Mathematics*, **35**, 415-423.
- Shibata, R. (1989). Statistical aspects of model selection, *From Data to Model* (ed. J. C. Willemsa), 215-240, Springer-Verlag, New York.
- Shimodaira, H. (1997). Assessing the error probability of the model selection test, *Annals of the Institute of Statistical Mathematics*, **49**, 395-410.
- 下平英寿 (2004). 情報量規準によるモデル選択とその信頼性評価, 『モデル選択—予測・検定・推定の交差点』(甘利俊一, 竹内 啓, 竹村彰通, 伊庭幸人 編), *統計科学のフロンティア* 3, 第 I 部, 1-76, 岩波書店, 東京.
- 下平英寿 (2007). モデル選択とブートストラップ, 『赤池情報量規準 AIC—モデリング・予測・知識発見—』, 第 II 編, 4 章, 133-156, 共立出版, 東京.
- Shimodaira, H. and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference, *Molecular Biology and Evolution*, **16**, 1114-1116.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society*, **B64**, 583-639.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2014). The deviance information

- criterion: 12 years on (with discussion), *Journal of the Royal Statistical Society*, **B76**, 485-493.
- Stone, M. (1974). Cross-validated choice and assessment of statistical predictions (with discussion), *Journal of the Royal Statistical Society*, **B39**, 111-147.
- Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, *Journal of the Royal Statistical Society*, **B39**, 44-47.
- Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections, *Communications in Statistics*, **A7**, 13-26.
- 竹内 啓 (1976). 情報統計量の分布とモデルの適切さの規準, *数理科学*, **153**, 12-18.
- Tateishi, S. and Konishi, S. (2011). Nonlinear regression modeling and detecting change points via the relevance vector machine, *Computational Statistics*, **26**, 477-490.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society*, **B58**, 267-288.
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association*, **81**, 82-86.
- Uchida, M. and Yoshida, N. (2001). Information criteria in model selection for mixing processes, *Statistical Inference and Stochastic Processes*, **4**, 73-98.
- Uchida, M. and Yoshida, N. (2004). Information criteria for small diffusions via the theory of Malliavin-Watanabe, *Statistical Inference and Stochastic Processes*, **7**, 35-67.
- Umezū, Y., Shimizu, Y., Masuda, H. and Ninomiya, Y. (2019). AIC for the non-concave penalized likelihood method, *Annals of the Institute of Statistical Mathematics*, **71**(2), 247-274.
- von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions, *Annals of Mathematical Statistics*, **18**, 309-348.
- Wang, H., Li, R. and Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method, *Biometrika*, **94**, 553-568.
- Wang, H., Li, B. and Leng, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters, *Journal of the Royal Statistical Society*, **B71**, 671-683.
- Watanabe, S. (2009). *Algebraic Geometry and Statistical Learning Theory*, Cambridge University Press, Cambridge, UK.
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory, *Journal of Machine Learning Research*, **11**, 3571-3594.
- Wood, S. N., Pya, N. and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models, *Journal of the American Statistical Association*, **111**, 1548-1575.
- Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation, *Biometrika*, **92**, 937-950.
- Ye, J. (1998). On measuring and correcting the effects of data mining and model selection, *Journal of the American Statistical Association*, **93**, 120-131.
- Yu, D. and Yau, K. K. (2012). Conditional Akaike information criterion for generalized linear mixed models, *Computational Statistics & Data Analysis*, **56**, 629-644.
- Zhang, Y., Li, R. and Tsai, C.-L. (2010). Regularization parameter selections via generalized information criterion, *Journal of the American Statistical Association*, **105**, 312-323.
- Zou, H., Hastie, T. and Tibshirani, R. (2007). On the "degrees of freedom" of the lasso, *Annals of Statistics*, **35**, 2173-2192.

The Role of Information Criterion AIC in Statistical Science

Sadanori Konishi

Department of Mathematics, Faculty of Science and Engineering, Chuo University;
Now at Faculty of Mathematics, Kyushu University

The Akaike information criterion (AIC) provides a useful tool for evaluating models estimated by the maximum likelihood method, and a number of successful applications of AIC have been reported in diverse fields of the natural and social sciences. AIC was essentially derived as an estimator of the Kullback-Leibler information from the predictive point of view, and it provided a new paradigm for model selection and evaluation problems in statistical science. The first objective of this paper is to provide a brief explanation of the concept and derivation of the AIC and related criteria.

With the development of modeling techniques such as regularization, sparse modeling, and Bayes modeling, it is necessary to present criteria that enable us to evaluate models constructed by various estimation procedures. The second objective of this paper is to review the AIC type of information criteria for evaluating models estimated by various techniques, with emphasis on the choice of the adjusted parameters, including a smoothing parameter. We review some advances in Bayesian information-theoretic criteria, where criteria were constructed, using the concept of the degrees of freedom as a bias-corrected adjustment. We also describe information-theoretic criteria for evaluating a Bayesian predictive distribution, derived from the fundamental principle behind AIC.

統計地震学の発展と地震活動予測： 個人的経験と展望

尾形 良彦[†]

(受付 2019 年 1 月 4 日；改訂 3 月 15 日；採択 3 月 26 日)

要 旨

地震予測の観点から統計地震学の発展と筆者の研究経験を説明する。点過程モデルによる地震活動の予測と統計的診断解析および地震活動の物理現象との関わりに焦点を当てる。

キーワード：地震活動，点過程モデル，ETAS モデル，階層ベイズ法，確率予測，前震の識別。

1. はじめに

筆者の統計科学や地震学の研究者に向けた研究紹介については既に尾形 (1993) や Ogata (2013, 2017a) などがある。なので、同様なことを繰り返すのは気が引けたが、本稿に似たタイトルの論文 (Vere-Jones, 2006) に倣って筆者も先人の成果を受け継いだ経緯の個人的経験と見解を述べたい。

筆者の折々の研究活動は、統計数理研究所学術研究レポジトリ <https://ismrepo.ism.ac.jp/> (最終閲覧日 2019 年 4 月 5 日) で検索でき、とくに統計地震学研究プロジェクトの成果と当時の目標をまとめた外部評価報告書 (英文和訳付き 統計数理研究所編 https://www.ism.ac.jp/evaluation/index_j/toukei_jisingaku2006.pdf, 最終閲覧日 2019 年 4 月 5 日) も参照できる。

本稿では、主に地震の短期・中期予測に関わる点過程解析について解説するが、予測全般については、筆者が編集した地震予測の特集号 (統計数理 63 巻 2015 年発行; 統計数理研究所編, <http://www.ism.ac.jp/editsec/toukei/tokeisuri-63j.html>, 最終閲覧日 2019 年 4 月 16 日) がある。これを本稿では「統計数理 63」と略記し、これに論文著者名を付加して引用する。

本稿では割愛するが、長きにわたり膨大化した地震発生様式のデータベースである震源カタログは、観測システムの発展に伴い、時空間的に不均質である。データの欠測推定や系統誤差の補正など、本質的な情報を抽出するために、ベイズモデルと解析法を考慮し応用した。これらの概要については Ogata (2017b, Sections 3.5, 3.6) を参照頂きたい。

本稿で述べる内容に関して全て引用するとなると、与えられたページ数を大幅に超えるので本稿では、以上の解説論文に含まれていない引用文献を記すに止め、その他は上記の解説論文の文献を通して検索されたい。特に、上記の統計学会誌 60 周年号解説論文、尾形 (1993) は本稿で割愛した内容を頻繁に引用し、Supplemental Material となるので文献欄に論文の PDF address を付けた。

[†] 統計数理研究所 名誉教授：〒190-8562 東京都立川市緑町 10-3

2. 統計地震学と地震活動の研究

2.1 地震活動

筆者が、点過程の応用問題を模索し、地震学会に出没して特に興味をもったのは地震活動の分科会であった。当時急速に蓄積された地震データの解析の報告が、相次いだものであった(例えば宇津, 1999, 参照)。

データに基づく地震の記述と理解のための研究分野としての統計地震学は、明治時代以来の地震学の主分野であった。一回の地震について分かることは「発生時刻, 規模, 場所」だけであったが、これら 3 つの性質を相手にして、地震発生に関する様々な性質を明らかにしようとした多くの研究がある。後に現代地震学の基礎を築いた安芸 (1956) の総合論文から分かる様に、昔からこの分野で議論されていたのは、地震発生の周期性, 気圧や海洋潮汐変化など地球物理データとの因果性の研究, 震源の移動, 地震活動パターン分類, 地震の規模(マグニチュード)の分布などであるが、それらの解析結果の妥当性をめぐり最大の争点は、余震や群発地震のような地震の続発性による、統計的検定や仮説の再現性の難しさにあった。

したがってアメリカの地震学界では余震の統計的研究は無視された。地震の規模(マグニチュード)を初めて考案したリヒター(Charles Richter)などは、余震は何の役にも立たないデータのゴミ(ノイズ)と言っていた。地震の予測にはゴミである余震を取り除いた本震の活動を見なければいけない、という考えだったようである。この思想は 20 世紀後半での「地震活動の静穏化」, 「地震発生の季節性や地球潮汐との同期」や「異常現象の大地震との因果性」などに関する統計的研究に受け継がれる。すなわち、地震活動のデータから「余震」を取り除く「除群(de-cluster)法」を使った研究である(尾形, 1993, 7.1 節参照)。先ず除群データの帰無仮説(定常ポアソン過程)を否定してから、焦眉の科学的主張をする。

しかし除群法は様々で、異なった結果を導いたりするだけでなく、余震活動自体の研究に興味をなさない。事実、地震活動は、時間的には逆べき型減衰の自己相関をもつ長記憶性、空間的にはフラクタル性をもつ自己相似性(尾形, 1993, 5.1, 6.5, 8.1 節; Ogata and Katsura, 1991; Guo and Ogata, 1997 参照)がみとめられる。このため、除群は首尾一貫したものでないといえる。とくにマグニチュード 7 以上の大地震といえども、履歴依存性や長記憶性がある(Ogata and Abe, 1991; Ogata, 2017b)。しかるに、大地震は定常ポアソン過程になるはずだ、という全く誤った帰無仮説による検定をし、地震波の原記録に基づいた阿部カタログのマグニチュードを雑に改竄した、20 世紀の地震カタログ(尾形, 1993, 5.2 節; Engdahl and Villasenor, 2002)が流布し、世界の被害地震の基礎データとして使われているのは深刻である。

2.2 余震の研究と統計地震学の発展

日本では、余震データを重要な情報として研究することで、地震活動の本質的な性質に迫ることになった。大きな地震の後に、より小さな地震が数多く発生するが、古来このような地震群を余震と呼び、そのきっかけとなった地震は本震と呼ばれている。

余震の発生は、本震直後は極めて頻繁で、その後、時間とともに徐々に低下するが、なかなか元の状態に戻らない。このような余震減衰の定量的な関係を初めて論じたのは大森房吉である。1894 年の濃尾地震の余震数の頻度の時間経過について、減衰のしかたを調べて「物理現象の減衰だから当然指数関数だろうと考えて当てはめてみたが良く合わない、然るに双曲線だとよく適合する」と述べている。さらに 1957 年、宇津徳治は日本や世界の余震を調べ、単位時間あたりの余震頻度の減衰率が

$$(2.1) \quad \nu(t) = K(t+c)^{-p}$$

の形になることを示した。ここで t は、本震の発生時刻からの経過時間である。宇津は、余震の頻度 $\nu(t)$ と経過時間 t を両対数方眼紙にプロットし、その減衰が漸近的に直線上に乗ることを示し、直線の傾きから指数 p の推定値を得た。そして、濃尾地震の余震は現在に至るまで、一世紀以上でも (2.1) 式の減衰が続いていることを示した。大森の言う上記の双曲線とは $p = 1$ のことである。一方、係数 c の解釈については議論がある (Utsu et al., 1995 参照)。宇津は (2.1) 式を「改良大森公式」と名付けたが、今日筆者らは「大森・宇津の公式」と呼んでいる。

さらに宇津は、多くの余震減衰のプロットで、単一の (2.1) 式で説明できない場合があることも示している。大きな余震の後、飛躍的な増加が起こって、再び減衰する二次余震(余震の余震)を発見し、その減衰も大森・宇津の式 (2.1) に従っていることを突き止めている。これは、余震が本震のみに誘発されるという当時の常識を覆したものである。このほか宇津は、余震群の空間的広がり和本震の規模(マグニチュード)との関係の定量的研究や、本震と最大余震のマグニチュード差の経験分布など、余震の多様な性質に関して多くの事例の統計的解析を行った。

さて、ここまでの研究では、余震の単位時間あたりの数として分析されてきたが、筆者は、(2.1) 式を、極短い時間内に1つの地震が起こる確率の微分(発生強度)と捉え、1つずつの地震の発生時刻の記録 $\{t_i; i = 1, 2, \dots\}$ をそのままデータとして使う、発生強度による推定法を提案した。これは余震列を非定常ポアソン過程として (2.1) 式のパラメータ K, c, p を最尤法(さいゆうほう)で推定する方法である。本震直後の欠測の問題や二次余震を含む場合も、赤池情報量規準(AIC)でモデル比較の上、偏りの無い推定ができる (Ogata, 1983)。今日、大地震が起きると、余震発生時刻データを基に (2.1) 式を推定して、余震の予報確率が計算されることになっている。背景として1960年代ごろからの、確率論や統計学の分野で発展した点過程とその統計的解析法の理論的整備が、これらを可能にしたのである。このことは次節で詳しく述べる。

同じく1960年代ごろからの地震の物理の飛躍的前進として、地震発生の応力蓄積の根源となるプレートテクトニクス理論が展開され、「地震の揺れの原因は断層面の急激な食い違い運動である」という弾性反発説が定着した。大地震ごとに断層運動モデルが、地震波の解析によって求められる。今世紀では、それは衛星観測の測地的変化から詳しく求められて、地震の発生様式や地震力学の物理的解明が進んだ。かくて地震カタログは、破壊の始まった時刻と位置、食い違った断層の面積とすべり量を表すデータと理解された。さらに断層面および直交面の方向や断層の重心などのデータも追加されるようになった。今日の地震活動は、これらのデータに基づいて研究されている。

しかし、小さい地震が増々捉えられ、データベースが豊富になるにしたがって、地震の発生様式の地域性や複雑系が顕著化して、詳細な統計的把握や予測も難しくなってくる。本質的で確かな情報を抽出するために、時間的空間的に非定常または不均質な予測モデルを考慮する必要があり、これらの統計モデルを用いた研究が避けられない様になってきた。時空間モデルなどの逆問題にかかわる大量の未知パラメータを含む大規模モデルを取り扱う階層的なベイズ法の助けが必要であり、この様な手法抜きでは地震予測の研究自体が難しくなっている。

3. 点過程の統計解析と統計地震学の研究の経過

3.1 点過程の地震統計解析の始まり

点過程は、事象(出来事)の発生時間と付置のマーク(ジャンプなどのスカラー量または震源の空間位置など)の時系列の数学的表現として、確率論的に深く研究されてきた。通常の連続確率過程の基礎が正規過程のホワイトノイズ(white noise)であるのに対して、点過程の基礎は定常ポアソン過程である。各種の応用に対応して点過程の一般化が進んで、1970年代は点過程の確率論、統計解析、応用の統一的な研究が進んだ時代であった (Lewis ed., 1972 参照)。

ヴェア・ジョーンズのトリガーモデル (Trigger model; 詳しくは尾形, 1993, 6.1, 6.4, 6.5 節参照) は, 地震活動の一般化点過程モデルの先駆けである. 本震は定常ポアソン過程に従い, 余震は本震のみに誘発され, 大森・宇津の減衰関数(2.1)の非定常ポアソン過程で, 地震活動はこれらを重ね合わせたものである. これは当時の地震学的常識を反映したものであるが, データのどれが本震であるか余震であるかの同定は明確でない. それらを設定するための組み合わせ論的複雑さを伴い, 尤度(ゆうど)関数で計算するのが極めて困難である. したがって, モーメント法に基づいた推論や解析が使われた. 特に, ホークス(A. G. Hawkes)と彼の学生 Adamopoulos (1976) は, トリガーモデルとホークスモデルを二次モーメントのスペクトル尤度(尾形, 1993, A3 節)で計算したが, 精度は良くない. 他方, 本来のトリガーモデルではないが, 筆者は発震時刻のみのデータから各地震の誘発地震数(クラスターサイズ)などを推定した(Ogata, 2001a).

3.2 統計数理研究所における点過程の光臨

1970年代当時, 赤池弘次は, 多変量時系列の統計的同定によってフィードバック効果を考慮した予測と制御の実用化を目指し, 外力変数を含む自己回帰モデルを状態空間表示し, 最小二乗法や最尤法による最良予測のモデル選択のため Final Prediction Error (FPE) や赤池情報量規準(AIC)を提案し, 世界の統計学界を牽引していた. 周囲の若手研究者に対しては, 新しい分野での各種統計モデルの構築を促していた.

ニュージーランドのヴェア・ジョーンズ(David Vere-Jones)は, 点過程の理論とモデルで地震データと取り組んだ先駆者であるが, 宇津を初めとする日本の地震学者との交流を求め, 1976年, 赤池に招待されて統計数理研究所を数か月間訪問した. その間, 点過程に関する一連の特別講義をした. このときの経緯が Vere-Jones (2006) や Ogata (2018) によって述べられている. 彼の講義を聞いて直ちに赤池は, 時系列解析と同様な予測の展開を, 点過程モデルの予測に関する中心概念である「条件付き強度関数」

$$(3.1) \quad \lambda(t|H_t) = P\{\text{an event occurs in } [t, t + \Delta] | H_t\} \Delta + o(\Delta)$$

に期待した. これは事象(点)が発生する条件付き確率の微分量である. ここで条件 H_t は時刻 t までの, 事象発生履歴および関連情報のことである. 単純な点過程なら H_t は時刻 t 直前までの発生時刻の系列 $\{t_1, \dots, t_n\}$ である. 条件付き強度関数をモデル化することによって, パラメータの推定や事象時刻のシミュレーションなどで, 予測問題を考えることができる. 例えば, 条件付き強度関数で, 最後の事象の時刻 t_n から次の事象が起こるまでの時間の確率分布 $F(t - t_n | H_{t_n})$ 及びその密度関数 $f(t - t_n | H_{t_n})$ との関係(Hazard 方程式)が得られ

$$(3.2) \quad \lambda(t|H_t) = f(t - t_n | H_{t_n}) / \{1 - F(t - t_n | H_{t_n})\}$$

となる. これを解くと隣り合う事象の時間間隔の分布が得られる.

3.3 点過程の尤度計算とシミュレーション法

特に, ヴェア・ジョーンズの講義でモデルの具体例として紹介された, 点過程の「自己回帰モデル」ともいえるべき, ホークス (Hawkes, 1971) の自己励起過程(self-exciting process)が注目された. このモデルの条件付き強度関数

$$(3.3) \quad \lambda(t|H_t) = \mu + \sum_{t_i < t} g(t - t_i) = \mu + \int_0^t g(t - s) dN_s$$

が, 事象の過去の発生時間の線形回帰形式になっているからである. 時系列研究グループ(ヴェ

ア・ジョーンズは Akaike School と呼んでいた)が、その後直ちに取り組んだのは、ホークス点過程の応答関数(response function) $g(\cdot)$ を AIC と最尤法によって推定することであった。

情報理論の分野では、条件付き強度関数とその尤度関数が理論的に導かれたばかりで、これによって最尤推定値(MLE)を数値的に求めることは当時挑戦するに足る課題であった。ヴェア・ジョーンズの助言で、赤池の第5研究部の最初の部下であった尾崎統(Ozaki, 1979)が、実行したのは $g(\cdot)$ を指数関数とし、(3.3)式でシミュレーションしたホークス過程の標本データから、対数尤度の最大化を Davidon-Fletcher-Powell 法のような効率的な準ニュートン法によって、計算し係数の MLE を得たことである。

このことに触発されて筆者が取り組んだのは3課題あった。以下の2課題は学術誌に掲載後、合わせて学位論文として提出した。もう一つの課題は次節で述べる。

(i) 上述のような、ホークス型モデルを含む、条件付き強度関数で特徴付けられた点過程の MLE や最大尤度の漸近理論。すなわち MLE の収束や誤差分布、尤度比統計量の漸近分布をマルチンゲールのエルゴード性や中心極限定理によって証明した。

(ii) 条件付き強度関数を直裁的に使った点過程のシミュレーション法の提案。尾崎が実行したシミュレーション法は Hazard 方程式(3.2)をニュートン法で反復して解くものであるが、多変量の場合など条件付き強度関数が複雑になると解が安定して求まらない。そこで筆者は Lewis and Shedler (1979)の間引き法(thinning method)に着目した。これは非定常ポアソン過程のシミュレーション法で、任意の密度関数から標本を生成する従来の rejection sampling 法と本質的に同じ方法である。筆者は間引き法を一般化し、多変量(マルチ・チャンネル)などにも拡張した。この正当性をマルチンゲール理論で証明し、多数の点過程モデルのシミュレーションの実例を示し、尤度比検定で正確さを実証した(Ogata, 1981)。この方法は現在、多分野で採用されている。

3.4 因果関係分析のための点過程モデリング

3つ目の課題は、除群法に代わるものとして、地震の統発効果を含む、以下の解析モデルの開発であった。すなわち、トレンド、季節性などの周期的な要素の解析、および他の外的データからの誘発効果の存否を確かめる因果関係の解析である。とくに因果関係の量的解析は、多くの科学分野でますます必要となっている。従来から一応、点過程系列間の相互相関統計量(cross Palm intensity; 尾形, 1993, 3.3 節)を使用することができるが、多くの統発事象を伴う場合や非定常性を含む場合から因果関係を定量的に導くことは困難である。

したがって課題はホークス型モデルを一般化して、トレンド、周期性、および因果関係を分析するモデル

$$(3.4) \quad \lambda(t|H_t, F_t) = \mu + f(t) + C(t; T_0) + \sum_{t_i < t} g(t - t_i | M_i) + \sum_{u_j < t} h(t - u_j) \xi(u_j)$$

を開発し実用化することだった。ここで H_t は解析対象の点過程の発生時間の履歴で、 F_t は外部入力データの時刻 u_j とマーク $\xi(u)$ を表す。時刻 u_j は等間隔でも不規則間隔でも良い。これで、日本や世界各地の地震活動の季節性や、地域間の地震活動の因果関係を導くことができた(尾形, 1993, 3.1, 3.3 節, 統計数理 63 尾形論文 8, 9 節参照)。(3.4)式右辺の第4項は、後ほど次節で述べる ETAS モデルを含む形にもなっている(例えば Kumazawa et al., 2016)。

このアプローチは AIC を採用し、誘発候補データの地震活動への因果性を調べる方法の原理である。今日まで、様々な地震活動の異常に基づいた大地震の警戒(alarm)情報が、電子メールなどで定期的に報知されているが、これらの有効性は議論の余地がある(Jordan et al., 2011)。そのような警戒情報には、客観的な有意性の評価が可能である(統計数理 63 庄・尾形論文)。他

方、警戒情報が確率予測として採用されるには、対象地域の標準的地震活動度と比べた優位性を示すために、警戒情報との因果関係の有意性および確率利得が、提供されるべきである。ここで「確率利得」とは「大地震の確率が平常時に比べて何倍高くなるのか」という量である。因果関係モデル(3.4)は、異常なデータや大きな地震との間の前兆性の統計的関係を調べるために使用することができる(統計数理 63 尾形論文 8 節参照)。

3.5 ETAS モデル

このモデルが考案される頃、統計数理研究所は大学共同利用研究所として改組された。筆者は「数理地震学」の共同研究会や複雑系・フラクタルの学習会を主催し、自己相似性やフラクタル次元(点群の集中度)の最尤法など、地震活動の特徴付けの研究(尾形, 1993, 5.1, 8.1 節; Ogata and Katsura, 1991)を行ったが、それらの単なる解釈に終始せず、震源データから直截的に地震活動を予測する実用的な点過程モデルを模索した。手掛りの多くは前節で述べたような、宇津によって探求・確立された余震の経験則であり、これらに基づいて、一定のマグニチュード M_c 以上の地震発生を予測するために epidemic-type aftershock sequence (ETAS) モデル

$$(3.5) \quad \lambda(t|H_t) = \mu + \sum_{t_i < t} \nu(t - t_i) e^{\alpha(M_i - M_c)}$$

を求めた。ここで H_t は過去の地震発生時とマグニチュードの系列の履歴データである。 $\nu(t)$ は大森・宇津の式(2.1)で、それに掛け算しているマグニチュードの指数関数の大小によって、大きい地震には多くの余震が誘発され、小さい地震もそれなりに余震を誘発する。時刻 t での条件付き強度は、線形の重ね合わせである。パラメータ α 値はマグニチュード効率で、大きければ典型的な本震・余震型の地震系列になり、小さければ群発地震のようになる。パラメータ μ は常時地震活動率と呼び、その地域特有の地震発生の強さを意味する。ただ、常時活動の一部は昔の地震の余震活動かもしれない。この様に、規模、常時発生率、余震の起き方など、それぞれの地域の特徴が定量化される。ETAS モデルの 5 つのパラメータ (μ, K, c, α, p) は最尤法で求められる。ETAS モデルは、シミュレーションで各地の標準的な地震活動を予測するのに良く使われている。

条件付き強度関数(3.5)の一定の期間の積分で、平均の地震数が与えられる。このことに基づいて、ある地域において或る期間に地震の数が ETAS の予測に対して有意に少なくなる現象に注目した。これを「相対的静穏化(relative quiescence)」という(尾形, 1993, 7.2, 7.3 節; 実例図としては <http://www.ism.ac.jp/editsec/toukei/pdf/45-1-139.pdf> 参照; 最終閲覧日 2019 年 4 月 5 日)。ある地域で地震活動が静穏化しているように見えても、それは単に以前の地震の余震活動の終息を意味するものかもしれない。ある程度の活動があっても、本来は以前の地震の余震活動がもっと盛んであるべきなのに、それに比べ、何らかの理由で不活発なのかもしれない。ETAS モデルを採用すると、このような効果を計算に入れて、減衰する余震活動でも、期待される活動レベルを基準に相対的な静穏化を論じることができる。この点で、従来の除群に基づく静穏化と違う概念である。同様に「相対的活発化」が定義できる (Ogata, 2005, 2007, 2011a; Kumazawa and Ogata, 2013; Kumazawa et al., 2010)。このような診断解析的な応用は多数例あり、下記 4.1 節で述べる。

3.6 階層時空間 ETAS モデル

「時空間 ETAS モデル」は ETAS モデル(3.5)を一般化して、地震の位置(震央)の変数を考慮したモデルで、リアルタイムで場所も地震の群の広がりも予測する (Ogata, 1998)。これによると今どこで危険度が高く、どこの場所が、これまでどの程度活発だったか一目瞭然である。こ

うしたモデルの様々な変型版が「地震予知可能性の研究のための共同研究」(Collaboratory for the Study of Earthquake Predictability; CSEP; <http://www.cseptesting.org/>; 最終閲覧日 2019 年 4 月 5 日)の各国のテストセンターに於いて、現在まで 10 年以上予測実験が継続中である。さらに、応力変化や発震機構など地殻内情報による追加情報をもとに時空間 ETAS モデルより良い適合度を目指す余震メカニズムの研究が盛んに追求されているが、今のところリアルタイム予測に結びついていない。

階層的時空間(hierarchical space-time)ETAS(HIST-ETAS)モデルは、地域の特徴を詳細に求めるために、パラメータも位置に依存させるものである(Ogata et al., 2003; Ogata, 2011b)。これらは隣接する地震を結んだドロネー 3 角形分割上の部分的線形関数で表現され、特に地震集中域では高精度な変化が表現できる。しかし HIST-ETAS モデルの総パラメータ係数は、データサイズの数倍となる。そこで、経験的ベイズ法と赤池ベイズ情報量規準を用いて、パラメータ係数間の制約の最適な事前分布を求める必要がある。

HIST-ETAS モデルの常時地震活動率 $\mu(x, y)$ の高い地域は、GPS で求めたストレス蓄積率変化の大きい地域と良く対応し、歴史的被害地震の場所もよく説明している。また、今世紀になってからの大地震の発生場所をよく言い当てており、将来の大きな地震の場所の確率予測の基礎となりうる(統計数理 63 尾形論文; 尾形, 2008, 2017)。

4. 点過程の診断解析と地震の物理

4.1 相対的静穏化と後続の大地震との関係

近傍で大地震が連発するのは珍しくない。日本に於ける各々の大地震について、その後起きた全ての大地震との時間差と距離を重ね合わせプロットすると、直後に近傍で別の大地震が起きる頻度は平時より数倍高くなり、離れるにしたがって逆比例することが分かる(Ogata, 2017b)。これには理由がある。断層が急激にずれるので周辺の応力が増加し、近傍の断層がずれ易くなるからである(統計数理 63 尾形 5 節参照)。

また著者は日本に於ける過去の大地震の余震列を 76 例調べた(Ogata, 2001b; 詳細記録は Ogata, 2001; <http://bemlar.ism.ac.jp/ogata/JGR01supplement/> 参照, 最終閲覧日 2019 年 4 月 9 日)。そのうち、相対的静穏化は 45% ほどの余震列で認められ特に珍しくないが、近辺で別の大きな地震が起こる確率は、前述の場合より更に 3~4 倍高くなる(Ogata, 2017b)。相対的静穏化の原因の一つとして、余震域で破壊応力の低下(stress shadow)をもたらすような「ゆっくりすべり(slow slip)」が近傍の断層内で起きる場合が考えられる。その際、そのような slow slip がきっかけとなって、新たな断層の破壊が誘発される確率があると考えられる。このような先行 slow slip を数例、GPS 測地データの異常変化の解析で確認することができた(Kumazawa et al., 2010; Ogata, 2007, 2010b, 2011a)が、更なる事例を数多く解析し、リアルタイムでモニターできるようになることが望まれる。

1990 年頃から現在まで、大地震と余震活動や広域地震活動の前駆的な、相対的静穏化や相対的活性化について、筆者は地震予知連絡会で報告した(地震予知連絡会報, 国土地理院; <http://cais.gsi.go.jp/YOCHIREN/report.html>; 最終閲覧日 2019 年 4 月 5 日; 報告内容は「統計数理研究所」として検索)。さらにそれらの大半の数十に亘る論文を地震関連学術誌(統計数理 63 尾形 7 節参照)に掲載したが、それらは一例を除いて事後報告である。したがって、今後 ETAS で相対的静穏化・活発化現象や、slow slip を地震の予測に繋げる為には、地震活動や地殻変動の異常検出を事前に探索し、効率的に検出できる統計的モデルや方法の研究を勧めたい。そのヒントの一部を以下に記す。

4.2 群発地震

プレート境界における群発地震と slow slip との関係については、数例の報告がある(たとえば Nishikawa and Ide, 2018). 従来から、群発的な地震活動は多様であることが指摘されている(宇津, 1970). 小さい α 値の ETAS モデルで表現できる群発地震(第 2 种群発地震)もあるが、火山性の群発地震のように ETAS モデルで良く表現できないものもある. これは(3.5)式の μ のパラメータが時間とともに変化する非定常 ETAS モデルが良い適合を示す(Kumazawa et al., 2010, 2016, 2017; Kumazawa and Ogata, 2013, 2014). パラメータ μ の時間変化は、マグマや熱水などの流体の断層系貫入に起因する固着断層の弱化と地震発生率との間に、量的な物理的関係を提供する可能性がある(統計数理 63 熊沢論文 4 節, 尾形論文 9 節参照).

4.3 余震の時空間診断解析

余震活動の大森・宇津式(2.1)の積分曲線で時間を伸縮変換し、余震群内の時空間的分布を詳しく見ると、余震域全域で一様な経緯を示す場合もあるが、各部分領域で異なったパターンのこともある(Ogata, 2010a). 特に、大きな余震が発生する前には局所の静穏化パターンを示すことが多い. このようなパターン異常は局所的な slow slip などが余震域の中でも起きている可能性がある.

5. 前震

5.1 マグニチュード系列について

地震のマグニチュード系列に関する研究は未だに発展途上である. 伝統的な標準モデル(Gutenberg-Richter 則; Gutenberg and Richter, 1944)は、一定のマグニチュード M_c 以上の指数分布

$$(5.1) \quad F(M|b) = 1 - 10^{-b(M-M_c)} = 1 - e^{-\beta(M-M_c)}; M \geq M_c, \beta = b \log_{10} e$$

に則り、独立同分布である. 係数 b は最尤法で決めると $M_i - M_c$ の標本平均の逆数に比例したものとなり、日本地域基準値では凡そ $b = 0.9$ である. ただし、気象庁などの多くの地震カタログの場合、マグニチュード値は四捨五入して 0.1 刻みで与えられているので、不偏推定の為には、下限のマグニチュードは 0.05 を差し引いたものにする. CSEP の予測実験に提出されているモデルは殆んど(5.1)を仮定している.

しかし、マグニチュードの時系列は、地震の発生パターンなどの履歴データに依存している可能性がある. また、マグニチュード分布の裾の部分を高めたものも考えられる. 例えばアメリカ合衆国地質調査所(USGS)の第三次カルフォルニア地震予測計画 UCERF3- ETAS (Field et al., 2017)では、サンアンドレアス断層などの近傍活断層への誘発可能性を含む、固有地震型マグニチュード分布を提案している.

様々な観測異常現象と大地震との因果関係をマグニチュード予測に組み込むことが考えられる. たとえば、次節で議論する「前震」の事前識別によって有効な予測を構成できる(Ogata, 2017b; Ogata et al., 2018). その予測の優劣は標準 Gutenberg-Richter 分布モデル(5.1)の予測結果に比べた対数尤度比基準(情報利得)で検証できる. このような研究は現時点で他に未だ余り見られない.

5.2 前震の統計的特徴

本節では、先ず本震を基準にした前震の相対的な研究(尾形, 1993, 8.3 節)を紹介する. これまで様々な前震の定義があるが、次節で予測を考慮するために、以下の定義で考える.

全ての地震($M \geq 4.0$)に関して、或る時空間的な距離より近いものを全て繋いで、孤立した地震や地震の群れに分離する(Single-link 法)。各群れの中の最大の地震を本震と呼ぶ。本震と、群内で先行する最大の地震とのマグニチュード差が、0.45 以上の場合、先行する地震を「前震」と呼ぶ。このマグニチュードの差は常識的に、なるべく大きくしたいが、そうすると全体の地震群に占める前震群の割合が少なくなるので、統計的な議論のために便宜上決めた。

余震に比べて前震は希少で、事例ごとに起こり方の相違が著しいので、全ての前震の時刻や位置座標を、本震が原点になるように移動して重ね合わせた統計的な特徴を述べると以下のようになる。(i)その時間頻度は、過去に向かって大森・宇津の(2.1)式に則っている。(ii)本震の位置から離れるとベキ則で頻度が減る。(i)と(ii)は言い換えると本震に集中することを示すが、それぞれの座標軸での周辺分布が一樣になるように座標変換をすると、(iii)あたかもドーナツ型が本震に向かって収束する様相になる。これは前震の本震へのマイグレーションを意味する(Ogata et al., 1995)。

他方、Gutenberg-Richter 独立分布でのマグニチュード系列から時空間 ETAS モデルでシミュレーションされた合成時空間データの前震からでも、重ね合わせの性質(i)-(iii)が再現できる(Helmstetter et al., 2003; Ogata and Katsura, 2014)。時空間 ETAS モデルは、単なる余震活動の重ね合わせなのに、なぜ前震の性質を再現するのか、この結果は、前震を物理学的に追究している研究者には意外に思われている。筆者は前震の物理的知見に基づく情報を持ち合わせていないが、この議論を深める鍵となるデータはマグニチュードの時系列である。その周辺分布は何れも指数分布(Gutenberg-Richter 則)であるが、その違いは、自然界の地震(気象庁カタログ)には自己相関があるのに、前述の合成カタログのものは独立である。

そこで、筆者らは、一方で、気象庁カタログのマグニチュード時系列を逐次 ETAS に入力し、発生時刻と震央座標を作った第 1 種の合成カタログを作製し、他方では、気象庁カタログのマグニチュードをブートストラップ(boot strapping)して独立性を確保し、それらを ETAS に入力することによって第 2 種の合成カタログを作製した。それぞれの合成カタログは、いずれも時空間的特徴(i)-(iii)の性質を満たすが、定量的には有意な違いがある。さらに、伝統的に知られている前震と余震の b 値の大小の関係については、本震の大きさなどの条件によって異なった結果を導くことを示した(Ogata and Katsura, 2014 参照)。

5.3 前震の事前識別

ある地域で新規の地震群が始まったとき、それらが「前震」である確率をリアルタイムで求めたい。すなわち、群れの中で、それまでの最大マグニチュードより 0.45 以上飛びぬけた、格段に大きな地震が起こりうる確率である。先ず、群れの先頭の地震の位置座標、群内の各々の地震の時間間隔、震央間距離、およびマグニチュード差を計算する。それらの履歴データを説明変数とし、最適な logit 関数のモデルによって、現在進行中の地震群が前震系列である確率、すなわち 1 か月以内にマグニチュードが、これまでより 0.45 以上の大きな地震が出現する確率を計算し予測する(統計数理 63 尾形論文 10 節参照)。

筆者らがそのような計算式を提案してから経過した 15 年間の予測の結果を検証した(Ogata and Katsura, 2012)。すなわち、経験的な一定確率での前震予測に対する対数尤度比基準(情報利得)によって比較した結果、予測成績が上回るということになった。特に、本震が大きい地震の群(例えば $M_0 \geq 6.5$)に限ると、識別予測の結果は極めて明瞭であった。

6. 多重確率予測式による大地震の予測

無情報の基での大地震の確率(永年確率)は極めて小さいが、何らかの異常現象が現れると、

その予測確率(確率利得)は相対的に増えることが期待される。しかし、前兆性や切迫性の識別には不確定さが伴う。なので、我々は経時観測データの異常を明瞭に定義し、それらが前兆として大地震に至る確率(「適中率, hit rate」)や確率利得を求める作業をする必要がある。また地震の見逃し(不意打ち)を避ける為には、低い適中率でもできるだけ多くの異常現象を発掘する必要がある。

大地震に至る高い適中率や確率利得を出すことは困難であろうが、それでも独立な幾つかの異常の予測因子が重なれば、「多重確率予測公式」(統計数理 63 尾形論文)によって、大地震予測の確率が実用的なものまで高まることもある。独立性が保証できない場合は、多重確率予測公式を一般化したロジット(logit)関数の展開モデルで算出し、AIC 比較でモデルの期待予測を高めることができる。

現状では、独立性を担保するような短期・中期・長期の予測確率を組み合わせるのが有望な策であり、先行研究では、宇津による 1978 年の伊豆大島近海地震、および安芸らによる中国の海城地震や唐山地震などに関して計算されている。最近著者も、M7.3 熊本地震に関して多重確率予測公式で試算した結果と予測確率のバラツキを示した(Ogata, 2017b)。

7. おわりに

何時だったか思い出せないが「統計モデルはデータ解析の望遠鏡や顕微鏡である」という赤池の言がある。古来、科学的発見を進める原動力は、適切な問題意識の上に、望遠鏡や顕微鏡などの発見や改良のように、方法論的革新が大きく絡んでいることはよく知られている。また、科学的仮説の実証は予測の結果によって決着がつく。同様に、統計科学の研究対象は誠に複雑系ではあるが、適切な統計モデルは予測をそれなりに全うするだけでなく、モデルを作るにあたって考慮されていなかった「異常」や「想定外」とも呼ばれる科学的な新事実を、現実のデータから露出することもある。統計モデルや、それに基づくグラフや画像による表現は、辛うじて見えるものや見えないものをはっきり見せ、新知見を導く科学的方法としての役割を果たしうる。

著者は地震統計にもとづく多くの経験則と、地震学の物理的仮説を、統計的点過程モデルとして表現して、統計的方法の有用性を示すように心がけてきた。たとえば ETAS モデルは地震のデータベースから短期発生率を予測するために、余震活動の経験則に基づいて構成されたが、地震活動の微妙な異常を検出する物差しとして使える可能性も提供する。地震活動研究と密接に関係する統計地震学の展開と、それらの予測における意義をお伝えできたなら幸いである。

参 考 文 献

- Adamopoulos, L. (1976). Cluster models for earthquakes: Regional comparisons, *Mathematical Geology*, **8**, 463-475.
- 安芸敬一 (1956). 統計地震学の現状, *地震*, **8**(4), 205-228, doi: 10.4294/zisin1948.8.4_205.
- Engdahl, E. R. and Villaseñor, A. (2002). Global seismicity: 1900-1999, *International Handbook of Earthquake and Engineering Seismology* (eds. H. K. Lee, H. Kanamori, P. C. Jennings, et al.), Part A, Academic Press, Amsterdam, <https://earthquake.usgs.gov/data/centennial/centennial.pdf>.
- Field, E. H., Milner, K. R., Hardebeck, J. L., Page, M., van der Elst, N., Jordan, T. H., Michael, A. J., Shaw, B. E. and Werner, M. J. (2017). A spatiotemporal clustering model for the third uniform California earthquake rupture forecast (UCERF3-ETAS): Toward an operational earthquake forecast, *Bulletin of the Seismological Society of America*, **107**(3) 1049-1081, doi:10.1785/

0120160173.

- Guo, Z. and Ogata, Y. (1997). Statistical relations between the parameters of aftershocks in time, space and magnitude, *Journal of Geophysical Research*, **102**(B2), 2857-2873.
- Gutenberg, B. and Richter, C. F. (1944). Frequency of earthquakes in California, *Bulletin of the Seismological Society of America*, **34**, 185-188.
- Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes, *Journal of the Royal Statistical Society, Series B*, **33**, 438-443.
- Helmstetter, A., Sornette, D. and Grasso, J. R. (2003). Mainshocks are aftershocks of conditional foreshocks: How do foreshock statistical properties emerge from aftershock laws, *Journal of Geophysical Research*, **108**(B1), 2046, doi:10.1029/2002JB001991.
- Jordan, T. H., Chen, Y.-T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., Papadopoulos, G., Sobolev, G., Yamaoka, K. and Zschau, J. (2011). Operational earthquake forecasting: State of knowledge and guidelines for implementation, Final report of the International Commission on Earthquake Forecasting for Civil Protection, *Annals of Geophysics*, **54**(4), 315-391, doi:10.4401/ag-5350, <http://www.annalsofgeophysics.eu/index.php/annals/article/view/5350/5371>.
- 熊澤貴雄 (2015). 地震活動の異常性とモデリング, *統計数理*, **63**(1), 45-64.
- Kumazawa, T. and Ogata, Y. (2013). Quantitative description of induced seismic activity before and after the 2011 Tohoku-Oki earthquake by non-stationary ETAS models, *Journal of Geophysical Research*, **118**(12), 6165-6182, doi:10.1002/2013JB010259.
- Kumazawa, T. and Ogata, Y. (2014). Nonstationary ETAS models for nonstandard earthquakes, *Annals of Applied Statistics*, **8**(3), 1825-1852, doi:10.1214/14-AOAS759, <http://projecteuclid.org/euclid.aoas/1414091236>.
- Kumazawa, T., Ogata, Y. and Toda, S. (2010). Precursory seismic anomalies and transient crustal deformation prior to the 2008 Mw = 6.9 Iwate-Miyagi Nairiku, Japan, earthquake, *Journal of Geophysical Research*, **115**, B10312, doi:10.1029/2010JB007567.
- Kumazawa, T., Ogata, Y., Kimura, K., Maeda, K. and Kobayashi, A. (2016). Background rates of swarm earthquakes that are synchronized with volumetric strain changes, *Earth and Planetary Science Letters*, **442**, 51-60, doi:10.1016/j.epsl.2016.02.049.
- Kumazawa, T., Ogata, Y. and Tsuruoka, H. (2017). Measuring seismicity diversity and anomalies by point process models: Case studies before and after the 2016 Kumamoto earthquakes in Kyushu, Japan, *Earth, Planets and Space*, **69**, Article 169.
- Lewis, P. A. W. (ed.) (1972). *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, Wiley, New York.
- Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning, *Naval Research Logistics Quarterly*, **26**(3), 403-413, <https://apps.dtic.mil/dtic/tr/fulltext/u2/a059904.pdf>.
- Nishikawa, T. and Ide, S. (2018). Recurring slow slip events and earthquake nucleation in the source region of the M7 Ibaraki-Oki earthquakes revealed by earthquake swarm and foreshock activity, *Journal of Geophysical Research*, doi:10.1029/2018JB015642, https://drive.google.com/file/d/16Ah30kGNuq3za038b2r_I-LNLSOvkGu5/view.
- Ogata, Y. (1981). On Lewis' simulation method for point processes, *IEEE Transaction of Information Theory*, **IT-27**, 23-31.
- Ogata, Y. (1983). Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure, *Journal of Physics of the Earth*, **31**, 115-124, https://www.jstage.jst.go.jp/article/jpe1952/31/2/31_2_115/_article/-char/en.
- 尾形良彦 (1993). 地震学とその周辺の地球学分野に於ける統計モデルと統計的手法, 日本統計学会誌,

- 22**(3), 413-463, doi: 10.11329/jjss1970.22.413.
- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences, *Annals of Institute of Statistical Mathematics*, **50**(2), 379-402.
- Ogata, Y. (2001a). Exploratory analysis of earthquake clusters by likelihood-based trigger models, Festschrift volume for Professor Vere-Jones, *Journal of Applied Probability*, **38A**, 202-212.
- Ogata, Y. (2001b). Increased probability of large earthquakes near aftershock regions with relative quiescence, *Journal of Geophysical Research*, **106**(B5), 8729-8744, doi: 10.1029/2000JB900400.
- Ogata, Y. (2005). Detection of anomalous seismicity as a stress change sensor, *Journal of Geophysical Research*, **110**(B5), B05S06, doi:10.1029/2004JB003245.
- Ogata, Y. (2007). Seismicity and geodetic anomalies in a wide area preceding the Niigata-Ken-Chuetsu earthquake of 23 October 2004, central Japan, *Journal of Geophysical Research*, **112**, B10301, doi:10.1029/2006JB004697.
- 尾形良彦 (2008). 最近 30 年の大地震発生と指定地域について, 地震予知連絡会会報, **79**(12-1), 623-625, http://cais.gsi.go.jp/YOCHIREN/report/kaihou79/12_01.pdf.
- Ogata, Y. (2010a). Space-time heterogeneity in aftershock activity, *Geophysical Journal International*, **181**(3), 1575-1592, doi:10.1111/j.1365-246X.2010.04542.x.
- Ogata, Y. (2010b). Anomalies of seismic activity and transient crustal deformations preceding the 2005 M7.0 earthquake west of Fukuoka, *Pure and Applied Geophysics*, **167**(8-9), doi:10.1007/s00024-010-0096-y.
- Ogata, Y. (2011a). Pre-seismic anomalies in seismicity and crustal deformation: Case studies of the 2007 Noto Hanto earthquake of M6.9 and the 2007 Chuetsu-oki earthquake of M6.8 after the 2004 Chuetsu earthquake of M6.8, *Geophysical Journal International*, **186**, doi:10.1111/j.1365-246X.2011.05033.x.
- Ogata, Y. (2011b). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity, *Earth, Planets and Space*, **63**(3), 217-229, doi:10.5047/eps.2010.09.001.
- Ogata, Y. (2013). A prospect of earthquake prediction research, *Statistical Science*, **28**, 521-541, doi:10.1214/13-STS439.
- 尾形良彦 (2015). 地震の確率予測の研究—その展望, 統計数理, **63**(1), 3-27.
- Ogata, Y. (2017a). Statistics of earthquake activity: Models and methods for earthquake predictability studies, *Annual Review of Earth and Planetary Sciences*, **45**, 497-527, doi:10.1146/annurev-earth-063016-015918.
- Ogata, Y. (2017b). Forecasting of a large earthquake: An outlook of the research, *Seismological Research Letters*, **88**(4), 1117-1126, doi:10.1785/0220170006.
- 尾形良彦 (2017). 日本列島内陸部の常時地震活動度について, 地震予知連絡会会報, **97**(9-12), http://cais.gsi.go.jp/YOCHIREN/report/kaihou97/1_3.pdf.
- Ogata, Y. (2018). Comment on “A review of self-exciting spatiotemporal point process and their applications” by Alex Reinhart, *Statistical Science*, **33**(3), 319-322, doi:10.1214/18-STS650, <https://projecteuclid.org/euclid.ss/1534147222>.
- Ogata, Y. and Abe, K. (1991). Some statistical features of the long-term variation of the global and regional seismic activity, *International Statistical Reviews*, **59**, 139-161, https://rmgsc.cr.usgs.gov/outgoing/threshold_articles/Ogata_Abe1991.pdf.
- Ogata, Y. and Katsura, K. (1991). Maximum likelihood estimates of the fractal dimension for random spatial patterns, *Biometrika*, **78**(3), 463-474.
- Ogata, Y. and Katsura, K. (2012). Prospective foreshock forecast experiment during the last 17 years, *Geophysical Journal International*, **191**(3), 1237-1244, doi:10.1111/j.1365-246X.2012.05645.x.
- Ogata, Y. and Katsura, K. (2014). Comparing foreshock characteristics and foreshock forecasting in observed and simulated earthquake catalogs, *Journal of Geophysical Research*, **119**(11), 8457-

- 8477, doi:10.1002/2014JB011250.
- Ogata, Y., Utsu, T. and Katsura, K. (1995). Statistical features of foreshocks in comparison with other earthquake clusters, *Geophysical Journal International*, **121**, 233-254.
- Ogata, Y., Katsura, K. and Tanemura, M. (2003). Modelling heterogeneous space-time occurrences of earthquakes and its residual analysis, *Applied Statistics (Journal of the Royal Statistical Society, Series C)*, **52**(4), 499-509.
- Ogata, Y., Katsura, K., Tsuruoka, H. and Hirata, N. (2018). Exploring magnitude forecasting of the next earthquake, *Seismological Research Letters*, doi:10.1785/0220180034.
- Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes, *Annals of the Institute of Statistical Mathematics*, **31**, 145-155.
- 庄建倉, 尾形良彦 (2015). 地震予測の評価法について, *統計数理*, **63**(1), 29-44.
- 宇津徳治 (1970). 5. 地震の時間的分布に関連する諸問題(その3): 余震, 前震, 群発地震の時間的性質, *北海道大学地球物理学研究報告*, **23**, 49-71, doi: 10.14943/gbh.23.49, <http://hdl.handle.net/2115/13973>.
- 宇津徳治 (1999). 『地震活動総覧』, 東大出版会, 東京.
- Utsu, T., Ogata, Y. and Matsu'ura, R. S. (1995). The centenary of the Omori formula for a decay law of the aftershock activity, *Journal of Physics of the Earth*, **43**, 1-33, https://www.jstage.jst.go.jp/article/jpe1952/43/1/43_1_1/_article/-char/en.
- Vere-Jones, D. (2006). The development of statistical seismology: A personal experience, *Tectonophysics*, **413**, 5-12.

The Development of Statistical Seismology: A Personal Experience and View

Yosihiko Ogata

Professor Emeritus, The Institute of Statistical Mathematics

I provide an overview of the development of statistical seismology in Japan and my research experience. Some focuses are placed on prediction of seismic activity by point process models, and statistical diagnostic analysis of anomalous seismic activities searching relations to physical phenomena.

チューブ法の理論・応用とその周辺

栗木 哲†

(受付 2018 年 12 月 28 日; 改訂 2019 年 3 月 20 日; 採択 3 月 26 日)

要 旨

チューブ法は, 正規確率場の最大値の上側裾確率を精度良く近似する積分幾何学的手法である. 同じ目的のための手法として, オイラー標数法がある. 本稿では, チューブ法とオイラー標数法の考え方を概観し, 統計学への応用, ならびに最近の発展について解説する.

キーワード: オイラー標数法, 同時信頼区間, どこでも効果, 射影追跡, 特異モデル, VBM データ解析.

1. はじめに

表題のチューブ法は, 正規確率場の最大値の上側裾確率を精度良く近似する積分幾何学的手法である. 同じ目的のための手法として, オイラー標数法がある. 本稿では, チューブ法とオイラー標数法の考え方を概観し, 統計学への応用, ならびに最近の発展について解説する.

チューブ法では, 集合の管状近傍(チューブ)の体積評価を通して確率分布を評価する. 統計学において, チューブの体積評価の概念が初めて現れたのは, Hotelling (1939) の非線形回帰モデルである:

$$y_i = \beta f(x_i, \theta) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n.$$

ここで x_i, y_i はそれぞれ説明変数, 目的変数, $\beta \in \mathbb{R}$ と $\theta \in \Theta \subset \mathbb{R}$ は未知パラメータ, f は θ を含む基底関数である. ここで帰無仮説 $H: \beta = 0$ の尤度比検定を考えると, $y = (y_1, \dots, y_n)^\top$, $f(\theta) = (f(x_1, \theta), \dots, f(x_n, \theta))^\top$ とおくと

$$(1.1) \quad \text{LRT} = \sup_{\theta \in \Theta} \left| \left\langle \frac{y}{\|y\|}, \frac{f(\theta)}{\|f(\theta)\|} \right\rangle \right|$$

が大きな値をとるとき棄却する検定となる. 帰無仮説 $H: \beta = 0$ のもとでは, $y \sim N(0, \sigma^2 I)$, $y/\|y\| \sim \text{Unif}(\mathbb{S}^{n-1})$ (n 次元ユークリッド空間の単位球面 \mathbb{S}^{n-1} 上の一様分布)となる. Hotelling (1939) は, 尤度比検定統計量 LRT の帰無分布を求めることと, \mathbb{S}^{n-1} において, 基準化回帰基底ベクトルの θ を動かした時の軌跡 $M = \{\pm f(\theta)/\|f(\theta)\| \in \mathbb{S}^{n-1} \mid \theta \in \Theta\}$ のまわりのチューブ(図 1)

$$\text{Tube}(M, \theta) = \{u \in \mathbb{S}^{n-1} \mid \text{dist}(u, M) \leq \theta\}, \quad \text{dist}(u, M) = \min_{v \in M} \cos \langle v, u \rangle \quad (\text{大円距離})$$

の体積を求めることは同値であることを, 以下の関係式のように指摘した:

$$\Pr(\text{LRT} \geq a) = \frac{\text{Vol}_{n-1}(\text{Tube}(M, \cos^{-1} a))}{\text{Vol}_{n-1}(\mathbb{S}^{n-1})}.$$

ここで Vol_{n-1} は $n-1$ 次元体積である.

† 統計数理研究所: 〒190-8562 東京都立川市緑町 10-3

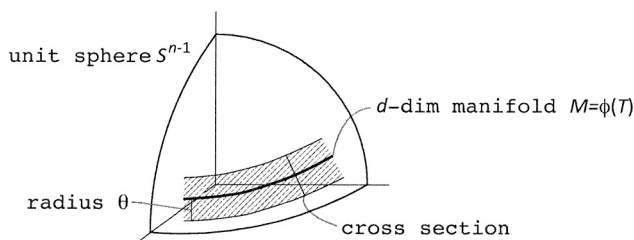


図 1. 球面チューブ.

Hotelling (1939) はさらに論文の中で M が 1 次元(曲線)の場合のチューブ体積公式を微分幾何の言葉で書き下している. この結果はプリンストン高等研究所の同僚であった H. Weyl によって M が一般次元閉多様体の場合に一般化され (Weyl, 1939), 現在では Hotelling-Weyl の定理とよばれている (丹野, 1976; 小林, 1997).

ところで, Hotelling の非線形回帰問題の尤度比検定統計量(1.1)は, 添字集合 θ の上で定義された確率場 $X(\theta) = \langle y/\|y\|, f(\theta)/\|f(\theta)\| \rangle$ の最大値 $\sup_{\theta \in \Theta} X(\theta)$ である. その視点より 1980 年代末から 90 年代にかけて, スタンフォード大で, チューブの体積と確率場の最大値分布に関するいくつかの研究が始まった (Knowles and Siegmund, 1989; Johansen and Johnstone, 1990; Sun, 1991, 1993). それがチューブ法 (tube method; volume-of-tube method) である.

なお近年 selective inference の名で, モデル選択後の条件付推論の方法論が, これもスタンフォード大の研究者を中心に発展している. そこではチューブ法におけるチューブ座標に沿って分解されたピボット統計量が利用される (Taylor et al., 2016 の Kac-Rice formula).

2. チューブ法の概要

$X(t), t \in T \subset \mathbb{R}^d$, を平均 0, 分散 1 のガウス確率場で, そのサンプルパスが滑らかであるものとする. チューブ法は, その最大値の上側裾確率

$$\Pr \left(\sup_{t \in T} X(t) \geq a \right) \quad (a \text{ が大きいとき})$$

を近似する方法である.

一般に平均 0 のガウス確率場 $X(t)$ は, 正則条件の下で以下のような Karhunen-Loève 展開 (KL 展開) を持つ.

$$X(t) = \sum_{i=1}^{\infty} \phi_i(t) \xi_i, \quad \xi_i \sim N(0, 1) \text{ i.i.d.}$$

ここで $\text{Var}(X(t)) = \sum \phi_i^2(t) = 1$ である. この表現は一意ではないが, 標準的な表現として, $\phi_i(\cdot)$ を正值カーネル $R(s, t) = \text{Cov}(X(s), X(t))$ が定義する再生核ヒルベルト空間の正規直交基底ととることができる. $R(s, t)$ の $s = t$ における微分可能性が高次であるほど, サンプルパスは滑らかとなる. チューブ法は, KL 展開が有限項 (n 項) で打ち切られている場合を対象とする. $\xi = (\xi_1, \dots, \xi_n)^\top \sim N_n(0, I_n)$, $\phi(t) = (\phi_1(t), \dots, \phi_n(t))^\top$, $M = \{\phi(t) \mid t \in T\}$ とおけば

$$X(t) = \langle \phi(t), \xi \rangle_{\mathbb{R}^n}, \quad \Pr \left(\sup_{t \in T} X(t) \geq a \right) = \Pr \left(\sup_{p \in M} \langle p, \xi \rangle_{\mathbb{R}^n} \geq a \right)$$

である. $\|\phi_i(t)\| = 1$ なので, $M \subset \mathbb{S}^{n-1}$ である. 以降確率場 $\{\langle p, \xi \rangle_{\mathbb{R}^n}\}_{p \in M}$ を扱う. このガウス確率場のモーメント構造は, $E[\langle p, \xi \rangle] = 0$, $\text{Var}[\langle p, \xi \rangle] = 1$, $\text{Cov}(\langle p, \xi \rangle, \langle q, \xi \rangle) = \langle p, q \rangle$ である.

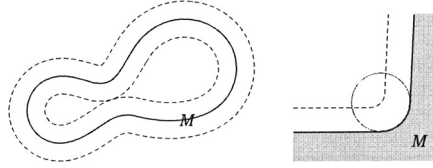


図 2. チューブの自己交差(左：大域的自己交差, 右：局所的自己交差).

M が \mathbb{S}^{n-1} の d 次元部分多様体であるとき, M が境界を持つ(閉多様体でない)場合を含めて, M のまわりのチューブの体積 $\text{Vol}_{n-1}(\text{Tube}(M, \theta))$ は, $0 \leq \theta \leq \exists \theta_c$ の範囲でベータ分布の上側確率の線形和で表すことができる (Takemura and Kuriki, 2002):

$$(2.1) \quad \frac{\text{Vol}_{n-1}(\text{Tube}(M, \theta))}{\text{Vol}_{n-1}(\mathbb{S}^{n-1})} = w_{d+1} \bar{B}_{\frac{d+1}{2}, \frac{n-d-1}{2}}(\cos^2 \theta) + w_d \bar{B}_{\frac{d}{2}, \frac{n-d}{2}}(\cos^2 \theta) + \cdots + w_1 \bar{B}_{\frac{1}{2}, \frac{n-1}{2}}(\cos^2 \theta).$$

ただし w_i は M のみに依存する定数で, 特に

$$w_{d+1} = \frac{\text{Vol}_d(M)}{\text{Vol}_d(\mathbb{S}^d)}, \quad w_d = \frac{\text{Vol}_{d-1}(\partial M)}{2\text{Vol}_d(\mathbb{S}^{d-1})},$$

$\bar{B}_{a,b}$ はパラメータ (a, b) のベータ分布の上側確率である. 定数 θ_c は M のみに依存する正定数で, 本節の最後で定義する.

ガウス確率ベクトル $\xi \sim N(0, I_n)$ の「長さ」 $\|\xi\|$ と「方位」 $\xi/\|\xi\|$ は独立に, それぞれ χ_n 分布(自由度 n のカイ 2 乗分布の平方根)と球面上一様分布 $\text{Unif}(\mathbb{S}^{n-1})$ に従うことから,

$$(2.2) \quad \Pr\left(\sup_{p \in M} \langle p, \xi \rangle_{\mathbb{R}^n} \geq a\right) = E\left[\Pr\left(\sup_{p \in M} \left\langle p, \frac{\xi}{\|\xi\|} \right\rangle_{\mathbb{R}^n} \geq \frac{a}{\|\xi\|} \mid \|\xi\|\right)\right] \\ = E\left[\frac{\text{Vol}_{n-1}\left(M, \cos^{-1}\left(\frac{a}{\|\xi\|}\right)\right)}{\text{Vol}_{n-1}(\mathbb{S}^{n-1})}\right] \\ \approx w_{d+1} \bar{G}_{d+1}(a^2) + w_d \bar{G}_d(a^2) + \cdots + w_1 \bar{G}_1(a^2).$$

ここで \bar{G}_ν は自由度 ν のカイ 2 乗分布の上側確率である. 最後の式は恒等式

$$E\left[\bar{B}_{\frac{\nu}{2}, \frac{n-\nu}{2}}\left(\frac{a}{\|\xi\|}\right)\right] = \bar{G}_\nu(a^2)$$

とチューブ体積公式 (2.1) の組合せで得られるが, (2.1) が θ が小さい ($\theta \leq \theta_c$) ときのみなりたっていたことに対応して, (2.2) は a が大きいときになりたつ近似式である. その誤差は $a \rightarrow \infty$ のとき $O(a^{n-2} e^{-\frac{1}{2}(1+\tan^2 \theta_c) a^2})$ である (Kuriki and Takemura, 2001).

誤差を定めている定数 θ_c はチューブの臨界半径とよばれ, 以下で定義される.

$$\theta_c = \sup\{\theta \mid \text{Tube}(M, \theta) \text{ は自己交差しない}\}.$$

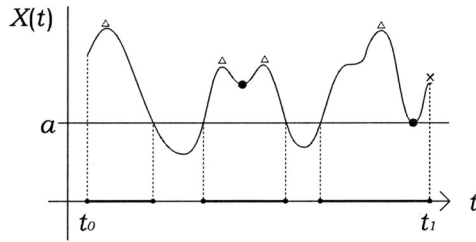
チューブの自己交差には, 大域的な自己交差と局所的な自己交差が考えられる (図 2).

M が与えられたとき, 臨界半径 θ_c を解析的ないしは数値的に評価することができる (栗木・竹村, 2008). 具体的な計算例は Kuriki and Takemura (2001) などにみられる.

3. オイラー標数法の概要

3.1 一次元オイラー標数法

チューブ法と同じ目的の手法に, オイラー標数法がある. ここではその方法を, 添字集合が



$$\chi(T_a) = 4(\triangle) - 2(\bullet) + 1(\times) = 3$$

図 3. エクスカーション集合とそのオイラー標数.

1次元の場合を中心に概観する.

$X(t)$ を, 添字集合 $T = [t_0, t_1]$ の上で定義された, 滑らかなサンプルパスを持つ実数値確率過程とする. ここで添字の一意性 " $X(s) = X(t)$ a.s. $\Rightarrow s = t$ " を仮定する.

与えられた閾値 a に対して, $X(t)$ が a 以上となるような添字の全体

$$T_a = \{t \in T \mid X(t) \geq a\}$$

をエクスカーション集合 (excursion set) という. エクスカーション集合 T_a のオイラー標数を $\chi(T_a)$ とおく. T_a は一次元集合なので, $\chi(T_a)$ は T_a の連結成分の個数である. $X(t)$ とエクスカーション集合の例を図 3 に与える. この例では $\chi(T_a) = 3$ である.

エクスカーション集合の定義から, $X(t)$ が a 以上となることは T_a が空集合とならないことと同値で, さらにそれは T_a の連結成分の個数が 0 でないことと同値である. すなわち

$$\sup_{t \in T} X(t) \geq a \Leftrightarrow T_a \neq \emptyset \Leftrightarrow \chi(T_a) \geq 1$$

である. ここで a の値が大きいとき, $X(t)$ が 2 つの離れた時点で a を超える事象は稀であることが予想される. すなわち事象 $\chi(T_a) = k$ ($k \geq 2$) が起こる確率は小さいことが期待できる. この近似のもとでは,

$$(3.1) \quad \mathbb{1}\left\{\sup_{t \in T} X(t) \geq a\right\} - \chi(T_a) = -\sum_{k=2}^{\infty} (k-1) \mathbb{1}(\chi(T_a) = k)$$

が無視できる量となり, 両辺の期待値をとることにより

$$(3.2) \quad \Pr\left(\sup_{t \in T} X(t) \geq a\right) \approx E[\chi(T_a)] \quad (a \text{ が大きいとき})$$

がなりたつことが期待される. この近似法を, オイラー標数法 (The Euler characteristic method) という (Adler and Hasofer, 1976; Worsley, 1995; Adler and Taylor, 2007). この方法は, 信号処理分野のライス公式 (Rice, 1944) の一般化に相当する. なお (3.1) は常に非正なので, 不等式

$$\Pr\left(\sup_{t \in T} X(t) \geq a\right) \leq E[\chi(T_a)] \quad (\text{すべての } a \text{ について})$$

がなりたつ (Naiman, 1986). この不等式は多重検定や同時信頼区間構成において, 近似が保守側であることを保証する.

次に (3.2) の右辺を具体的に評価する.

定義 1. 区間 $T = [t_0, t_1]$ で定義された滑らかな関数 $X(t)$ の、(拡張された)臨界点 (augmented critical point) とは

$$\dot{X}(t) = 0 \quad (t \in (t_0, t_1)), \quad < 0 \quad (t = t_0), \quad > 0 \quad (t = t_1)$$

をみたす点 $t = t^*$ である。(拡張された)臨界点集合を T^* とおく。

仮定 1. $a > 0$ を所与とする。 t^* を $X(t)$ の臨界点で $X(t^*) \geq a$ であるような点とする。確率 1 で $\dot{X}(t^*) \neq 0$ であること、また t^* は孤立点であることを仮定する。

チューブ法で扱う確率場では、本仮定がなりたつことは容易に確認できる。一般には Worsley (1995) の正則条件 C1-C3 の下でなりたつ。このとき、 $\mathbb{1}_{\{\cdot\}}$ を事象の指示関数とすると、図を描けば分かるように (図 3)

$$(3.3) \quad \chi(T_a) = \sum_{t \in T^* \cap \text{int}(T)} \mathbb{1}_{\{X(t) \geq a\}} \text{sgn}(-\ddot{X}(t)) + \sum_{t \in T^* \cap \partial T} \mathbb{1}_{\{X(t) \geq a\}}$$

がなりたつ。これが一次元版のモースの定理である。

以下では、式(3.3)の第 1 項、第 2 項の期待値を別々にとる。最初に第 1 項を考える。 $t^* \in \text{int}(T)$ とする。まず

$$\lim_{\varepsilon \rightarrow 0} \int_{\text{small region } (\ni t^*)} f(t) \frac{\mathbb{1}_{\{\dot{X}(t) \in (-\varepsilon, \varepsilon)\}}}{2\varepsilon} |\ddot{X}(t)| dt = f(t^*)$$

である。なぜならば、 $\dot{X}(t) = y$ とおき、 $\ddot{X}(t^*) > 0$ の場合は左辺は $\lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} f((\dot{X})^{-1}(y)) dy = f(t^*)$ 、 $\ddot{X}(t^*) < 0$ の場合は左辺は $\lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_{\varepsilon}^{-\varepsilon} f((\dot{X})^{-1}(y)) d(-y) = f(t^*)$ 。これより

$$\begin{aligned} E \left[\sum_{t \in T^* \cap \text{int}(T)} \mathbb{1}_{\{X(t) \geq a\}} \text{sgn}(-\ddot{X}(t)) \right] &= E \left[\lim_{\varepsilon \rightarrow 0} \int_{t_0}^{t_1} \mathbb{1}_{\{X(t) \geq a\}} \text{sgn}(-\ddot{X}(t)) \frac{\mathbb{1}_{\{\dot{X}(t) \in (-\varepsilon, \varepsilon)\}}}{2\varepsilon} |\ddot{X}(t)| dt \right] \\ &= \lim_{\varepsilon \rightarrow 0} \int_{t_0}^{t_1} E \left[\mathbb{1}_{\{X(t) \geq a\}} (-\ddot{X}(t)) \frac{\mathbb{1}_{\{\dot{X}(t) \in (-\varepsilon, \varepsilon)\}}}{2\varepsilon} \right] dt \\ &= \lim_{\varepsilon \rightarrow 0} \int_{t_0}^{t_1} \frac{E[\mathbb{1}_{\{X(t) \geq a\}} (-\ddot{X}(t)) \mathbb{1}_{\{\dot{X}(t) \in (-\varepsilon, \varepsilon)\}}]}{E[\mathbb{1}_{\{\dot{X}(t) \in (-\varepsilon, \varepsilon)\}}]} \\ &\quad \times \frac{E[\mathbb{1}_{\{\dot{X}(t) \in (-\varepsilon, \varepsilon)\}}]}{2\varepsilon} dt \\ &= \int_{t_0}^{t_1} E[\mathbb{1}_{\{X(t) \geq a\}} (-\ddot{X}(t)) | \dot{X}(t) = 0] \theta_{\dot{X}(t)}(0) dt, \end{aligned}$$

ただし

$$\theta_{\dot{X}(t)}(0) = \lim_{\varepsilon \rightarrow 0} \frac{E[\mathbb{1}_{\{\dot{X}(t) \in (-\varepsilon, \varepsilon)\}}]}{2\varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{\Pr(\dot{X}(t) \in (-\varepsilon, \varepsilon))}{2\varepsilon}$$

は $\dot{X}(t)$ の密度関数の 0 での値である。(3.3)の第 2 項の期待値評価は容易であり、

$$E[\chi(T_a)] = \int_{t_0}^{t_1} \rho(t) dt + \Pr(X(t_0) \geq a, \dot{X}(t_0) < 0) + \Pr(X(t_1) \geq a, \dot{X}(t_1) > 0),$$

$$\rho(t) = E[\mathbb{1}_{\{X(t) \geq a\}} (-\ddot{X}(t)) | \dot{X}(t) = 0] \theta_{\dot{X}(t)}(0)$$

を得る。これは Worsley (1995) の公式の一次元版である。

3.2 平均 0, 分散 1 のガウス場の場合

以下ではチューブ法と同じ設定 $X(t) = \phi(t)^\top \xi$, $\xi \sim N_n(0, I_n)$ で $\rho(t)$ の具体形をもとめる。 $\phi(t)$ を $n \times 1$ 単位ベクトルで t について C^3 級とする。また $\phi: T \rightarrow \mathbb{S}^{n-1}$ は単射とする。

$X(t)$ の 1 階, 2 階微分は $\dot{X}(t) = \dot{\phi}(t)^\top \xi$, $\ddot{X}(t) = \ddot{\phi}(t)^\top \xi$. これより t を固定して考えると $(X(t), \dot{X}(t), \ddot{X}(t))$ は 3 変量正規分布に従う。正規分布の標準的な計算から $g(t) = \|\dot{\phi}(t)\|^2$ とおくと $\theta_{\dot{X}(t)}(0) = 1/\sqrt{2\pi g(t)}$, $\rho(t) = g(t)^{\frac{1}{2}} e^{-\frac{1}{2}a^2}/(2\pi)$ が得られ,

$$E[\chi(T_a)] = \frac{1}{2\pi} \int_{t_0}^{t_1} g(t)^{\frac{1}{2}} dt \times \Pr(\chi_2^2 \geq a^2) + \frac{1}{2} \Pr(\chi_1^2 \geq a^2)$$

を得る。ここで $\int_{t_0}^{t_1} g(t)^{\frac{1}{2}} dt$ は $M = \{\phi(t) \mid t \in [t_0, t_1]\}$ の一次元体積であるので, (2.2) の $d = 1$ の場合に一致する。

添字集合が多次元 (d 次元) の場合もオイラー標数法は (3.2) で定義される。その場合の期待値 $E[\chi(T_a)]$ の評価は, 拡張された臨界点とモースの定理の一般次元への一般化が必要となる。

定理 1. (Takemura and Kuriki, 2002) $X(t)$, $t \in T \subset \mathbb{R}^d$ を平均 0, 分散 1 の滑らかなサンプルパスを持つガウス確率場とする。エクスカージョン集合 $T_a = \{t \in T \mid X(t) \geq a\}$ のオイラー標数の期待値は

$$(3.4) \quad E[\chi(T_a)] = \frac{\text{Vol}_d(T)}{\text{Vol}(\mathbb{S}^d)} \Pr(\chi_{d+1}^2 \geq a^2) + \frac{\text{Vol}_{d-1}(\partial T)}{2\text{Vol}(\mathbb{S}^{d-1})} \Pr(\chi_d^2 \geq a^2) + \dots$$

ここで \mathbb{S}^d は d 次元単位球面, 添字集合 T は計量

$$g_{ij} = \text{Cov} \left(\frac{\partial X(t)}{\partial t_i}, \frac{\partial X(t)}{\partial t_j} \right) = \left\langle \frac{\partial \phi(t)}{\partial t_i}, \frac{\partial \phi(t)}{\partial t_j} \right\rangle_{\mathbb{R}^n}$$

が付与されたリーマン多様体と考えている。

(3.4) は (2.2) と全く同じ式であり, オイラー標数法はチューブ法と同等であることが分かる。

4. 統計学への応用

4.1 同時信頼区間の構成

本節では, チューブ法の統計学への応用を紹介する。最初の例は Naiman (1986) による回帰モデルの同時信頼区間構成である。

観測値 $\{(x_i, y_i)\}$ に対して回帰モデル

$$y_i = b^\top f(x_i) + \varepsilon_i, \quad x_i \in \mathcal{X} \subset \mathbb{R}, \quad \varepsilon_i \sim N(0, 1) \text{ i.i.d.}$$

を想定する。ここで b は未知パラメータベクトル, $f(x)$ は所与の基底関数ベクトル f である。例えば $f(x) = (1, x, x^2, \dots)^\top$ とおくと多項式回帰モデルになる。簡単のため, 誤差項の分散は既知で 1 とする。

同時信頼区間とは, 真の回帰曲線 $\{(x, b^\top f(x)) \mid x \in \mathcal{X}\}$ を確率 $1 - \alpha$ 以上で含む集合である。その集合の取り方には任意性があるが, ここでは \hat{b} を b の最小二乗推定量 (OLS) とし,

$$\{(x, y) \mid y \in \hat{b}^\top f(x) \pm \text{s.d.}(\hat{b}^\top f(x)) c_\alpha, x \in \mathcal{X}\}$$

の形を考える。ここで $A \pm B$ は区間 $(A - B, A + B)$ であり, c_α は α に応じて定める定数である。情報行列の逆行列を $\Sigma = (\sum_i f(x_i) f(x_i)^\top)^{-1}$ とおくと, c_α は

$$1 - \alpha \leq \Pr(b^\top f(x) \in \widehat{b}^\top f(x) \pm \|\Sigma^{\frac{1}{2}} f(x)\| \cdot c_\alpha, \forall x \in \mathcal{X}) = \Pr\left(\max_{x \in \mathcal{X}} |\phi(x)^\top \xi| \leq c_\alpha\right),$$

ただし $\xi = \Sigma^{-\frac{1}{2}}(\widehat{b} - b)$, $\phi(x) = \Sigma^{\frac{1}{2}} f(x) / \|\Sigma^{\frac{1}{2}} f(x)\|$ をみたくし閾値として選ぶことができる。ここで $\xi \sim N_n(0, I_n)$, $\phi(x) \in \mathbb{S}^{n-1}$ なので、チューブ法が利用可能である。Naiman (1986) の不等式により、チューブ法で構成した同時信頼区間はわずかに保守側の近似となる。

また Lu and Kuriki (2017) は、複数の回帰曲線

$$\{(x, b_{(h)}^\top f(x)) \mid x \in \mathcal{X}\}, \quad h = 1, \dots, k$$

のコントラスト

$$\left\{ \sum_{h=1}^k c_{(h)} b_{(h)}^\top f(x) \mid x \in \mathcal{X}, (c_{(h)}) \in \mathcal{C} \right\}$$

の保守側同時信頼区間を与えている。ここで \mathcal{C} は $\sum_{h=1}^k c_{(h)} = 0$ をみたくし係数ベクトル $(c_{(h)}) = (c_{(1)}, \dots, c_{(k)})$ の全体である。この問題は複数の成長曲線の多重比較として現れる。

なお、回帰曲線の同時信頼区間構成法として、コーシー・シュワルツの不等式

$$\max_{x \in \mathcal{X}} |\phi(x)^\top \xi| \leq \|\phi(x)\| \cdot \|\xi\| = \|\xi\|$$

を用いる方法 (Scheffé 法) がある。本方法はラオ (1986) で紹介されていることからしばしば使われるが、全実軸を定義域とする単回帰モデル $(f(x) = (1, x)^\top, x \in \mathcal{X} = \mathbb{R})$ という特別の場合を除いてコーシー・シュワルツの不等式の上界はタイトにならないため、しばしば非常に保守的となり、使うべきではない。

4.2 Look-Elsewhere Effect (LEE, どこでも効果)

実験物理、特に新素粒子探索において、探索する領域を広くすると見せかけの新発見が起きることが知られている。この現象は、Look-Elsewhere Effect (LEE, どこでも効果) とよばれているが、統計学の言葉では検定の多重性に他ならない (Gross and Vitells, 2010)。

正規性のもとでは標準偏差の 2 倍が 95% に相当するため、通常、実験結果の図示には 2 シグマのエラーバーをつけることが慣習的に行われている。しかしながら素粒子探索においては 5 倍の偏差の検出をもって、新発見と認められる (村山, 2013)。この 5 シグマルールの正当化はオイラー標数法に基づいてなされている (van Dyk, 2014)。

4.3 ランダム行列の最大特異値

成分が独立に標準正規分布に従う $n \times p$ ランダム行列 $\Xi = (\xi_{ij}) \in \mathbb{R}^{n \times p}$, $\xi_{ij} \sim N(0, 1)$ i.i.d. を考える。その最大特異値 $\sigma_{\max}(\Xi)$ は $g \in \mathbb{R}^n, h \in \mathbb{R}^p$ を長さ 1 のベクトルとするときの双線形式 $g^\top \Xi h = \text{tr}((hg^\top)^\top \Xi)$ の最大値である。 hg^\top の値はノルムが 1 でランク 1 の $n \times p$ 行列全体を取りうるので

$$\sigma_{\max}(\Xi) = \max_{H \in M} \text{tr}(H^\top \Xi), \quad M = \{H \in \mathbb{R}^{n \times p} \mid \text{tr}(H^\top H) = 1, \text{rank}(H) = 1\}.$$

M は $\mathbb{R}^{n \times p}$ の単位球面 $\mathbb{S}^{n \times p - 1} = \{H \in \mathbb{R}^{n \times p} \mid \text{tr}(H^\top H) = 1\}$ の部分多様体なので、その分布はチューブ法で近似できる (Kuriki and Takemura, 2001, 2008b) :

$$\Pr(\sigma_{\max}(\Xi)^2 \geq a) \approx \frac{2^{p+n-2} \Gamma(\frac{p+1}{2}) \Gamma(\frac{n+1}{2})}{\sqrt{\pi}} \sum_{j=0}^{p-1} \left(-\frac{1}{2}\right)^j \frac{\Gamma(\frac{n+p-1-2j}{2})}{\Gamma(p-j) \Gamma(n-j) j!} \Pr(\chi_{n+p-1-2j}^2 \geq a).$$

$\sigma_{\max}(\Xi)^2$ は自由度 n の $p \times p$ ウィンシャート行列 $\Xi^\top \Xi$ の最大固有値である。

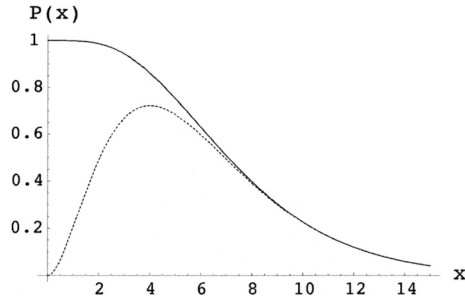


図 4. 2×2 自由度 5 Wishart 分布の最大固有値の上側確率
(実線：真値，点線：チューブ法).

図 4 に， $n = 5$ ， $p = 2$ の場合の σ_{\max}^2 の上側確率とそのチューブ法近似を図示する．なお最近の研究結果として，ランダム行列の手法を用いて，チューブ法近似の n と p が発散するときの極限を論じることができる (栗木, 2019)．

4.4 多次元正規性検定/探索的射影追跡の有意性

q 次元 i.i.d. ベクトル値観測系列 $x_t = (x_t^1, \dots, x_t^q) \in \mathbb{R}^q$ ($t = 1, \dots, n$) が得られているとする．このサンプルが q 次元正規分布に従っているかどうかを検定し，正規分布に従っていない場合にはどのような正規性からの乖離が見られるかを検出したいとする．

サンプルの方向ベクトル $h \in \mathbb{S}^{q-1}$ 方向の成分 $h^\top x_t \in \mathbb{R}$ ($t = 1, \dots, n$) の j 次サンプルキウムラントを $K_j(h)$ ($j = 2, 3, 4$) とする．これらから，射影データの歪度と尖度が

$$B_1(h) = \frac{K_3(h)}{K_2(h)^{\frac{3}{2}}}, \quad B_2(h) = \frac{K_4(h)}{K_2(h)^2}$$

と計算される．正規性の仮定のもとでは，全ての h に対して $K_3(h)$ と $K_4(h)$ の絶対値は小さくなるという性質に基づいて，Jones and Sibson (1987) は射影追跡のためのモーメント指標

$$I_{JS}(h) = \frac{1}{6}B_1(h)^2 + \frac{1}{24}B_2(h)^2$$

を提案した．その最大値 $\max_{h \in \mathbb{S}^{q-1}} I_{JS}(h)$ は多変量正規性の検定統計量として用いることができ，また最大値をとる h が，正規性の乖離がもっとも大きい方向と解釈することができる．

多変量正規性検定の有意水準は，帰無分布のもとでの上側裾確率から評価される：

$$\Pr\left(\max_{h \in \mathbb{S}^{q-1}} I_{JS}(h) \geq a \mid H_0\right), \quad H_0 : x_t \sim q \text{ 次元正規分布.}$$

この確率を，サンプル数 n が大きいときの漸近近似で考える． \mathbb{S}^{q-1} 上で定義される連続関数の空間 $C(\mathbb{S}^{q-1})$ の中心極限定理により，法収束

$$\sqrt{\frac{n}{6}}B_1(\cdot) \xrightarrow{d} Z_1(\cdot), \quad \sqrt{\frac{n}{24}}B_2(\cdot) \xrightarrow{d} Z_2(\cdot) \quad (\text{indep.})$$

が示される．ただし $Z_1(\cdot)$ と $Z_2(\cdot)$ は独立な \mathbb{S}^{q-1} 上正規確率場で，それらの平均は 0，分散 1，相関関数は $\text{corr}[Z_1(u), Z_1(v)] = (u^\top v)^3$ ， $\text{corr}[Z_2(u), Z_2(v)] = (u^\top v)^4$ である．チューブ法によって，極限分布

$$\max_{h \in \mathbb{S}^{q-1}} nI_{JS}(h) \xrightarrow{d} \max_{h \in \mathbb{S}^{q-1}} [Z_1(h)^2 + Z_2(h)^2]$$

の上側裾確率を近似できる．その際，左辺の最大値は確率場 $u_1 Z_1(h) + u_2 Z_2(h)$ ， $(h, u) \in \mathbb{S}^{q-1} \times \mathbb{S}^1$

の最大値であることを用いる (Kuriki and Takemura, 2008a).

4.5 特異モデル：有限混合モデルのコンポーネント数検定

正則な統計モデルでは、モデルの各点で、スコア関数が有限次元の線形空間を張る。そうでない点を特異点といい、特異点を含む統計モデルを「特異モデル」という (福水 他, 2004)。

特異モデルでは、その特異点を帰無仮説にする尤度比は、漸近カイ 2 乗性がなりたたない。一つの典型例は有限混合モデルである。最も簡単な例として、2 つの 2 項分布 $\text{Bin}(k, p)$ と $\text{Bin}(k, p)$ の混合分布を考える：

$$x_1, \dots, x_n \sim \alpha \text{Bin}(k, p) + (1 - \alpha) \text{Bin}(k, 1/2) \quad \text{i.i.d.}$$

ここで k は定数、 α, p は推定対象のパラメータである。これは $p = 1/2$ のとき α は推定不能になり、 $\alpha = 0$ のとき p が推定不能となるという特異モデルである。

このモデルにおいて、コンポーネント数の検定

$$\begin{aligned} H_0 : \alpha = 0 \quad \text{or} \quad p = 1/2 \quad (1 \text{ コンポーネント}), \\ H_1 : \alpha \neq 0 \quad \text{and} \quad p \neq 1/2 \quad (2 \text{ コンポーネント}) \end{aligned}$$

を考える。 x_t の取り得る値は $\{0, \dots, k\}$ であり、 x_t をダミー変数表示したベクトルは多項分布

$$(\delta_{x_t,0}, \dots, \delta_{x_t,k}) \sim \text{Mult}_{k+1}(1; q_0, \dots, q_k), \quad t = 1, \dots, n, \quad \text{i.i.d.}$$

に埋め込むことができる。ここで

$$q_i = \frac{1}{2^k} \binom{k}{i} \{ \alpha(1 + \phi)^i (1 - \phi)^{k-i} + (1 - \alpha) \} \quad (\phi = 2p - 1).$$

特異モデルの一般論より、 H_0 vs. H_1 の尤度比検定統計量は H_0 の下で正規確率場の最大値

$$\max\{0, \sup_{|\phi| \leq 1} X(\phi)\}^2, \quad X(\phi) = \frac{2^{k/2} \sum_{i=0}^k \{(1 + \phi)^i (1 - \phi)^{k-i} - 1\} \xi_i}{\sqrt{\sum_{i=0}^k \binom{k}{i} \{(1 + \phi)^i (1 - \phi)^{k-i} - 1\}^2}}, \quad \xi_i \sim N(0, 1) \quad \text{i.i.d.}$$

に分布収束する。この極限分布に基づいて、チューブ法により p 値の近似評価式が得られる。

$$\Pr(\sup_{|\phi| \leq 1} X(\phi) \geq a) \approx \frac{\text{Vol}(M)}{2\pi} \Pr(\chi_2^2 \geq a^2) + \frac{1}{2} \Pr(\chi_1^2 \geq a^2),$$

$$\text{Vol}(M) = 2 \int_0^1 \frac{\sqrt{k \{(1 + \phi^2)^k - 1 - k\phi^2\} (1 + \phi^2)^{k-2}}}{(1 + \phi^2)^k - 1} d\phi.$$

4.6 特異モデル：順序分割表解析

$a \times b$ 2 元表データ $N = (n_{ij})_{a \times b}$ を、 (i, j) セルの確率が p_{ij} の多項分布の実現値としてモデル化する。 p_{ij} のモデリングとして、対応分析モデル (CA モデル; 数量化 3 類)

$$p_{ij} = p_i \cdot p_j (1 + \phi \mu_i \nu_j)$$

と RC 連関モデル

$$\log p_{ij} = \alpha_i + \beta_j + \phi \mu_i \nu_j$$

がよく知られている (Kateri, 2014)。ここで行と列が順序尺度の場合、不等式制約

$$\mu_1 \leq \dots \leq \mu_a, \quad \nu_1 \leq \dots \leq \nu_b$$

を仮定したモデルを考える．このモデルのもとでは局所対数オッズ比は非負

$$\log\left(\frac{p_{i+1,j+1}p_{i,j}}{p_{i,j+1}p_{i+1,j}}\right) = \phi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j) \geq 0, \quad \forall i, j$$

となる．このモデルにおいて，独立性の仮説 $H: \phi = 0$ は， μ_i, ν_j が識別不能となるモデルの特異点である．この特異点を帰無仮説とする尤度比検定の極限帰無分布は，正規確率場の最大値の分布となり，その裾確率の近似にチューブ法を使うことができる (Kuriki, 2005)．

4.7 脳画像データ解析

K. Worsley と K. Friston は一連の研究で，fMRI データをガウス確率場とみなし，そのピークの値の統計的有意性をオイラー標数法によって評価する方法を提案した (Friston et al., 1994; Ashby, 2011)．彼らの方法は，VBM (Voxel based morphometry) データなどの脳画像解析の標準ソフト SPM に実装されている (The FIL Methods Group, 2018; 根本, 2014; 川口, 2017)．

5. おわりに

本項では滑らかなサンプルパスを持つガウス確率場の最大値の分布を求めるためのチューブ法とオイラー標数法について，その考え方と歴史を概観した．また統計学，データ解析への応用をいくつか紹介した．いままで述べたように，チューブ法は，(i) 多くの最大値型検定統計量，特に尤度比検定統計量 p 値の計算に用いることができる．(ii) 近似公式は a の有限項の漸近展開で与えられ計算が容易．(iii) 実用的な範囲 (例えば p 値 ≤ 0.3) で非常に精度がよい．(iv) $d = 1$ のとき，近似公式は保守側のバウンドを与える，といった実用的にも優れた特徴を持つ．

一方で理論的に残された問題として，添字集合の次元や体積が無限に発散する場合のチューブ法の性質は，まだほとんど解明されていない (栗木, 2019 を参照)．数学的な定式化を含めて，今後研究すべき研究分野である．

謝 辞

本稿執筆の機会を与えてくださった田村義保名誉教授，チューブ法研究の共著者である竹村彰通教授，Henry Wynn 教授，加藤直広博士，Xiaolei Lu 博士，4.7 節の内容についてご教示いただいた川口淳教授に感謝いたします．また詳細に論文を査読いただいたレフリーの方に感謝いたします．本稿は 2016 年度日本統計学受賞者記念講演での同名の講演に基づきます．関係者の方々に感謝いたします．

参 考 文 献

- Adler, R. J. and Hasofer, A. M. (1976). Level crossings for random fields, *The Annals of Probability*, **4**(1), 1–12.
- Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*, Springer, New York.
- Ashby, F. G. (2011). *Statistical Analysis of fMRI Data*, The MIT Press, Cambridge, Massachusetts.
- Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C. and Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent, *Human Brain Mapping*, **1**, 214–220.
- 福水健次, 栗木哲, 竹内啓, 赤平昌文 (2004). 『特異モデルの統計学—未解決問題への新しい視点』, 統計科学のフロンティア 7, 岩波書店, 東京.
- Gross, E. and Vitells, O. (2010). Trial factors for the look elsewhere effect in high energy physics, *The European Physical Journal C*, **70**(1–2), 525–530.

- Hotelling, H. (1939). Tubes and Spheres in n -spaces, and a class of statistical problems, *American Journal of Mathematics*, **61**(2), 440–460.
- Johansen, S. and Johnstone, I. M. (1990). Hotelling’s theorem on the volume of tubes: Some illustrations in simultaneous inference and data analysis, *The Annals of Statistics*, **18**(2), 652–684.
- Jones, M. C. and Sibson, R. (1987). What is projection pursuit?, *Journal of the Royal Statistical Society, Series A*, **150**(1), 1–37.
- Kateri, M. (2014). *Contingency Table Analysis: Methods and Implementation Using R*, Springer.
- 川口淳 (2017). 脳画像解析における統計学的クラスター推測, 日本磁気共鳴医学会雑誌, **37**(2), 39–49.
- Knowles, M. and Siegmund, D. (1989). On Hotelling’s geometric approach to testing for a nonlinear parameter in regression, *International Statistical Review*, **57**(3), 205–220.
- 小林昭七 (1997). 1940 年代, 50 年代の日本の微分幾何, 数学, **49**(3), 225–234.
- Kuriki, S. (2005). Asymptotic distribution of inequality-restricted canonical correlation with application to tests for independence in ordered contingency tables, *Journal of Multivariate Analysis*, **94**(2), 420–449.
- 栗木哲 (2019). オイラー標数法によるウィシャート行列の最大固有値分布の近似, 統計数理, **67**(2), 366–367.
- Kuriki, S. and Takemura, A. (2001). Tail probabilities of the maxima of multilinear forms and their applications, *The Annals of Statistics*, **29**(2), 328–371.
- Kuriki, S. and Takemura, A. (2008a). The tube method for the moment index in projection pursuit, *Journal of Statistical Planning and Inference*, **138**(9), 2749–2762.
- Kuriki, S. and Takemura, A. (2008b). Euler characteristic heuristic for approximating the distribution of the largest eigenvalue of an orthogonally invariant random matrix, *Journal of Statistical Planning and Inference*, **138**(11), 3357–3378.
- 栗木哲, 竹村彰通 (2008). チューブの体積と正規確率場の最大値の分布, 数学, **60**(2), 134–155.
- Lu, X. and Kuriki, S. (2017). Simultaneous confidence bands for contrasts between several nonlinear regression curves, *Journal of Multivariate Analysis*, **155**, 83–104.
- 村山齊 (2013). インタビュー「ファビオラ・ジャノッティ博士に聞く」, *Kavli IPMU News*, **24**, 40–47.
- Naiman, D. Q. (1986). Conservative confidence bands in curvilinear regression, *The Annals of Statistics*, **14**(3), 896–906.
- 根本清貴 (2014). 『すぐできる VBM : 精神・神経疾患の脳画像解析』, 学研メディカル秀潤社, 東京.
- ラオ, C. R. (1986). 『統計的推測とその応用』 (奥野忠一 監訳), 東京図書, 東京.
- Rice, S. O. (1944). Mathematical analysis of random noise, *The Bell System Technical Journal*, **23**, 282–332.
- Sun, J. (1991). Significance levels in exploratory projection pursuit, *Biometrika*, **78**(4), 759–769.
- Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields, *The Annals of Probability*, **21**(1), 34–71.
- Takemura, A. and Kuriki, S. (2002). On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains, *The Annals of Applied Probability*, **12**(2), 768–796.
- 丹野修吉 (1976). 『多様体の微分幾何学』, 実教出版, 東京.
- Taylor, J. E., Loftus, J. R. and Tibshirani, R. J. (2016). Inference in adaptive regression via the Kac-Rice formula, *The Annals of Statistics*, **44**(2), 743–770.
- The FIL Methods Group (2018). SPM12 Manual, <https://www.fil.ion.ucl.ac.uk/spm/>.
- van Dyk, D. A. (2014). The role of statistics in the discovery of a Higgs boson, *Annual Review of Statistics and Its Application*, **1**, 41–59.
- Weyl, H. (1939). On the volume of tubes, *American Journal of Mathematics*, **61**(2), 461–472.
- Worsley, K. J. (1995). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics, *Advances in Applied Probability*, **27**(4), 943–959.

The Tube Method: Theory and Applications

Satoshi Kuriki

The Institute of Statistical Mathematics

The tube method is an integral geometric method that is used to estimate the upper tail probabilities of the maxima of Gaussian random fields. The Euler characteristic method is used for the same purpose. In this paper, the concepts underlying the tube method and Euler characteristic method are investigated, and their associated applications in statistics and recent technical developments are discussed.

粒子フィルタとデータ同化

上野 玄太^{1,2}

(受付 2019 年 1 月 4 日; 改訂 6 月 17 日; 採択 6 月 21 日)

要 旨

粒子フィルタによる状態推定には、粒子と呼ばれる状態ベクトルの実現値が多数必要である。一方で、数値シミュレーションモデルを基盤とするデータ同化では、時間積分の計算コストが大きいことから、データ同化のアルゴリズムの計算コストを抑えることが要求される。多数の粒子を生成することと計算コストを抑えることは両立が難しい要求であるため、粒子フィルタによるデータ同化の方法は、現在のところ現業段階には至っておらず、研究段階にあると見てよい。粒子フィルタを適用する際の課題の一つは、各粒子に割り当てられる重みが 1 粒子に集中する、いわゆる退化の問題を限られた数の粒子でいかに克服するかである。

本稿では、重点サンプリング法により対処する方法として、陰粒子フィルタ (implicit particle filter) および等重粒子フィルタ (equivalent-weights particle filter) の考え方とアルゴリズムを解説した。

キーワード：粒子フィルタ，データ同化，重点サンプリング，陰粒子フィルタ，等重粒子フィルタ。

1. はじめに

数値シミュレーションモデルに基づくデータ同化システムの最大の特徴は、変数が多いことである。典型的には、状態ベクトル x_t の次元は $k = 10^5 \sim 10^8$ 、観測ベクトル y_t の次元は $\ell = 10^3 \sim 10^7$ である。このため、計算アルゴリズムに気を配る必要があり、典型的な例を挙げると、 $k \times k$ 次元の密行列の数値計算を行うことはご法度とされる。これが意味することは、状態空間モデルの推定で最も有名なアルゴリズムであるカルマンフィルタ、また、その延長にある拡張カルマンフィルタは適用できないということである。そのような状況で、データ同化の方法の研究の方向性は大きく二つある。第 1 の方向性は、推定誤差を求めないことである。その代表的なアルゴリズムは 4 次元変分法 (アジョイント法) である。4 次元変分法は、事後分布最大解 (MAP 解) を求めることに注力し、解の誤差評価は行わないという計算コストの選択と集中を実施することで、大規模なシミュレーションモデルに対してもデータ同化を可能にしている。第 2 の方向性は、数学的な厳密さに多少目をつぶることである。その代表はアンサンブルカルマンフィルタである。アンサンブルカルマンフィルタは、推定誤差は求めることができる一方で、厳密性は欠くが変数の分布をガウス分布とみなして推定を行うアルゴリズムである。

一方で、急激な変化を伴う現象を幅をつけて予報するなど、状態変数の非ガウス分布を忠実に扱いたい場合には、第 3 の方向性として、粒子フィルタ (Kitagawa, 1996) の適用が検討され

¹ 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

² 総合研究大学院大学 複合科学研究科統計科学専攻：〒190-8562 東京都立川市緑町 10-3

ている．ところが，データ同化と粒子フィルタは実際のところ相性が悪い．粒子フィルタにより適切に状態推定を行うためには，多くの粒子，具体的には粒子数を N としたときに， $N \gg k$ を満たす粒子数が必要であることが予想される．この条件は，単に k 次元ガウス分布を推定する場合であっても，正定性を満たすサンプル分散共分散行列を得るためには $k+1$ 個のサンプルが必要になることを考えれば，より複雑な非ガウス分布を対象にしたい以上は妥当なものであると考えられる．しかし，各粒子を生成するためにシミュレーションモデルの時間積分の実施が必要であること，その時間積分には大きな計算コストが要求されるのが一般的であることから， $N \gg k$ はもちろん， $N > k$ を要求するアルゴリズムの実施も不可能である．そのため，たとえば $N = 2k + 1$ であるアンセンテッドカルマンフィルタ (Julier and Uhlmann, 2004) も適用不可である．

すべての状態変数間の共分散の正確な推定は難しいかもしれないが，みなしガウス分布のような便宜的な操作を施さず，シミュレーションモデルによる状態変数の非線形発展をそのまま生かす粒子フィルタの特徴は魅力である．実際，アルゴリズム的には粒子数 N は小さくても粒子フィルタの実行は可能であるから，粒子のばらつきから幅を持った予測もできるはずである．ところが，粒子フィルタを実際に適用してみると，観測ベクトルの次元が高いことに起因して，リサンプリングの際の重みが 1 粒子に集中してしまう，いわゆる退化の問題が深刻であることが明らかになる．1 粒子では確率分布の推定はもちろん，幅を表現することすら不可能である．

そのため，データ同化の分野においては，第 3 の方向性として，粒子フィルタの適用というよりも，粒子フィルタの改良の研究が進められているという段階である．改良の方向は二つに大別でき，一つは粒子フィルタによる退化を緩和させ，粒子の多様性を失わせない処理を付け加えるものである (例えば融合粒子フィルタは Nakano et al., 2007, レビューは van Leeuwen, 2009)．近年は，変数間の相関距離を短く打ち切る，いわゆる局所化の処理を加えた粒子フィルタの研究が盛んである (Farchi and Bocquet, 2018; van Leeuwen et al., 2019)．融合にせよ局所化にせよ，付け加える処理は，状態空間モデルが表現する変数の数学的な構造に，多少目をつぶった結果の操作である．

もう一つの改良の方向は，重点サンプリングの方法をデータ同化の問題向けに展開するものである．ここに分類されるアルゴリズムは，変数の構造の数学的な厳密さを失わずに構築されていることが特徴である．ただし，粒子数 N をそれほど大きくはとれない以上，数学的厳密性の有無が，推定・予測性能に直結するかどうかは自明ではないことは注意しておきたい．本稿では，後者に分類されるアルゴリズムとして，陰粒子フィルタおよび等重粒子フィルタを解説する．現時点で筆者の知る限り，大規模なデータ同化システムに適した，重点サンプリングを基礎としたアルゴリズムはこの二つのフィルタに限られる．

2. モデルと粒子フィルタ

非線形のシステムモデル，線形の観測モデルからなる状態空間モデルを考える．システムノイズ，観測ノイズはガウス分布に従うとする．

$$(2.1) \quad \mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}) + \mathbf{v}_t$$

$$(2.2) \quad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{Q}_t)$$

$$(2.3) \quad \mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{w}_t$$

$$(2.4) \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{R}_t)$$

ここで， $\mathbf{x}_t, \mathbf{x}_{t-1}, \mathbf{v}_t$ は k 次元ベクトル， $\mathbf{y}_t, \mathbf{w}_t$ は ℓ 次元ベクトル， \mathbf{f}_t は k 次元非線形関数，

H_t は $\ell \times k$ の観測行列であるとする。 Q_t, R_t はそれぞれ、 $k \times k, \ell \times \ell$ の正定値対称行列で、既知であるとする。

状態空間モデルで表現される状態の統計的推測に関しては、カルマンフィルタや粒子フィルタに代表される逐次的な状態推定アルゴリズムが知られている。このアルゴリズムでは、一期先予測分布 $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ 、およびフィルタ分布 $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ が逐次的に得られる。すなわち、初期状態 \mathbf{x}_0 の確率分布 $p(\mathbf{x}_0)$ を与えたとき、以降の時刻での状態 \mathbf{x}_t の一期先予測分布、フィルタ分布はそれぞれ

$$(2.5) \quad p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1}$$

$$(2.6) \quad p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t}$$

と与えられる。

粒子フィルタでは、 \mathbf{x} の確率分布を多数 (N 個とする) の実現値 $\{\mathbf{x}^{(n)}\}_{n=1}^N$ を用いて

$$p(\mathbf{x}) \doteq \frac{1}{N} \sum_{n=1}^N \delta(\mathbf{x} - \mathbf{x}^{(n)})$$

とモンテカルロ近似により表現する。ここで、 \doteq はモンテカルロ近似等号である。個々の実現値 $\mathbf{x}^{(n)}$ を粒子という。以降、一期先予測分布 $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ 、フィルタ分布 $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ を近似する粒子をそれぞれ $\mathbf{x}_{t|t-1}^{(n)}$ 、 $\mathbf{x}_{t|t}^{(n)}$ と書き、

$$(2.7) \quad p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \doteq \frac{1}{N} \sum_{n=1}^N \delta(\mathbf{x}_t - \mathbf{x}_{t|t-1}^{(n)})$$

$$(2.8) \quad p(\mathbf{x}_t|\mathbf{y}_{1:t}) \doteq \frac{1}{N} \sum_{n=1}^N \delta(\mathbf{x}_t - \mathbf{x}_{t|t}^{(n)})$$

と近似表現する。

フィルタ分布 (2.6) にアンサンブル近似 (2.7) を用いると、

$$(2.9) \quad p(\mathbf{x}_t|\mathbf{y}_{1:t}) \doteq \sum_{n=1}^N \beta_t^{(n)} \delta(\mathbf{x}_t - \mathbf{x}_{t|t-1}^{(n)})$$

$$(2.10) \quad \beta_t^{(n)} \stackrel{\text{def}}{=} \frac{p(\mathbf{y}_t|\mathbf{x}_{t|t-1}^{(n)})}{\sum_{n=1}^N p(\mathbf{y}_t|\mathbf{x}_{t|t-1}^{(n)})}$$

となる。この式を (2.8) 式のように、各粒子で等しい重み ($1/N$) を持つ粒子 $\mathbf{x}_{t|t}^{(n)}$ で表現するには、重み $\beta_t^{(n)}$ ($n = 1, \dots, N$) で $\{\mathbf{x}_{t|t-1}^{(n)}\}_{n=1}^N$ から計 N 回のリサンプリング (復元抽出) を行い、得られた N 個のサンプルを $\{\mathbf{x}_{t|t}^{(n)}\}_{n=1}^N$ とすることで実現できる。

3. 重点サンプリングの導入

フィルタ分布 (2.6) に一期先予測分布 (2.5) を代入し、 \mathbf{x}_{t-1} のフィルタ分布の粒子近似 ((2.8) 式で $t-1$ としたもの) を行くと、次のようになる。

$$(3.1) \quad p(\mathbf{x}_t|\mathbf{y}_{1:t}) \doteq \frac{\sum_{n=1}^N p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1|t-1}^{(n)})}{\int \sum_{n=1}^N p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1|t-1}^{(n)})d\mathbf{x}_t}$$

(3.1) 式を重点サンプリングの方法に基づき近似する. $\mathbf{x}_{t-1|t-1}^{(n)}$ をパラメータとして持つ, \mathbf{x}_t のある提案分布 $q(\mathbf{x}_t; \mathbf{x}_{t-1|t-1}^{(n)})$ からの M 個の独立なサンプルを $\mathbf{x}_t^{(n,1)}, \mathbf{x}_t^{(n,2)}, \dots, \mathbf{x}_t^{(n,M)}$ とすると, (3.1) 式の分子および分母に現れる項は

$$\begin{aligned} p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1|t-1}^{(n)}) &= \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1|t-1}^{(n)})}{q(\mathbf{x}_t; \mathbf{x}_{t-1|t-1}^{(n)})} q(\mathbf{x}_t; \mathbf{x}_{t-1|t-1}^{(n)}) \\ &\doteq \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1|t-1}^{(n)})}{q(\mathbf{x}_t; \mathbf{x}_{t-1|t-1}^{(n)})} \frac{1}{M} \sum_{m=1}^M \delta(\mathbf{x}_t - \mathbf{x}_t^{(n,m)}) \\ &= \frac{1}{M} \sum_{m=1}^M \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(n,m)})p(\mathbf{x}_t^{(n,m)}|\mathbf{x}_{t-1|t-1}^{(n)})}{q(\mathbf{x}_t^{(n,m)}; \mathbf{x}_{t-1|t-1}^{(n)})} \delta(\mathbf{x}_t - \mathbf{x}_t^{(n,m)}) \end{aligned}$$

と表される. これを (3.1) 式に代入すると,

$$(3.2) \quad p(\mathbf{x}_t|\mathbf{y}_{1:t}) \doteq \sum_{n=1}^N \sum_{m=1}^M \beta_t^{(n,m)} \delta(\mathbf{x}_t - \mathbf{x}_t^{(n,m)})$$

となる. ここで,

$$(3.3) \quad \beta_t^{(n,m)} \stackrel{\text{def}}{=} \frac{\lambda_t^{(n,m)}}{\sum_{n=1}^N \sum_{m=1}^M \lambda_t^{(n,m)}}$$

$$(3.4) \quad \lambda_t^{(n,m)} \stackrel{\text{def}}{=} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(n,m)})p(\mathbf{x}_t^{(n,m)}|\mathbf{x}_{t-1|t-1}^{(n)})}{q(\mathbf{x}_t^{(n,m)}; \mathbf{x}_{t-1|t-1}^{(n)})}$$

と置いた. (3.2) 式は, n と m に関する二重の和で, NM 個の粒子によるモンテカルロ近似になっていることに注意されたい.

粒子フィルタ (2 節) は, 提案分布を $q(\mathbf{x}_t; \mathbf{x}_{t-1|t-1}^{(n)}) = p(\mathbf{x}_t|\mathbf{x}_{t-1|t-1}^{(n)})$, 各提案分布からの独立なサンプル数を $M = 1$ として, 合計 N 個の粒子でフィルタ分布を表現したものである. 一方, 4 次元変分法に基づき最適解を一つだけ ($\mathbf{x}_{t-1|t-1}^{(1)}$, $N = 1$ に相当) 得ている状況においても, 提案分布 $q(\mathbf{x}_t; \mathbf{x}_{t-1|t-1}^{(1)})$ から M 個の粒子を生成し, フィルタ分布の表現が可能である. 具体的には陰粒子フィルタ (5 節) の方法による (Atkins et al., 2013).

4. 最適提案分布

提案分布として

$$(4.1) \quad q(\mathbf{x}_t; \mathbf{x}_{t-1|t-1}^{(n)}) = p(\mathbf{x}_t|\mathbf{y}_t, \mathbf{x}_{t-1|t-1}^{(n)})$$

を選ぶ. ベイズの定理

$$(4.2) \quad p(\mathbf{x}_t|\mathbf{y}_t, \mathbf{x}_{t-1|t-1}^{(n)}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1|t-1}^{(n)})}{p(\mathbf{y}_t|\mathbf{x}_{t-1|t-1}^{(n)})}$$

を用いると, (3.3), (3.4) 式より,

$$(4.3) \quad \beta_t^{(n,m)} = \frac{p(\mathbf{y}_t|\mathbf{x}_{t-1|t-1}^{(n)})}{M \sum_{n=1}^N p(\mathbf{y}_t|\mathbf{x}_{t-1|t-1}^{(n)})}$$

となる. このときの重み $\beta_t^{(n,m)}$ は m に依存しない. これは, n を固定したもとは,

$m = 1, \dots, M$ に関する重みの分散は最小値であるゼロとなる, と言い換えることができる. 等しい重みであれば, リサンプリングによる退化は完全に克服できる. この分散が最小になるという性質から, (4.1)を最適提案分布という (Doucet et al., 2000, Proposition 2). ただし, n を $n = 1, \dots, N$ と変化させ, $M = 1$ とする粒子フィルタのアルゴリズムにおいては $n = 1, \dots, N$ に関する重みの分散は最小というわけではない.

状態空間モデル(2.1)–(2.4)のもとでは, (4.1), (4.3)式に現れる確率分布が解析的に表現でき,

$$(4.4) \quad p(\mathbf{x}_t | \mathbf{y}_t, \mathbf{x}_{t-1|t-1}^{(n)}) = \phi(\mathbf{x}_t; \zeta_t^{(n)}, U_t)$$

$$(4.5) \quad p(\mathbf{y}_t | \mathbf{x}_{t-1|t-1}^{(n)}) = \phi(\mathbf{y}_t; H_t \mathbf{f}_t(\mathbf{x}_{t-1|t-1}^{(n)}), H_t Q_t H_t' + R_t)$$

となる. ここで, $\phi(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$ は平均ベクトル $\boldsymbol{\mu}$, 分散共分散行列 Σ の正規分布に従う確率変数 \mathbf{x} の密度関数とし, 以下のようにおいている.

$$(4.6) \quad \zeta_t^{(n)} \stackrel{\text{def}}{=} \mathbf{f}_t(\mathbf{x}_{t-1|t-1}^{(n)}) + K_t(\mathbf{y}_t - H_t \mathbf{f}_t(\mathbf{x}_{t-1|t-1}^{(n)}))$$

$$(4.7) \quad U_t \stackrel{\text{def}}{=} (I - K_t H_t) Q_t$$

$$(4.8) \quad K_t \stackrel{\text{def}}{=} Q_t H_t' (H_t Q_t H_t' + R_t)^{-1}$$

導出方法は, 樋口 他 (2011, A.3 節) を参照のこと. 提案分布(4.4)からのサンプルは, U_t のコレスキー分解によって得られる $LL' = U_t$ を満たす $k \times k$ 下三角行列 L を用いれば,

$$(4.9) \quad \mathbf{x}_t^{(n,m)} \stackrel{\text{def}}{=} \zeta_t^{(n)} + L \boldsymbol{\xi}_t^{(n,m)} \quad n = 1, \dots, N; m = 1, \dots, M$$

として生成できる (北川, 2005, 16.3 節). ただし, $\boldsymbol{\xi}_t^{(n,m)}$ は k 次元標準正規分布に従う確率変数

$$(4.10) \quad \boldsymbol{\xi}_t \sim N(\mathbf{0}, I)$$

の実現値である.

5. 陰粒子フィルタ

ところが, (4.9)式に基づいて $\mathbf{x}_t^{(n,m)}$ を生成するには非常に計算コストがかかる. 最も計算コストがかかる部分は, (4.9)式の右辺で用いる L を得るための, コレスキー分解 $LL' = U_t$ である. $k \times k$ 次元の密行列のコレスキー分解はデータ同化分野ではご法度とされている演算である. そこで, 計算コストを抑えることを意識して, (4.9)式と同値な関係式を導くことを考える.

5.1 サンプリングの方法

(4.9)式を変形すると, $\mathbf{x}_t^{(n,m)}$ は, $\boldsymbol{\xi}_t^{(n,m)}$ を与えたもとで, 次の関係式を満たすものと考えることができる.

$$(5.1) \quad \frac{1}{2}(\mathbf{x}_t^{(n,m)} - \zeta_t^{(n)})' U_t (\mathbf{x}_t^{(n,m)} - \zeta_t^{(n)}) = \frac{1}{2}(\boldsymbol{\xi}_t^{(n,m)})' \boldsymbol{\xi}_t^{(n,m)}$$

ここで, 左辺は

$$\begin{aligned} & \frac{k}{2} \log 2\pi + \frac{1}{2} \log |U_t| + \frac{1}{2}(\mathbf{x}_t^{(n,m)} - \zeta_t^{(n)})' U_t (\mathbf{x}_t^{(n,m)} - \zeta_t^{(n)}) - \frac{k}{2} \log 2\pi - \frac{1}{2} \log |U_t| \\ &= -\log \phi(\mathbf{x}_t^{(n,m)}; \zeta_t^{(n)}, U_t) - [-\log \phi(\zeta_t^{(n)}; \zeta_t^{(n)}, U_t)] \\ &= -\log \phi(\mathbf{x}_t^{(n,m)}; \zeta_t^{(n)}, U_t) - \min_{\mathbf{x}_t} [-\log \phi(\mathbf{x}_t; \zeta_t^{(n)}, U_t)] \end{aligned}$$

と書けることに着目する. (4.4)式およびバイズの定理(4.2)を用いると, この式はさらに,

$$\begin{aligned} & -\log p(\mathbf{x}_t^{(n,m)}|\mathbf{y}_t, \mathbf{x}_{t-1|t-1}^{(n)}) - \min_{\mathbf{x}_t}[-\log p(\mathbf{x}_t|\mathbf{y}_t, \mathbf{x}_{t-1|t-1}^{(n)})] \\ (5.2) \quad & = F^{(n)}(\mathbf{x}_t^{(n,m)}) - F_{\min}^{(n)} \end{aligned}$$

となる. ここで

$$(5.3) \quad F^{(n)}(\mathbf{x}_t) \stackrel{\text{def}}{=} -\log p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1|t-1}^{(n)})$$

$$(5.4) \quad F_{\min}^{(n)} \stackrel{\text{def}}{=} \min_{\mathbf{x}_t} F^{(n)}(\mathbf{x}_t)$$

と定義した. したがって, (5.1), (5.2)式より, k 次元標準正規分布に従う確率変数の実現値 $\xi_t^{(n,m)}$ ($n = 1, \dots, N; m = 1, \dots, M$) を与えたもとの,

$$(5.5) \quad F^{(n)}(\mathbf{x}_t^{(n,m)}) - F_{\min}^{(n)} = \frac{1}{2}(\xi_t^{(n,m)})' \xi_t^{(n,m)}$$

を解くことで, $\mathbf{x}_t^{(n,m)}$ を生成することができ. この方法を陰粒子フィルタ (implicit particle filter) という (Chorin et al., 2010, 2013; Atkins et al., 2013). (5.5)式を計算コストを抑えて解く方法が, Morzfeld and Chorin (2012) にある.

状態空間モデル(2.1)–(2.4)のもとでは, (5.3)式は

$$(5.6) \quad F^{(n)}(\mathbf{x}_t) = \frac{\ell}{2} \log 2\pi + \frac{1}{2} \log |R_t| + \frac{1}{2}(\mathbf{y}_t - H_t \mathbf{x}_t)' R_t^{-1}(\mathbf{y}_t - H_t \mathbf{x}_t) \\ + \frac{k}{2} \log 2\pi + \frac{1}{2} \log |Q_t| + \frac{1}{2}(\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1|t-1}^{(n)}))' Q_t^{-1}(\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1|t-1}^{(n)}))$$

となる. 第 3 項, 第 6 項以外は \mathbf{x}_t に依存しないので, 変分法で通常用いる目的関数

$$(5.7) \quad J^{(n)}(\mathbf{x}_t) \stackrel{\text{def}}{=} \frac{1}{2}(\mathbf{y}_t - H_t \mathbf{x}_t)' R_t^{-1}(\mathbf{y}_t - H_t \mathbf{x}_t) \\ + \frac{1}{2}(\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1|t-1}^{(n)}))' Q_t^{-1}(\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1|t-1}^{(n)}))$$

$$(5.8) \quad J_{\min}^{(n)} \stackrel{\text{def}}{=} \min_{\mathbf{x}_t} J^{(n)}(\mathbf{x}_t)$$

を用いると, (5.5)式の代わりに

$$(5.9) \quad J^{(n)}(\mathbf{x}_t^{(n,m)}) - J_{\min}^{(n)} = \frac{1}{2}(\xi_t^{(n,m)})' \xi_t^{(n,m)}$$

を解いて $\mathbf{x}_t^{(n,m)}$ を求めればよい.

5.2 重みの計算

(3.3), (3.4)式に基づき, 重みを求める.

(3.4)式の分母に現れる提案分布 $q(\mathbf{x}_t; \mathbf{x}_{t-1|t-1}^{(n)})$ を考える. (4.9)式より, \mathbf{x}_t は標準正規分布(4.10)に従う確率変数 ξ_t から, 次の関係を経て生成される確率変数であると解釈できる.

$$(5.10) \quad \mathbf{x}_t = \zeta_t^{(n)} + L \xi_t$$

(5.10)式により生成される \mathbf{x}_t の確率分布, すなわちサンプリングを行う提案分布 $q(\mathbf{x}_t; \mathbf{x}_{t-1|t-1}^{(n)})$ は

$$\phi(\xi_t; \mathbf{0}, I) = q(\mathbf{x}_t(\xi_t); \mathbf{x}_{t-1|t-1}^{(n)}) \left\| \frac{\partial \mathbf{x}_t(\xi_t)}{\partial \xi_t'} \right\|$$

$$(5.11) \quad = q(\mathbf{x}_t(\boldsymbol{\xi}_t); \mathbf{x}_{t-1|t-1}^{(n)}) \|L\|$$

を満たす．(5.11)式，および(5.5)式を用いると，(3.4)，(3.3)式より，重みは次のように表される．

$$\begin{aligned} \lambda_t^{(n,m)} &= \frac{\exp\{-[-\log p(\mathbf{y}_t|\mathbf{x}_t^{(n,m)})p(\mathbf{x}_t^{(n,m)}|\mathbf{x}_{t-1|t-1}^{(n)})]\}}{q(\mathbf{x}_t^{(n,m)}; \mathbf{x}_{t-1|t-1}^{(n)})} \\ &= \frac{\exp[-F^{(n)}(\mathbf{x}_t^{(n,m)})]}{\frac{\phi(\boldsymbol{\xi}_t^{(n,m)}; \mathbf{0}, I)}{\|L\|}} \\ &= \frac{\exp\left[-F_{\min}^{(n)} - \frac{1}{2}(\boldsymbol{\xi}_t^{(n,m)})'\boldsymbol{\xi}_t^{(n,m)}\right]}{\left(\frac{1}{2\pi}\right)^{k/2} \exp\left[-\frac{1}{2}(\boldsymbol{\xi}_t^{(n,m)})'\boldsymbol{\xi}_t^{(n,m)}\right]} \|L\| \\ (5.12) \quad &= (2\pi)^{k/2} \exp(-F_{\min}^{(n)}) \|L\| \end{aligned}$$

$$(5.13) \quad \beta_t^{(n,m)} = \frac{\exp(-F_{\min}^{(n)})}{M \sum_{n=1}^N \exp(-F_{\min}^{(n)})}$$

状態空間モデル(2.1)–(2.4)のもとでは，重み(5.13)は

$$(5.14) \quad \beta_t^{(n,m)} = \frac{\exp(-J_{\min}^{(n)})}{M \sum_{n=1}^N \exp(-J_{\min}^{(n)})}$$

となる．

特に，4次元変分法のように $N = 1$ の場合で， $m = 1, \dots, M$ によって多数の粒子を生成する場合には，重み(5.13)，(5.14)は m によらない定数になるため，生成した粒子 M 個をリサンプリングなしでそのままフィルタ分布を表現する粒子として扱うことができる．

6. 等重粒子フィルタ

(4.3)，(5.13)式で与えられるように，最適提案分布ならびに陰粒子フィルタでは，重み $\beta_t^{(n,m)}$ は m に依存しないが， n を変えれば重みは $\mathbf{x}_{t-1|t-1}^{(n)}$ に依存して変化する．そのため， $M = 1$ として N 個の粒子を扱う粒子フィルタでは，最適提案分布から受ける恩恵は限定的である．実際，観測ベクトル \mathbf{y}_t の次元 ℓ が大きい場合には，最適提案分布を用いた重みであっても退化が起こる．等重粒子フィルタ (equivalent-weights particle filter, van Leeuwen, 2010; Ades and van Leeuwen, 2013, 2015) は，重み $\beta_t^{(n,m)}$ を n にも依存しない定数になるように粒子を生成する方法である．以降の節で見ると，得られる重みは厳密に等しくなるわけではなく，ほとんど等しい重みとなる．そのため，「ほとんど等重粒子フィルタ」と呼ぶのが適切で，提案当初は almost が添えられていた (van Leeuwen, 2010) が，最近は単に等重粒子フィルタと呼ばれている．

等重粒子フィルタも粒子フィルタと同様に， $M = 1$ として N 個の粒子を扱う．そのため，本節の以下の説明では $m = 1$ の場合のみを考えているとご理解頂きたい．

6.1 サンプリングの考察

重みの構成要素である $\lambda_t^{(n,m)}$ を表現する(3.4)式は，提案分布にもとづき生成される粒子 $\mathbf{x}_t^{(n,m)}$ の関数の分数として表される．そこで，提案分布をうまく設計して，分子と分母それ

それを粒子に依存しない値にすることを考える。 $\lambda_t^{(n,m)}$ の値が粒子に依存しなければ、重み $\beta_t^{(n,m)}$ も n, m によらない一定値となる。

以下で見るように、まず、分子が一定値となるために、各粒子が満たすべき条件を求める。提案分布はその条件を満たす粒子を生成しやすいもの、すなわち、条件を満たす粒子の確率密度の値が大きいものを設定する。次に、分母も一定値とするために、これらの条件を満たす粒子の確率密度は共通の値をとるとする。

(3.4)式の分子を、ベイズの定理(4.2)および(4.4)、(4.5)式を用いて変形すると、

$$\begin{aligned}
 & p(\mathbf{y}_t | \mathbf{x}_t^{(n,m)}) p(\mathbf{x}_t^{(n,m)} | \mathbf{x}_{t-1}^{(n)}) \\
 &= p(\mathbf{x}_t^{(n,m)} | \mathbf{y}_t, \mathbf{x}_{t-1}^{(n)}) p(\mathbf{y}_t | \mathbf{x}_{t-1}^{(n)}) \\
 &= \phi(\mathbf{x}_t^{(n,m)}; \zeta_t^{(n)}, U_t) \phi(\mathbf{y}_t; H_t \mathbf{f}_t(\mathbf{x}_{t-1}^{(n)}), H_t Q_t H_t' + R_t) \\
 (6.1) \quad &= \left(\frac{1}{2\pi}\right)^{k/2} \frac{1}{\sqrt{|U_t|}} \left(\frac{1}{2\pi}\right)^{\ell/2} \frac{1}{\sqrt{|H_t Q_t H_t' + R_t|}} \exp[-J^{(n)}(\mathbf{x}_t^{(n,m)})]
 \end{aligned}$$

となる。ここで、

$$(6.2) \quad J^{(n)}(\mathbf{x}_t) = \frac{1}{2}(\mathbf{x}_t - \zeta_t^{(n)})' U_t^{-1} (\mathbf{x}_t - \zeta_t^{(n)}) + \frac{1}{2}(\boldsymbol{\eta}_t^{(n)})' (H_t Q_t H_t' + R_t)^{-1} \boldsymbol{\eta}_t^{(n)}$$

$$(6.3) \quad \boldsymbol{\eta}_t^{(n)} \stackrel{\text{def}}{=} \mathbf{y}_t - H_t \mathbf{f}_t(\mathbf{x}_{t-1}^{(n)})$$

と置いた。(6.2)式で(5.7)式と同じ $J^{(n)}(\mathbf{x}_t)$ を用いているのは、これらの2式が恒等式であることによる。(6.2)式より、 $J^{(n)}(\mathbf{x}_t)$ は $\mathbf{x}_t = \zeta_t^{(n)}$ のときに最小値 $\frac{1}{2}(\boldsymbol{\eta}_t^{(n)})' (H_t Q_t H_t' + R_t)^{-1} \boldsymbol{\eta}_t^{(n)}$ をとることが分かる。また、 \mathbf{x}_t をうまく選ぶことで、 $J^{(n)}(\mathbf{x}_t)$ はこの最小値以上の任意の値をとることが可能であることに注意しておく。

ここでは、(6.1)式に含まれる $J^{(n)}(\mathbf{x}_t^{(n,m)})$ がある定数 C となるように $\mathbf{x}_t^{(n,m)}$ を生成することを考える。 $\mathbf{x}_t^{(n,m)}$ の表現方法には任意性があるが、ある実数 $\alpha^{(n)}$ を導入して

$$(6.4) \quad \mathbf{x}_t^{(n,m)} = \mathbf{f}_t(\mathbf{x}_{t-1}^{(n)}) + \alpha^{(n)} K_t \boldsymbol{\eta}_t^{(n)}$$

と表すことにする。(6.2)、(6.4)式を用いると、 $J^{(n)}(\mathbf{x}_t^{(n,m)}) = C$ となるために $\alpha^{(n)}$ が満たすべき条件は、

$$(6.5) \quad C = \frac{1}{2}(\alpha^{(n)} - 1)^2 (\boldsymbol{\eta}_t^{(n)})' K_t' U_t^{-1} K_t \boldsymbol{\eta}_t^{(n)} + \frac{1}{2}(\boldsymbol{\eta}_t^{(n)})' (H_t Q_t H_t' + R_t)^{-1} \boldsymbol{\eta}_t^{(n)}$$

となる。さらに、逆行列補題(例えば樋口他, 2011, A.1 節)により、 $K_t = (Q_t^{-1} + H_t' R_t^{-1} H_t)^{-1} H_t' R_t^{-1}$ 、 $U_t = (Q_t^{-1} + H_t' R_t^{-1} H_t)^{-1}$ と表されることを用いると、 $K_t' U_t^{-1} = R_t^{-1} H_t$ となることから、(6.5)式は、

$$(6.6) \quad C = \frac{1}{2}(\alpha^{(n)})^2 (\boldsymbol{\eta}_t^{(n)})' R_t^{-1} H_t K_t \boldsymbol{\eta}_t^{(n)} - \alpha^{(n)} (\boldsymbol{\eta}_t^{(n)})' R_t^{-1} H_t K_t \boldsymbol{\eta}_t^{(n)} + \frac{1}{2}(\boldsymbol{\eta}_t^{(n)})' R_t^{-1} \boldsymbol{\eta}_t^{(n)}$$

となる。ここで、右辺第3項の導出においては、次の変形を行った。

$$\begin{aligned}
 R_t^{-1} H_t K_t + (H_t Q_t H_t' + R_t)^{-1} &= R_t^{-1} H_t Q H_t' (H_t Q_t H_t' + R_t)^{-1} + (H_t Q_t H_t' + R_t)^{-1} \\
 &= (R_t^{-1} H_t Q H_t' + I) (H_t Q_t H_t' + R_t)^{-1} \\
 &= R_t^{-1} (H_t Q H_t' + R_t) (H_t Q_t H_t' + R_t)^{-1} \\
 &= R_t^{-1}
 \end{aligned}$$

(6.6)式は $\alpha^{(n)}$ についての2次方程式なので、厳密解が得られ、

$$(6.7) \quad \alpha^{(n)} = 1 \pm \sqrt{1 - \frac{(\boldsymbol{\eta}_t^{(n)})' R_t^{-1} \boldsymbol{\eta}_t^{(n)} - 2C}{(\boldsymbol{\eta}_t^{(n)})' R_t^{-1} H_t K_t \boldsymbol{\eta}_t^{(n)}}}$$

となる．(6.7)式には複号で示される 2 根があるが，正符号の根を採用する． $\alpha^{(n)}$ が確実に正の値となることで各粒子が観測 \mathbf{y}_t に近づくからである (Ades and van Leeuwen, 2013, Section 3.3)．

このようにして，(6.7)式を用いて粒子(6.4)を生成すれば，(6.2)式は $J^{(n)}(\mathbf{x}_t^{(n,m)}) = C$ となり，(6.1)式は

$$(6.8) \quad \left(\frac{1}{2\pi}\right)^{k/2} \frac{1}{\sqrt{|U_t|}} \left(\frac{1}{2\pi}\right)^{\ell/2} \frac{1}{\sqrt{|H_t Q_t H_t' + R_t|}} \exp(-C)$$

となる．すなわち，(3.4)式の分子は $\mathbf{x}_t^{(n,m)}$ に依存しない値をとる．

定数 C は， N 個の粒子のうちいくつをフィルタ分布のアンサンブル近似に残したいかというユーザーの意向に応じて決めることができる．例えば， $N = 100$ で，80 個の粒子を残したいならば， $J^{(n)}(\mathbf{x}_t)$ 最小値の集合 $\{\min_{\mathbf{x}_t} J^{(n)}(\mathbf{x}_t)\}_{n=1}^{100} = \{\frac{1}{2}(\boldsymbol{\eta}_t^{(n)})'(H_t Q_t H_t' + R_t)^{-1} \boldsymbol{\eta}_t^{(n)}\}_{n=1}^{100}$ から，値が 80 番目に小さい要素を選び，その値を C_{80} とする． $\min_{\mathbf{x}_t} J^{(n)}(\mathbf{x}_t)$ の値が C_{80} 以下の 80 個の n に対しては，(6.7)式により $\alpha^{(n)}$ を選び，(6.4)式を計算することで， $J^{(n)}(\mathbf{x}_t^{(n,m)}) = C_{80}$ となる $\mathbf{x}_t^{(n,m)}$ が生成される． $\min_{\mathbf{x}_t} J^{(n)}(\mathbf{x}_t)$ の値が C_{80} より大きい残り 20 個の n に対しては，その 20 個の予測粒子 $\mathbf{f}_t(\mathbf{x}_{t-1|t-1}^{(n)})$ ，生成された 80 個の $\mathbf{x}_t^{(n,m)}$ (重みは C_{80} から決まる共通の値)を合わせた計 100 個の粒子から 20 個をリサンプリングして補充する．

6.2 提案分布の設計とサンプリング

つづいて，(3.4)式の分母を考える．分母には提案分布が入るが，その分布の性質として期待されることは，(6.4)で与えられる $\mathbf{x}_t^{(n,m)}$ が生成されやすいこと，そのときの確率密度の値が n によらず一定になることである．素朴には，デルタ関数を用いた

$$(6.9) \quad q(\mathbf{x}_t; \mathbf{x}_{t-1|t-1}^{(n)}) = \delta(\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1|t-1}^{(n)}) - \alpha^{(n)} K_t \boldsymbol{\eta}_t^{(n)})$$

がこの条件にふさわしそうである．提案分布(6.9)を用いれば，必ず \mathbf{x}_t は(6.4)で与えられる $\mathbf{x}_t^{(n,m)}$ となり，(6.8)式から決まる各粒子の重みは等しくなる．

しかし，(6.9)式は，重点サンプリングの基本的な条件として，提案分布の台が分子の台を含むという条件を満たしていないという点で問題がある．そこで，重みが厳密に等しくなるといふ条件を緩和した次善策として，

$$(6.10) \quad \mathbf{x}_t^{(n,m)} = \mathbf{f}_t(\mathbf{x}_{t-1|t-1}^{(n)}) + \alpha^{(n)} K_t \boldsymbol{\eta}_t^{(n)} + Q^{1/2} \boldsymbol{\xi}_t^{(n,m)}$$

と $\mathbf{x}_t^{(n,m)}$ を生成することにする．ここで， $\boldsymbol{\xi}_t^{(n,m)}$ は次の密度関数を持つ確率変数 $\boldsymbol{\xi}_t$ の実現値であるとする．

$$(6.11) \quad p(\boldsymbol{\xi}_t; \epsilon, \gamma_U, \gamma_N) \stackrel{\text{def}}{=} (1 - \epsilon) \prod_{i=1}^k \left(\frac{1}{2\gamma_U} \right) \mathbb{1}(|(\boldsymbol{\xi}_t)_i| \leq \gamma_U) + \epsilon \phi(\boldsymbol{\xi}_t; \mathbf{0}, \gamma_N^2 I)$$

ここで， $(\boldsymbol{\xi}_t)_i$ は $\boldsymbol{\xi}_t$ の i 番目の要素を表す．(6.10)式の右辺第 3 項として，ノイズ $Q^{1/2} \boldsymbol{\xi}_t^{(n,m)}$ が導入されているところが，台の問題を回避するポイントである．

(6.11)式で与えられる $\boldsymbol{\xi}_t$ の密度関数は， $\boldsymbol{\xi}_t = \mathbf{0}$ 周りに限定した $[-\gamma_U, \gamma_U]^k$ に台を持つ一様分布と，平均が $\boldsymbol{\xi}_t = \mathbf{0}$ である正規分布の混合分布である．混合比を表すパラメータ $0 < \epsilon < 1$ を小さく設定することで，確率密度の値は一様分布で与えられる n に依存しない一定の値

$(1-\epsilon)/(2\gamma_U)^k$ でほとんど決まることになる. Ades and van Leeuwen (2013) では, $\epsilon = 10^{-3}/N$ を採用している. $\gamma_U > 0, \gamma_N > 0$ も小さい数を設定することで, 確率密度を $\boldsymbol{\xi}_t = \mathbf{0}$ の近傍に集中させることができる. 具体的な値は 6.3 節で紹介する. また, 懸案であった台の大きさは, 正規分布の台 $(-\infty, \infty)^k$ が分子の台を確実に含むことで解決している. (6.10)での $Q_t^{1/2}$ の乗算は, 各成分が独立な $\boldsymbol{\xi}_t$ に相関を持たせ, 状態変数間の物理的なバランスを考慮したい意図がある.

6.3 重みの計算

(3.4)式に基づき, 重みを求めよう.

まず, (3.4)式の分子は(6.1)式で与えられ, (6.1)式に含まれる $J^{(n)}(\boldsymbol{x}_t^{(n,m)})$ は(6.2)式で与えられる. 粒子(6.10)を(6.2)式に代入し, (6.2)と(5.7)が恒等式であることを用いると,

$$(6.12) \quad \begin{aligned} J^{(n)}(\boldsymbol{x}_t^{(n,m)}) &= \frac{1}{2}(\boldsymbol{y}_t - H_t \boldsymbol{x}_t^{(n,m)})' R_t^{-1} (\boldsymbol{y}_t - H_t \boldsymbol{x}_t^{(n,m)}) \\ &\quad + \frac{1}{2}[\alpha^{(n)} Q_t^{1/2} H_t' (H_t Q_t H_t' + R_t)^{-1} \boldsymbol{\eta}_t^{(n)} + \boldsymbol{\xi}^{(n,m)}]' \\ &\quad \times [\alpha^{(n)} Q_t^{1/2} H_t' (H_t Q_t H_t' + R_t)^{-1} \boldsymbol{\eta}_t^{(n)} + \boldsymbol{\xi}^{(n,m)}] \end{aligned}$$

となる. (6.10)式で与えられる粒子 $\boldsymbol{x}_t^{(n,m)}$ は, (6.12)式が C に近いほぼ定数値になるように設計されている. その結果, (6.12)式を用いる(6.1)式もほぼ定数, すなわち(3.4)式の分子もほぼ定数になる.

次に, (3.4)式の分母を考える. (5.10)式を得たときと同様に考えると, 粒子の生成(6.10)は, 確率変数 \boldsymbol{x}_t と $\boldsymbol{\xi}_t$ の関係

$$(6.13) \quad \boldsymbol{x}_t = \boldsymbol{f}_t(\boldsymbol{x}_{t-1|t-1}^{(n)}) + \alpha^{(n)} K_t \boldsymbol{\eta}_t^{(n)} + Q^{1/2} \boldsymbol{\xi}_t$$

に対応し, $\partial \boldsymbol{x}_t / \partial \boldsymbol{\xi}_t' = Q^{1/2}$ となる. よって, $\boldsymbol{\xi}_t$ の密度関数(6.11)を用いると, (3.4)式の分母は次のようになる.

$$(6.14) \quad \begin{aligned} q(\boldsymbol{x}_t^{(n,m)}; \boldsymbol{x}_{t-1|t-1}^{(n)}) &= \frac{p(\boldsymbol{\xi}_t^{(n,m)}; \epsilon, \gamma_U, \gamma_N)}{\left\| \frac{\partial \boldsymbol{x}_t}{\partial \boldsymbol{\xi}_t'}(\boldsymbol{\xi}_t^{(n,m)}) \right\|} \\ &= \frac{1-\epsilon}{(2\gamma_U)^k} \frac{1}{\sqrt{|Q|}} \left\{ \mathbb{1}(\boldsymbol{\xi}_t^{(n,m)} \in [-\gamma_U, \gamma_U]^k) \right. \\ &\quad \left. + \frac{\epsilon}{1-\epsilon} \left(\frac{2}{\pi}\right)^{k/2} \left(\frac{\gamma_U}{\gamma_N}\right)^k \exp\left[-\frac{1}{2\gamma_N^2} (\boldsymbol{\xi}_t^{(n,m)})' \boldsymbol{\xi}_t^{(n,m)}\right] \right\} \end{aligned}$$

よって, 重みは(3.3), (3.4)式より, $\lambda_t^{(n,m)}$ の定数部分を除いて定義する $\tilde{\lambda}_t^{(n,m)}$ を用いて, 次のように表される.

$$(6.15) \quad \beta_t^{(n,m)} = \frac{\tilde{\lambda}_t^{(n,m)}}{\sum_{n=1}^N \sum_{m=1}^M \tilde{\lambda}_t^{(n,m)}}$$

$$(6.16) \quad \tilde{\lambda}^{(n,m)} \stackrel{\text{def}}{=} \frac{\exp[-J^{(n)}(\boldsymbol{x}_t^{(n,m)})]}{\mathbb{1}(\boldsymbol{\xi}^{(n,m)} \in [-\gamma_U, \gamma_U]^k) + \frac{\epsilon}{1-\epsilon} \left(\frac{2}{\pi}\right)^{k/2} \left(\frac{\gamma_U}{\gamma_N}\right)^k \exp\left[-\frac{1}{2\gamma_N^2} (\boldsymbol{\xi}^{(n,m)})' \boldsymbol{\xi}^{(n,m)}\right]}$$

ただし, (6.16)式の分子において, $J^{(n)}(\mathbf{x}_t^{(n,m)})$ は(6.12)式で, $\mathbf{x}_t^{(n,m)}$ は(6.10)式でそれぞれ与えられるものとする.

γ_U, γ_N の設定値について述べておく. (6.16)式に見るように, γ_U は一様分布の台を, また, 比 γ_U/γ_N が正規分布の寄与を制御している. Ades and van Leeuwen (2013) では, $\gamma_U = 10^{-5}$ と設定している. 正規分布の寄与は重みの値の違いをもたらすので, 退化を起こさないようにしたいという観点からは, できるだけ小さいほうが望ましい. すでに ϵ は小さい値を設定しているので($\epsilon = 10^{-3}/N$ など, 6.2 節), $(2/\pi)^{k/2}(\gamma_U/\gamma_N)^k = 1$ であれば十分である. これより, $\gamma_N = (2/\pi)\gamma_U$ と設定することができる.

謝 辞

査読者からの有益なコメントに感謝いたします. 本研究は統計数理研究所共同研究プログラム(27-共研-2007)の助成を受けたものです.

参 考 文 献

- Ades, M. and van Leeuwen, P. J. (2013). An exploration of the equivalent weights particle filter, *Quarterly Journal of the Royal Meteorological Society*, **139**(672), 820–840, DOI:10.1002/qj.1995.
- Ades, M. and van Leeuwen, P. J. (2015). The equivalent-weights particle filter in a high-dimensional system, *Quarterly Journal of the Royal Meteorological Society*, **141**(687), 484–503, DOI:10.1002/qj2370.
- Atkins, E., Morzfeld, M. and Chorin, A. J. (2013). Implicit particle methods and their connection with variational data assimilation, *Monthly Weather Review*, **141**, 1786–1803, DOI:10.1175/MWR-D-12-00145.1.
- Chorin, A., Morzfeld, M. and Tu, X. (2010). Implicit particle filters for data assimilation, *Communications in Applied Mathematics and Computational Science*, **5**(2), 221–240.
- Chorin, A. J., Morzfeld, M. and Tu, X. (2013). A survey of implicit particle filters for data assimilation, *State-Space Models: Applications in Economics and Finance* (eds. Y. Zeng and S. Wu), Statistics and Econometrics for Finance 1, 63–88, Springer, New York, DOI:10.1007/978-1-4614-7789-1_3.
- Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing*, **10**(3), 197–208.
- Farchi, A. and Bocquet, M. (2018). Review article: Comparison of local particle filters and new implementations, *Nonlinear Processes in Geophysics*, **25**(4), 765–807, <https://www.nonlin-processes-geophys.net/25/765/2018/>, DOI:10.5194/npg-25-765-2018.
- 樋口知之, 上野玄太, 中野慎也, 中村和幸, 吉田 亮 (2011). 『データ同化入門』, 予測と発見の科学, **6**, 朝倉書店, 東京.
- Julier, S. J. and Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation, *Proceedings of the IEEE*, **92**, 401–422, DOI:10.1109/JPROC.2003.823141.
- Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics*, **5**(1), 1–25.
- 北川源四郎 (2005). 『時系列解析入門』, 岩波書店, 東京.
- Morzfeld, M. and Chorin, A. J. (2012). Implicit particle filtering for models with partial noise, and an application to geomagnetic data assimilation, *Nonlinear Processes in Geophysics*, **19**(3), 365–382, <https://www.nonlin-processes-geophys.net/19/365/2012/>, DOI:10.5194/npg-19-365-2012.
- Nakano, S., Ueno, G. and Higuchi, T. (2007). Merging particle filter for sequential data assimilation, *Nonlinear Processes in Geophysics*, **14**, 395–408.
- van Leeuwen, P. J. (2009). Particle filtering in geophysical systems, *Monthly Weather Review*, **137**(12), 4089–4114, DOI:10.1175/2009MWR2835.1.

- van Leeuwen, P. J. (2010). Nonlinear data assimilation in geosciences: An extremely efficient particle filter, *Quarterly Journal of the Royal Meteorological Society*, **136**(653), 1991–1999, DOI:10.1002/qj.699.
- van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R. and Reich, S. (2019). Particle filters for high-dimensional geoscience applications: A review, *Quarterly Journal of the Royal Meteorological Society*, **145**, 2335–2365, DOI:10.1002/qj.3551.

Particle Filter and Data Assimilation

Genta Ueno^{1,2}

¹The Institute of the Statistical Mathematics

²Department of Statistical Science, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies, SOKENDAI

State estimation by the particle filter requires many realizations of the state vector, called particles, for representing non-Gaussian distributions. On the other hand, data assimilation based on a numerical simulation model cannot adopt an assimilation algorithm that is computationally expensive because computational resources need to be assigned to time integration of the state vector by the simulation model. Therefore, the methodology for data assimilation by the particle filter is considered to still be in the research stage. One important issue is how to avoid so-called the filter degeneracy, in which weights for resampling concentrate to a single particle. This article reviews particle filter algorithms on the basis of importance sampling, specifically the implicit particle filter and the equivalent-weights particle filter.

統計数理研究所における最適化研究

土谷 隆^{1,2}

(受付 2019 年 5 月 7 日; 改訂 9 月 10 日; 採択 9 月 20 日)

要 旨

最適化は統計科学における重要な方法論である。最適化アルゴリズムの発展は新しいモデルの実用化を可能とし、新しいモデルの追及は、最適化アルゴリズムの進展をもたらす。統計科学の研究所としての統計数理研究所では、さまざまな形で最適化の研究が行われてきた。本論説では、統計数理研究所における最適化法の研究の展開について概説する。

キーワード：無限次元最適化、最急降下法、Kaczmarz 法、非線形最適化、内点法、情報幾何。

1. はじめに

最適化は統計科学における重要な方法論である。統計科学は、データの背後にそれを生み出す確率的機構、すなわち統計モデルを想定することで、現象の予測や説明を目指す。その過程では、最小二乗法や最尤法に見られるように、適切な定量的規準を最適化してモデルのパラメータを推定することが行われる。最適化手法が深化し発展することで、より多くの統計モデルが実用化される。遡ってみれば、線形計画法の原型は、 L_1 回帰として現れたともいわれている。また、ネイマン・ピアソンの基本定理にも見られるように、推定量や検定量の最適性についての解析が、古くからの統計学と最適化の接点としてあった。そこでは、より抽象的な、無限次元の最適化問題が扱われ、最適性条件等が問題となることが多かった。このように、最適化法は、さまざまな形で統計科学に古くからかかわってきた。

第 2 次大戦後の電子計算機の登場は、現代科学技術の性格を大きく変えることとなった。実際に計算を行い現実問題を解くことができる可能性が拓けたことを契機とし、多くの数学的・数理的分野が「モデリングとアルゴリズム」を意識した新しい方向に力強く発展していった。統計学についても然り、また最適化についても然りである。特に最適化においては、1947 年、Dantzig による線形計画法の実用化とともに、数理計画法という新しい学問が勃興し、最適化手法の開発とそれを意識したモデリングと数理に重点を置いた分野として研究が活発に進められるようになってきた。

数理計画の分野では、連続最適化と離散最適化の 2 つの分野が互いに影響を及ぼしあいながら研究が展開された。1950 年代から 70 年代にかけては、連続最適化では、数理的には、双対理論の構築や最適性条件の解析などの研究が進展し、アルゴリズム研究においては、無制約最適化問題に対する最急降下法、ニュートン法の解析や、準ニュートン法などの開発、制約付最適化問題については、罰金関数法や障壁関数法、乗数法などの開発が行われた。離散的最適化

¹ 政策研究大学院大学 政策研究科：〒106-8677 東京都港区六本木 7-22-1

² 統計数理研究所 客員：〒190-8562 東京都立川市緑町 10-3

では、ネットワーク最適化問題の解法や整数計画問題に対する分枝限定法などが開発されてきた。また、モデリングの分野では、線形計画問題がさまざまな分野に適用され、問題解決のための有力な道具として活用されるようになった。

1970年代の後半に計算複雑度の理論が勃興し、最適化問題を解くための計算量が離散的最適化、連続的最適化の両分野で重要な研究課題となった。そして、線形計画問題に対する最初の多項式時間解法である楕円体法が Khachiyan によって 1979 年により提案され、さらに、実用的多項式時間解法である Karmarkar 法の 1984 年の登場により、そして分野全体を巻き込んだ大きな変革をもたらした内点法の研究へと進んでいった。

計算機の急速な発展と相俟って、凸 2 次計画問題や半正定値計画問題、2 次錐計画問題などが新たな最適化問題として解けるようになった。これは、線形計画法が数理モデリングの世界で果たしてきた役割をさらに広げ、最適化モデリングの可能性を大きく拡大するものであった。多くの実際問題が「最適化問題を解くことに帰着される」のではなく「線形計画問題、凸 2 次計画問題、半正定値計画問題や 2 次錐計画問題という『最適化モデル』に帰着することにより」解かれるようになったのである。凸 2 次計画法が実用化されたのちに、サポートベクターマシンが現れ、また、圧縮センシングが線形計画問題の思いがけない応用として登場し、半正定値計画法は制御理論の世界を大きく変え、さらに、統計学、信号処理や種々の分野で統計的モデリングに活用されている。

そして、2000 年代に入り、ビッグデータの時代の到来とともに、機械学習における基本的道具立てとしての連続的最適化が改めて注目され、超大規模問題に対する最適化手法が精力的に研究され、また、離散的最適化においては、数千から数万変数の整数計画法の問題が日常的に解かれるようになり、様々な分野で活用されている。

このような数理計画分野の研究の歴史の中で、統計数理研究所においては、1955 年の改組において第 3 研究部が「OR および統計解析理論を研究する部署」として位置づけられ(統計数理研究所, 1994)、最適化とモデリングの学問としての数理計画法の研究が統計数理研究の一環として行われ、国際的にも通用する重要な研究成果を発信してきた。これらの成果の中から、本論説では、石井による、無限次元線形計画問題とチェビシェフ限界に関する研究、赤池による最急降下法の解析、田邊による Kaczmarz 法の解析、非線形最適化に関する微分幾何学的接近法、主双対内点法、そして、筆者を含む研究所の最適化研究者が国内外の共同研究者とともに行ってきた、内点法や最適化の情報幾何についての研究成果を中心として紹介する。なお、本論に登場する研究者の中で、統計数理研究所員については、以下で初めて登場した時に在職時期を付記している。

本論説には、「最適化」という言葉と「数理計画」という言葉が現れる。本論に進む前に、二つの言葉の意味するところの微妙な違いについて触れておきたい。最適化は数理の世界で古くからある言葉であるが、数理計画は Dantzig による線形計画法の登場後、使われるようになった言葉である。数理計画においては数理やアルゴリズムのみならず、モデリングが重要な研究対象となる。実際、数理計画の草創期には、線形計画問題が解けることを前提として、いかにして現実の問題を線形計画問題に落とし込むかというモデリングの手法が一つの重要な研究課題であった。より俯瞰的に見れば、豊かな構造を持つ基本的モデルがあり、実際の問題をそのモデルに帰着できる形にモデリングすることは、戦後勃興した数理科学諸分野において共通に見られた研究の流れであった。このように、数理計画法においては「最適化の数理やアルゴリズムを意識した」モデリングの研究が不可欠なものであり、数理計画法は最適化に関する「モデリング・数理・アルゴリズム」の 3 本の柱からなる学問分野である、と考えられる。このことは、Dantzig の古典的名著や Nesterov による定評ある教科書の序文でも強調されており、また、統計数理研究所で長年開催されてきた最適化の研究集会「最適化：モデリングとアルゴリズム」の

名称にはそのような気持ちが込められている。また、近年盛んに研究されている、機械学習分野における「最適化研究」も、これら3つの分野に跨っていることが見て取れる。

2. チェビシエフ型のバウンドと無限次元線形計画法

モーメントが与えられた時にそのモーメントを実現するような確率分布が存在するかどうかはモーメント問題と呼ばれ、古典的かつ重要な問題である。石井恵一(1955年-1964年在職)は1950年代後半から1960年代にかけて、この問題に連なる一連の数学的問題に取り組んだ(Isii, 1960, 1962, 1964)。その典型的な成果を以下に紹介する(Isii, 1962)。

Ω を全体集合、 \mathcal{U} をその上の σ 加法族とし、可測空間 (Ω, \mathcal{U}) 上での非負測度の族 \mathcal{B} が与えられているものとする。 $f_0(\omega), \dots, f_m(\omega)$ を Ω から R^{m+1} へのボレル可測写像で \mathcal{B} の各要素について積分可能とする。また、 $g(\omega)$ を \mathcal{B} の各要素について積分可能な実関数とする。

この設定の下で、次の条件を満たす (Ω, \mathcal{U}) 上の非負測度 $P \in \mathcal{B}$ を考える。即ち M_0, \dots, M_m を与えられた実数として、

$$(2.1) \quad E_P(f_i) \equiv \int f_i(\omega) dP(\omega) = M_i,$$

がすべての $i = 0, \dots, m$ について成立すると仮定する。この条件を満たす P の内、積分

$$\int g(\omega) dP(\omega)$$

の値をできるだけ大きくすることを考える。すなわち、

$$\mathcal{B}(M) = \{P | P \in \mathcal{B}, E_P(f_i) = M_i, \quad i = 0, \dots, m\}$$

として、次の最適化問題を考える：

$$(2.2) \quad \sup \int g(\omega) dP(\omega), \quad \text{subject to } P \in \mathcal{B}(M).$$

この問題(2.2)は、無限次元線形計画問題と捉えることができる。例えば、確率変数の平均や分散が与えられた時に、ある特定の領域にどの位まで大きな確率を割り当てることが可能であるか？といった、確率分布についての minimax 型の限界を与えるような問題(チェビシエフ型不等式)が(2.2)に帰着できる。

石井の得た主結果は下記の通りである。

定理 1. もし、 \mathcal{B} が凸ならば、集合

$$\mathcal{M} = \left\{ \int f dP(\omega) \mid P \in \mathcal{B} \right\} \quad (f = (f_0, \dots, f_m)^T)$$

は R^{m+1} 上の凸集合である。もし、 (M_0, \dots, M_m) が \mathcal{M} の内点であれば、(2.2)の最適値は、

$$(2.3) \quad \inf_{a_0, \dots, a_m} \left\{ \sum_{i=0}^m a_i M_i + \sup_{P \in \mathcal{B}} \int \left(g - \sum_i a_i f_i \right) dP \right\}$$

と書くことができる。さらに、もし、最適値が有限であれば、実際に最適値を達成する a_0, \dots, a_m が存在する。(2.2)には最適解(最適な測度 P)が存在するとは限らないが、もし、それが存在するならば、(2.3)を達成するものである。

石井は Isii (1964) でこの結果を拡張し、さらに、目的関数を非線形化することで、Cramér-Rao

の不等式を導出している。石井の成果はモーメント問題の古典的な結果として (Lasserre, 2010; Wolkowicz et al., 2000) 等良く知られている専門書に引用されている。また、最近でも, Minimax Probability Machine (Lanckriet et al., 2002) の構築など機械学習の文脈でも活用されており, 息の長い基礎的な成果である。

3. 最急降下法について

最急降下法は, 微分可能な関数 $f(x)$ を最小化するために,

$$x^+ = x - t\nabla f(x)$$

という反復を繰り返す基本的最適化手法である。この手法について, 赤池弘次(1952年-1994年在職)は, Akaike (1959)において, 現在では最適化分野では古典的な結果として広く知られ, 標準的な教科書 (Fletcher, 2000; Noceda and Wright, 2000) 等にも掲載されている, 以下の結果を示した。

「目的関数 f が(そのヘッセ行列の固有値がすべて異なるような)狭義凸 2 次関数の場合, 正確な直線探索を行うと, 最急降下法によって生成される反復列 x^k は, 漸近的に, f のヘッセ行列の最大固有値と最小固有値に対応する 2 つの固有ベクトルからなる 2 次元空間の方向から最適解に近づく。」

より正確には, 赤池は, 正則な係数行列を持つ連立一次方程式 $Ax = b$ を解く問題を, 正定値対称行列 P を用いて,

$$f(x) = (Ax - b)^T P(Ax - b)$$

の最小化問題と捉え, この問題に最急降下法を適用した時の反復過程を解析している。この時, 正確な直線探索を与える t は, $t = 2(Ax - b)^T P g / g^T P g$, $g = A^T P(Ax - b)$ と書ける。

赤池は, ヘッセ行列の固有ベクトルと関連づけて適切な座標を取り直して最急降下法の反復を解析した。新たな座標系で見た誤差ベクトル $(\alpha_1, \dots, \alpha_n)$ をもとに生成される, $(\alpha_1^2, \dots, \alpha_n^2) / \sum \alpha_i^2$ の各要素を生起確率とする多項分布を考え, 各反復でこの多項分布のパラメータが受ける変化を解析することで, 上記の結果を得ている。

赤池は, のちに, 本研究を回想する中で「この問題の取り扱いを通じて, 筆者は数値計算法全般に対する興味と, 関数の局地近傍の形状の二次曲面による近似の実用性の感覚を得た。これが後に筆者を統計モデルの評価に関する情報量規準の導入に誘うことになる。」(赤池, 1999) と記している。一見関係なさそうに見える, 最急降下法と AIC を関連づける形で赤池の研究が進められているのは印象的である。また, この論文の参考文献を見ると, Chernoff や Forsythe, Kantorovich ら統計学, 数値解析, 最適化の一流研究者の論文(のみ)が参照されており, 統計学・最適化・数値解析各分野の研究者が互いの問題に大きな関心を寄せていた, 数理科学が細分化される前の学際的な雰囲気を感じとれる。また, 数値実験が富士通 FACOM-128 というリレー計算機で行われていたことも記されており, その点でも当時の雰囲気を感じさせるものとして興味深い。

4. Kaczmarz 法の収束性

最適化と数値解析の境界領域となるが, 田邊國士(1967-2005年在職)は, 論文(Tanabe, 1971)において, 連立一次方程式に対する Kaczmarz 法の解析を行った。Kaczmarz 法は, 連立一次方程式に対する緩和法の一種であり, 元来正則な連立一次方程式系のための解法であったが, 田邊は変数と方程式の数異なるような場合まで含めた形で拡張して解析し, この方法に関す

る古典的結果を導びいた。田邊は複素数の場合について解析を行っているが、ここでは簡単のため、以下、係数行列や右辺は実数とする。

A を $m \times n$ 行列、 b を m 次元ベクトルとして、 n 変数の m 次元連立一次方程式

$$Ax = b$$

を考える。連立一次方程式 $Ax = b$ の解法として、次の反復を考えるのが Kaczmarz 法である。

まず、内部反復として、各方程式が定義する超平面への射影を逐次行う。 x^k が与えられたとしよう。内部反復は下記の通りとなる。

- (1) $x := x^k$.
- (2) $i = 1, \dots, m$ について、

$$x^+ := \operatorname{argmin}_{a_i^T y - b = 0} \|y - x\|_2; \quad x := x^+;$$

内部反復が終了したあとの x^+ を x^{k+1} として、同様にして反復を続ける。この時、適当な行列 Q, R を用いて、

$$x^{k+1} = Qx^k + Rb$$

と書ける。

Kaczmarz 法は、元々は、行列 A が正方で正則な場合の連立一次方程式の解法として提案され、解析されたが、田邊は、この方法の収束性を変数と方程式の次元が異なる場合を含め、ごく一般的な場合に解析し、以下の結果を得た。

「初期点 x^0 から出発した Kaczmarz 法の反復列は、次の点に収束する：

$$x^\infty = P_{\mathcal{K}}x^0 + Gb.$$

ここで、 $G = (I - Q)^{-1}R$ である。 G は行列 A のムーア・ペンローズ一般化逆行列となっている。 $P_{\mathcal{K}}$ は、 A の核 $\mathcal{K} = \{u | Au = 0\}$ への直交射影である。」

この結果は、方程式 $Ax = b$ が解を持つ場合には、任意の x^0 について x^∞ は解となり、さらに、 $x^0 = 0$ ととると、方程式が解を持たなくても、点列が最小二乗解に収束することを示している。

Kaczmarz 法は緩和法と呼ばれる方法の一種であり、この考え方は、凸集合の共通部分を求めるなどの手法に一般化されている。田邊の解析した方法は、コンピュータトモグラフィの画像再構成法にも活用され、現在では、機械学習の文脈で再び脚光を集めている。コンピュータトモグラフィがノーベル賞を受賞した時期には、多くの別刷り送付の依頼が寄せられた、とのことである。

5. 非線形計画問題に対する微分幾何的接近法

1960年代、連続的最適化の分野で、制約付き最適化問題に対する典型的接近法は、罰金関数法や障壁関数法と呼ばれるものであった。これは、目的関数に対して、罰金関数や障壁関数と呼ばれる、制約条件が満たされない点においては無限大かあるいは極端に大きな値をとるような補助関数を足したものを考え、この関数、つまり目的関数+補助関数、を無制約最適化することで、制約付き最適化問題を解くものである。罰金関数や障壁関数は、制約条件を破ることに對して罰金を科す補助的関数で、この関数を導入することで、制約付き最適化問題を無制約最適化問題に直すことができる。現在でもよく使われる方法であるが、最適解の近くで数値的

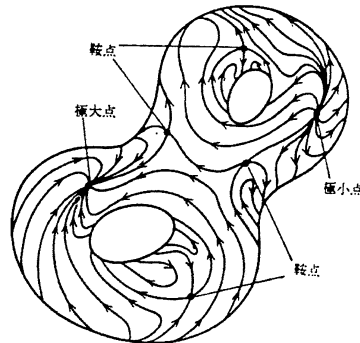


図 1. 射影勾配流の概念図(田邊, 1976, 図 4 を引用).

に悪条件となることが欠点とされていた。そのために、1970 年代に入り、拡大ラグランジュ関数などを用いて悪条件性を緩和した、乗数法などの解法が提案され、研究が進められた。

田邊は、同時期に、非線形計画問題に対し、微分幾何学的(力学系的)アプローチを試みていた。等号制約付き最適化問題

$$\max f(x), \text{ subject to } g_i(x) = 0, i = 1, \dots, p, x \in R^m$$

を考える。(不等号制約 $h(x) \leq 0$ は、新しい変数 s を導入して、 $h(x) + s^2 = 0$ と書き直せる。)以下、 ∇ は x に関する勾配を表すものとする。

目的関数 f の勾配を制約領域の接平面に射影したベクトル場を射影勾配と呼ぶことにする。制約条件を満たす領域上での f の射影勾配流(射影勾配の積分曲線)は概念的には図 1 のようになり、最適解は射影勾配流の停留点となる。射影勾配流を定義する微分方程式を数値的に解くことで、最適解を求めることができる。田邊はこの微分方程式の漸近的な挙動等について解析した(Tanabe, 1974)。これは、Rosen の射影勾配法の連続版である。

射影勾配流自身は、制約領域を満たさない点においても、 $g_i(x)$ の右辺を適切に(定数分)変更することで同様のものが定義できる。しかし、そのような初期値から出発して射影勾配流を数値的に積分しても制約条件を満たさないので、最適解を求めることはできない。一方、一般には、制約領域を満たさない点から出発しても、制約領域に近づくことができるように、制約条件を満たすべき方程式系 $g_i(x) = 0, i = 1, \dots, p$ 上の点を求める「ニュートン法」のベクトル場を定義することができる。この方程式系は変数の方が方程式の数よりも多い過少決定系であるが、適当な正則条件の下で、一般逆行列を用いることで、ニュートン法のベクトル場を定義できる。したがって、(いわばシフトした)射影勾配と制約領域を求めるニュートン法のベクトル場を加えたベクトル場を定義し、それを数値的に積分して最適解に至る解法を考えることができる(Yamashita, 1980)。田邊は(Tanabe, 1980)等でこれらのベクトル場が定義する力学系の漸近的な性質を詳しく解析した。

さらに、田邊は、

$$\min f(x), \text{ subject to } g_i(x) = 0, i = 1, \dots, p, x \in R^m$$

に対するカルシュ・キューン・タッカー条件(一次の最適性条件)

$$\nabla f + J(x)^T \lambda = 0, \quad g(x) = 0, \quad \left(J_{ij}(x) = \frac{\partial g_i}{\partial x_j} \right)$$

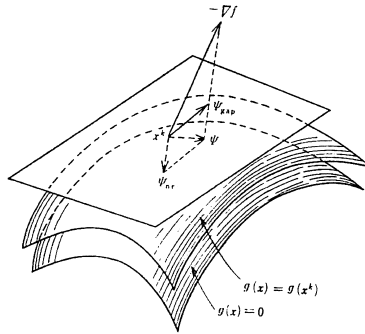


図 2. FIGAP 法 の 概 念 図 (田 辺, 1981, 図 4 を 引 用 ; 図 中 Ψ は 探 索 ベ ク ト ル, Ψ_{nr} は 制 約 条 件 を 満 た す た め の ニ ュ ー ト ン 方 向, Ψ_{gap} は 目 的 関 数 を 最 適 化 す る 方 向 で $\Psi = \Psi_{nr} + \Psi_{gap}$ と な っ て い る).

に対するニュートン法の反復式

$$(5.1) \quad \begin{pmatrix} \nabla^2 L(x, \lambda) & J^T(x) \\ J(x) & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \lambda^+ \end{pmatrix} = - \begin{pmatrix} \nabla f \\ g(x) \end{pmatrix}, \quad (L(x, \lambda) = f(x) + \lambda^T g(x))$$

において、(5.1)の左辺係数行列の $\nabla^2 L(x, \lambda)$ をさまざまな行列に置き換えることで、非線形最適化アルゴリズム設計の統一的指針を得ることができることを見出し、この考え方に基づいてFIGAP (Feasibility-Improving-Gradient-Accute-Projection)法を提案した(図 2) (Tanabe, 1981; 田辺, 1981).

これらのアルゴリズムの背後にある考え方は、罰金法や障壁関数法のように問題を無制約最適化に落とし込むのではなく、カルシュ・キューン・タッカー条件を満たす点が漸近安定点となるようなベクトル場を最急降下法やニュートン法などを考慮しつつ適切に定義し、そのベクトル場の積分曲線を離散的に追跡することにより最適化アルゴリズムを構築するというものである。田邊は折に触れてニュートン法の重要性を強調している。この思想は、ある意味でその後登場する内点法の考え方に通じるものである。

6. 内点法を巡って

線形計画法は、1947年、Dantzigによって創始されて以来、現在にいたるまで、数理計画法の中心にありつづけている。線形計画法の解法としてはDantzigによる単体法が定番であったが、1970年代に計算複雑度の理論が勃興し、線形計画問題に対して所謂多項式時間解法が存在するか否かが一つの大きな未解決問題となった。この問題は、Khachiyanにより、初めての多項式時間解法である楕円体法の登場によって肯定的に解かれた。しかし、楕円体法は理論的には数理計画法全体に大きく貢献したものの、実用解法ではなかった。1984年に射影変換と最急降下法を繰り返して問題を解く射影変換法が多項式性を有する実用解法としてKarmarkarによって提案され、数理計画全体に大きな影響をおよぼす潮流を20年余りに渡り作り出した。Karmarkarに端を発する一連のアルゴリズムは、実行可能領域の内部から最適解に近づいていく解法なので、内点法と呼ばれる。統計数理研究所では、内点法についての研究が活発に行われた。以下、成果を紹介する。

次の線形計画問題 (P)

$$\min c^T x, \text{ subject to } Ax = b, x \geq 0$$

と双対問題 (D)

$$\max b^T y, \text{ subject to } s = c - A^T y, s \geq 0$$

を考えよう. x, s の次元を n とし, A の行ベクトルは一次独立とする.

主問題 (P) の実行可能解 x における目的関数は, 双対問題の実行可能解 (s, y) における目的関数の値よりも大きいか等しい. これは, 以下のようにして簡単にわかる.

$$c^T x - b^T y = x^T c - x^T A^T y = x^T s \geq 0.$$

上式の左辺の量を双対ギャップという. 双対ギャップが 0 の場合には, その時の x と (s, y) は, それぞれの問題の最適解となる. $x^T s = 0$ は, $x \geq 0, s \geq 0$ の条件の下では, $x_i s_i = 0$ と等価であるから, 主問題と双対問題を解くことは, 次のような双線形等式・不等式系を解くことと等価である.

$$(6.1) \quad x \circ s = 0, Ax = b, s = c - A^T y, x \geq 0, s \geq 0.$$

ここで, $x \circ s$ はその各要素が x と s の各要素の積であるベクトルを表す. $x > 0, s > 0$ であるような実行可能解を内点実行可能解という. 以下, 主問題と双対問題に内点実行可能解があることを想定する.

内点法の研究は, Karmarkar による射影変換法に端を発する. (P) に対する射影変換法は, 実行可能領域上で, Karmarkar ポテンシャル関数

$$f(x) = (n+1) \log(c^T x - c_{\text{opt}}) - \sum_{i=1}^{n+1} \log x_i = \log \frac{(c^T x - c_{\text{opt}})^{n+1}}{\prod x_i}$$

が減少する方向に進むことを繰り返すことで (P) を解く反復解法である. ここで c_{opt} は (P) の最適値である. 実行可能領域上で $f(x)$ が十分に小さくなると, $c^T x$ は c_{opt} に十分に近い値となる. 反復の各ステップでポテンシャルの減少方向を求めるのに射影変換と最急降下法を用いることで, 多項式時間アルゴリズムを構築できる. Karmarkar ポテンシャル関数は, $-\sum \log x_i$ を含む点に特色がある. すなわち, この項があることにより, (最適解以外の) 実行可能領域の境界に近づくとき $f(x)$ はその値が無限に発散するため, $f(x)$ は, 実行可能領域の内側を通過して効率よく最適解に近づくためのガイドの役割を果たすことができる. この関数

$$\psi(x) = -\sum \log x_i,$$

を対数障壁関数という. 対数障壁関数は第一象限の境界に近づくにつれてその値が無限大に発散する凸関数である. 制約領域において, 対数障壁関数を最小化する点を解析的中心と呼ぶ. 対数障壁関数は内点法の理論において重要な役割を果たす.

正のパラメータ ν を導入し, 対数障壁関数を用いて, 次のような最適化問題を考える.

$$(6.2) \quad \min c^T x + \nu \psi(x) \text{ subject to } Ax = b, x \geq 0.$$

この問題の最適解は ν に依存する. これを $x(\nu)$ と記す. この問題は, $\psi(x)$ の項が不等式制約領域の境界に近づくにつれて $(x_i \rightarrow 0)$ で無限大に発散するため, 事実上不等式条件を無視できる最適化問題である. ν を小さくすれば, より境界に近いところに $x(\nu)$ は移動し, 直観的には $\nu \rightarrow 0$ で (P) の最適解に収束する. $x(\nu)$ の満たすべき条件は, 任意の $A\Delta x = 0$ なる Δx について,

$$\nabla(c^T x + \nu \psi(x))^T \Delta x = (c - \nu \text{diag}(x)^{-1} \mathbf{1})^T \Delta x = 0$$

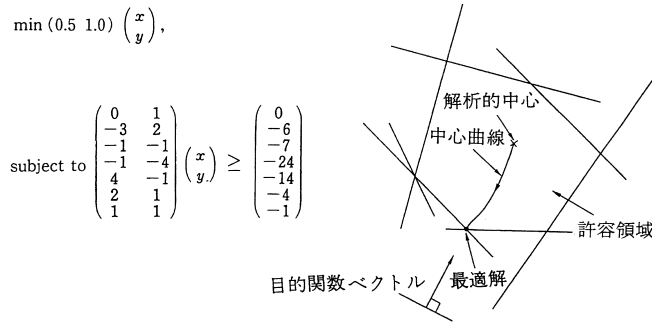


図 3. 内点法の概念図.

が成立することであるが、これは、

$$(6.3) \quad c - \nu \text{diag}(x)^{-1} \mathbf{1} = A^T y$$

がある y について成立することと同値である。ここで、ベクトル u に対し、 $\text{diag}(u)$ は、 u の各成分をその対角要素とする対角行列である。 ν を動かした時、 $x(\nu)$ は実行可能領域の内部を通り $\nu \rightarrow 0$ で (P) の最適解に収束する曲線となる。これを主中心曲線と呼ぶ。

さて、同様のことを双対問題について考えてみると、(6.2) に対応する問題は、

$$(6.4) \quad \max b^T y - \nu \phi(s), \quad s = c - A^T y$$

となる。双対問題においては、 $\phi(s) = -\sum \log(c_i - a_i^T y)$ が制約条件の境界に近づくにつれてその値が無限大に発散する障壁関数となる。パラメータ ν の時の最適解を $(s(\nu), y(\nu))$ と記すこととする。 $(s(\nu), y(\nu))$ の満たすべき最適性条件は、

$$(6.5) \quad b = \nu A \text{diag}^{-1}(s) \mathbf{1}, \quad s = c - A^T y$$

である。 ν を (正の範囲で) 動かした時、 $(s(\nu), y(\nu))$ は実行可能領域の内部を通り $\nu \rightarrow 0$ で (D) の最適解に収束する曲線となる。これを双対中心曲線と呼ぶ。

そこで、(P) を解くために、 ν を固定しておいて、 $x(\nu)$ を近似的にニュートン法によって求め、 ν の値を減らす、という手続きを繰り返して (P) の最適解を求める方法、あるいは、同様のことを (D) について行って (D) の最適解を求める線形計画問題の解法が考えられる。これが内点法の基本的な考え方である。前者を主内点法、後者を双対内点法という。図 3 に上で説明してきた内点法の概念を示す。

内点法の登場は衝撃的であった。次第に理論的にも実用的にも優れた解法であることが判明してくるにつれ、その研究のうねりは 80 年代後半から、20 年にわたり、最適化のみならず、隣接分野にも影響を及ぼす大きな潮流となっていった。後述するように、機械学習分野でサポートベクターマシンや圧縮センシングなどの新しいモデルを生み出す原動力ともなった。本節冒頭でも述べたように、統計数理研究所では、種々の内点法の解析や情報幾何との繋がりなどの研究が活発に行われた。

7. 主双対内点法

初期の内点法は、これまで説明してきたように、主問題あるいは双対問題のどちらかに点列を生成するものであったが、新しい立場の内点法アルゴリズムが、Megiddo によって示唆され、

田邊と小島・水野・吉瀬の 3 人によって、1987 年 2 月、統計数理研究所の研究集会「線形計画問題の新解法」において、世界で始めて、同時に提案された(小島 他, 1987; Tanabe, 1987a). 以下、その内容を述べる.

パラメータ $\nu > 0$ に対して、主中心曲線上でパラメータ値が ν である点 $x(\nu)$ と双対中心曲線上のパラメータ値が ν である点 $(s(\nu), y(\nu))$ は、実は、下記のような方程式の解となっていることが Megiddo によって指摘された(Megiddo, 1989).

$$x \circ s = \nu \mathbf{1}, \quad Ax = b, \quad s = c - A^T y, \quad x \geq 0, \quad s \geq 0.$$

この事実は(6.3)と(6.5)が満たしている関係を上式に代入することで簡単にわかる. この方程式は(6.1)の最初の方程式の右辺を 0 から各 i について等しく ν に緩めたものであるとも考えられる.

この事実は、主問題と双対問題を対等に扱う対称な解法が存在すること、そして、 $x \circ s$ の値を揃えながら減少させることで、効率良い解法が得られることを示唆する. つまり、障壁関数を用いて実行可能領域の境界から離れることにより、効率的な最適化を実現する、というメカニズムが、相補性条件 $x \circ s = 0$ を緩め、相補性に関する残差を揃えながら減少させる、という形でも実現しうることを示唆している.

このような背景の下で、田邊は、(Tanabe, 1987a, 1988)において、前節で発展させた、非線形最適化に関する微分幾何学的方法に「中心化」という、各方程式の残差を揃えながら減少させていく、という考え方を加えているいろいろな形に一般化し「中心化ニュートン法」という一連の解法を提案した. 特に、線形計画問題と凸 2 次計画問題については、現在主双対内点法として知られているものの具体形を与えた. そして、最適解への近さを計る尺度として、主双対ポテンシャル関数

$$f_T(x, s) = \rho \log(x^T s) - \sum_i \log(x_i s_i)$$

を導入した. ただし $\rho > n$ である. この関数は、

$$f_T(x, s) = (\rho - n) \log(x^T s) + g(x, s), \quad g(x, s) = n \log(x^T s) - \sum_i \log(x_i s_i)$$

と書き直すことができる. 簡単な解析により、 $g(x, s)$ は実行可能領域上では中心曲線上でのみ最小値 $n \log n$ をとることがわかる. そこで、田邊は $g(x, s)$ の値で中心曲線の近傍を定義することを提案した. また、上の表現より、 $\rho > n$ であれば、 $f_T(x, s)$ を $-\infty$ に減少させることで、双対ギャップ $x^T s$ を 0 に近づけることができ、(P) と (D) を解くことができることもわかる. したがって、主双対内点法の枠組みで、このポテンシャル関数を減少させるというアルゴリズムが考えられる. ポテンシャル関数 $f_T(x, s)$ は、現在では、Tanabe-Todd-Ye ポテンシャル関数、と呼ばれている. 田邊はさらに、主双対内点法の挙動の解析を連続ニュートン法や力学系の立場から行った. 図 4 に中心化ニュートン法の概念図を示す. 一方、田邊の発表と同時に終わった、小島・水野・吉瀬の発表は、線形計画問題に対する主双対内点法の計算複雑度を解析し、多項式性の証明を与える点に力点があった. 後年、現在広く使われている非実行可能点列内点法をニュートン法の立場から一早く提案したことも、田邊の注目すべき業績の一つである(田邊, 1989).

論文(Kojima et al., 1989)の著者の一人である水野真治(1990 年-1999 年在職)は研究所に合流後、さらに主双対内点法の研究を推進し、Todd や Ye 等と国際共同研究を行い、水野・Todd・Ye 予測子・修正子法(Mizuno et al., 1993)、非実行可能点列内点法に対する予測子・修正子法(Mizuno, 1996)、同次自己双対内点法(Ye et al., 1994)等の良く知られている成果を得た. 土

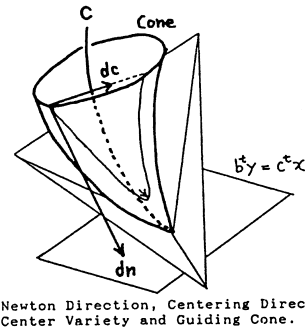


図 4. 中心化ニュートン法概念図(Tanabe, 1987a, Fig. 6 を引用).

谷隆(1986年-2010年在職)も、総合研究大学院大学学生であった村松正和と共に、基本的な内点法であるアフィンスケーリング法の大域的収束性を示し(Tsuchiya and Muramatsu, 1995)、このことをきっかけとして、Monteiro, Faybusovich, Megiddo, Terlaky, Roos 等と国際共同研究を展開した。また、El-Bakry, Tapia, Zhang と共に、一般の非線形最適化に対する主双対内点法の拡張についての論文を執筆した(El-Bakry et al., 1996)。論文(Mizuno, 1996; Mizuno et al., 1993; Ye et al., 1994; El-Bakry et al., 1996)は今でも折に触れて引用される標準的文献となっている。なお、本節で参照した田邊の中心化ニュートン法、5節および9節で紹介している最適化の微分幾何学に関する一連の研究は、共同研究レポート(田辺, 1996)にまとめられている。

8. 半正定値計画問題、2次錐線形計画への主双対内点法の拡張

線形計画問題に対する内点法の研究の進展が一段落した1990年代中盤より、線形計画問題における「変数の非負条件」を「変数がある凸錐に属している」という条件に拡張した最適化問題についての研究が進展した。その中でも典型的なものが、半正定値計画問題と2次錐計画問題であり、これらは以下のようにかけられる問題である。半正定値計画問題は、正定値対称行列上の最適化問題で、主問題は

$$\min \text{tr}(CX), \text{ subject to } \text{tr}(A_i X) = b_i, i = 1, \dots, n, X \text{ は半正定値対称行列}$$

と書ける。この問題の双対問題は、

$$\max \sum b_i y_i, \text{ subject to } S = C - \sum_i A_i y_i, i = 1, \dots, n, S \text{ は半正定値対称行列}$$

となる。また、2次錐計画問題は、線形計画問題における変数の非負条件を、変数がいくつかの2次錐の直積に属する、という条件に置き換えた、線形計画問題の拡張である。ここで2次錐とは、3次元円錐の高次元版で、一般にn次元2次錐は $\{(u_0, u_1) \in R \times R^{n-1} \mid \|u_1\| \leq u_0\}$ とかける集合である。具体的には、2次錐計画問題は下記のようにかける問題である。

$$\min c^T x, \text{ subject to } Ax = b, x \text{ はいくつかの2次錐の直積に属する}$$

この問題(主問題)の双対問題は、

$$\max b^T y, \text{ subject to } s = c - A^T y, s \text{ は上と同じ2次錐の直積に属する}$$

という問題である。線形計画問題と同様、主問題の任意の実行可能解での目的関数値は双対問

題の任意の実行可能解での目的関数値よりも大きいか等しく、等しければ、それぞれの問題の最適解となっている。また、緩い正則条件の下で主問題と双対問題の目的関数値が等しい最適解の存在が保証される。

半正定値計画問題や 2 次錐計画問題は、線形計画問題よりもより複雑な条件をモデリングすることができる。統計学、機械学習、制御、信号処理、最適設計、量子化学等多くの分野に応用を持ち、凸最適化の適用範囲を大きく広げた。線形計画問題の場合の主双対内点法を半正定値計画問題や 2 次錐計画問題に拡張することは、1990 年代後半の興味深い研究課題であった。土谷は、Monteiro とともに半正定値計画問題への主双対内点法について共同研究を進めるとともに (Monteiro and Tsuchiya, 1999), ユークリッド的ジョルダン代数を用いて 2 次錐計画問題への多項式時間主双対内点法の拡張を行った (Tsuchiya, 1999; Monteiro and Tsuchiya, 2000)。また、Faybusovich とともに、2 次錐計画問題に対する主双対内点法を無限次元に拡張した (Faybusovich and Tsuchiya, 2003)。Tsuchiya (1999), Monteiro and Tsuchiya (2000) は 2 次錐計画問題に関する標準的文献として折に触れて引用されている。

9. 内点法の情報幾何

情報幾何は、甘利と長岡によってその枠組みが作られた、統計学や機械学習等の解析に有効な微分幾何である。統計数理研究所は、情報幾何と内点法の両方に関心を有する研究者がいる世界でも数少ない場であったこともあり、内点法の情報幾何の研究が活発に展開された。

情報幾何では、凸ポテンシャル関数とそのヘッセ行列として定まるリーマン計量、そして、2 つの互いに双対な測地線が基本的な役割を果たす。さて、山下によって、主内点法や双対内点法の探索方向が、対数障壁関数を最小化するためのニュートン方向と、アフィンスケーリング方向という、アフィンスケーリング法の探索方向から成っていることが指摘されていた (Yamashita, 1986)。アフィンスケーリング方向は、対数障壁関数のヘッセ行列を計量とする目的関数の最急降下方向である。アフィンスケーリング法はカーマー法よりも 20 年近く前にツ連の Dikin によって提案されていた世界最初の内点法である。Bayer と Lagarias は、Legendre 変換によってこれらの内点法の探索方向が直線化されることを見出した (Bayer and Lagarias, 1989)。田邊は双対内点法に現れるベクトル場の積分曲線の挙動を解析し、その立場から同様の事実を見出し、この変換を Center Flattening Transformation と名づけた (Tanabe, 1987b)。さらに、田邊と土谷は双対内点法に以下のような情報幾何的な構造があることを指摘した (田辺・土谷, 1988)。すなわち、多面体上の対数障壁関数をポテンシャル関数として導入される双対平坦空間の構造を考えると、対数障壁関数のヘッセ行列によりリーマン計量が定義され、アフィンスケーリング法の連続版の軌跡は、情報幾何における双対測地線となっている。この時、主測地線は、元の空間での直線である。特に、中心曲線は、特別な双対測地線であり、これを離散的に近似的に辿ることによって、最適解を効率良く求めることができる。図 5 に図 3 の例題についてリーマン計量による“半径”1 の楕円(体)を、図 6 に双対測地線(アフィンスケーリング法の連続版の軌跡)を示す。図 5 の楕円体は Dikin 楕円体と呼ばれている。このような視点から、田邊は双対問題 (D) に対する中心化ニュートン法を提案し、関連する連続版の漸近的な挙動の解析を行った (田辺, 1990)。この線形計画問題に対する情報幾何の枠組みは、のちに小原敦美によって、半正定値計画問題に拡張された (小原, 1998)。そして、2000 年前後から、小原、土谷は柿原聡と共に凸最適化の情報幾何と内点法についての共同研究を行い、以下の成果を得た (Kakihara et al., 2013, 2014)。

線形計画問題の場合、凸最適化の情報幾何構造は以下ようになる。(P) の空間の第一象限の内部を Ω と記すことにして、 Ω に対する対数障壁関数 $\psi(x) = -\sum_{i=1}^n \log x_i$ を考える。これ

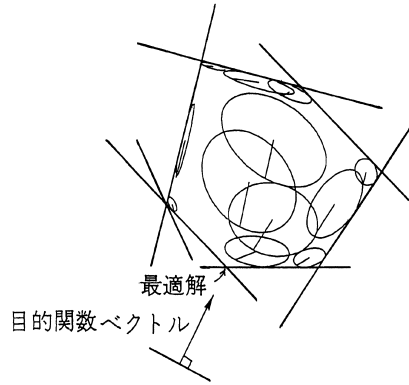


図 5. Dikin 楕円体.

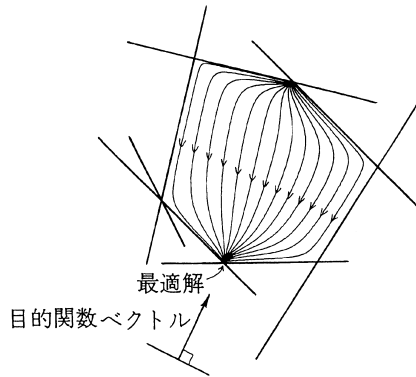


図 6. 双対測地線(連続版アフィンスケーリング法の軌跡).

が情報幾何の双対平坦空間を定めるポテンシャルとなる.

ルジャンドル変換 $s(x) = -\partial\psi(x)/\partial x$ を導入すると, $s(x)$ は Ω における非線形大域座標系となり, $s(\Omega) = \Omega$ となる. x を ∇ 接続に対するアファイン座標系, $s(x)$ を ∇^* 接続に対するアファイン座標系として, 双対平坦空間の構造が導入される(ここで ∇, ∇^* は勾配を意味する数学記号と異なる情報幾何の用語である). この空間のリーマン計量は $\partial^2\psi/\partial x^2$ で与えられる. 主問題の実行可能領域の内部 \mathcal{P} は Ω の部分多様体だが, 双対問題の実行可能領域の内部 \mathcal{D} も $s(x)$ を用いると Ω の部分多様体と考えることができる. x 座標系では \mathcal{P} は平坦だが \mathcal{D} は曲がっており, s 座標系では \mathcal{D} は平坦だが \mathcal{P} は曲がっている(図 7). この立場では, 中心曲線のパラメトライゼーションとして $t = 1/\nu$ を導入して, 中心曲線を

$$Ax = b, \quad s = c - A^T y, \quad x \circ s = \frac{1}{t} \mathbf{1}$$

と表すのが自然となる. この時, 主双対中心曲線を $(x(t), s(t), y(t))$ と記すとすると, $h_P(t)$ を主実行可能領域への主中心曲線 $x(t)$ の ∇^* 接続に関する埋め込み曲率の(リーマン計量の意味での)ノルム, $h_D(t)$ を, 双対実行可能領域への双対中心曲線の ∇ 接続の埋め込み曲率のノルムとすると, 中心曲線を追跡する, パス追跡型の主内点法の反復回数や双対内点法の反復回数が, 中心曲線上の積分

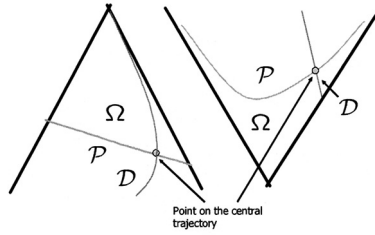


図 7. 最適化の情報幾何の概念図(両者は同じ問題を x と s による座標系で表現したものである).

$$I_P(\nu_1, \nu_2) = \frac{1}{\sqrt{2}} \int_{1/\nu_1}^{1/\nu_2} \sqrt{h_P(t)} dt,$$

$$I_D(\nu_1, \nu_2) = \frac{1}{\sqrt{2}} \int_{1/\nu_1}^{1/\nu_2} \sqrt{h_D(t)} dt,$$

を用いて, $I_P/\sqrt{\beta}$, $I_D/\sqrt{\beta}$ と書けることが分かる. ここで, β は, アルゴリズムが中心曲線を追跡する時にその中に反復列が生成される中心曲線の近傍の大きさを示すパラメータで, その値が小さいほど, より狭い近傍に留まりながら中心曲線を追跡することになり, 精密に曲線を追跡するため, 反復回数が大きくなる傾向がある. 定義からもわかるように, I_P , I_D は, 主中心曲線, 双対中心曲線の幾何学的性質を表す量である.

一方, パス追跡型の主双対内点法の反復回数についても, 同様の積分表示

$$I_{PD}(\nu_1, \nu_2) = \int_{1/\nu_1}^{1/\nu_2} \sqrt{h_{PD}(t)} dt$$

が求められ, 反復回数が $I_{PD}/\sqrt{\beta}$ とかけることが知られていた. この事実自身は (Monteiro and Tsuchiya, 2008) において情報幾何とは独立に導出されたことであるが, 一見情報幾何とは無関係に導出された量であった h_{PD} が実は, h_P と h_D を用いて定義される, 下記のような情報幾何学的量

$$(9.1) \quad h_{PD}^2 = h_P^2 + h_D^2$$

として書けるといふ少々驚くべき結果が (Kakihara et al., 2013) において得られた. これは「ピタゴラスの定理(関係)」を想起させる興味深い関係である. このように, 主双対内点法の反復回数を表す積分も情報幾何的な量として書くことができることが明らかとなった.

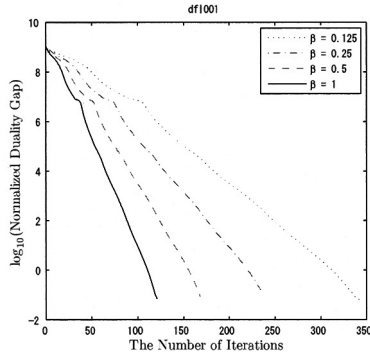
また, 論文 (Monteiro and Tsuchiya, 2008) では, L_A を (P), (D) の係数行列 A の入力ビット数として, 中心曲線全体に渡る I_{PD} の積分値が存在し, $I_{PD}(0, \infty) = O(n^{3.5} L_A)$ であることが示されている. この結果と上のピタゴラス関係を用いると, $I_P(0, \infty)$, $I_D(0, \infty)$, $I_{PD}(0, \infty)$ について,

$$\max[I_P(0, \infty), I_D(0, \infty), I_{PD}(0, \infty)] = O(n^{3.5} L_A)$$

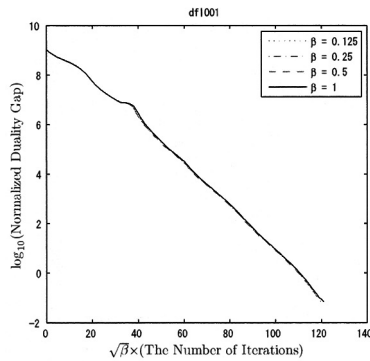
という評価が成立し, 特に, ネットワークフロー問題については, $L_A = mn$ で抑えられることより,

$$\max[I_P(0, \infty), I_D(0, \infty), I_{PD}(0, \infty)] = O(n^{4.5} m)$$

となる. これらは, 中心曲線の大域的な幾何学的性質を明らかにした結果である.



(a) 反復回数と正規化双対ギャップ ν



(b) (反復回数 $\times \sqrt{\beta}$) と正規化双対ギャップ ν

図 8. DFL001 に対する主双対内点法の挙動.

この積分が反復回数を非常によく近似していることを実用レベルの著名なベンチマーク問題 DFL001 を例にして紹介する (Kakihara et al., 2014). DFL001 は制約式の数が 6072, 変数の数が 12230 の線形計画問題である. この問題に対して主双対内点法を行ったときの反復回数と双対ギャップの減少の関係が図 8(a) である. β を変えると, 反復回数と ν の値の関係は, 似た形をとり, 横軸を適当にスケールすれば, 重なりそうである. 反復回数を $\sqrt{\beta}$ 倍したものは積分値そのものになるので, 重なることが期待される. 実際, 反復回数に $\sqrt{\beta}$ かけたものを横軸に取り直してグラフをプロットすると重なることがわかる (図 8(b)). 積分そのものは $\beta = 1$ とした時の反復回数であり, このように, 実験的にも「内点法の反復回数は情報幾何的積分量である」ことが示されている. この結果は半正定値計画問題や 2 次錐計画問題にも拡張された (Kakihara et al., 2014).

10. 最適化モデリング, さまざまな研究

統計数理研究所で行われてきた最適化モデリングのひとつの典型例は, パラメータ推定問題を,

$$(10.1) \quad \min f(x, \theta) + \lambda g(\theta)$$

の形に帰着させるものである. x がデータ, θ がパラメータである. λ はハイパーパラメータ

である。直観的には、 f がデータへのモデルのあてはまりをあらわす(対数)尤度関数、 g はパラメータについての事前情報を表現する正則化項である。これは、現在では、Lasso などのスパースモデリングの形式として幅広く研究されている枠組みであるが、パラメータ数がデータよりも多いモデルも取り扱うことができ、柔軟なモデリングが可能となる点が長所である。統計数理研究所では、このタイプのモデリングが、赤池を中心に、1980 年代にベイズモデルの枠組みで精力的に研究されてきた。田邊は、共同研究者とともに、最適化を意識した立場から、地震波による地球内部構造推定や肺換気能の解析、肺気道の導通度の解析等にこの枠組みを積極的にもちいたモデリングを行い、先駆的な成果を得ている(Inoue et al., 1990; Kobayashi et al., 1987; Kawashiro et al., 1991)。特に Inoue et al. (1990) は地震波によるトモグラフィの基本的文献として今でも参照され続けている。

2 次錐計画問題に対する主双対内点法が開発された時には、リアモーターカーの磁気シールドの最適設計問題が 2 次錐計画問題に定式化され、同手法で解かれている(Sasakawa and Tsuchiya, 2003)。これは、世界で最初に大規模な 2 次錐計画問題に対して主双対内点法を適用した例となった。解法の高速度により、多数回、ランダムに摂動を加えた問題を解きなおすことによって、設計されたシールドのロバストネスを評価することなども可能となった(土谷・笹川, 2005)。

ガウシアングラフィカルモデルの最尤推定は、半正定値計画問題に対する内点法と密接に関連した凸最適化問題に帰着する。この事実を用いて、上野玄太(2003 年より在職)と土谷は、スーパーコンピュータを用いて、海洋データ同化に現れる行列サイズ約 3 万 \times 約 3 万、パラメータ数 10 万強の大規模分散共分散行列の推定を行った(Ueno and Tsuchiya, 2009)。

北川源一郎(1974 年-2011 年在職)が提案した粒子フィルタ(モンテカルロフィルタ, Kitagawa, 1996)は、統計数理研究所で開発され現在も幅広く使われている統計手法の一つである。土谷と中村和幸(当時総合研究大学院在学)は、関数の勾配を計算するための自動微分法のメモリ節約の手法が粒子フィルタのスムージングへ適用できることを指摘し、スムージングに使える粒子数を格段に増やすことができることを示した(Nakamura and Tsuchiya, 2007)。一風変わった形ではあるが、これも最適化と統計科学の交わった研究であるといえよう。

また、上田澄江(1970 年-2010 年在職)・牧野久実・伊藤栄明(1972 年-2007 年在職)と土谷は、2000 枚程度現在存在する、メソポタミア粘土板データに登場する人物の索引から、当時の有力人物の生存年代や人口動態、文書の成立した順番等を推定する問題を最適化モデリングの立場から扱った。内点法を用いて大規模凸 2 次計画問題を解くことで推定が可能となり、現在、さらに解析が進められている(Ueda et al., 2014)。

2000 年代の後半から、機械学習への最適化の適用が活発に行われてきた。この文脈では、伏木忠義(2003 年-2014 年在職)・土谷・堀内によって半正定値計画問題を用いた密度関数推定の手法が開発されている(Fushiki et al., 2006)。田邊は Kernel 法に触発され、双対理論の枠組みを活用する形で Dualized-Penalized-Logistic-Regression-Machine を提案し、松井知子(2003 年より在職)らと共にさまざまな分野への適用を試みている(田邊, 2007; Matsui and Tanabe, 2006)。また、武田朗子(2016 年-2018 年在職)や田中未来(2017 年より在職)等によって、DC(Difference of Convex functions)計画法の機械学習への適用や、非負行列分解の応用などの研究が進められている(Fujiwara et al., 2017; Mizutani and Tanaka, 2018)。

ここまで、幾つか焦点を絞って、最適化に関連する成果を中心に述べてきたが、統計数理研究所では、他にもさまざまな形で最適化に関する研究が進められてきた。例えば、半無限最適化計画問題は、有限次元のパラメータの最適化を無限個の制約の下で最適化する問題であり、凸錐上の線形計画問題を含む興味深い最適化問題のクラスであるが、この問題について、伊藤聡はしばしば国際共同研究を行い、さまざまな手法の開発に取り組んできた(Ito et al., 2000)。

本稿ではあまり取り上げられていなかったが、離散的最適化問題、特に整数計画問題は、重要な分野である。この分野については、モデリングの観点から、吉本敦(1999年-2003年在職、2008年より再度在職)によって、森林管理の最適計画などの研究が進められている(Yoshimoto et al., 2017)。また、伊藤聡(1991年より在職)らによってスポーツのリーグ戦におけるクリンチナンバーなどの計算もいち早く進められ、実用に供されている(伊藤, 2012; Ito and Shinano, 2018)。

11. 研究集会等

1944年に創立された統計数理研究所は1985年に改組され、国立大学共同利用機関となり、その後、大学共同利用機関となった。これに伴い、共同研究集会や、共同研究レポートの発行など、共同利用の体制が整えられた。先に述べた、内点法の研究が契機となって、1986年度に「線形計画問題の新解法」という研究集会が開催された。先に述べたように、この研究集会において、その後世界中で使われることとなった主双対内点法が、田邊と小島・水野・吉瀬より世界で初めて提案された。研究集会「線形計画問題の新解法」は、1988年度まで開催され、1989年度には「非線形最適化：モデリングとアルゴリズム」と名前を変えて引き継がれ、1990年度以降は「最適化：モデリングとアルゴリズム」の名前で、東日本大震災のあった2010年度を除いてずっと継続して開催されてきた。長年続いてきた最適化分野での研究集会として、若手を中心に多くの興味深い優れた研究成果が発表されてきた。各研究集会での発表は、共同研究レポートとしてまとめられている。第1節でも述べたように、最適化を意識したモデリングが重要である、ということが、本研究集会の開催趣旨の一つである。その考えを反映した形で「シリーズ：最適化モデリング」が立ち上げられ、現在刊行が進められている(赤池弘次 他, 2015; 山下浩 他, 2016; 池上敦子, 2018)。

これらの国内研究集会と同期する形で、国際研究集会が、1996年と2001年の2回、文部科学省からの予算を得て開催された。1996年に開催された国際研究集会“International Symposium on Optimization and Computation”は、湘南国際村で開催され、最適化に関する一流研究者が世界中から集い、研究成果を発表し、議論・情報交換を行った。2001年に京都の京大会館にて開催された研究集会“New Trends on Optimization and Computational Algorithms”では、最適化・機械学習・統計科学の一流研究者が集い、学際的交流を図った。これは、最適化と機械学習の交流を意識して開催された、最初期の研究集会の一つとして認識され、成果がMathematical Programming Series B, そして、Annals of the Institute of Statistical Mathematicsの特集として出版されている。また、2017年度より、統計数理研究所は九州大学のInstitute of Mathematics for Industry, ドイツのZuse Institute Berlin (ZIB)と共同して、最適化に力点を置いたワークショップISM-ZIB-IMI MODAL Workshopをもち回りで開催している。

1988年に、Mathematical Programming Societyにより3年に一度開催される、世界で最大の最適化関係の集会International Symposium on Mathematical Programmingが東京春日の中央大学で開催された。この時に、田邊は国際プログラム委員長の重責を担い、シンポジウムの成功に貢献した。28年後の2016年、Mathematical Optimization Society(旧Mathematical Programming Society)による、連続的最適化に関する最大の国際研究集会International Conference on Continuous Optimizationが、東京六本木の政策研究大学院大学で開催された。この時に、組織委員長を水野が、共同実行委員長を水野と土谷と村松が努めた。また、伊藤、武田と田中が実行委員会のメンバーとなった。統計数理研究所で若手時代に最適化の研究を行ってきた研究者が第一級の国際学会を日本に招致しそれを開催する中心となったことは、人材育成を担ってきた研究所の役割として、記憶に留められてもよいことではないかと思う。

12. 将来

最適化は、統計数理の新しい潮流と密接な関係を保ち続けている。1990 年代の後半から、サポートベクターマシン (SVM) が新しい判別機械として導入され、現在では完全に定着しているが、SVM の開発には、凸 2 次計画法が実用化されることが重要であった。主双対内点法によって大規模な凸 2 次計画問題が解けるようになって、はじめて SVM が実用化されたといえよう。SVM で重要な役割を果たしたスパース性の概念は、その後、Candes, Tao らによる圧縮センシング等、スパースモデリングの展開につながっていく。そこでは、線形計画法が重要な役割を果たし、2 次錐計画や半正定値計画も大きく寄与している。スパースモデリングは (10.1) に帰着するため、多くの最適化アルゴリズムの研究がさまざまな立場から行われた。ここに、1950 年代の赤池の論文から読み取れるような、数理計画と統計科学の実り多い交流を再び見出すことができるのである。アルゴリズムの進展が新しいモデルの実用化を促す好例である。スパースモデリングは、ブラックホールの撮像にも重要な役割を果たした。

一方、統計科学や機械学習の現場で使われるモデルが最適化研究を促すこともある。まず、機械学習における最適化問題は超大規模で、ヘッセ行列を必要とする、ニュートン法系統の解法は困難である。そのことにより、Nesterov に代表される、最急降下法を加速する手法や、確率的降下法などが提案されている。最近急激に研究が進んでいる深層学習においては、ReLU 関数の合成関数により関数が表現され、バックプロパゲーション (最急降下法) によってパラメータ最適化が行われている。ところが、ReLU 関数は原点で微分不可能であり、勾配を計算することが困難である。非凸関数であるために劣微分の理論も簡単ではない。このような問題についても、新しい光があたりつつある、といってもよいであろう。

21 世紀に入り、ビッグデータの時代を迎え、社会現象、経済現象、気象現象、災害、環境問題、マーケティング、社会保障等、科学技術の新しいフロンティアとチャレンジは、より複雑で不確実性を含んだ問題となりつつある。その中で予測とモデリングの科学としての統計科学は大きな貢献を成し得るものと考えられる。この文脈では、予測のみならず、モデルに基づいた制度設計や制御が必須となる。そこで重要な役割を果たすと考えられるのが、整数計画法と確率計画法である。これらの分野にも目配りをしながら、短期的な成果を求めず、腹を括って長期的な視点に立って総合的な立場で視野の広い人材を育成していくことが統計数理研究所における最適化の研究の展開にとっては重要であろう、と考えるし、そのような方向で研究所が発展することを望みたい。特に研究所の若手研究者諸賢におかれては、地に足のついた問題意識を持ち、統計数理研究所の伝統と文化を受け継ぎつつ、自らの力で少しずつでも着実に、統計数理の可能性を拓げる研究成果を積み上げていかれるよう、研究所で良き時代を過ごした一在職経験者として、心より願っている。

参 考 文 献

- Akaike, H. (1959). On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method, *Annals of the Institute of Statistical Mathematics*, **11**, 1-16.
- 赤池弘次 (1999). 応用数理の遊歩道 (17): 統計的思考と応用数理-形式的な確率的構造の利用, *応用数理*, **9**, 169-171.
- 赤池弘次 他 (2015). 『モデリング—広い視野を求めて—』 (室田一雄, 池上敦子, 土谷隆 編), 最適化モデリング, 第 1 巻, 近代科学社, 東京.
- Bayer, D. A. and Lagarias, J. C. (1989). The nonlinear geometry of linear programming II. Legendre transform coordinates and central trajectories, *Transactions of the American mathematical society*, **314**, 527-581.

- El-Bakry, A., Tapia, R. A., Tsuchiya, T. and Zhang, Y. (1996). On the formulation and theory of the Newton interior-point method for nonlinear programming, *Journal of Optimization Theory and Applications*, **89**, 507-541.
- Faybusovich, L. and Tsuchiya, T. (2003). Primal-dual Algorithms and Infinite-dimensional Jordan Algebras, *Mathematical Programming*, **97**, 471-493.
- Fletcher, R. (2000). *Practical Methods of Optimization*, John Wiley & Sons, Chichester.
- Fujiwara, S., Takeda, A. and Kanamori, T. (2017). DC algorithm for extended robust support vector machine, *Neural Computation*, **29**, 1406-1438.
- Fushiki, T., Horiuchi, S. and Tsuchiya, T. (2006). A maximum likelihood approach to density estimation with semidefinite programming, *Neural Computation*, **18**, 2777-2812.
- 池上敦子 (2018). 『ナーススケジューリング』(室田一雄, 池上敦子, 土谷隆 編), 最適化モデリング, 第3巻, 近代科学社, 東京.
- Inoue, H., Fukao, Y., Tanabe, K. and Ogata, Y. (1990). Whole mantle P-wave travel time tomography, *Physics of the Earth and Planetary Interiors*, **59**(4), 294-328.
- Isii, K. (1960). The extrema of probability determined by generalized moments (I) bounded random variables, *Annals of the Institute of Statistical Mathematics*, **12**(2), 119-134.
- Isii, K. (1962). On sharpness of Tchebycheff-type inequalities, *Annals of the Institute of Statistical Mathematics*, **14**(1), 185-197.
- Isii, K. (1964). Inequalities of the types of Chebyshev and Cramér-Rao and mathematical programming, *Annals of the Institute of Statistical Mathematics*, **16**(1), 277-293.
- 伊藤聡 (2012). クリンチとエリミネーションの数理〜リーグスポーツにおけるロバスト最適化, システム/制御/情報, **56**(7).
- Ito, S. and Shinano, Y. (2018). Calculation of clinch and elimination numbers for sports leagues with multiple tiebreaking criteria, *ZIB Report* 18-51, Zuse Institute Berlin.
- Ito, S., Liu, Y. and Teo, K. L. (2000). A dual parametrization method for convex semi-infinite programming, *Annals of Operations Research*, **98**, 189-213.
- Kakihara, S., Ohara, A. and Tsuchiya, T. (2013). Information geometry and interior-point algorithms in semidefinite programs and symmetric cone programs, *Journal of Optimization Theory and Applications*, **157**, 749-780.
- Kakihara, S., Ohara, A. and Tsuchiya, T. (2014). Curvature integrals and iteration complexities in SDP and symmetric cone programs, *Computational Optimization and Applications*, **57**, 623-665.
- Kawashiro, T., Yamasawa, F., Okada, Y., Kobayashi, H. and Tanabe, K. (1991). Uneven distribution of ventilation-perfusion ratios in lungs estimated by a modified Newton method, *Mathematical Programming*, **52**(1-3), 1-9.
- Kitagawa, G. (1996). Monte Carlo filter and smoother for Non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics*, **5**(1), 1-25.
- Kobayashi, H., Abe, T., Kawashiro, T., Tanabe, K. and Yokoyama, T. (1987). Estimation of the distribution profile of airway resistance in the lungs, *Computers and Biomedical Research*, **20**(6), 5070-5525.
- 小島正和・水野眞治・吉瀬章子 (1987). 多項式オーダの主対内点法. 統計数理研究所共同研究レポート 5, “線形計画問題の新解法,” 13-24.
- Kojima, M., Mizuno, S. and Yoshise, A. (1989). A primal-dual interior point algorithm for linear programming, *Progress in Mathematical Programming* (ed. N. Megiddo), 29-47, Springer-Verlag, New York.
- Lanckriet, G. R. G., Ghaoui, L. El, Bhattacharyya, C. and Jordan, M. I. (2002). A robust minimax approach to classification, *Journal of Machine Learning Research*, **3**, 555-582.
- Lasserre, J.-B. (2010). *Moments, positive polynomials and their applications*, Imperial College Press, London, UK.
- Matsui, T. and Tanabe, K. (2006). Comparative study of speaker identification methods: DPLRM,

- SVM and GMM, *IEICE Transactions on Information and Systems*, **E89-D**, 1066-1073.
- Megiddo, N. (1989). Pathways to the optimal set in linear programming. *Progress in Mathematical Programming* (ed. N. Megiddo), 131-158, Springer-Verlag, New York.
- Mizuno, S. (1996). Polynomiality of infeasible-interior-point algorithms for linear programming, *Mathematical Programming*, **67**(1-3), 109-119.
- Mizuno, S., Todd, M. J. and Ye, Y. (1993). On adaptive-step primal-dual interior-point algorithms for linear programming, *Mathematics of Operations Research*, **18**, 964-981.
- Mizutani, T. and Tanaka, M. (2018). Efficient preconditioning for noisy separable nonnegative matrix factorization problems by successive projection based low-rank approximations, *Machine Learning*, **107**, 643-673.
- Monteiro, R. D. C. and Tsuchiya, T. (1999). Polynomial convergence of a new family of primal-dual algorithms for semidefinite programming, *SIAM Journal on Optimization*, **9**, 551-577.
- Monteiro, R. D. C. and Tsuchiya, T. (2000). Polynomial convergence of primal-dual algorithms for the second-order cone program based on the MZ-Family of directions, *Mathematical Programming*, **88**, 61-83.
- Monteiro, R. D. C. and Tsuchiya, T. (2008). A strong bound on the integral of the central path curvature and its relationship with the iteration complexity of primal-dual path-following LP algorithms, *Mathematical Programming*, **115**, 105-149.
- Nakamura, K. and Tsuchiya, T. (2007). A recursive recomputation approach for smoothing in nonlinear state-space modeling: An attempt for reducing space complexity, *IEEE Transactions on Signal Processing*, Vol. 55, 5167-5178.
- Nesterov, Y. (2004). *Introductory Lectures on Convex Optimization*, Boston, Dordrecht, London.
- Nocedal, J. and Wright, S. (2000). *Numerical Optimization*, Springer-Verlag.
- 小原敦美 (1998). 半正定値計画問題に関する情報幾何を用いた考察, 統計数理, **46**(2), 317-334.
- Sasakawa, T. and Tsuchiya, T. (2003). Optimal magnetic shield design with second-order cone programming, *SIAM Journal on Scientific Computing*, **24**(6), 1930-1950.
- Tanabe, K. (1971). Projection method for solving a singular system of linear equations and its applications, *Numerische Mathematik*, **17**(3), 203-214.
- Tanabe, K. (1974). An algorithm for the constrained maximization in nonlinear programming, *Journal of the Operations Research Society of Japan*, **14**, 184-201.
- 田辺國士 (1976). アナログ型非線形計画法, 数理科学, **157**, 45-51.
- Tanabe, K. (1980). Geometric method in nonlinear programming, *Journal of Optimization Theory and Applications*, **30**, 181-210.
- Tanabe, K. (1981). Feasibility-improving gradient-acute-projection methods: A unified approach to nonlinear programming, *Lecture Notes in Numerical and Applied Analysis*, **3**, 57-76.
- 田辺國士 (1981). 最適化問題とニュートン法, 数理科学, **218**, 71-76.
- Tanabe, K. (1987a). Complementarity-enforcing centered Newton method for linear programming, 「線形計画問題の新解法」, 統計数理研究所共同研究リポート, No. 5, 118-144. 統計数理研究所, 東京.
- Tanabe, K. (1987b). Center flattening transformation and a centered Newton method for linear programming, 日本オペレーションズ・リサーチ学会数理計画研究部会 1987 年 7 月部会発表資料, ((田辺國士, 1996) 統計数理研究所共同研究リポート, No. 89, 中心化ニュートン法と最適化の微分幾何, 1996 年, 5-18.), 統計数理研究所, 東京.
- Tanabe, K. (1988). Centered Newton method for mathematical programming, *System Modelling and Optimization* (eds. M. Iri and K. Yajima), Lecture Notes in Control and Information Sciences, Vol. 113, 197-206, Springer, Berlin, Heidelberg.
- 田邊國士 (1989). Centered Newton method for linear programming: Exterior point method, 統計数理, **37**(1), 203-204.
- 田辺國士 (1990). 最適化の微分幾何学, 応用数理, 創刊準備第 3 号, 39-50.
- 田辺國士 (1996). 中心化ニュートン法と最適化の微分幾何, 統計数理研究所共同研究リポート, No. 89,

- 統計数理研究所, 東京.
- 田邊國士 (2007). 帰納推論機械 PLRM と dPLRM : 方法論, モデル, アルゴリズムおよび応用, システム/制御/情報, **51**(2), 87-95.
- 田辺國士, 土谷隆 (1988). 線形計画の新しい幾何学, 数理科学, **303**, 32-37.
- 統計数理研究所 (1994). 統計数理研究所 50 年のあゆみ, 統計数理研究所, 東京.
- Tsuchiya, T. (1999). A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming, *Optimization Methods and Software*, **11, 12**, 141-182.
- Tsuchiya, T. and Muramatsu, M. (1995). Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems, *SIAM Journal on Optimization*, **5**(3), 525-551.
- 土谷隆, 笹川卓 (2005). 2 次錐計画問題による磁気シールドのロバスト最適化, 統計数理, **53**, 297-315.
- Ueda, S., Makino, K., Itoh, Y. and Tsuchiya, T. (2014). Logistic growth for the Nuzi cuneiform tablets: Analyzing family networks in ancient Mesopotamia, *Physica A*, **421**, 223-232.
- Ueno, G. and Tsuchiya, T. (2009). Covariance regularization in inverse space, *Quarterly Journal of the Royal meteorological Society*, **135**(642), 1133-1156.
- Wolkowicz, H., Saigal, R., Vandenberghe, L. (eds.) (2000). *Handbook of Semidefinite Programming*, Kluwer, Boston, Dordrecht, London.
- Yamashita, H. (1980). A differential equation approach to nonlinear programming, *Mathematical Programming*, **18**, 155-168.
- Yamashita, H. (1986). *A Polynomially and Quadratically Convergent Method for Linear Programming*, Mathematical Systems Inc., Tokyo.
- 山下浩 他 (2016). 『モデリングの諸相—OR と数理科学の交叉点—』(室田一雄, 池上敦子, 土谷隆 編), 最適化モデリング, 第 5 巻, 近代科学社, 東京.
- Ye, Y., Todd, M. J. and Mizuno, S. (1994). An $O(\sqrt{n}L)$ -iteration homogeneous and self-dual linear programming algorithm, *Mathematics of Operations Research*, **19**(1), 53-67.
- Yoshimoto, A., Asante, P., Konoshima, M. and Surovy, P. (2017). Integer programming approach to control invasive species spread based on cellular automaton model, *Natural Resource Modeling*, **30**(2), NRM12101.

A Historic Review of Research on Optimization at the Institute of Statistical Mathematics

Takashi Tsuchiya^{1,2}

¹National Graduate Institute for Policy Studies

²The Institute of Statistical Mathematics

Optimization is an important discipline in statistical science. Broadening the frontier of tractable optimization problems directly leads to innovation of statistical science. This is a main reason why optimization research has been conducted at the Institute of Statistical Mathematics. In this paper, we review historic development of optimization research at the institute, and its contribution to statistical sciences. The topics include duality theory of infinite-dimensional linear programming and its application to statistics, convergence analysis of the steepest descent algorithm and Kaczmarz method, differential-geometric approach to nonlinear programming, interior-point algorithms and information geometry for linear programming, second-order cone programming and semidefinite programming, and various applications.

標準コウホート表のコウホート分析モデルの デザイン行列について

中村 隆†

(受付 2019 年 5 月 17 日；改訂 10 月 10 日；採択 10 月 17 日)

要 旨

継続調査で得られる年齢区分×調査時点形式の標準コウホート表データから年齢・時代・世代(コウホート)要因の効果を分離するコウホート分析モデルについて、3 要因のデザイン行列を陽な表現により与えた。デザイン行列を、目視により与えるとともに、コウホート表のセルに 1 対 1 で対応するセルパラメータに等値制約と同値のゼロ和制約を課すことにより導出した。

キーワード：APC モデル，年齢・時代・世代効果，セルパラメータ，等値制約，ゼロ和制約，継続(反復横断)調査。

1. はじめに

コウホート分析(cohort analysis or age-period-cohort analysis)は、継続調査(反復横断調査)から得られる何らかの調査項目の年齢区分×調査時点形式の集計表データから、年齢・時代・コウホート(世代)効果を分離しようとする統計的方法である。3 効果がうまく分離できれば、社会の変化の構造に関する知見が得られ、将来の予測に資することができる(Ryder, 1965; Mason and Fienberg, 1985; Glenn, 2005)。

しかしながらコウホート分析には、何らかの付加条件を与えなければ 3 効果を分離できないという識別問題が存在し、付加条件の与え方とその評価を巡って長年に渡り議論が続けられてきた。近年の成書に、Yang and Land (2013), O'Brien (2015), Fu (2018)がある。

中村 (1982), Nakamura (1986)は、統計数理研究所が 1953 年以降 5 年ごとに継続実施している日本人の国民性調査データへの適用を目的として、コウホート分析の識別問題を克服するパラメータの漸進的変化の条件(1 次階差制約)を取り入れ、赤池ベイズ型情報量規準(ABIC)最小化により最適モデルを選択するベイズ型コウホートモデルを開発した。中村 (2005)では、年齢×時代の交互作用効果をもつコウホートモデルに拡張している。

さて、コウホート分析モデルのデザイン行列を具体的に例示したものは散見される(たとえば、Fienberg and Mason, 1979, p.21)ものの、行列表示によって明示的に示したのは坂口・中村 (2019)が最初である。ただし、これは目視によるものである。

以下では、陽には示されてこなかったコウホート要因のデザイン行列の導出を行う。2 節では、年齢・時代・世代 3 要因のデザイン行列を目視により与える。3 節では年齢と時代要因のデザイン行列を、4 節ではコウホート要因のデザイン行列を、コウホート表のセルに 1 対 1 で

† 統計数理研究所 名誉教授：〒190-8562 東京都立川市緑町 10-3

対応するセルパラメータに等値制約を課すことにより導出する。付録に、導出の過程に有用なベクトルと行列の定義と操作についてまとめる。

2. コウホートモデル

2.1 標準コウホート表

継続調査から得られる年齢区分×調査時点形式の集計表は、コウホート分析の観点から「コウホート表」と呼ばれる。年齢区分幅と調査実施間隔が等しい場合(たとえば、5歳幅と5年間隔)は特に「標準コウホート表」と言う。

標準以外の表は「一般コウホート表」と呼ばれ、年齢区分幅が調査実施間隔と一致しなかったり、調査実施間隔が不規則だったり、年齢区分幅が調査年によって変化したりする場合に得られる。コウホート分析の対象は標準表データに限らないが、本稿では以下、標準コウホート表を想定する。

年齢区分数を I 、調査時点数を J とし、第 j 調査時点の第 i 年齢区分を (i, j) セルと呼ぶ ($i = 1, \dots, I; j = 1, \dots, J$)。表 1 は、 $I = 4$ 、 $J = 3$ の場合の模式的な標準コウホート表である。調査実施間隔を 10 年、年齢区分幅を 10 歳としている。表のセルにはデータではなく、セルと 1 対 1 に対応するセルパラメータ η_{ij} を配した。

標準コウホート表について特徴的なことは、表に自然に現れるコウホート区分の数を K とすると、 $K = I + J - 1$ となることである(表 1 では $K = 6$)。また、 (i, j) セルに対応するコウホートを第 k コウホート区分とすると、 i, j, k の間に、

$$(2.1) \quad k = k(i, j) = j - i + I,$$

という関係(「出生年 = 調査年 - 年齢」のインデックス版)がある ($k = 1, \dots, K$)。この関係 $k(i, j)$ がコウホート分析における識別問題の源泉である(本稿では識別問題については 4 節で簡単に触れるに留める。詳しくは、たとえば、中村, 2005, 坂口・中村, 2019 を参照)。

2.2 コウホートモデル

標準コウホート表における (i, j) セルの何らかの集計項目の期待値をリンク関数で変換したセルパラメータ η_{ij} を、コウホートモデルは次のように分解する。

$$(2.2) \quad \eta_{ij} = \beta^G + \beta_i^A + \beta_j^P + \beta_k^C,$$

ここで、 β^G は総平均効果、 β_i^A 、 β_j^P 、 β_k^C はそれぞれ年齢、時代、コウホート効果のパラメータである。

表 1. 模式的な標準コウホート表のセルパラメータ (η_{ij})。

年齢区分	調査時点		
	2000年 ($j = 1$)	2010年 ($j = 2$)	2020年 ($j = 3$)
20歳台 ($i = 1$)	η_{11}	η_{12}	η_{13}
30歳台 ($i = 2$)	η_{21}	η_{22}	η_{23}
40歳台 ($i = 3$)	η_{31}	η_{32}	η_{33}
50歳台 ($i = 4$)	η_{41}	η_{42}	η_{43}

パラメータたち $\{\eta_{ij}\}$, $\{\beta_i^A\}$, $\{\beta_j^P\}$, $\{\beta_k^C\}$ をベクトルに配したものをそれぞれ η , β^A , β^P , β^C とし, パラメータ β^A , β^P , β^C に対応するデザイン行列をそれぞれ X_A , X_P , X_C とすると, モデル (2.2) は,

$$\eta = \beta^G \mathbf{1}_{IJ} + X_A \beta^A + X_P \beta^P + X_C \beta^C,$$

と書くことができる. ここで, $\mathbf{1}_{IJ}$ はすべての要素が1のワンズベクトル(付録 A 参照)である.

2.3 3 要因のデザイン行列

まず, コホート要因のデザイン行列 X_C を目視により与える.

表 2 は, 表 1 と同じ仕様の標準コホート表であるが, コホート区分のインデックス k を配したものである. 表において, 一番古いコホート区分 ($k=1$) は, 一番古い調査時点 ($j=1$ [2000 年]) の最高齢年齢区分 ($i=I=4$ [50 歳台]) で 1940 年代生まれ, 一方, 一番新しいコホート区分 ($k=K=6$) は, 一番新しい調査時点 ($j=J=3$ [2020 年]) の最若齢年齢区分 ($i=1$ [20 歳台]) で 1990 年代生まれである. 両者の間に同一コホート区分は斜め右下方向を辿るセル群に現れ, 右上にいくほど新しいコホート区分 ($k=2, \dots, 5$) となって並んでいる.

表 3 は, 表側に表 2 のセルを縦に並べ, 表頭にコホート効果パラメータを横に並べて, コホート区分のインデックスが対応するところに 1 を立てたものである. 調査時点ごとのブロックに付録 (A.1) の反単位行列 \check{E}_I が現れている. しかも調査時点が新しくなるにつれ右方向にずれていく. これを表現するには (A.2) の拡張反単位行列 $\check{E}_{I \times K}$ に (A.3) のシフト行列のベキ

表 2. 標準コホート表におけるコホート効果のインデックス.

年齢区分	調査時点		
	2000年 ($j=1$)	2010年 ($j=2$)	2020年 ($j=3$)
20歳台 ($i=1$)	($k=4$)	($k=5$)	($k=6$)
30歳台 ($i=2$)	($k=3$)	($k=4$)	($k=5$)
40歳台 ($i=3$)	($k=2$)	($k=3$)	($k=4$)
50歳台 ($i=4$)	($k=1$)	($k=2$)	($k=3$)

表 3. 標準コホート表セルとコホート効果パラメータの対応.

セル			コホート効果パラメータ (β_k^C)					
調査時点	年齢区分	コホート区分	β_1^C	β_2^C	β_3^C	β_4^C	β_5^C	β_6^C
$j=1$	$i=1$	$k=4$				1		
	$i=2$	$k=3$			1			
	$i=3$	$k=2$		1				
	$i=4$	$k=1$	1					
$j=2$	$i=1$	$k=5$					1	
	$i=2$	$k=4$				1		
	$i=3$	$k=3$			1			
	$i=4$	$k=2$		1				
$j=3$	$i=1$	$k=6$						1
	$i=2$	$k=5$					1	
	$i=3$	$k=4$				1		
	$i=4$	$k=3$			1			

乗 \mathbf{N}_K^{j-1} ($j = 1, \dots, J$) を右からかけて順に積み上げれば得られる。すなわち、コウホート要因のデザイン行列 \mathbf{X}_C は、付録(B.3)に従って、

$$(2.3) \quad \mathbf{X}_C = \begin{bmatrix} \check{\mathbf{E}}_{I \times K} \mathbf{N}_K^0 \\ \check{\mathbf{E}}_{I \times K} \mathbf{N}_K^1 \\ \vdots \\ \check{\mathbf{E}}_{I \times K} \mathbf{N}_K^{J-1} \end{bmatrix} = \left(\bigoplus_{j=1}^J \check{\mathbf{E}}_{I \times K} \mathbf{N}_K^{j-1} \right) (\mathbf{1}_J \otimes \mathbf{E}_K),$$

とすればよい。

時代要因、年齢要因のデザイン行列についても、表 2 にそれぞれのインデックスを配して同様に考察することにより、

$$(2.4) \quad \mathbf{X}_P = \begin{bmatrix} \mathbf{1}_I & & & \\ & \mathbf{1}_I & & \\ & & \ddots & \\ & & & \mathbf{1}_I \end{bmatrix} = \mathbf{E}_J \otimes \mathbf{1}_I, \quad \mathbf{X}_A = \begin{bmatrix} \mathbf{E}_I \\ \mathbf{E}_I \\ \vdots \\ \mathbf{E}_I \end{bmatrix} = \mathbf{1}_J \otimes \mathbf{E}_I,$$

と得られる。

3. 年齢と時代要因のデザイン行列の導出

3.1 セルパラメータ

コウホート表のセルと 1 対 1 に対応するセルパラメータ $\{\eta_{ij}\}$ を配したベクトルと行列を、

$$\boldsymbol{\eta}_j = [\eta_{1j}, \dots, \eta_{Ij}]' \in \mathbb{R}^I \quad (j = 1, \dots, J), \\ \mathbf{H} = [\boldsymbol{\eta}_1 \ \cdots \ \boldsymbol{\eta}_J] \in \mathbb{R}^{I \times J}, \quad \boldsymbol{\eta} = \text{vec} \mathbf{H} \in \mathbb{R}^{IJ},$$

とおく ($\boldsymbol{\eta}$ については先に 2.2 節で言及した)。ここで、 vec は行列の列ベクトルを順に縦に並べる操作である。

同時に、同じ仕様のパラメータ β_{ij} を、以下での説明の都合上用意しておく。

$$(3.1) \quad \boldsymbol{\beta}_j = [\beta_{1j}, \dots, \beta_{Ij}]', \quad \mathbf{B} = [\boldsymbol{\beta}_1 \ \cdots \ \boldsymbol{\beta}_J], \quad \boldsymbol{\beta} = \text{vec} \mathbf{B},$$

である。また、 \mathbf{B} の行ベクトルを転置したものを

$$\boldsymbol{\gamma}_i = [\beta_{i1}, \dots, \beta_{iJ}]' \in \mathbb{R}^J \quad (i = 1, \dots, I),$$

とおく。

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\gamma}_1' \\ \vdots \\ \boldsymbol{\gamma}_I' \end{bmatrix},$$

である。

3.2 セルパラメータの等値制約・ゼロ和制約とデザイン行列

セルパラメータに対する等値制約と同値のゼロ和制約を齊次連立 1 次方程式として解くことによりデザイン行列を求める方法について説明する。

セルパラメータ・ベクトルの IJ 個の要素の内, ある要因のある水準が対応する L 個の要素を取り出して $\eta_* = [\eta_1, \dots, \eta_L]'$ とし, 残りを η_{**} とおいてあらためて $\eta = [\eta_*, \eta_{**}]'$ とする.

等値制約 $\eta_1 = \dots = \eta_L$ は, 等しい値を β とおけば, $\eta_* = \beta \mathbf{1}_L$ と表わすことができる. これより

$$\eta = \begin{bmatrix} \beta \mathbf{1}_L \\ \eta_{**} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_L & \mathbf{0} \\ \mathbf{O} & \mathbf{E} \end{bmatrix} \begin{bmatrix} \beta \\ \eta_{**} \end{bmatrix},$$

であるから, この水準のデザイン行列は $\begin{bmatrix} \mathbf{1}_L \\ \mathbf{0} \end{bmatrix}$ とすればよい(ただし, 元の η に対するデザイン行列とするためには, 適宜行を並べ戻す必要がある). これは目視によるデザイン行列の作成にあたる.

等値制約と同値のゼロ和制約 $\eta_1 - \eta_2 = \dots = \eta_{L-1} - \eta_L = 0$ は, 付録(A.22)の階差行列 D_L を用いれば,

$$D_L \eta_* = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_L \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}_L,$$

と表わすことができ, η 全体に対しては,

$$\mathbf{R} = \begin{bmatrix} D_L & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix},$$

とにおいて, $\mathbf{R}\eta = \mathbf{0}$ である. これを斉次連立1次方程式として解く. 付録Cと(A.24)より,

$$\mathbf{R}^- \mathbf{R} = \begin{bmatrix} D_L^- D_L & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}, \quad \mathbf{E} - \mathbf{R}^- \mathbf{R} = \begin{bmatrix} \mathbf{1}_L e_L' & \mathbf{O} \\ \mathbf{O} & \mathbf{E} \end{bmatrix},$$

に注意すれば, 任意の $\beta = [\beta_*, \beta_{**}]'$, $\beta_* = [\beta_1, \dots, \beta_L]'$ に対して,

$$\eta = (\mathbf{E} - \mathbf{R}^- \mathbf{R})\beta = \begin{bmatrix} \mathbf{1}_L e_L' \beta_* \\ \beta_{**} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_L & \mathbf{O} \\ \mathbf{O} & \mathbf{E} \end{bmatrix} \begin{bmatrix} \beta_L \\ \beta_{**} \end{bmatrix},$$

となり, デザイン行列は同じ $\begin{bmatrix} \mathbf{1}_L \\ \mathbf{0} \end{bmatrix}$ となる. 元の η の並び替えないゼロ和制約の表現を与えることができれば, 直接デザイン行列の表現が得られる.

3.3 時代要因のデザイン行列

時代要因のデザイン行列を得るために, セルパラメータ $\{\eta_{ij}\}$ に同一時点のすべての年齢区分の値が等しいという等値制約と同値のゼロ和制約

$$\eta_{ij} - \eta_{i+1,j} = 0 \quad (i = 1, \dots, I-1; j = 1, \dots, J),$$

を課す. 階差行列を用いれば $D_I \eta_j = \mathbf{0}_{I-1}$ ($j = 1, \dots, J$), あるいは, $D_I \mathbf{H} = \mathbf{O}_{(I-1) \times J}$ となる. vec を施せば, η に対する線形制約

$$\mathbf{0} = \text{vec } D_I \mathbf{H} = (\mathbf{E}_J \otimes D_I) \text{vec } \mathbf{H} = (\mathbf{E}_J \otimes D_I) \eta,$$

となる(付録(B.1)を使った).

上記制約を斉次連立1次方程式として η について解く. 付録(C.1)で, $\mathbf{E}_b = \mathbf{E}_J \otimes \mathbf{E}_I$,

$A = E_J \otimes D_I$ ととればよい. $E_J \otimes D_I$ の一般逆行列の1つは $E_J \otimes D_I^-$ であり, (A.24)より,

$$(E_J \otimes E_I) - (E_J \otimes D_I^-)(E_J \otimes D_I) = E_J \otimes (E_I - D_I^- D_I) = E_J \otimes \mathbf{1}_I e'_{I,I},$$

であるから, 一般解は, (3.1)で用意した任意の β について,

$$\eta = (E_J \otimes \mathbf{1}_I e'_I) \beta = \text{vec}(\mathbf{1}_I e'_I B) = \text{vec}(\mathbf{1}_I \gamma'_I),$$

と得られる. 最右辺の γ'_I は B の最下行であるが, 要素が時代のインデックスに依存するだけであるから時代効果ベクトルとしてよく, $\beta_j^P = \beta_{Ij}$ として $\beta^P = [\beta_1^P, \dots, \beta_J^P]' (= [\beta_{I1}, \dots, \beta_{IJ}]' = \gamma_I)$ とおくと,

$$\eta = \text{vec}\{\mathbf{1}_I(\beta^P)'\} = (E_J \otimes \mathbf{1}_I) \text{vec}(\beta^P)' = (E_J \otimes \mathbf{1}_I) \beta^P,$$

となって, (2.4)で示した時代要因のデザイン行列 $X_P = E_J \otimes \mathbf{1}_I$ が得られる.

3.4 年齢要因のデザイン行列

年齢要因のデザイン行列を得るために, セルパラメータ $\{\eta_{ij}\}$ に異なる時点の同じ年齢区分の値が等しいというゼロ和制約を課す. すなわち,

$$\eta_{ij} - \eta_{i,j+1} = 0 \quad (i = 1, \dots, I; j = 1, \dots, J-1),$$

である. これはゼロ和制約 $HD'_j = O_{I \times (J-1)}$ となり, vec を施せば, η に対する線形制約

$$\mathbf{0} = \text{vec}HD'_j = (D_J \otimes E_I) \text{vec}H = (D_J \otimes E_I) \eta,$$

となる.

上記制約を斉次連立1次方程式として η について解く.

$$(E_J \otimes E_I) - (D_J^- \otimes E_I)(D_J \otimes E_I) = (E_J - D_J^- D_J) \otimes E_I = \mathbf{1}_J e'_{J,J} \otimes E_I,$$

であるから, 一般解は, 任意の β について,

$$\eta = (\mathbf{1}_J e'_J \otimes E_I) \beta = \text{vec}(B e_J \mathbf{1}'_J) = \text{vec}(\beta_J \mathbf{1}'_J),$$

であり, β_J は要素が年齢区分のインデックスに依存するだけであるから年齢効果ベクトルとしてよく, $\beta_i^A = \beta_{iJ}$ として $\beta^A = [\beta_1^A, \dots, \beta_I^A]' (= [\beta_{1J}, \dots, \beta_{IJ}]' = \beta_J)$ とおくことにすれば,

$$\eta = \text{vec}(\beta^A \mathbf{1}'_J) = (\mathbf{1}_J \otimes E_I) \text{vec} \beta^A = (\mathbf{1}_J \otimes E_I) \beta^A,$$

となって, (2.4)で示した年齢要因のデザイン行列 $X_A = \mathbf{1}_J \otimes E_I$ が得られる.

4. コウホート要因のデザイン行列の導出

前節の時代要因および年齢要因のデザイン行列を導出したときと同じ考え方にに基づき, 以下でセルパラメータのゼロ和制約によりコウホート要因のデザイン行列を導出する.

4.1 反転セルパラメータ

η_j の要素は, 新しいコウホートから古いコウホートに対応するものが並んでいる. これを逆順にして古いコウホートから新しいコウホートに対応するようにしておくのが後のために便利である. (A.1)の反転行列 \check{E}_I を使い, 反転セルパラメータ・ベクトルとして,

$$\check{\eta}_j = \check{\mathbf{E}}_I \eta_j = [\eta_{Ij}, \dots, \eta_{1j}]',$$

とし,

$$\begin{aligned} \check{\mathbf{H}} &= \check{\mathbf{E}}_I \mathbf{H} = [\check{\mathbf{E}}_I \eta_1 \quad \cdots \quad \check{\mathbf{E}}_I \eta_J] = [\check{\eta}_1 \quad \cdots \quad \check{\eta}_J] \in \mathbb{R}^{I \times J}, \\ \check{\eta} &= \text{vec } \check{\mathbf{H}} = \text{vec}(\check{\mathbf{E}}_I \mathbf{H}) = (\mathbf{E}_J \otimes \check{\mathbf{E}}_I) \text{vec } \mathbf{H} = (\mathbf{E}_J \otimes \check{\mathbf{E}}_I) \eta \in \mathbb{R}^{IJ}, \end{aligned}$$

とおく. $\check{\eta}$ から η に戻すには, $\check{\mathbf{E}}_a^{-1} = \check{\mathbf{E}}_a$ だから,

$$(4.1) \quad \eta = (\mathbf{E}_J \otimes \check{\mathbf{E}}_I) \check{\eta} = \left(\bigoplus_{j=1}^J \check{\mathbf{E}}_I \right) \check{\eta},$$

とすればよい.

4.2 同一コウホートの等値制約

隣り合う時点の $\check{\eta}_j$ と $\check{\eta}_{j+1}$ の要素を次のように対比する. コウホート区分のインデックスが (2.1) の $k = j - i + I$ であることに注意して,

$$\check{\eta}_j = \begin{bmatrix} \eta_{Ij} \\ \eta_{I-1,j} \\ \vdots \\ \eta_{2j} \\ \eta_{1j} \end{bmatrix} \rightarrow \begin{bmatrix} j\text{th cohort} \\ (j+1)\text{th c.} \\ \vdots \\ (j+I-2)\text{th c.} \\ (j+I-1)\text{th c.} \end{bmatrix}, \quad \begin{bmatrix} (j+1)\text{th c.} \\ (j+2)\text{th c.} \\ \vdots \\ (j+I-1)\text{th c.} \\ (j+I)\text{th c.} \end{bmatrix} \leftarrow \begin{bmatrix} \eta_{I,j+1} \\ \eta_{I-1,j+1} \\ \vdots \\ \eta_{2,j+1} \\ \eta_{1,j+1} \end{bmatrix} = \check{\eta}_{j+1}.$$

$\check{\eta}_j$ の第 i 要素と $\check{\eta}_{j+1}$ の第 $i-1$ 要素が同じコウホートに対応するから, $\check{\eta}_j$ の第 1 要素と $\check{\eta}_{j+1}$ の最終要素を次のように削除して揃え, すなわち,

$$\mathbf{F}_I \mathbf{N}_I \check{\eta}_j = \begin{bmatrix} \eta_{I-1,j} \\ \vdots \\ \eta_{2j} \\ \eta_{1j} \end{bmatrix} \rightarrow \begin{bmatrix} (j+1)\text{th c.} \\ \vdots \\ (j+I-2)\text{th c.} \\ (j+I-1)\text{th c.} \end{bmatrix} = \begin{bmatrix} (j+1)\text{th c.} \\ (j+2)\text{th c.} \\ \vdots \\ (j+I-1)\text{th c.} \end{bmatrix} \leftarrow \begin{bmatrix} \eta_{I,j+1} \\ \eta_{I-1,j+1} \\ \vdots \\ \eta_{2,j+1} \end{bmatrix} = \mathbf{F}_I \check{\eta}_{j+1},$$

とする (ここで, $\check{\eta}_j$ と $\check{\eta}_{j+1}$ のそれぞれに (A.13) と (A.10) を右からかけた). これに基づき,

$$\mathbf{0} = \mathbf{F}_I \mathbf{N}_I \check{\eta}_j - \mathbf{F}_I \check{\eta}_{j+1} = [\mathbf{F}_I \mathbf{N}_I \quad -\mathbf{F}_I] \begin{bmatrix} \check{\eta}_j \\ \check{\eta}_{j+1} \end{bmatrix} \quad (j = 1, \dots, J-1),$$

のように等値制約と同値のゼロ和制約を課す. 全セルについては,

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} \mathbf{F}_I \mathbf{N}_I & -\mathbf{F}_I & & & & & \\ & \mathbf{F}_I \mathbf{N}_I & -\mathbf{F}_I & & & & \\ & & \mathbf{F}_I \mathbf{N}_I & \ddots & & & \\ & & & \ddots & -\mathbf{F}_I & & \\ & & & & \mathbf{F}_I \mathbf{N}_I & -\mathbf{F}_I & \\ & & & & & & \end{bmatrix} \\ &= (\mathbf{F}_J \otimes \mathbf{F}_I \mathbf{N}_I) - (\mathbf{F}_J \mathbf{N}_J \otimes \mathbf{F}_I) \in \mathbb{R}^{(I-1)(J-1) \times IJ}, \end{aligned}$$

とおいて, $\check{\eta}$ に対する線形制約

$$(4.2) \quad R\check{\eta} = \mathbf{0},$$

となる.

4.3 R の一般逆行列

R の一般逆行列の候補として,

$$G = \begin{bmatrix} O & & & & & \\ -F'_I & O & & & & \\ -N_I F'_I & -F'_I & O & & & \\ -N_I^2 F'_I & -N_I F'_I & -F'_I & O & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \\ -N_I^{J-2} F'_I & -N_I^{J-3} F'_I & \cdots & -N_I F'_I & -F'_I & \end{bmatrix} \in \mathbb{R}^{IJ \times (I-1)(J-1)},$$

ととる. 積 RG の対角ブロックが E_{I-1} に, 非対角ブロックが O_{I-1} になるように考えると
い付く.

$$\begin{aligned} G &= -(N'_J F'_J \otimes F'_I) - \{(N'_J)^2 F'_J \otimes N_I F'_I\} - \cdots - \{(N'_J)^{J-1} F'_J \otimes N_I^{J-2} F'_I\} \\ &= -\sum_{j=1}^{J-1} \{(N'_J)^j F'_J \otimes N_I^{j-1} F'_I\}, \end{aligned}$$

であるから,

$$(4.3) \quad RG = -\sum_{j=1}^{J-1} \{F_J (N'_J)^j F'_J \otimes F_I N_I^j F'_I\} + \sum_{j=1}^{J-1} \{F_J N_J (N'_J)^j F'_J \otimes F_I N_I^{j-1} F'_I\},$$

である. 右辺第 1 項 $\text{RHS}_1^{(4.3)}$ は, 和の終項が (A.19) より $F_J (N'_J)^{J-1} F'_J = O$ (ゼロズ行列) であるから,

$$\text{RHS}_1^{(4.3)} = -\sum_{j=1}^{J-2} \{F_J (N'_J)^j F'_J \otimes F_I N_I^j F'_I\} = -\sum_{j=2}^{J-1} \{F_J (N'_J)^{j-1} F'_J \otimes F_I N_I^{j-1} F'_I\},$$

となる. 右辺第 2 項 $\text{RHS}_2^{(4.3)}$ は, 和の初項を分離し, (A.11) より $F_I F'_I = E_{I-1}$, (A.13) より $F_J N_J N'_J F'_J = E_{J-1}$, (A.18) より $F_J N_J N'_J = F_J$ であるから,

$$\begin{aligned} \text{RHS}_2^{(4.3)} &= (F_J N_J N'_J F'_J \otimes F_I F'_I) + \sum_{j=2}^{J-1} \{F_J N_J N'_J (N'_J)^{j-1} F'_J \otimes F_I N_I^{j-1} F'_I\} \\ &= (E_{J-1} \otimes E_{I-1}) + \sum_{j=2}^{J-1} \{F_J (N'_J)^{j-1} F'_J \otimes F_I N_I^{j-1} F'_I\}, \end{aligned}$$

となる. したがって, 右辺第 1 項 $\text{RHS}_1^{(4.3)}$ と, 右辺第 2 項 $\text{RHS}_2^{(4.3)}$ の第 2 項が打ち消し合って,

$$RG = (E_{J-1} \otimes E_{I-1}) = E_{(I-1) \times (J-1)},$$

と単位行列になり, 確かに G が R の一般逆行列の 1 つであることがわかる. 以下, $R^- = G$ とおく.

4.4 $R^{-}R$ を求める

(A.17) より $j \geq 1$ で $F'_a F_a N_a^j = N_a^j$ であることに注意して,

$$(4.4) \quad \begin{aligned} R^{-}R &= - \sum_{j=1}^{J-1} \{ (N_j^j)' F'_j F_j \otimes N_j^{j-1} F'_I F_I N_I \} + \sum_{j=1}^{J-1} \{ (N_j^j)' F'_j F_j N_J \otimes N_I^{j-1} F'_I F_I \} \\ &= - \sum_{j=1}^{J-1} \{ (N_j^j)' \otimes N_I^j \} + \sum_{j=1}^{J-1} \{ (N_j^j)' N_J \otimes N_I^{j-1} F'_I F_I \}, \end{aligned}$$

である. 右辺第 1 項 $\text{RHS}_1^{(4.4)}$ は,

$$\begin{aligned} \text{RHS}_1^{(4.4)} &= - \sum_{j=1}^{J-1} \{ (N_j^j)' E_J \otimes N_I^j \} \\ &= - \sum_{j=1}^{J-1} \left\{ (N_j^j)' \begin{bmatrix} 1 & \\ & O_{J-1} \end{bmatrix} \otimes N_I^j \right\} - \sum_{j=1}^{J-1} \left\{ (N_j^j)' \begin{bmatrix} 0 & \\ & E_{J-1} \end{bmatrix} \otimes N_I^j \right\} \\ &= - \sum_{j=1}^{J-1} \{ (N_j^j)' e_{1,J} e'_{J,J} \otimes N_I^j \} - \sum_{j=1}^{J-1} \{ (N_j^j)' N'_J N_J \otimes N_I^j \} \\ &= - \sum_{j=1}^{J-1} \{ e_{j+1} e'_j \otimes N_I^j \} - \sum_{j=1}^{J-2} \{ (N_j^{j+1})' N_J \otimes N_I^j \} \\ &= - \sum_{j=1}^{J-1} \{ e_{j+1} e'_j \otimes N_I^j \} - \sum_{j=2}^{J-1} \{ (N_j^j)' N_J \otimes N_I^{j-1} \}, \end{aligned}$$

となる. 最後から 2 番目の等式では $N_J^J = O_J$ であることを使った.

一方, (4.4) の右辺第 2 項 $\text{RHS}_2^{(4.4)}$ は, 和の初項を分離して,

$$\text{RHS}_2^{(4.4)} = (N'_J N_J \otimes F'_I F_I) + \sum_{j=2}^{J-1} \{ (N_j^j)' N_J \otimes N_I^{j-1} F'_I F_I \},$$

となる.

$\text{RHS}_1^{(4.4)}$ と $\text{RHS}_2^{(4.4)}$ それぞれの第 2 項を足すと, (A.16) より $F'_I F_I - E_I = -e_I e'_I$ だから,

$$\begin{aligned} \sum_{j=2}^{J-1} \{ (N_j^j)' N_J \otimes N_I^{j-1} (F'_I F_I - E_I) \} &= - \sum_{j=2}^{J-1} \{ (N_j^j)' N_J \otimes N_I^{j-1} e_I e'_I \} \\ &= - \sum_{j=2}^{J-1} \{ (N_j^j)' N_J \otimes e_{I-j+1} e'_I \}, \end{aligned}$$

となって,

$$R^{-}R = (N'_J N_J \otimes F'_I F_I) - \sum_{j=1}^{J-1} \{ e_{j+1} e'_j \otimes N_I^j \} - \sum_{j=2}^{J-1} \{ (N_j^j)' N_J \otimes e_{I-j+1} e'_I \},$$

と得られる.

であり, これより一般に,

$$(4.7) \quad N_K^{j-1} = \begin{bmatrix} N_I^{j-1} & A_j \\ & N_{J-1}^{j-1} \end{bmatrix}, \quad A_j = \sum_{\ell=1}^{j-1} e_{I-(j-1)+\ell, I} e'_{\ell, J-1},$$

であることを数学的帰納法により示すことができる. これより,

$$(4.8) \quad \check{E}_{I \times K} N_K^{j-1} = [E_I \quad O_{I \times (J-1)}] \begin{bmatrix} N_I^{j-1} & A_j \\ & N_{J-1}^{j-1} \end{bmatrix} = [N_I^{j-1} \quad A_j],$$

である.

4.7 コウホート要因のデザイン行列

3.3, 3.4節と同様に斉次連立1次方程式(4.2)を $\check{\eta}$ について解くと, 任意の $\check{\beta} = (E_J \otimes \check{E}_I)\beta$ について, 一般解は

$$\check{\eta} = (E - R^{-1}R) \check{\beta},$$

である. まず, (4.5)の右辺第1項 $\text{RHS}_1^{(4.6)}$ の和の中について $\check{\beta}$ をかけると,

$$(e_j e'_1 \otimes N_I^{j-1}) \check{\beta} = \text{vec}(N_I^{j-1} \check{B} e_1 e'_j) = \text{vec}(N_I^{j-1} \check{\beta}_1 e'_j) = (e_j \otimes N_I^{j-1}) \check{\beta}_1,$$

であり, 右辺第2項 $\text{RHS}_2^{(4.6)}$ の和の中についても同様にすると, (A.15)より,

$$\begin{aligned} \{(N_J^{j-1})' N_J \otimes e_{I-j+2} e'_1\} \check{\beta} &= \text{vec}(e_{I-j+2} e'_1 \check{B} N_J' F_J' N_J^{j-2} F_J N_J) \\ &= \text{vec}(e_{I-j+2} \gamma_1' N_J' F_J' N_J^{j-2} F_J N_J) \\ &= \{N_J' F_J' (N_{J-1}^{j-2})' \otimes e_{I-j+2}\} \check{\beta}_*, \end{aligned}$$

となる. ここで, \check{B} の最下行の第1要素を除いた行 $\check{\beta}_* = F_J N_J \gamma_1 = [\beta_{12}, \dots, \beta_{1J}]'$ とおいた.

以上より,

$$\begin{aligned} \check{\eta} = (E - R^{-1}R) \check{\beta} &= \left\{ \sum_{j=1}^J (e_j \otimes N_I^{j-1}) \right\} \check{\beta}_1 + \left\{ \sum_{j=2}^J (N_J' F_J' (N_{J-1}^{j-2})' \otimes e_{I-j+2}) \right\} \check{\beta}_* \\ &= Z \begin{bmatrix} \check{\beta}_1 \\ \check{\beta}_* \end{bmatrix}, \end{aligned}$$

となる. ここで,

$$Z = \left[\sum_{j=1}^J (e_j \otimes N_I^{j-1}) \quad \sum_{j=2}^J (N_J' F_J' (N_{J-1}^{j-2})' \otimes e_{I-j+2}) \right] \in \mathbb{R}^{IJ \times (I+J-1)},$$

とおいた.

Z の第 j^* 行ブロックは, $e_{j^*}' \otimes E_I$ を左からかけると得られる. すなわち,

$$(e_{j^*}' \otimes E_I) Z = \left[\sum_{j=1}^J (e_{j^*}' e_j \otimes N_I^{j-1}) \quad \sum_{j=2}^J \{e_{j^*}' N_J' F_J' (N_{J-1}^{j-2})' \otimes e_{I-j+2}\} \right]$$

$$= \left[N_I^{j^*-1} \sum_{j=2}^J \{e'_{j^*-1, J-1} (N_{j-1}^{j-2})' \otimes e_{I-j+2}\} \right],$$

である。まず、第 1 行ブロックは、 $j^* = 1$ として、

$$(e'_1 \otimes E_I)Z = \left[N_I^0 \sum_{j=2}^J \{e'_{0, J-1} (N_{j-1}^{j-2})' \otimes e_{I-j+2}\} \right] = [E_I \quad O_{I \times (J-1)}],$$

である。

次に、 $j^* \geq 2$ について考える。(A.8)より、下記は、 $j^* - 1 = \ell + (j - 2)$ 、すなわち、 $\ell = j^* - j + 1$ のときだけ残り、

$$e'_{j^*-1, J-1} (N_{j-1}^{j-2})' = e'_{j^*-1, J-1} \left(\sum_{\ell=1}^{j-1} e_{\ell+(j-2)} e'_\ell \right) = e'_{j^*-j+1, J-1},$$

となる。 $j^* - j + 1 \geq 1$ でなければ $e'_{j^*-j+1, J-1} = \mathbf{0}_{J-1}$ であるから、 $j = 2, \dots, j^*$ と渡ればよく、

$$\sum_{j=2}^J \{e'_{j^*-1, J-1} (N_{j-1}^{j-2})' \otimes e_{I-j+2}\} = \sum_{j=2}^{j^*} (e'_{j^*-j+1} \otimes e_{I-j+2}) = \sum_{j=1}^{j^*-1} (e_{I-j+1} e'_{j^*-j}),$$

と簡単になる。さらに $j^* - j = m$ とおけば、 $j = 1, \dots, j^* - 1$ に対して $m = j^* - 1, \dots, 1$ であり、

$$\sum_{j=1}^{j^*-1} (e_{I-j+1} e'_{j^*-j}) = \sum_{j=1}^{j^*-1} (e_{I-j^*+m+1} e'_m) = \mathbf{A}_{j^*},$$

となって、(4.7)で $j = j^*$ とおいた \mathbf{A}_j と一致することがわかる。結局、第 2 以降の行ブロックは、(4.8)より、

$$(e'_{j^*} \otimes E_I)Z = [N_I^{j^*-1} \quad \mathbf{A}_{j^*}] = \check{E}_{I \times K} N_K^{j^*-1},$$

と得られる。 Z はより具体的には、

$$\begin{aligned} Z &= \begin{bmatrix} E_I & & & & & \\ N_I & e_I & & & & \\ N_I^2 & e_{I-1} & e_I & & & \\ \vdots & \ddots & \ddots & \ddots & & \\ N_I^{J-1} & e_{I-J+2} & \cdots & e_{I-1} & e_I & \end{bmatrix} = \begin{bmatrix} E_{I \times K} \\ E_{I \times K} N_K \\ E_{I \times K} N_K^2 \\ \vdots \\ E_{I \times K} N_K^{J-1} \end{bmatrix} \\ &= \left(\bigoplus_{j=1}^J E_{I \times K} N_K^{j-1} \right) (\mathbf{1}_J \otimes E_K), \end{aligned}$$

となっている。

一方、 $\beta_k^C = \beta_{I-i+j}^C = \beta_{ij}$ であるから、

$$\begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_* \end{bmatrix} = \begin{bmatrix} \beta_{Ii} \\ \beta_{I-1,i} \\ \vdots \\ \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1J} \end{bmatrix} = \begin{bmatrix} \beta_1^C \\ \beta_2^C \\ \vdots \\ \beta_I^C \\ \beta_{I-1}^C \\ \vdots \\ \beta_{I+J-1}^C \end{bmatrix} = \beta^C,$$

とにおいて、コウホート効果ベクトルとしてよい。

以上より、 $\tilde{\eta} = Z\beta^C$ であるが、 η について戻せば、(4.1) より、

$$\begin{aligned} \eta &= \left(\bigoplus_{j=1}^J \tilde{E}_I \right) \left(\bigoplus_{j=1}^J E_{I \times K} N_K^{j-1} \right) (\mathbf{1}_J \otimes E_K) \beta^C \\ &= \left(\bigoplus_{j=1}^J \tilde{E}_{I \times K} N_K^{j-1} \right) (\mathbf{1}_J \otimes E_K) \beta^C, \end{aligned}$$

となり、(2.3) で示したコウホート要因のデザイン行列 X_C が得られる。

5. おわりに

コウホート分析モデルにおける3要因のデザイン行列を、セルパラメータに等値制約と同値のゼロ和制約を課すことにより導出し、デザイン行列の性格付けを行った。また、その過程に有用な行列や行列の操作について付録にまとめた。

デザイン行列の直和やクロネッカ積を用いた陽な表現は、コウホートモデルの識別問題を理論的に議論する際に役に立つ。たとえば、インデックスベクトルを $n_a = [1, 2, \dots, a]'$ で定義するとき、ここでは詳細は省くが、

$$X_A n_I - X_P n_J + X_C n_K = I \mathbf{1}_{IJ},$$

を示すことができる(中村, 2019)。これは、インデックス間の関係(2.1)による $i - j + k = I$ のインデックスベクトル版であり、コウホートモデル全体のデザイン行列 $X = [\mathbf{1}_{IJ} \ X_A \ X_P \ X_C]$ の列ベクトルが1次従属であることを示す。コウホートモデルの識別問題の源泉である。

デザイン行列 X は、直上の1次従属関係の他に後3つの関係 $X_A \mathbf{1}_I = X_P \mathbf{1}_J = X_C \mathbf{1}_K = \mathbf{1}_{IJ}$ により、列数 $I + J + K + 1$ より4つ以上ランクが落ちていることはわかるが、それがちょうど4つであることは数値的には示すことができるものの数式によっては示されていない。この点については別の機会に譲ることにしたい。

謝 辞

統計数理研究所国民性調査委員会の先達および同僚諸氏、また調査に協力して下さった対象者および関係者の方々は、60年以上にもわたる貴重な調査データの蓄積を通してコウホート分析のための豊かな研究土壌を与えて下さいました。レフェリーの方には重要かつ適切なコメントをいただきました。ここに記して心より感謝いたします。

付録 A. 特別なベクトルと行列の準備

ベクトルと行列の大きさを、それぞれ a 次元実ベクトルの集合 \mathbb{R}^a と大きさ $a \times b$ の実行列の集合 $\mathbb{R}^{a \times b}$ への帰属として示す。

ゼロズベクトル $\mathbf{0}_a$ / ゼロズ行列 $\mathbf{O}_{a \times b}$: すべての要素が 0 のゼロズベクトルを $\mathbf{0}_a \in \mathbb{R}^a$, ゼロズ行列を $\mathbf{O}_{a \times b} \in \mathbb{R}^{a \times b}$ で表す. 正方行列の場合は単に \mathbf{O}_a と書く.

ワンズベクトル $\mathbf{1}_a$: すべての要素が 1 のワンズベクトルを $\mathbf{1}_a \in \mathbb{R}^a$ で表す.

単位ベクトル $e_{j,a}$: 第 j 要素のみが 1 でその他の要素が 0 の単位ベクトルを $e_{j,a} \in \mathbb{R}^a$ で表す ($j = 1, \dots, a$). $j < 1$ または $j > a$ のときは, $e_{j,a} = \mathbf{0}_a$ と定める. 後述のシフト行列 \mathbf{N}_a を使えば, $e_{j,a} = (\mathbf{N}_a^{j-1})' e_{1,a}$ であるが, $j > a$ ならば $\mathbf{N}_a^{j-1} = \mathbf{O}$ なので正当化できる ($j < 1$ の場合も同様). 大きさが明らかな場合は単に e_j とすることがある.

単位行列 \mathbf{E}_a / 拡張単位行列 $\mathbf{E}_{a \times b}$: 単位行列を $\mathbf{E}_a \in \mathbb{R}^{a \times a}$ で表す.

$$\mathbf{E}_a \equiv [e_{1,a} \quad e_{2,a} \quad \cdots \quad e_{a,a}] = \sum_{j=1}^a e_j e_j' = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix},$$

である. 単位行列の右側にゼロズ行列を補った拡張単位行列を $\mathbf{E}_{a \times b} \in \mathbb{R}^{a \times b}$ と書く.

$$\mathbf{E}_{a \times b} \equiv [\mathbf{E}_a \quad \mathbf{O}_{a \times (b-a)}] = \sum_{j=1}^a e_{j,a} e_{j,b}' \quad (b \geq a),$$

である.

反単位(反転)行列 $\check{\mathbf{E}}_a$ / 拡張反単位行列 $\check{\mathbf{E}}_{a \times b}$: 左からかけてベクトルの要素あるいは行列の列を反転する反単位行列 $\check{\mathbf{E}}_a \in \mathbb{R}^{a \times a}$ を,

$$(A.1) \quad \check{\mathbf{E}}_a \equiv [e_{a,a} \quad e_{a-1,a} \quad \cdots \quad e_{1,a}] = \sum_{j=1}^a e_{j,a} e_{a-j+1,a}' = \begin{bmatrix} & & & 1 \\ & & \ddots & \\ & & & \\ 1 & & & \end{bmatrix},$$

で定義する. $\check{\mathbf{E}}_a \check{\mathbf{E}}_a = \mathbf{E}_a$, $\check{\mathbf{E}}_a^{-1} = \check{\mathbf{E}}_a$ である.

反単位行列の右側にゼロズ行列を補った拡張反単位行列を $\check{\mathbf{E}}_{a \times b}$ と書く.

$$(A.2) \quad \check{\mathbf{E}}_{a \times b} \equiv [\check{\mathbf{E}}_a \quad \mathbf{O}_{a \times (b-a)}] = \sum_{j=1}^a e_{j,a} e_{a-j+1,b}' \quad (b \geq a),$$

である.

シフト行列 \mathbf{N}_a : 左からかけてベクトルの要素あるいは行列の行を 1 つだけ上にシフトする行列 \mathbf{N}_a を,

$$(A.3) \quad \mathbf{N}_a \equiv \begin{bmatrix} \mathbf{0}_{a-1} & \mathbf{E}_{a-1} \\ 0 & \mathbf{0}'_{a-1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix} \in \mathbb{R}^{a \times a},$$

で定義する. \mathbf{N}_a を右からかければ右にシフト, 転置した \mathbf{N}_a' を左からかければ下にシフト, \mathbf{N}_a' を右からかければ左にシフトとなる.

$$N_a^j = \begin{bmatrix} & \mathbf{E}_{a-j} \\ \mathbf{O}_j & \end{bmatrix} = \begin{bmatrix} N_{a-1}^{j-1} \\ 0 \end{bmatrix} \quad (j = 1, \dots, a-1),$$

であり, $j \geq a$ で $N_a^j = \mathbf{O}_a$ となるから N_a はベキ零行列の 1 つである. $N_a^0 = \mathbf{E}_a$ と定める.

以下, 単位ベクトルを $e_\ell = e_{\ell,a}$ と略記する.

$$(A.4) \quad N_a' N_a = \begin{bmatrix} & 0 \\ \mathbf{E}_{a-1} & \end{bmatrix} N_a = \begin{bmatrix} 0 \\ \mathbf{E}_{a-1} \end{bmatrix} = \mathbf{E}_a - e_1 e_1',$$

$$(A.5) \quad N_a N_a' = N_a \begin{bmatrix} & 0 \\ \mathbf{E}_{a-1} & \end{bmatrix} = \begin{bmatrix} \mathbf{E}_{a-1} \\ 0 \end{bmatrix} = \mathbf{E}_a - e_a e_a',$$

で対称であり, $N_a N_a' N_a = N_a$, $N_a' N_a N_a' = N_a'$ であるから, N_a のムーア=ペンローズ逆行列 N_a^+ は,

$$(A.6) \quad N_a^+ = N_a',$$

である. また,

$$(A.7) \quad (N_a^j)' N_a = \begin{bmatrix} & 0 \\ (N_{a-1}^{j-1})' & \end{bmatrix} N_a = \begin{bmatrix} 0 \\ (N_{a-1}^{j-1})' \end{bmatrix},$$

となる.

ところで,

$$N_a = \sum_{\ell=1}^{a-1} e_\ell e_{\ell+1}' = \sum_{\ell=1}^a e_\ell e_{\ell+1}',$$

とも書ける. $e_{a+1}' = \mathbf{0}'$ であるから, 和の範囲を a までとしてもかまわない. $e_\ell' e_m = 0$ ($\ell \neq m$), $e_\ell' e_\ell = 1$ であるから,

$$(A.8) \quad N_a^2 = \left(\sum_{\ell=1}^a e_\ell e_{\ell+1}' \right) \left(\sum_{m=1}^a e_m e_{m+1}' \right) = \sum_{\ell=1}^a e_\ell e_{\ell+1}' e_{\ell+1} e_{\ell+2}' = \sum_{\ell=1}^a e_\ell e_{\ell+2}',$$

$$N_a^j = \sum_{\ell=1}^a e_\ell e_{\ell+j}' = \sum_{\ell=1}^{a-j} e_\ell e_{\ell+j}' \quad (j < a),$$

であり,

$$(N_a^i)' N_a^j = \left(\sum_{\ell=1}^a e_{\ell+i} e_\ell' \right) \left(\sum_{m=1}^a e_m e_{m+j}' \right) = \sum_{\ell=1}^a e_{\ell+i} e_{\ell+j}',$$

である.

ある上三角行列 U_a : シフト行列 N_a の次のベキ和を U_a とおく.

$$U_a \equiv \sum_{j=1}^a N_a^{j-1} = \mathbf{E}_a + N_a + N_a^2 + \cdots + N_a^{a-1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ & 1 & \ddots & 1 \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix} \in \mathbb{R}^{a \times a},$$

であり, 右上のすべての要素が 1 の上三角行列になる.

$$(\mathbf{E}_a - N_a) U_a = (\mathbf{E}_a - N_a) (\mathbf{E}_a + N_a + N_a^2 + \cdots + N_a^{a-1}) = \mathbf{E}_a,$$

であることから.

$$(A.9) \quad U_a(\mathbf{E}_a - N_a) = \mathbf{E}_a, \quad (\mathbf{E}_a - N_a)^{-1} = U_a,$$

である.

削除行列 F_a : 左からかけてベクトルの最終要素あるいは行列の最終行を削除する行列 F_a を,

$$(A.10) \quad F_a \equiv [\mathbf{E}_{a-1} \quad \mathbf{0}_{a-1}] (= \mathbf{E}_{(a-1) \times a}) \in \mathbb{R}^{(a-1) \times a},$$

で定義する.

$$(A.11) \quad F_a F_a' = \mathbf{E}_{a-1},$$

で $F_a F_a' F_a = F_a$ を満たすから, F_a の一般逆行列の 1 つは,

$$F_a^- = F_a' \in \mathbb{R}^{a \times (a-1)},$$

である.

転置した F_a' を左から行列にかければゼロズ行ベクトルの追加になり, 右から行列にかければ最終列の削除となる. F_a を右から行列にかければゼロズ(列)ベクトルの追加でもある.

$$(A.12) \quad \mathbf{1}_a = \begin{bmatrix} \mathbf{1}_{a-1} \\ 1 \end{bmatrix} = F_a' \mathbf{1}_{a-1} + e_a,$$

と分解できる.

シフト行列 N_a と組み合わせると,

$$(A.13) \quad F_a N_a = [\mathbf{0}_{a-1} \quad \mathbf{E}_{a-1}] \in \mathbb{R}^{(a-1) \times a}, \quad F_a N_a N_a' F_a' = \mathbf{E}_{a-1},$$

となり, $F_a N_a$ は左からかけてベクトルの先頭要素あるいは行列の先頭行を削除する行列となる.

$$(A.14) \quad F_a N_a F_a' = N_{a-1},$$

である. $F_a N_a$ を右からかければ先頭列としてゼロズ列ベクトルを追加する.

転置した $(F_a N_a)'$ を左からかければ先頭行としてゼロズ行ベクトルを追加し, 右からかければ先頭列を削除する. これより (A.7) は,

$$(A.15) \quad (N_a^j)' N_a = N_a' F_a' (N_{a-1}^{j-1})' F_a N_a,$$

と表すこともできる.

さて, (A.5), (A.6) より,

$$(A.16) \quad F_a' F_a = \begin{bmatrix} \mathbf{E}_{a-1} \\ 0 \end{bmatrix} = \mathbf{E}_a - e_a e_a' = N_a N_a' = N_a N_a^+,$$

であることから,

$$(A.17) \quad F_a' F_a N_a = N_a N_a^+ N_a = N_a, \quad F_a' F_a N_a^j = N_a^j \quad (j \geq 1), \\ N_a' F_a' F_a N_a = N_a' N_a = N_a^+ N_a,$$

であり,

$$(A.18) \quad F_a N_a N_a' = F_a F_a' F_a = F_a F_a^- F_a = F_a,$$

$$(A.19) \quad \begin{aligned} \mathbf{F}_a \mathbf{N}'_a \mathbf{F}'_a &= [\mathbf{N}'_{a-1} \quad \mathbf{0}] \mathbf{F}'_a = \mathbf{N}'_{a-1}, \\ \mathbf{F}_a (\mathbf{N}_a^{a-1})' \mathbf{F}'_a &= \mathbf{F}_a \begin{bmatrix} \mathbf{O}_{a-1} \\ \mathbf{1} \end{bmatrix} \mathbf{F}'_a = \mathbf{O}_{a-1}, \end{aligned}$$

である.

さらに, (A.9)より単位行列を分離すれば $\mathbf{U}_{a-1} = \mathbf{E}_{a-1} + \mathbf{U}_{a-1} \mathbf{N}_{a-1}$ であり, (A.14)を使えば,

$$\mathbf{U}_{a-1} \mathbf{F}_a = (\mathbf{E}_{a-1} + \mathbf{U}_{a-1} \mathbf{N}_{a-1}) \mathbf{F}_a = \mathbf{F}_a + \mathbf{U}_{a-1} \mathbf{F}_a \mathbf{N}_a \mathbf{F}'_a \mathbf{F}_a,$$

であるから, (A.16)を使って,

$$(A.20) \quad \begin{aligned} \mathbf{U}_{a-1} \mathbf{F}_a (\mathbf{E}_a - \mathbf{N}_a) &= \mathbf{F}_a + \mathbf{U}_{a-1} \mathbf{F}_a \mathbf{N}_a (\mathbf{F}'_a \mathbf{F}_a - \mathbf{E}_a) = \mathbf{F}_a - \mathbf{U}_{a-1} \mathbf{F}_a \mathbf{N}_a \mathbf{e}_a \mathbf{e}'_a \\ &= \mathbf{F}_a - \mathbf{U}_{a-1} \mathbf{e}_{a-1, a-1} \mathbf{e}'_a = \mathbf{F}_a - \mathbf{1}_{a-1} \mathbf{e}'_a = [\mathbf{E}_{a-1} \quad -\mathbf{1}_{a-1}], \end{aligned}$$

となる. また,

$$(A.21) \quad \mathbf{U}_a = \mathbf{F}'_a \mathbf{U}_{a-1} \mathbf{F}_a + \mathbf{e}_a \mathbf{e}'_a,$$

とも分解できる.

階差行列 \mathbf{D}_a : 左からかけてベクトルの要素の1次階差を作る行列 \mathbf{D}_a を,

$$(A.22) \quad \mathbf{D}_a \equiv [\mathbf{E}_{a-1} \quad \mathbf{0}_{a-1}] - [\mathbf{0}_{a-1} \quad \mathbf{E}_{a-1}] = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(a-1) \times a},$$

で定義する. $\mathbf{D}_a \mathbf{1}_a = \mathbf{0}_a$ であり, $\beta = [\beta_1, \beta_2, \dots, \beta_a]'$ に対して,

$$\mathbf{D}_a \beta = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_2 - \beta_3 \\ \vdots \\ \beta_{a-1} - \beta_a \end{bmatrix},$$

である. 削除行列とシフト行列を使えば,

$$\mathbf{D}_a = \mathbf{F}_a - \mathbf{F}_a \mathbf{N}_a = \mathbf{F}_a (\mathbf{E}_a - \mathbf{N}_a),$$

である.

階差行列 \mathbf{D}_a の1つの一般逆行列を与える. (A.9)より,

$$(A.23) \quad \mathbf{D}_a^- = \mathbf{U}_a \mathbf{F}'_a \in \mathbb{R}^{a \times (a-1)},$$

が予想され, 確かに,

$$\mathbf{D}_a \mathbf{D}_a^- = \mathbf{F}_a (\mathbf{E}_a - \mathbf{N}_a) \mathbf{U}_a \mathbf{F}'_a = \mathbf{F}_a \mathbf{E}_a \mathbf{F}'_a = \mathbf{E}_{a-1},$$

となって, 一般逆行列の条件 $\mathbf{D}_a \mathbf{D}_a^- \mathbf{D}_a = \mathbf{D}_a$ を満たす. (A.23)は具体的には, \mathbf{U}_a が上三角行列で, その右から \mathbf{F}'_a をかけているから最終列が削除され,

$$D_a^- = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ & 1 & \ddots & 1 \\ & & \ddots & \vdots \\ & & & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

である。(A.21)を利用すれば,

$$D_a^- = U_a F_a' = F_a' U_{a-1} F_a F_a' + e_a e_a' F_a' = F_a' U_{a-1},$$

でもある。(A.20)より,

$$D_a^- D_a = F_a' U_{a-1} F_a (E_a - N_a) = F_a' (F_a - \mathbf{1}_{a-1} e_a'),$$

であり,(A.12)も使って,

$$(A.24) \quad E_a - D_a^- D_a = F_a' F_a + e_a e_a' - F_a' F_a + F_a' \mathbf{1}_{a-1} e_a' = (e_a + F_a' \mathbf{1}_{a-1}) e_a' = \mathbf{1}_a e_a',$$

と簡単になることがわかる。

付録 B. 行列の積み上げ一直和とクロネッカ積

行列 $X_j (j = 1, \dots, b)$ の直和は,

$$X_1 \oplus X_2 \oplus \cdots \oplus X_b = \bigoplus_{j=1}^b X_j = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_b \end{bmatrix},$$

と定義される。一般に行列 X_j の大きさは異なってもよい。行列 $Y_j (j = 1, \dots, b)$ があるとき、行列積 $X_j Y_j$ が定義されれば,

$$(\bigoplus_j X_j)(\bigoplus_j Y_j) = \bigoplus_j X_j Y_j,$$

である。 X_j がすべて同一のときは($X_j = Z$ として),

$$\bigoplus_j Z = \begin{bmatrix} Z & & \\ & \ddots & \\ & & Z \end{bmatrix} = E_b \otimes Z,$$

とクロネッカ積を用いても書ける。

クロネッカ積は左側の行列の要素に右側の行列を埋め込んで大きな行列を作る操作と考えてよい。行列の大きさが整合していれば, $(A \otimes B)(C \otimes D) = AC \otimes BD$ である。行列をベクトル化する vec 演算子との間に,

$$(B.1) \quad \begin{aligned} \text{vec} ABC &= (C' \otimes A) \text{vec} B, \\ \text{vec} AB &= (E \otimes A) \text{vec} B = (B' \otimes E) \text{vec} A, \end{aligned}$$

という関係がある (Magnus and Neudecker, 1999, p.30)。

以下, 行列 $X_j \in \mathbb{R}^{a \times c} (j = 1, \dots, b)$ とし, これら b 個を縦方向に積み上げた行列

$$\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_b \end{bmatrix} \in \mathbb{R}^{ab \times c},$$

を考える。まず,

$$(B.2) \quad \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_b \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & & \\ & \ddots & \\ & & \mathbf{X}_b \end{bmatrix} \begin{bmatrix} \mathbf{E}_c \\ \vdots \\ \mathbf{E}_c \end{bmatrix} = (\oplus_j \mathbf{X}_j)(\mathbf{1}_b \otimes \mathbf{E}_c),$$

と行列の直和とクロネッカ積を使って表すことができる。 \mathbf{X}_j がすべて同一のときは ($\mathbf{X}_j = \mathbf{Z}$ として),

$$\begin{bmatrix} \mathbf{Z} \\ \vdots \\ \mathbf{Z} \end{bmatrix} = \mathbf{1}_b \otimes \mathbf{Z} = (\mathbf{E}_b \otimes \mathbf{Z})(\mathbf{1}_b \otimes \mathbf{E}_c) = (\oplus_j \mathbf{Z})(\mathbf{1}_b \otimes \mathbf{E}_c),$$

である。

つぎに, 行列 $\mathbf{A} \in \mathbb{R}^{d \times c}$ に対しては,

$$(B.3) \quad \begin{bmatrix} \mathbf{A}\mathbf{X}_1 \\ \vdots \\ \mathbf{A}\mathbf{X}_b \end{bmatrix} = \begin{bmatrix} \mathbf{A} & & \\ & \ddots & \\ & & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_b \end{bmatrix} = (\mathbf{E}_b \otimes \mathbf{A})(\oplus_j \mathbf{X}_j)(\mathbf{1}_b \otimes \mathbf{E}_c) \\ = (\oplus_j \mathbf{A})(\oplus_j \mathbf{X}_j)(\mathbf{1}_b \otimes \mathbf{E}_c) \\ = (\oplus_j \mathbf{A}\mathbf{X}_j)(\mathbf{1}_b \otimes \mathbf{E}_c) \in \mathbb{R}^{bd \times c}$$

となり, (B.2) から直接得られるものに当然ながら等しい。

また, 行列 $\mathbf{B} \in \mathbb{R}^{c \times e}$ を右からかけると, 縦積みのブロック行列の表記からは一目瞭然ではあるが,

$$\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_b \end{bmatrix} \mathbf{B} = (\oplus_j \mathbf{X}_j)(\mathbf{1}_b \otimes \mathbf{E}_c) \mathbf{B} = (\oplus_j \mathbf{X}_j)(\mathbf{1}_b \otimes \mathbf{E}_c)(\mathbf{1} \otimes \mathbf{B}) \\ = (\oplus_j \mathbf{X}_j)(\mathbf{1}_b \otimes \mathbf{B}) = (\oplus_j \mathbf{X}_j)(\mathbf{E}_b \otimes \mathbf{B})(\mathbf{1}_b \otimes \mathbf{E}_e) \\ = (\oplus_j \mathbf{X}_j)(\oplus_j \mathbf{B})(\mathbf{1}_b \otimes \mathbf{E}_e) = (\oplus_j \mathbf{X}_j \mathbf{B})(\mathbf{1}_b \otimes \mathbf{E}_e) \\ = \begin{bmatrix} \mathbf{X}_1 \mathbf{B} \\ \vdots \\ \mathbf{X}_b \mathbf{B} \end{bmatrix} \in \mathbb{R}^{ab \times e},$$

と \mathbf{B} が潜り込む。

あわせて,

$$\begin{bmatrix} \mathbf{A}\mathbf{X}_1 \mathbf{B} \\ \vdots \\ \mathbf{A}\mathbf{X}_b \mathbf{B} \end{bmatrix} = (\oplus_j \mathbf{A})(\oplus_j \mathbf{X}_j)(\oplus_j \mathbf{B})(\mathbf{1}_b \otimes \mathbf{E}_e) = (\oplus_j \mathbf{A}\mathbf{X}_j \mathbf{B})(\mathbf{1}_b \otimes \mathbf{E}_e),$$

である。この形に整理されれば、積み上げた全体の上から j 番目のブロックの行列 $\mathbf{A}\mathbf{X}_j\mathbf{B}$ を取り出すのは容易である。

付録 C. 一般逆行列と斉次連立 1 次方程式の一般解

行列 $\mathbf{A} \in \mathbb{R}^{a \times b}$ に対して $\mathbf{A}\mathbf{G}\mathbf{A} = \mathbf{A}$ を満たす行列 $\mathbf{G} \in \mathbb{R}^{b \times a}$ は \mathbf{A} の一般逆行列と呼ばれ、 \mathbf{A}^- で表記される。 \mathbf{A}^- は唯一とは限らない。

斉次連立 1 次方程式 $\mathbf{A}\mathbf{x} = \mathbf{0}_a$ ($\mathbf{x} \in \mathbb{R}^b$) の一般解は、任意の $\mathbf{z} \in \mathbb{R}^b$ について、

$$(C.1) \quad \mathbf{x} = (\mathbf{E}_b - \mathbf{A}^- \mathbf{A})\mathbf{z},$$

で与えられる (Harville, 1997, p.140)。

参 考 文 献

- Fienberg, S. E. and Mason, W. M. (1979). Identification and estimation of age-period-cohort models in the analysis of discrete archival data, *Sociological Methodology 1979* (ed. K. F. Shuessler), 1-67, Jossey-Bass, San Francisco.
- Fu, W. (2018). *A Practical Guide to Age-Period-Cohort Analysis: The Identification Problem and Beyond*, CRC Press, Boca Raton.
- Glenn, N. D. (2005). *Cohort Analysis*, 2nd ed., Sage Publications, Thousand Oaks, London, New Delhi.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*, Springer, New York.
- Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edition, John Wiley & Sons, New York.
- Mason, W. M. and Fienberg, S. E. (eds.) (1985). *Cohort Analysis in Social Research: Beyond the Identification Problem*, Springer-Verlag, New York.
- 中村 隆 (1982). バイズ型コウホート・モデル—標準コウホート表への適用—, 統計数理研究所彙報, **29**, 77-97.
- Nakamura, T. (1986). Bayesian cohort models for general cohort table analyses, *Annals of the Institute of Statistical Mathematics*, **38**, 353-370.
- 中村 隆 (2005). コウホート分析における交互作用効果モデル再考, 統計数理, **53**, 103-132.
- 中村 隆 (2019). コウホート分析の古今—一般化線形混合モデルとしてのコウホートモデル, 社会と調査, **32**, 106-114.
- O'Brien, R. M. (2015). *Age-Period-Cohort Models: Approaches and Analyses with Aggregated Data*, CRC Press, Boca Raton.
- Ryder, N. B. (1965). The cohort as a concept in the study of social change, *American Sociological Review*, **30**, 843-861.
- 坂口尚文, 中村 隆 (2019). 混合効果モデルとしてみたコウホート分析モデル, 理論と方法, **33**, 3-17.
- Yang, Y. and Land, K.C. (2013). *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications*, CRC Press, Boca Raton.

The Design Matrices of an Age-Period-Cohort Model for a Standard Cohort Table

Takashi Nakamura

Professor Emeritus, The Institute of Statistical Mathematics

This paper gives the explicit expressions of the design matrices of an age-period-cohort model for standard cohort table data classified by age group and survey period, obtained from repeated cross-sectional surveys, to separate the effects of age, period, and cohort factors. The matrices are derived from imposing equality or zero-sum constraints on the cell parameters in a cohort table as well as from visual inspection.

統計数理研究所 75 周年記念 研究業績紹介

目次

- | | | |
|----|------------|--|
| 1 | 樋口知之 | モデリング変革の四半世紀を振り返って |
| 2 | 川崎能典 | スパース正則化法によるリスク要因の探索 |
| 3 | 宮里義彦 | 統計科学における制御理論の研究 |
| 4 | 吉本敦 | 森林資源管理のための数理最適化モデリング |
| 5 | 庄建倉 | 地震間の対話の理解と地震活動確率予測の展望 |
| 6 | 瀧澤由美 | 神経系の動作の解明と時空間推定システムへの適用 |
| 7 | 三分一史和 | 時空間データ解析による生体メカニズムの解明 |
| 8 | 伊庭幸人 | マルコフ連鎖モンテカルロ法の世界を広げる
-高次元の多峰性分布からのサンプリング- |
| 9 | 福水健次 | 正定値カーネルによる統計的機械学習 |
| 10 | 南和宏 | 位置情報軌跡の匿名化技術 |
| 11 | 小山慎介 | 神経スパイク時系列パターンの特徴解析 |
| 12 | 日野英逸 | ノンパラメトリックエントロピー推定とその応用 |
| 13 | 早水桃子 | 組合せ論的系統学における最近の話題-グラフ理論が拓く系統解析の新展開- |
| 14 | 中野純司 | 統計数理研究所で開発された R パッケージ |
| 15 | 上野玄太 | 混合分布モデルとデータ同化 |
| 16 | 中野慎也 | 地球磁気圏の撮像観測とそのデータ同化 |
| 17 | 野村俊一 | 前震識別とその予測可能性 |
| 18 | 吉野諒三 | 「日本人の国民性調査」から「文化多様体解析」へ |
| 19 | 前田忠彦 | 社会調査方法論の実践的研究 |
| 20 | 朴堯星 | 組織規範継承を可能にする目的指向型成果評価と相互依存性の計量分析 |
| 21 | 廣瀬雅代 | 立川市町丁目別住民意識調査分析追記-小地域推定モデル活用に向けて- |
| 22 | 山下智志 | 制度的制約下におけるデータベース構造化、モデリング、モデル評価 |
| 23 | 金藤浩司 | 極値分布と指数逆ガウス型分布に関するある一般化について |
| 24 | 伊藤陽一 | 医師主導治験による医療機器開発の経験 |
| 25 | 船渡川伊久子 | 経時データ解析と健康指標の長期推移 |
| 26 | 野間久史 | 先端医学研究の発展を支える統計数理とデータサイエンス |
| 27 | 清水信夫 | 集約的シンボリックデータ解析 |
| 28 | 松井知子 | ツイートを利用した都市気温の解析-都市インテリジェンス向上をめざして- |
| 29 | 吉田亮 | 機械学習による新物質の発掘 |
| 30 | 島谷健一郎 | 方向統計学と動物移動生態学 |
| 31 | Stephen Wu | 機械学習に基づく高熱伝導率ポリマーの設計事例 |
| 32 | 村上大輔 | 大規模空間データのための空間可変係数モデリング |
| 33 | 栗木哲 | オイラー標数法によるウィシャート行列の最大固有値分布の近似 |
| 34 | 二宮嘉行 | 変化点モデルのための AIC |
| 35 | 間野修平 | 抽出法と計算代数 |
| 36 | 加藤昇吾 | 角度の観測を含むデータのための統計モデル |
| 37 | 志村隆彰 | 極値分布の吸引領域と離散性 |
| 38 | 荻原哲平 | 拡散過程の非同期・ノイズ付観測データに対する最尤型推定法 |
| 39 | 江口真透 | 一般化平均による統計モデル |
| 40 | 藤澤洋徳 | ガンマ・ダイバージェンスに基づいたロバスト統計 |

- 41 持橋大地 統計的自然言語処理と統計学
- 42 逸見昌之 メタアナリシスにおける公表バイアスの最悪評価に基づく感度解析
- 43 坂田綾香 スピングラス理論による制限等長定数評価
- 44 伊藤聡 多層整数計画に基づくクリンチ/エリミネーション数の計算
- 45 池田思朗 天文学とデータ科学
- 46 田中未来 制約付き非凸スパース最適化問題に対する DC アルゴリズム
- 47 今泉允聡 関数推定の理論に基づく深層学習の原理解析

モデリング変革の四半世紀を振り返って

25 years of a paradigm shift in a statistical modeling technology

前所長 樋口 知之 (Tomoyuki Higuchi) *

キーワード：ベイズモデリング、粒子フィルタ、カーネル法、深層学習

私が研究所に入ったのは博士号取得の直後である平成元年4月、また所長を退任すると同時に研究所を退職するのが平成が終わる直前の平成31年3月であるため、平成の30年間は私にとって研究者人生そのものである。私が入所した当時は、その10年ほど前から研究が活発化していたベイズモデリングの成熟期にあった。事前分布および尤度関数が線形・ガウス型の理論は確立し、応用の観点からも、いくつかの課題では本研究所の研究成果が標準的方法として認知されるなど、研究進展も著しかった。現象や既存知識のモデリングにおいては、計算上からくる線形・ガウス性の制約は、研究者の自然な思考を妨げる。そのため、その制約を“人手”でもって解析的に緩める論文が数多く出版されていた。その流れと平行して、計算機集約的な方法の開発も活発化していた。高い次元の潜在変数がつくる超高次元の分布から、マルコフ連鎖でもって直接的に実現値を得るマルコフ連鎖モンテカルロ法や、その分布を低次元の分布に分解し（いわゆる逐次ベイズフィルタ）、低次元の分布の逐次更新を数値的に得る「非線形フィルタ」などである。私は学位を宇宙プラズマ物理の領域で得たため、統計力学の自由エネルギーや転送積分とほぼ同一の発想にもとづく両手法を自然に理解でき、大いに魅惑された。

平成にはいつしばらくして北川元所長は、「非線形フィルタ」の適用範囲を大幅に拡張するモンテカルロフィルタを提案された（1993,1996年）。この手法は今日において、粒子フィルタの手法群の中で、原始的ではあるがレガシーとして高く評価されている。北川先生のすぐ近くで研究していた自分は、そのアルゴリズムの大規模並列コンピュータへの実装容易性に衝撃を受け、さまざまな非線形・非ガウス型の時系列モデルへの適用を試みた。また、世界同時的に多くの研究者が、アルゴリズムの高度化と現実問題への応用をすすめ、21世紀早々（2001年）にシュプリングァーから書籍を出版できた。私も著者の一人として参加するチャンスを得、当時新進気鋭の海外の若い研究者と交流できたことは、その後の自分の研究者人生にとって大きな糧となった。この友人の多くは現在、後述する深層学習に研究の軸足を移しているが、そのことは偶然でなく、むしろモデリング技術の発展を考えれば極めて自然なことと私は考える。粒子フィルタは、時系列モデルが所与であれば、パラメータや状態変数の推定に関して原理的には万能である。もちろん、モンテカルロ誤差（有限サンプルによる表現限界）からくる、分布の表現能力の喪失（いわゆる退化問題）や、尤度値の不安定性など、数値的問題は避けられないが、問題や経験に即して“人”が対象を自由にモデリングできるようになった点は、モデリング技術の発展において大きな革新であった。

2000年代にはいつし、非線形モデリングの観点からは、別の形で大きな進歩があった。それはカーネル法の登場である。1998年にGoogleが、また2004年にはFacebookが創業されるなど、ビッグデータを新しい情報サービスという価値に転換する企業が当時、続々と誕生していた。ビッグデータを用いた容易なタスクは、識別関数の構築を通じた判別や分別である。カーネル法は、データ空間で複雑な識別境界面を同定するのではなく、あえて超高次元の特徴

* 中央大学理理工学部：〒112-8551 東京都文京区春日 1-13-27

ベクトル空間を構成し、そこで豊富に蓄積された線形のモデリング技術を適用するものである。もちろん、データ間の類似度を規定するカーネル関数は“人”が設定しなければならないが、データ空間での非線形モデリングに求められる巧みの技の多くを、カーネルトリックに押しつけられた利点は大きい。また、パターン認識手法の多くはデータどうしの内積計算を含むため、カーネルトリックによって既存の線形モデリングの諸手法が非線形版に自然に拡張できた。2000年代はビッグデータの登場により、統計的モデリングの主たる興味である生成モデルの構築から、カーネル法と最適化の活用による、複雑な識別関数（識別モデル）の自動構築に、研究のトレンドがシフトした。それと同時に、機械学習の言葉で代表される研究者コミュニティが育っていったのもこの時代の特徴である。

2000年代半ばから2010年代にかけて、非ガウス型モデリングにおいても大きな発展がスパースモデリングの普及によってもたらされた。応用現場における統計的モデリングの成否は、膨大な説明変数群の中からタスクの解決に有効な特徴ベクトルの構築にあると断言できる。なお、画像、音声、自然言語の処理においては、現在、その課題は深層学習によって大幅に自動化されている。カーネル法は、データ空間から（暗に）高次元の特徴ベクトルを構成する方法なので、情報圧縮の文脈では逆センスの手法である。今、興味ある対象を膨大な説明変数の中から少数個で線形回帰表現する問題を考える。何次の回帰モデルとするのか、さらにどの変数を使うかで、膨大な数の回帰モデルが存在する。モデル数は組み合わせ爆発しており、AICなどの情報量規準による最適モデルの探索（変数選択）は機能しない。一方スパースモデリングでは、回帰係数にL1正則化を加えた上で推定を行う。最適化関数の形がL1であることから、L1正則化は非ガウスモデリングと言える。この最適化の結果として、重要な説明変数のセットが自動的に浮きあがる。このように、説明変数の選択といった、限定されたモデリング技術に関しては、スパースモデリングにより線形・非ガウス型モデリングが実現された。

2010年代にはいつの最大の衝撃は、特定のタスクでの深層学習の圧倒的パフォーマンスである。その性能の高さは、入出国時の自動顔画像判別、スマートスピーカーでの音声認識、多言語自動翻訳など、私たちの生活に身近な製品として既に具現化している。これらの技術は、人の働き方や社会の構造そのものにも直接的に影響を与えていることから、深層学習はこの四半世紀の情報科学技術における最大のブレイクスルーであると言わざるを得ない。深層学習で使われるニューラルネットワークは、層数が大幅に増えた以外、第二次ニューロブームの時のものと違いは無く、そのパラメータ学習アルゴリズムも、Back propagationを基本とする以前のものとは大差はない。層数が大幅に増えた結果、パラメータ数も爆発的に増え、学習アルゴリズムもさほど賢くなっているわけでもないのに、必然と計算リソースはこれまでとは桁違いに必要となる。ただし、上述したように、それまでの特徴ベクトル構成法は機械学習の『匠の技』と言え、“機械”学習にもかかわらず、人間の判断が最も性能向上に大切という羊頭狗肉の面もあったが、その問題を特定領域では基本的に解決した点は画期的である。深層学習では、通常、最適化関数はL2であるため、誤差にガウス性を仮定している。よって、非線形・ガウス型モデリング技術は、深層学習により、少なくとも予測・判別性能の観点からはほぼ極みに達したと言える。今後は、深層学習の大きな成功事例である、入力データが画像、音声、テキストなどに、一般的な問題での入力データをあえて変換すれば、特徴ベクトルの選択問題を回避可能となる。

非ガウス・非線形モデリングを自動化する試みは、この四半世紀に大きな飛躍を遂げた。その一方で、ビッグデータ×AI（人工知能）を土台に大きな存在感を示す米中のプラットフォームに対する風当たりも強くなってきている。今後は、モデリング技術の向上を計る上で、モデルの“人”による解釈可能性、帰納法で構築された意思決定システムの説明責任、データのバイアスと社会的偏見の分離など、人に寄り添った視点が極めて重要になってくる。

スパース正則化法によるリスク要因の探索

Identifying Risk Factors by Sparse Regularization

モデリング研究系 川崎 能典 (Yoshinori Kawasaki)

1. 円滑閾値型推定方程式による変数選択・グルーピング

情報通信技術やデータ計測技術等の発展に伴い、大規模なデータが蓄積されるに至って久しい。統計科学においては高次元データとしばしば言及される状況は、典型的には個体に付随して観測される属性が多数に及んでいる場合を指し、しばしば個体数に比べて属性数が遙かに上回る。こうした状況下で回帰分析を行う場合、「説明変数候補は多数得られているが、目的変数と関連性のある要因はごく少数である」という制約を置くのが現実的である。それを実現したのが LASSO (Least Absolute Shrinkage and Selection Operator) とその変種であり、一般にスパース正則化、 L^1 正則化、スパース推定法などと言われる。

ペナルティつき損失関数を一般に $L(\theta) + \sum_{j=1}^d \rho_j(|\theta_j|)$ と記す。ここで ρ_j は j 番目のパラメータ θ_j に関する非負のペナルティ関数であり、 $\rho_j(|\theta_j|) = \lambda_j |\theta_j|$ なら adaptive LASSO である。ここで $\rho_j(|\theta_j|) = w_j \theta_j^2 / 2$ としつつも、 $\delta_j \in [0, 1]$ によって $w_j = \delta_j / (1 - \delta_j)$ と取り直すことで、解くべき推定方程式は $(1 - \delta_j) \partial L(\theta) / \partial \theta_j + \delta_j \theta_j = 0$ ($j = 1, \dots, d$) となる。 $\delta_j = 1$ は $w_j = \infty$ に対応し、このとき $\theta_j = 0$ に帰着する。これを円滑閾値型推定方程式 (Smooth-Threshold Estimating Equation, Ueki (2009)) と呼び、以下 STEE と略す。

Ueki and Kawasaki (2011) は、STEE の方法論を変数のグルーピングにも拡張し、スパース変数選択とグルーピングを半自動的に行う方法を提案した。応用例としては、気管支疾患の判別や与信スコアリングの問題を取り上げている。Kawasaki and Ueki (2015) は、電話によるマーケティングデータを例に、STEE 法の性能を他のスパース正則化法と比較している。

2. 多重共線性と標準化更新度

STEE 法は、初期推定量の良さに依存している。初期推定量は飽和重回帰やリッジ回帰から構成するが、データ間に多重共線性が強ければ、有意な変数は偶発的に生き残っているに過ぎない可能性がある。そこで、ひとつのモデルを信頼するのではなく、たまたま選ばれなかったかもしれない真の因果変数も拾い上げる方法の構築が必要となる。

Ueki and Kawasaki (2013) は、線形回帰モデルの枠組みにおいて、飽和回帰モデルからの相対差で適合度基準を設定し、その基準を満たすモデルを前進選択法で探索することで、複数の「説明力同等」なモデルを手元に残す方法を提案した。手順は以下の通り。1) まず各変数を 1 つだけ含んだ p 個の単回帰モデルから出発し並列的に探索、2) 適合度基準を満たしたモデルにはお墨付きを与えて終了、3) 満たさないモデルについては、それ以外の変数を各ステップで 1 つずつ取り込み、適合度基準を満たすまで深掘り、4) 現在のモデルの変数添字集合を C とし、いま変数 k を加えたときのあてはまりの改善度を標準化更新度 (Standardized Update)

$$SU_{k,C} = \frac{\|y - X_C \hat{\beta}_C\|^2 - \|y - X_{C \cup \{k\}} \hat{\beta}_{C \cup \{k\}}\|^2}{\|y - \bar{y}1\|^2 - \|y - X \hat{\beta}\|^2}, k \notin C$$

で測り、 SU がある閾値を超えた時に変数 k を採用する。

数値実験の結果, 完全多重共線性の下でも SU を使う提案手法は偽陽性・偽陰性ともに小さく, 真の回帰関数に関連している変数の組合せを高精度で発見できた. 一方 Elastic Net は全般的に偽陽性率が高く, 完全多重共線性の下では偽陰性も高いことが示された.

3. 効果がマスクされた変数の探索

多重共線性が深刻な説明変数群を使って推定された線形飽和重回帰モデルでは, 有意な説明変数は偶発的に有意になっている過ぎない可能性がある. 一方, 飽和重回帰モデルで有意にならず, かつ目的変数との周辺相関がないと思われる説明変数でも, 特定の説明変数集合を伴って推定されれば有意になることがある.

Ueki, Kawasaki and Tamiya (2017) では, 効果が他の変数にマスクされている説明変数を効率的に探索する方法を提案している. 飽和回帰モデルが推定可能な状況であっても, 部分回帰的全探索は計算負荷が高い. 提案手法は, 説明変数・被説明変数を合わせた変数群内の全てのペアで計算した相関係数を基に変数間の接続性を双方向グラフで表現し, 特定の説明変数から目的変数までの最短経路をダイクストラ法で求めることで, 効果を浮かび上がらせる共変量集合を特定する. この方法は, ケースの次元より説明変数の次元が大きい状況下で, かつ説明変数群に多重共線性が潜んでいる状況でも適用可能である. (ただし上掲論文中の応用例では, 対象に関する知見から, 変数群を独立なブロックに分けることが妥当性を持つので, 各ブロックごとに計算している.)

謝 辞

本稿で紹介した一連の研究は, 植木優夫博士 (執筆時点では理化学研究所革新知能統合研究センター所属) との共同研究である. リスク解析戦略研究センター草創期の 2008 年に特任研究員として着任した同氏から, Ueki (2009) の草稿について討論させて頂いたのを契機に, その後は公募型共同利用 (25-共研-1018, 26-共研-1014, 27-共研-1013, 28-共研-1011, 29-共研-1009, 30-共研-1013) が一連の研究成果につながっていった. ここに記して感謝申し上げます.

参 考 文 献

- Ueki, M. (2009). A note on automatic variable selection using smooth-threshold estimating equations, *Biometrika*, **96**, 1005–1011.
- Ueki, M. and Kawasaki, Y. (2011). Automatic grouping using smooth-threshold estimating equations, *Electronic Journal of Statistics*, **5**, 309–328.
- Ueki, M. and Kawasaki, Y. (2013). Multiple choice from competing regression models under multicollinearity based on standardized update, *Computational Statistics and Data Analysis*, **63**, 31–41.
- Kawasaki, Y. and Ueki, M. (2015). Sparse predictive modeling for bank telemarketing success using smooth-threshold estimating equations, *Journal of Japanese Society of Computational Statistics*, **28**, 53–66.
- Ueki, M., Kawasaki, Y. and Tamiya, G. (2017). Detecting genetic association through shortest paths in a bi-directed graph, *Genetic Epidemiology*, **41**, 481–497.

統計科学における制御理論の研究

Research on Control Theory in Statistical Science

モデリング研究系 宮里 義彦 (Yoshihiko Miyasato)

1. 制御理論の背景

制御理論の歴史は伝達関数と周波数領域の評価に基づく古典制御に始まり、状態空間と時間領域の評価に基づく現代制御の時代を通過し、周波数領域と時間領域、伝達関数と状態空間の双方の評価方式と表現方式を統合したポスト現代制御の時代を経て現在に至っている。いずれの手法においても制御系を設計するに当たって、制御対象の適正なモデルを求める必要があるが、制御対象をどのようにモデリングするか、またそのモデルに含まれる不確定要因をどのように評価するかによって、適用される制御手法や達成される制御性能が規定される。従って高性能の制御系を実現するために、制御を意識したモデリングやモデルの不確定要因を考慮した制御という視点が重要で、モデリングと制御は切り離して考えられない。

2. 制御理論の研究

このような制御理論とモデリングの関係を考慮して、モデリングと制御を同時に行う適応学習制御を中心とした研究を行っている。特に適応制御は制御器の実時間調整のために制御系全体の安定解析が困難で、適用上様々な制約を受けるが、その制約を緩和し適応制御の適用範囲を広げる研究を進めてきて、現在は適応制御に関連する非線形制御と線形制御の立場から、およびそれらの多体系（マルチエージェント系）への適用の立場から研究に従事している。また応用研究の分野でも、車両のセミアクティブサスペンションの制御系設計を行い、実機（高速バス）の走行試験から良好な結果を得ることが出来た。

3. 研究事例 I：非線形制御の立場

未知の対象を制御する場合、対象のパラメトリックモデルを求めて、モデルのパラメータの推定値を正しいと見なして制御器の設計をする（certainty equivalence の立場）。ところがその推定値は常に正しいとは限らず、またパラメトリックモデル自体も対象の近似表現の一つにすぎないので、所望の制御性能を達成するためには、対象のさまざまな不確定要因も考慮に入れて、制御系を構成しなければならない。このような問題に対して、パラメータの推定誤差を制御問題の外乱と見なして、外乱の影響を抑制する非線形適応制御系を設計する手法を開発し、パラメータが時間的に変動したり不確定要素が存在する場合でも、所望の制御性能を達成できるようになった。この手法は非線形パラメトリックモデルの一種であるニューラルネットを含む制御系の設計にも適用が可能である。同じ考え方は高次振動モードを有する複雑なシステム（非線形プロセス、柔軟構造物、弾性アームなど）のモデリングと制御にも適用可能で、低次元モデルに含まれるモデリング誤差（スピルオーバー）の影響を低減化して、実用的な有限次元（低次元）の制御器でシステムの振動抑制と制御を実現することができる。

4. 研究事例 II：線形制御の立場

パラメータの上下限が規定された線形システムとして記述されるプロセスは、その上下限を端点とするポリトープの内点として表現することができる（ポリトピックモデル）。そのような対象についてはポリトープの内点を定めるパラメータをスケジューリングパラメータ（SP）と見なして、SP に応じて制御器の特性を変化させることで、より高性能の制御結果が得られる。この制御方式はゲインスケジューリング（GS）制御と呼ばれていて、制御理論においては線形行列不等式に基づく制約下における設計手順として定式化される。GS 制御は SP が正確に求められれば、ポリトープ内の変動に対してシステムの安定化と外乱抑制が達成されるが、SP が正確でないと安定性も保証されない。このような場合に、プロセスの操業データを用いて制御誤差を観測して SP を実時間で再調整する適応型の GS 制御方式を開発した。ポリトピックモデルに基づく適応制御は、制御のためのモデリングにも多大の影響を与えらると思われる。

5. 研究事例 III：多体系（マルチエージェント系）への適用の立場

未知パラメータを含む複数の動的システムを個々のエージェントとするマルチエージェント系に対して、適応的に速度追従型の群生行動を実現するフォーメーション制御やリーダーフォロワー型の追従行動を実現するコンセンサス（合意形成）制御の研究を行っている。フォーメーション制御においては、未知パラメータの推定誤差と群生行動に関するポテンシャル関数の誤差を等価的な外乱と見なした H_∞ 制御問題の解としてフォーメーション制御機構を導出し、コンセンサス制御においては限定された通信構造に対応するネットワークグラフに着目し、未知パラメータの推定誤差とネットワーク密度に関する不確定性を等価的な外乱と見なした H_∞ 制御問題の解としてコンセンサス制御機構を導出している。実際の問題としては高速道路における自動車の群制御（スマートハイウェイ）や複数のロボットマニピュレータによる協調動作、ドローンの群制御や宇宙機のランデブー問題などが該当し、これらの調和行动の実現のための基本原理を解明する研究である。

6. 最後に

統計科学では、有限時間の現象を再現する開ループ的なモデルを構築することに主眼があり、制御はモデルの有効性を検証する項目の 1 つと見なされることが多い。しかし制御系を良好に動作させるには、既存モデルの統計的な当てはめに留まらず対象の物理的特性や原理にも目配りを行い、モデルと制御の総合的な考察が必要となる。それには制御科学の深い知見が必要であり、制御科学と統計科学の緊密な関係が必要不可欠である。また制御理論がこれまで取り扱ってきた時間軸上の動的システムという枠組みを越えて、より広いクラスの離散事象システム、生産システム、通信ネットワークシステム等も対象として発展していくためには、システム科学や情報科学全般も含めた横断的な研究が今後ますます重要になっていくと思われる。

参 考 文 献

- 宮里義彦 (2013). 適応制御の回顧と展望, 計測と制御, **52** (4), 361-367.
- 宮里義彦 (2014). 分布定数系の適応制御, システム/制御/情報, **58** (9) 365-370.
- 宮里義彦 (2017). 無限次元系の協調制御, 計測と制御, **56** (12), 925-930.
- 宮里義彦 (2018). 『適応制御』, コロナ社, 東京.

森林資源管理のための数理最適化モデリング

Mathematical Optimization for Forest Resource Management

モデリング研究系 吉本 敦 (Atsushi Yoshimoto)

キーワード：森林資源管理, 数理最適化, 最適制御, 動的計画法, 0-1 整数計画法.

森林資源管理において、森林を「いつ、どこから、どのくらい」伐採するかを決めることは、古くから取り組まれてきた主要な課題の一つである。このような意思決定をサポートするツールとして、資源の伐採量あるいはそこから得られる利益の最大化と言った資源管理の目的に対して、与えられた条件下で最適な伐採量・方法の在り方を探索できる最適化モデルが構築されてきた。そして、地域的な政策や経済的な要求に対応すべく広くその開発・応用がすすめられてきた。特に、最適な伐採の時空間的配置を考慮した管理の探索ができる整数計画法は、大規模な資源開発、あるいは土地利用の変化に伴う環境への負荷を評価するアプローチとして注目され、例えば、空間的に隣接し合う場所での伐採を同時期に行うことができないという条件を加えることで、伐採が空間的に分散され、大規模な伐採域の創出を避けることができる。森林資源管理は基本的に植林と伐採という単純な制御で対応できるが、それらの時空間的な組み合わせの違いは、生態系のみならず、地域的な社会経済へ異なる影響を及ぼす。森林資源管理問題の研究では、個々の管理ユニット（森林経営上の単位）である林分を対象とする林分レベルと、それらの集合体である森林全体を対象とする森林レベルに大きく分類することができ、それぞれに対し効率的な数理最適化モデルの開発が進められている。

林分レベルでは主に最適な間伐戦略の探求が課題となり、その解の探求では、まず経営から得られる総利益などの最適化を最適制御あるいは変分法の枠組みで連続的に捉え、目的関数を定式化する。次に間伐行動が離散的であることから、動的計画法のフレームワークに変換し、繰り返し演算を行うことにより最適解を探求するというものである。仮に $x(t)$ を森林の状態を表す時間依存の状態変数、 $u(t)$ を森林の成長に影響を与える間伐などの程度を表す制御変数 (control variable) とする。ここで現時点の森林の状態および制御により得られるであろう微小時間間隔の利益の現在価値あるいはパフォーマンス指標を $\dot{I}(x(t), u(t))$ とすることで、下記のように目的を時間 t_0 から t_n までの積分値の最大化として、最適間伐戦略を探求することができる。

$$\begin{aligned} \max_{\{u(t)\}} J &= \int_{t_0}^{t_n} \dot{I}(x(t), u(t)) dt \\ \text{subject to} \\ \dot{x}(t) &= f(x(t), u(t)), \quad x(t_0) = x_0. \end{aligned}$$

なお、 $f(\cdot)$ は森林の状態変化を表す $x(t)$ および $u(t)$ の関数であり、 x_0 は初期状態を表す。上記のように最適制御の枠組みにて定式化できれば、離散的に発生する間伐制御 $u(t)$ による動的計画法への変換が可能になり、林分レベルの数理最適化モデルが構築できる (Yoshimoto *et al.*, 2016)。

一方、森林レベルの数理最適化モデルの構築は 1970 年代に線形計画法の応用から始まり、Johnson and Scheurman (1977) により最適な伐採時期・伐採量を探求するモデルとして Model I および Model II が開発され、伐採計画問題の基本構造が定式化された。Model I が各林分に対し計画期間内で実施可能な伐採パターン（施業）を決定変数 (decision variable) として定式

化するのに対し、Model II は各林分の各期における伐採量を決定変数として定式化している。Model I を例に、 x_{ij} を決定変数とし、第 i 番目の林分において第 j 番目の施業を実施する面積比率を表すこととする。また、 c_{ij} を x_{ij} を実施することにより得られる利益の現在価値とし、森林全体から得られる総利益の現在価値の最大化を目的とすれば、下記のように面積比率制約と各期伐採量制約を伴う伐採計画問題が定式化できる。

$$\begin{aligned} Z &= \max_{\{x_{ij}\}} \sum_{i=1}^m \sum_{j=1}^n c_{ij} \cdot x_{ij} \\ \text{(Model I)} \quad &\text{subject to} \\ &\sum_{j=0}^n x_{ij} = 1 \quad \forall i, \quad (1 - \alpha) \cdot v_0 \leq \sum_{i=1}^m \sum_{j=1}^n v_{ij}^p \cdot x_{ij} \leq (1 + \alpha) \cdot v_0 \quad p = 1, \dots, T. \end{aligned}$$

なお、 m は林分数、 n は 1 つの林分に対する施業数、 α は每期伐採量変化の許容率、 v_{ij}^p は x_{ij} の実行に伴う期間 p で得られる伐採量、 T は計画期間、そして v_0 は伐採量変化の基準となる基準伐採量である。ここで x_{ij} を 0-1 決定変数とすれば上記問題は伐採を施す林分も特定できる数理最適化モデルとなる。

森林レベルの数理最適化モデルの応用はさらに拡大し、1980 年代後半から広大な皆伐域を回避することによる野鳥獣の生息域などの保護、森林景観の確保等の環境保全といった環境配慮型志向に伴い、新たに「隣接しあった林分を同時期に伐採することはできない」という空間的制約条件（隣接林分制約条件）を課すようになってきた。ここで A^S を空間的な林分同士の隣接行列、 A^V を施業間の同時期伐採関係を表す伐採隣接行列とし、 $\text{vec}(X') = (x_{11}, \dots, x_{1n}, \dots, x_{m1}, \dots, x_{mn})'$ とすれば下記により隣接林分制約条件を定式化することができ、上記の Model I に加えることにより解の探求が可能になる (Yoshimoto and Konoshima, 2016)。

$$[\tilde{A} + \text{diag}(\tilde{A} \cdot 1_{mn})] \cdot \text{vec}(X') \leq \tilde{A} \cdot 1_{mn} \quad \text{ただし} \quad \tilde{A} = A^S \otimes A^V.$$

昨今ではこれらのモデルの拡張により、林分の集約を伴う最大伐採域許容問題 (Yoshimoto and Asante, 2018a) や老齢林バッファゾーン形成問題 (Yoshimoto and Asante, 2018b) など様々な空間的制約条件を伴う森林資源管理の数理最適化モデルの開発が盛んに行われている。また生態系保護の観点から病害虫・外来種などの拡散防止に向けた数理最適化モデルも開発されている (Yoshimoto *et al.*, 2017)。

参 考 文 献

- Johnson, K.N. and Scheurman, H.L. (1977). Techniques for prescribing optimal timber harvest and investment under different objectives—discussion and synthesis. *Forest Sci. Mono.* **573**:18p.
- Yoshimoto, A. and Asante, P. (2018a). A new optimization model for spatially constrained harvest scheduling under area restrictions through maximum flow problem. *Forest Sci.* **64**(4):392–406.
- Yoshimoto, A. and Asante, P. (2018b). Focal-point aggregation under area restrictions through spatially constrained optimal harvest scheduling. *Forest Sci.*, in press, DOI:10.1093/forsci/fxy044.
- Yoshimoto, A., Asante, P. and Konoshima, M. (2016). Stand-level forest management planning approaches. *Curr. Forestry. Rep.* **2**:163–176.
- Yoshimoto, A., Asante, P., Konoshima, M., Surový, P. (2017). Integer programming approach to control invasive species spread based on cellular automaton model, *Nat. Resour. Model.* **30**(2), DOI:10.1111/nrm.12101.
- Yoshimoto, A. and Konoshima, M. (2016). Spatially constrained harvest scheduling for multiple harvests by exact formulation with common matrix algebra. *J. Forest Res.* **21**:15–22.

地震間の対話の理解と地震活動確率予測の展望

Understanding the conversations among earthquakes and developing probability earthquake forecasting models

モデリング研究系 庄 建倉 (Jianchang Zhuang)

地震は地震波形記録図の中では観測地での地震動を示す連続なパルス波形である。地震カタログの中では、各地震は地震波形を解析して生成された地震の位置、深さ、大きさ、震源機構、記録精度などの情報を記録した一列の数字で表示される。しかし、現実ではこれらの地震、特に陸上または沖合の大地震は人類に生命や財産の莫大な損失をもたらす。正確な地震予知を達成することは長年にわたる人類の願いである。

現代の地震学的研究結果によれば、地震の発生は自己組織化された臨界現象であり、単一の地震の規模とタイミングを正確に予測することは非常に困難である [1]。しかし、個々の地震は完全に独立していない。地震と地震の間には「対話」がある。地震の前にはしばしば前震と余震がある。同時に、これらの大地震は元の断層のコロンボ応力分布を変化させ、余震の分布に影響を与え、その後の大地震の発生を遅らせたり、早めたりする。同時に、地震断層では高応力状態にあるとき様々な異常が発生する。したがって、地震活動は完全にランダム、予測不可能ではない。これらの予測可能なコンポーネントの解明は、将来の地震リスクの推定、防災政策の制定、震災後救助に大きな影響を与える。

地震活動において、どの成分を予測でき、どの成分を予測できないのか？地震活動の予測可能な最大の成分は地震のクラスタリングである。地震学者はこれらのクラスターの特徴について深く研究し、多く経験則を得ることができた。ETAS モデルは、これらの経験則の統合に基づいて提案した [2]。このモデルでは各地震は特定の確率ルールに従って自らのクラスターを誘発する。これらの確率ルールは既存の経験則である [2,3]。今日既に、地震活動の最も大きな予測成分を定量化する確率モデル、特に地震活動の仮説検定および他のモデルの予測評価の標準相場モデルはかなり成功している。同時に ETAS モデルは地震予測の確率モデルでもあり、カリフォルニア地震予知プログラム (UCERF3) で採択された。

ETAS モデルの予測レベルを改善する、すなわちその予測の確率利得を改善させるには多く方法が存在する。例えば、本来の ETAS モデルを 2 次元震央から 3 次元震源版 [4,5] に、点震源から有限断層源 [6,7] に拡張すると、モデルはより詳細になり、より高い確率利得予測を得ることができる。地震活動データの以外の GPS 変形データ、ULF 地電位信号、地磁気信号など他の地球物理観測データと組み合わせると、より高い確率利得予測が得られる。これらの現象の大部分が地震の前兆であるかどうかは論争があり、客観的な評価が非常に重要である。ETAS モデルは現在までの地震活動の最良の定量化モデルであることが証明され、外部前兆の励起効果を組み合わせた新しい ETAS モデルが提案されている [8,9]。

ETAS モデルの推定手法も大きく発展した。確率除群法、確率再構築法、ベイズノンパラメトリック推定手法の開発は、犯罪、テロ事件、森林火災などの地震以外の自然現象や社会現象の研究のため、より広い Hawkes モデルにも適用される [10]。他の応用分野における予測理論や手法の開発は、地震の確率予測に応用することもできる。

参 考 文 献

- [1]Zhuang, J., D. Wang, and M. Matsu'ura (2016) Features of the earthquake source process simulated by Vere-Jones' branching crack model. *Bulletin of the Seismological Society of America*. Volume 106. doi:10.1785/0120150337.
- [2]Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.*, 83(401), 9-27.
- [3]Zhuang J., Ogata Y. and Vere-Jones D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97: 369-380.
- [4]Guo, Y., Zhuang, J., and Zhou, S. (2015) A hypocentral version of the space-time ETAS model. *Geophysical Journal International*, 203: 366-372. doi: 10.1093/gji/ggv319.
- [5]Guo, Y., J. Zhuang, and N. Hirata (2018) Modeling and forecasting 3D-hypocenter seismicity in the Kanto region. *Geophysical Journal International*, 214: 520-530. doi:10.1093/gji/ggy154.
- [6] Guo, Y., Zhuang J., Hirata N., Zhou S. (2017) Heterogeneity of direct aftershock productivity of the main shock rupture. *Journal of Geophysical Research Solid Earth*, 122, 5288-5305 doi:10.1002/2017JB014064.
- [7] Zhuang, J., M. Murru, G. Falcone, Y. Guo (2019) An extensive study of clustering features of seismicity in Italy from 2005 to 2016. *Geophysical Journal International*. 216:302-318. doi:10.1093/gji/ggy428.
- [8] Zhuang J., (2011) , Next-day earthquake forecasts for the Japan region generated by the ETAS model. *Earth Planets Space*, 63, 207-216. doi:10.5047/eps.2010.12.010. [pdf] doi:10.1029/2003JB002879.
- [9] Zhuang, J., M. Matsu' ura, P. Han (2019) Critical zone of the branching crack model for earthquakes: inherent randomness, earthquake predictability, and precursor modelling. Submitted to *Europhysics*, in Revision
- [10] Zhuang, J., J. Mateu (2018) A semi-parametric spatiotemporal Hawkes-type point process model with periodic background for crime data. Accepted by *Journal of the Royal Statistical Society, Ser. A*.

神経系の動作の解明と時空間推定システムへの適用

Study in Neural Operation and its Application to Time-Space Estimation

モデリング系 瀧澤 由美 (Yumi Takizawai)

1. 神経系の動作の研究

生物の神経系は、実際に生起した事象の時刻と場所（時間・空間）を知覚する能力を有する。脳は情報を知覚する広範で高度な機能をもつが、それらの原理は現在においてもほとんど明らかではない。筆者は多くの機能の中で特に基本的な機能として、時間・空間知覚能力に着目し、その動作原理を電気物理学的に解明することを試みた。脳の基本動作は多数の神経細胞とその結合である神経システムよりなる。従来の研究では、神経システムの動作の電気物理的観測、心理学による考察などの機能面からの解明、または刺激（入力）と応答（出力）を模擬する人工システムとしてとらえ、数理的データの面からの解明が試みられてきている。しかし、この取り組みは人工物に焦点をあて、生体としての神経細胞、神経システムの実態をとらえるものとは異なる。本研究では神経細胞（ニューロン）単体を能動的な電気信号（パルスまたはプラトー）の発振器としてとらえる。次に、神経細胞群（ニューロン集合体）をニューロン群により自律的同期システムとしてとらえる。安定な同期信号の発生は、正確

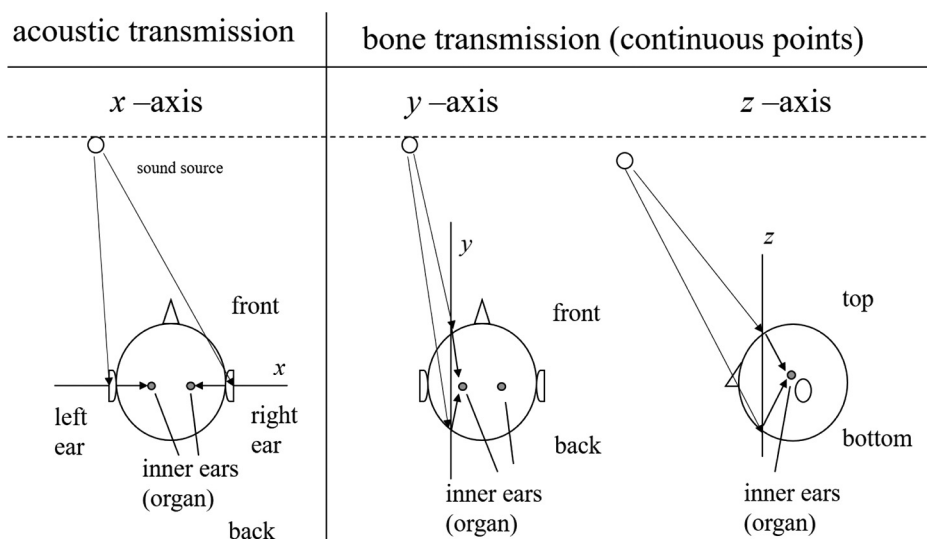


図 1. 両耳と骨伝導による 3次元位置の推定.

な時刻の知覚を可能にし、同時に場所（空間）の知覚を実現する。脳の高度な情報処理は、この同期パルスに基づくデジタル処理と緩やかで安定なプラトー信号に基づくアナログ処理とにより達成される。脳内における時空間知覚と処理は脳内マッピングとして動物および人間の脳において観測されている。筆者はこれの電気物理的モデルとその解析により、それが実現されることを示した。筆者による上記の研究成果は国際学会において注目され、2012 年学会より Best Paper Award を授与された。

2. 時空間推定の実システムへの応用研究

上記の研究成果の適用例として、電磁波を用いた位置計測システムを研究し、実用化を進めている。このシステムでは電磁波の送信と受信時刻差から、対称物までの距離を精密に測定する。この方式は、液化天然ガス（LNG）タンカーにおいて電磁波（マイクロ波）を用いた積載量計量システムを実現し、資源の貯蔵、輸送に用いられる。一方、航空機の搭載により、資源探査、植生観測による農業管理への適用が進められている。この研究では特にマイクロ波を用いた小型高性能な円偏波アンテナ（送信／受信）を実現し、実用化を進めている。

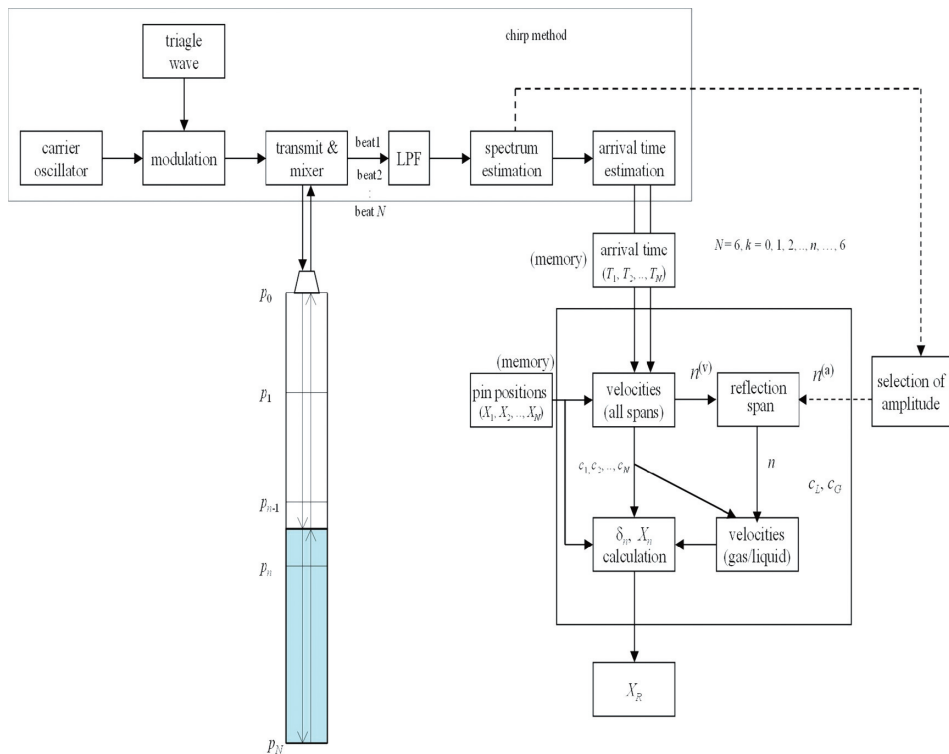


図 2. 液面計測システムの一例。

時空間データ解析による生体メカニズムの解明

Elucidation of biological mechanism by spatiotemporal data analysis

モデリング研究系 三分一 史和 (Fumikazu Miwakeichi)

1. 時空間解析によるニューロン間の因果的結合性とネットワーク構造の推定

このプロジェクトでは、兵庫医科大学、ドイツのゲッティンゲン大学との共同研究として呼吸リズムを形成するニューロンネットワーク機構の解明を目指している。具体的には、マウスの脳幹のスライス標本に蛍光色素を添加し、ニューロンの活動による細胞内カルシウム濃度の変化をイメージングデータ（動画）で記録するカルシウムイメージング法を用いて研究を行っている。自発的な呼吸を引き起こすニューロンの活性化タイミングを特定するために、標本表面に留置された電極により数十～数百個の細胞の局所場電位データも同時計測しており、呼吸性のバースト波形が出現したときの画像を解析すれば、呼吸に係るニューロンの検出と活性化タイミングを調べることができる（図 1）。

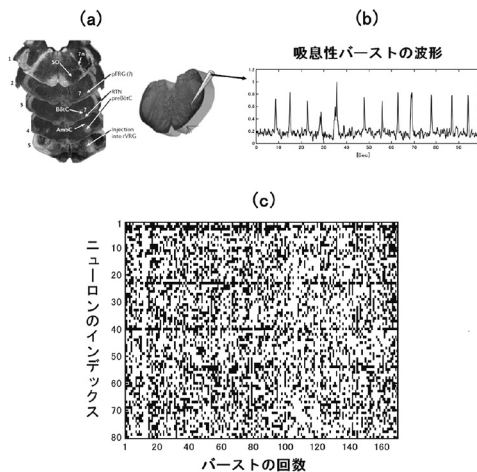


図 1. (a) ラットの脳幹のスライスの模式図
 (b) 計測される吸息性バーストの波形 (局所場電位)
 (c) ニューロンの活性化タイミングを記録したラスタープロット

これまでの研究において、動画のブレ補正法、ノイズ低減のための時空間フィルタリング法、などの事前処理法を開発した。そして、事前処理を行ったイメージングデータに遅延時間を考慮した相互相関解析を適用することにより、呼吸に関連する興奮性ニューロン、グリシン抑制性ニューロン、GABA 抑制性ニューロンの 3 種類のニューロンを特定することが可能となり、呼吸性ニューロン間の活性化順序は呼吸サイクルごとに変化するが、同じニューロンタイプで

は活性化順序に一定の規則性が存在するという重要な発見をした。

さらに進んで、呼吸リズムを生成するニューロンネットワーク構造の推定と同調メカニズムの解明を目指すには時空間解析の方法による因果的結合性の推定を行う必要がある。しかし、相互相関解析では、呼吸と関係するニューロンとの活性化タイミングの検出は可能であるが、ニューロン同士の因果的結合性を推定することができない。そこで、ニューロンの自励活動のみを考慮した自己回帰モデルと他のニューロンからの入力 を考慮した外生変数型自己回帰モデルをカルシウムイメージングデータに適用し、これら 2 つのモデルの赤池情報量規準量の差を調べることにより、特定の興奮性ニューロン Y (図 2 ☆印) と因果的結合性を持つニューロンの推定を試みた。例えば、図 2(b) と (c) の両方に検出されているニューロン A と C はニューロン Y と双方向に結合しているが、ニューロン B は図 2(c) では検出されていないので、ニューロン Y からの因果性はあるが、ニューロン Y へ因果性はないという非対称的な結合性であるということが分かった。現在は、薬剤の添加や物理的な方法で脳領域やニューロン間の結合を離断させる阻害実験の前と後のデータ時系列モデルを適用し、パラメータを定量比較することにより、ニューロンネットワークの実在性の検証を行っている。

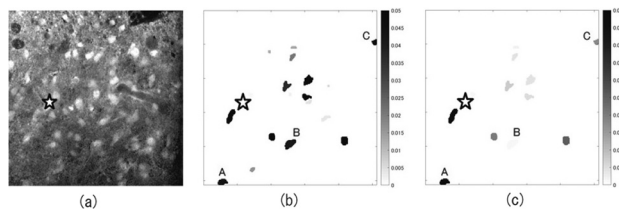


図 2. (a) 活動性ニューロンの空間分布
(b) ☆印の位置にある興奮性ニューロン(Y)が影響を及ぼしているニューロン
(c) ☆印の位置にある興奮性ニューロン(Y)に影響を及ぼしているニューロン

2. 時系列データにおける変化点の検出

人や動物の健康状態の把握や異常状態の検出のために生体システムのモニタリングが必要な場合が数多くある。複数のセンサで生体信号をリアルタイム計測し、信号間の相互相関解析あるいは回帰分析などでデータを逐次処理し、パラメータの変化から異常値を検出するというアプローチが一般的であるが、異常状態にあるときはデータの定常性が崩れていたり、非線形性が生じていたりして、一般的な処理方法の適用条件を満たさなくなり、解析そのものが困難になる。本研究では正常状態にある生体信号を用いて時系列モデルを同定し、その同定したモデルでデータを逐次フィルタリングを行うイノベーションアプローチに基づいた異常値検出方法の開発を行っている。具体的には、牛の体温データを用いた出産時刻の予測法の開発を目指している。体温はサーカディアンリズムに従った変動をしているが、出産直前に体温が低下することが知られており、この現象を利用して出産時刻の予測を行うことは原理的には可能である。しかし、体温データには気温や給餌タイミングなどの外部要因の影響も重畳しており、これらの影響を考慮する必要がある。これまでに、季節調整法に外生変数型自己回帰モデルを組み込んだ時系列モデルを用いたモニタリングシステムを構築し、実データを用いた検証を行っている。

マルコフ連鎖モンテカルロ法の世界を拓げる

— 高次元の多峰性分布からのサンプリング —

Expanding world of Markov chain Monte Carlo

— Methods for sampling high-dimensional multimodal distributions —

モデリング研究系 伊庭 幸人 (Yukito Iba)

要 旨

多峰性分布に強いマルコフ連鎖モンテカルロ法とその応用について述べた。

キーワード：レプリカ交換モンテカルロ法，マルチカノニカル法，多峰性分布，レアイベントサンプリング

1. はじめに

マルコフ連鎖モンテカルロ法 (MCMC) は高次元・多変量の確率分布からの乱数生成 (サンプリング) の手法である。当初は統計物理の世界で使われていたのが、1990 年代以降に統計学や機械学習の世界で利用されるようになり、大きな影響をもたらしたことはよく知られている。著者は長年マルコフ連鎖モンテカルロ法の応用に興味を持っているが、その間、以下の 2 つの点にこだわりを持ち続けてきた。

1. 通常の MCMC ではサンプリングしにくい多峰性のある分布 (確率の高い領域が複数に分かれている分布) に対して有効な手法の開発。
2. 統計データ解析でも統計物理でもない分野への応用。

MCMC というと、ベイズ統計やそのソフトウェアを思い浮かべる方が多いかもしれないが、それとはかなり違う問題意識で研究を進めてきたといえると思う。以下、その軌跡を紹介したい。

2. 拡張アンサンブル MCMC とその応用

多峰性の分布に対して MCMC を適用する手法として Simulated Annealing 法が知られている。Simulated Annealing 法では「温度」に対応するパラメータを最初は「高温」に設定し次第に「低温」にすることで、重要でない分布の山 (局所的極大) にトラップされることを防ぐ。しかし、これはあくまで最適化の範疇であって、多峰性の分布を正しくサンプルすることができるとは限らない。統計物理での有限温度の分布、ベイズ統計での事後分布からのサンプリングには別のアイデアが必要である。1993 年ごろにこの問題に興味を持った著者は、今日「レプリカ交換モンテカルロ法」(パラレルテンパリング) と呼ばれている方法を独自に思いついたが、あまり知られていない先行研究 (木村・瀧 (1990) 及び Geyer (1991)) の存在に気づいたため、査読つき論文を出すには至らなかった (短報は「統計数理」の研究報告会要旨に掲載され、オンラインでも見られる。伊庭 (1993))。それから数年たってから考えたのは、温度に

限らず、取束を悪くしている特定の制約条件を緩めることでよりよい結果が得られるということで、この考え方をタンパク質の格子モデルに適用して開発した手法 (MSOE 法) は当該分野で高く評価されている (第 2 報の論文 Chikenji et al. (1999) は web of science で引用数 74)。また、それまでに考えたことを踏まえて、関連する一連の手法を「拡張アンサンブル法」としてまとめたレビュー論文 Iba (2001) は 2018 年現在 web of science で 142 件引用されている。

3. さまざまな分野での高次元分布からのサンプリング

MCMC が有用な分野としては統計物理とベイズ統計が代表格であるが、これらの分野では、それぞれの主役である「ギブス分布 (カノニカル分布)」「事後分布」が高次元の確率分布であるという点が共通である。そう考えると、統計物理やベイズ統計以外でも MCMC によって面白いことができる分野はもっとあるのではないか。「最適化からサンプリングへ」というスローガンは、単に MCMC の応用ということを超えて、いろいろな学問の世界で「確率構造」「測度構造」を考える糸口になるかもしれない。過去 10 年ほど行ってきた研究は、こうした発想に基づくものである。

具体的な応用としては、ランダム行列やランダムグラフの大偏差の数値計算、力学系の珍しい軌道のサンプリング、ランダム系の不純物に関する平均 (クエンチ平均) の効率の計算、ラテン方阵や分割表の数の計算、複雑な帰無仮説下での検定統計量の分布の計算 (サロゲーション法への応用) などがある。これらの例の多くでは、与えられた分布の極端な裾からのサンプリングが必要とされるが、そこで多峰性分布からのサンプリングの手法が重要となる。マルチカノニカル法、レプリカ交換モンテカルロ法などを利用することで、与えられた確率分布のもとでの生起確率が 10^{-200} といった極端なレアイベント (大偏差事象) のサンプリング及び確率計算ができることが示されている。興味を持たれた方は、解説 Iba et al. (2014) の中の事例と引用文献を参照されたい。

4. 今後の展望

マルチカノニカル法やレプリカ交換モンテカルロ法 (パラレル・テンパリング) など多峰性分布に対応した MCMC 手法は、ベイズ統計にもとづくデータ解析でも有用なはずであり、研究レベルではレプリカ交換モンテカルロ法を中心にいろいろな適用例がある。しかし、現場での応用が MCMC ソフトウェア中心に展開しているため、十分活用されていないのが現状である。これらの手法に対応する統計ソフトウェアが登場することで、より自由な統計モデリングが行えるようになることを望みたい。

参 考 文 献

- Chikenji, G., Kikuchi, M. and Iba, Y. (1999). Multi-self-overlap ensemble for protein folding: ground state search and thermodynamics, *Physical Review Letters*, **83** (9), 1886–1889.
- Geyer, C. J. (1991). *Computing science and statistics: Proceedings of 23rd Symposium on the Interface* (ed. E. Keramidis), 156–163, Interface Foundation, Fairfax Station.
- Iba, Y. (2001). Extended ensemble Monte Carlo, *International Journal of Modern Physics C*, **12** (05), 623–656.
- Iba, Y., Saito, N. and Kitajima, A. (2014). Multicanonical MCMC for sampling rare events: an illustrative review, *Annals of the Institute of Statistical Mathematics*, **66** (3), 611–645.
- 伊庭幸人 (1993). メトロポリスのモンテカルロ法の緩和について (統計数理研究所研究活動 (平成 4 年度 研究報告会要旨)), 統計数理, **41** (1), 65–67.
- 木村宏一, 瀧 和男 (1990). 時間的一様な並列アニーリングアルゴリズム, 電子情報通信学会 NC-90-1, 1–8.

正定値カーネルによる統計的機械学習

Statistical Machine Learning by Positive Definite Kernels

数理・推論研究系 福水 健次 (Kenji Fukumizu)

要 旨

正定値カーネルとそれが定める再生核ヒルベルト空間を用いて、確率分布を表現する方法論を確立し、さまざまな統計的問題に適用してきた。その一連の研究に関して概説する。

キーワード：正定値カーネル，再生核ヒルベルト空間，統計的推論，検定，ベイズ推定

1. はじめに

カーネル法は、データを（非線形）写像することによってデータの高次モーメントを扱う方法論であり、サポートベクターマシンの提案以来、機械学習の主要技術の一つとして発展してきた(福水, 2010)。データに変換を施してから解析する手法は古くから存在するが、カーネル法の特徴は、特殊な内積を持つ関数空間への写像を用いることにより、写像後のデータに対する線形の処理が効率的に行える点にある。

正定値カーネルとは、集合 Ω （データが存在する空間）上に定義された対称な 2 変数関数 $k: \Omega \times \Omega \rightarrow \mathbb{R}$ で、任意の点 $x_1, \dots, x_n \in \Omega$ に対しグラム行列 $(k(x_i, x_j))$ が半正定値性を満たすものである。 Ω 上の正定値カーネル k に対し、 Ω 上の関数からなるヒルベルト空間 H が定まり、カーネル法ではこれを「特徴空間」と呼ぶ。このヒルベルト空間は特別な内積を有しており、第 2 変数を $x \in \Omega$ に固定して第 1 変数に関する関数とみなした $k(\cdot, x)$ と任意の関数 $f \in H$ の内積が、 $\langle f, k(\cdot, x) \rangle_H = f(x)$ と関数値に一致する。この性質を再生性といい、ヒルベルト空間 H を再生核ヒルベルト空間と呼ぶ。

データ解析に正定値カーネルを用いる際には、データの存在する空間 Ω に正定値カーネル k を定め、次の「特徴写像」によって特徴ベクトル $\phi(x) \in H$ を仮想的に作成する。

$$\phi: \Omega \rightarrow H, \quad x \mapsto k(\cdot, x).$$

再生性を用いると、

$$\langle \phi(x), \phi(y) \rangle_H = k(x, y)$$

が得られるが、これは 2 つの特徴ベクトル $\phi(x), \phi(y)$ のヒルベルト空間 H における内積が、正定値カーネルの値の評価によって容易に計算されることを意味しており、カーネル法の鍵となる。データ x_1, \dots, x_n に対して特徴ベクトル $\phi(x_1), \dots, \phi(x_n)$ を想定し、これらに線形回帰、主成分分析など様々な既存のデータ解析手法を適用したものがカーネル法として総称されている。

2. 分布の表現とその応用

Ω 上の確率分布 P に対し $\mu_P := \int k(\cdot, x) dP(x) \in H$ という関数を、分布 P の表現として

用いることが可能である。これをカーネル平均と呼ぶ。 k が非線形カーネルの場合、カーネル平均は分布 P の高次モーメントの情報を有している。特にガウスカーネルなどは、分布 P から再生核ヒルベルト空間 H への 1 対 1 写像 $P \mapsto \mu_P \in H$ を定めることが知られており、確率分布の特性関数と同様の役割を持つ。このようなカーネルを特性的カーネルと呼ぶ (Fukumizu et al., 2004)。確率 P に従う有限個の i.i.d. データ X_1, \dots, X_n が与えられた場合、 μ_P を $(1/n) \sum_{i=1}^n k(\cdot, X_i)$ により推定することが可能である。

同時分布 P を持つ確率変数 (X, Y) に対して (非心) 共分散作用素を $C_{YX} = E_P[k_X(\cdot, X)k_Y(\cdot, Y)]$ により定める。ここで、 k_X, k_Y はそれぞれ X, Y が値をとる空間に定義された正定値カーネルである。 C_{YX} は有限次元確率ベクトルの共分散行列の一般化であり、 X と Y の統計的関係を表現している。これも、グラム行列を用いて有限サンプルから容易に推定可能である。

近年、筆者を含むグループらにより、カーネル平均と共分散作用素を様々な統計的問題へ適用する研究がなされてきた。例えば、Fukumizu et al. (2004) では、特性的なカーネルを導入し、確率変数の条件付き独立性を共分散作用素によって特徴づけ、それを回帰問題における次元削減に応用している。また、2つの分布の距離をカーネル平均の距離により定義し、2標本問題に応用する研究や、独立性、条件付き独立性の検定にカーネル法を応用する研究などが発展している。これらカーネル平均を用いた方法に関しては Song et al. (2013) や Muandet et al. (2017) などを見ていただくとうい。

また、Fukumizu et al. (2013) は、カーネル平均と共分散作用素を用いてベイズ事後確率を推定する方法を提案した。この方法は状態空間モデルにおけるフィルタリングの問題などに応用されている。

カーネル平均や共分散作用素を用いたデータ解析は一般的なノンパラメトリック推論の方法論であり、最近では、近似ベイズ計算への応用、因果推論への応用、さらに深層学習と組み合わせた生成モデルへの応用など、大きな広がりを見せている。本稿によって興味を持たれた方がさらに研究を発展させていくことを期待している

参 考 文 献

- 福水健次 (2010) 『カーネル法入門 – 正定値カーネルによるデータ解析』, 朝倉書店, 東京.
- Fukumizu, K., Bach, F.R. and Jordan, M. I. (2004). Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces, *Journal of Machine Learning Research*, 5, 73–99.
- Song, L., Gretton, A., and Fukumizu, K. (2013) Kernel Embeddings of Conditional Distributions. *IEEE Signal Processing Magazine* 30(4), pp. 98–111.
- Muandet, K., Fukumizu, K., Sriperumbudur, B. and Schölkopf, B. (2017), Kernel Mean Embedding of Distributions: A Review and Beyond, *Foundations and Trends in Machine Learning*, Vol. 10: No. 1–2, pp 1–141.
- Fukumizu, K., Song, L., and Gretton, A. (2013) Kernel Bayes' Rule: Bayesian Inference with Positive Definite Kernels. *Journal of Machine Learning Research*, 14, 3753–4783.

位置情報軌跡の匿名化技術 Anonymization Techniques for Trajectory Location Datasets

モデリング研究系 南 和宏 (Kazuhiro Minami)

1. 位置情報の匿名化における課題

スマートフォンの普及に伴い、我々の位置情報の取得が容易になり、多くのユーザーの移動履歴は、交通情報の提供、都市設計といった社会サービス、また商圏分析等の商用ビジネスにも活用されている。その一方、位置情報から、個人の興味に関するプライバシーな情報が漏洩する危険性が懸念される。よって位置情報の安全な2次利用には、匿名化と呼ばれる個人の識別情報を取り除くデータ処理が不可欠である。一般に、位置情報の匿名化処理では、氏名等の個人の識別子を削除するだけでは不十分である。なぜなら目撃情報、名簿等の外部知識から特定の日時、場所の位置情報が識別され、その結果、その位置情報を含む軌跡全体が特定されるリスクが存在するからである。したがって、位置情報から k 未満のユーザーへの絞り込みを防ぐための k -匿名化処理 (Sweeney, 2002)が必要となる。

しかし通常の k -匿名化の手法を位置情報軌跡に適用する場合、2つの課題が存在する。1つは、位置情報軌跡のような時系列データの場合、 k -匿名化を実施するとデータの有用性が著しく劣化する問題である。長期の位置情報軌跡を匿名化する場合、 k -匿名化の前提となる軌跡群へのグループ化が困難である。そのため、 k -匿名化を実現するための一般化処理による情報損失は大きくなり、有益なデータ分析に堪えなくなる。2つめは、位置情報軌跡のデータ間に時空間の相関性が存在し、匿名化した位置情報から統計的推論により元の軌跡情報が復元される問題である。位置情報軌跡には、人の移動に関する物理的制約が反映し、車、電車といった交通手段により移動経路は限定される。また長期的な移動軌跡には通勤、病院への通院といった個人の生活習慣を反映した特徴的なパターンが現れる。そのような移動パターンに関する外部知識を用いると匿名化された位置情報から元の位置情報が復元される危険性がある。

近年、著者はこれらの課題を解決するための2つの匿名化技術に取り組んできた。1つは位置情報軌跡を複数のセグメントに分割する動的仮名交換手法 (Tanjo et al., 2014)であり、ミックスゾーンと呼ばれる複数ユーザーの集積点でのランダムな仮名の再割当により移動先の不確定性を確保する手法である。もう1つは、ユーザーの移動パターンをマルコフ過程でモデル化し、隠れマルコフモデルにおける内部状態の推定問題として匿名化データの安全性の評価を行う手法 (Minami, 2014)である。

2. ミックスゾーンにおける動的仮名割当

個人の行動パターンが顕著に現れる位置情報軌跡の場合、その中の幾つかの点に過ぎない外部知識を用いて軌跡全体の識別が可能であり、情報漏えいリスクが非常に高い。位置情報軌跡の開示リスクを局所するため、Mano et al. (2013)では位置情報軌跡に紐付けられる仮名を動的に更新し長期間の軌跡データを複数の軌跡セグメントに分割する方式を提案した。この仮名の更新は複数のユーザーが同一の時間、場所に存在する「ミックスゾーン」と呼ばれる領域でラ

ンダムな仮名交換の形式で実施し、ミックスゾーン前後の軌跡セグメント間の関連性を分断する。個人の位置情報はミックスゾーンを経由することで代替経路が増大するので、その不確実性に着目して位置情報軌跡の仮名化データに対するプライバシー指標を定式化した。ただし、攻撃者の外部知識と整合性を保持する代替経路の列挙には、ミックスゾーンを頂点、位置情報の軌跡セグメントを辺とするグラフにおける排他的辺素パス問題を解く必要がある。一般の排他的辺素パス問題は NP 困難であるため、(Tanjo et al., 2014)では排他的辺素パス問題を制約充足問題に変換する効率的な安全性評価手法を開発した。

3. 隠れマルコフモデルに基づく匿名化処理の安全性評価

仮名更新による位置情報軌跡の分割は位置情報軌跡の識別リスクを局所化する手法である。しかし位置情報が識別されると同じ軌跡セグメント内の位置情報は依然として漏洩してしまう。したがって分割して次元を削減した軌跡セグメント単位に k -匿名化を実施することが望ましい。ただし、位置情報軌跡には時空間の相関性が存在するので、通常の k -匿名化では不十分な場合が多い。

Minami (2014)では、ユーザーの移動パターンをマルコフ過程でモデル化し、匿名化データの安全性を隠れマルコフモデルにおける観測情報から内部状態の推定問題として定式化した。モデルにおける観測情報は匿名化データ、内部状態遷移は秘匿すべき元の位置情報に対応し、匿名化アルゴリズムは、内部状態から観測情報への確率的な変換を定義する記号出力行列として記述される。さらに匿名化データの安全性は、観測情報、記号出力行列、内部状態のマルコフ過程が与えられたときに正しく内部状態を推定する条件付き確率として定式化した。実データを用いた評価実験では、通常の k -匿名化処理では想定した安全性が確保できず、追加の秘匿処理の必要性を明らかにし、統計モデルに基づく安全性評価の有用性を実証的に示すことができた。

参 考 文 献

- Mano, K., Minami, K. and Maruyama, H. (2013). Protecting Location Privacy with K-Confusing Paths Based on Dynamic Pseudonyms, *5th IEEE International Workshop on SSecurity and SOCial Networking*, March.
- Minami, K. (2014). Preventing denial-of-request inference attacks in location-sharing services, *2014 Seventh International Conference on Mobile Computing and Ubiquitous Networking*, 50–55, January.
- Sweeney, L. (2002). k -anonymity: a model for protecting privacy, *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, **10** (5), 557–570.
- Tanjo, T., Minami, K., Mano, K. and Maruyama, H. (2014). Evaluating data utility of privacy-preserving pseudonymized location datasets, *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, **5** (3), 63–78, September.

神経スパイク時系列パターンの特徴解析

Statistical analysis of neural spike trains

モデリング研究系 小山 慎介 (Shinsuke Koyama)

要 旨

本研究では、神経スパイク時系列の統計パターンを記述する方法を提案する。提案方法の新規性は、スパイク発火時刻のゆらぎ（分散）を平均のべき関数で表すところにある。べき関数のパラメータを調整することで、脳内で観測されるさまざまなスパイク発火ゆらぎを再現することができる。この枠組みに基づきスパイク時系列の点過程モデルを構築し、データからパラメータを推定する方法を提案する。

キーワード：スパイク時系列，点過程，ゆらぎのスケーリング則

1. はじめに

高度な情報処理を行う脳は、異なる働きを持つ部分が有機的に組み合わせられたヘテロ構造体である。このような見方は、およそ 100 年前にブロードマンが解剖学的・細胞構築学的観点から大脳皮質を区分して「脳地図」を描いたことに始まる。今日、その区分が脳の機能と密接に関係していることも明らかになっている。

近年 Shinomoto et al.(2009) は、神経スパイク発火パターンの大脳皮質全体にわたる非一様性に着目した。スパイク発火の不規則性を測るために「局所変動係数 (local variation) L_V 」を提案し、大脳皮質のさまざまな領野のデータから L_V を求めたところ、細胞ごとに固有の発火パターンがあり、それが大脳皮質の機能とも相関している、という事実を見いだした。この発見は、大脳皮質が機能的に区分できるという事実が神経細胞が用いる信号レベルにも反映されているということを示唆している。本研究の目的は、スパイク発火パターンをより系統的に特徴付けるための統計的方法を構築することである。

2. アイデアのスケッチ

Shinomoto らが提案した局所変動係数 L_V は、時系列の非定常性の影響を取り除いたスパイク間隔 (ISI) の変動係数 (coefficient of variation $C_V = \text{標準偏差}/\text{平均}$) とみなせる。したがって彼らが見いだしたことは、局所的に定常と見なせる短い時間スケールで

$$(2.1) \quad \text{Var}(\text{ISI}) = \phi E(\text{ISI})^2$$

という関係を持つことを仮定すると、領野・皮質層ごとに固有な ϕ の値をとり、神経細胞を ϕ の値で大まかに分類できると理解できる。

一方で Troy and Robson(1992) は、定常な光輝度刺激に対する網膜ガングリオン細胞のスパイク間隔の C_V を推定したところ、おおよそ $C_V^2 \propto E(\text{ISI})$ という関係があることを発見した。これを書き換えると

$$(2.2) \quad \text{Var}(\text{ISI}) = \phi E(\text{ISI})^3$$

という平均と分散の関係式が得られる。ここで式 (2.1) と比べてべき指数が異なることに着目しよう。

式 (2.1) と (2.2) を一般化してべき指数 α を導入すれば、少なくとも局所的に定常と見なせる時間スケールで、スパイク間隔の平均と分散の間に

$$(2.3) \quad \text{Var}(\text{ISI}) = \phi E(\text{ISI})^\alpha$$

というスケーリング則が得られる。これが提案する方法の根幹をなす仮定である。本研究の目的は、式 (2.3) を仮定してスケール因子 ϕ とべき指数 α を用いてスパイク発火パターンを特徴付けるための統計的方法を構築することである。

まず定常リニューアル過程を考えよう。ISI の平均を $\mu = E(\text{ISI})$ として $\text{ISI} \rightarrow \mu^{-1}\text{ISI}$ とスケール変換すると、式 (2.3) より $E(\text{ISI}) \rightarrow 1$, $\phi \rightarrow \mu^{\alpha-2}\phi$ とリスケールされる。したがって、平均と分散がそれぞれ μ と $\phi\mu^\alpha$ である確率密度関数 $f(x; \mu, \phi)$ でこのスケール変換に対して不変なものは

$$(2.4) \quad f(x; \mu, \phi) = \mu^{-1} f(\mu^{-1}x; \mu^{\alpha-2}\phi)$$

を満たす。ここで $f(x; \phi) := f(x; 1, \phi)$ である。すなわち、平均が 1 で分散が ϕ である任意の確率密度関数を式 (2.4) でスケール変換することによって式 (2.3) を満たす確率密度関数を作ることができる。

Koyama (2015) では、このアイデアを非定常リニューアル過程に拡張することで、ゆらぎのスケーリング則をもつスパイク時系列モデルを提案した。また、このモデルに基づいてパラメータを推定する方法を提案し、実験データに適用して有効性を確認した。

3. まとめと展望

本研究では、スパイク発火時刻のゆらぎを記述するためのスケーリング則を定式化し、これに基づいてスパイク時系列の統計モデルとパラメータ推定方法を提案した。この方法を脳の幅広い部位から記録したデータに適用してスパイク発火パターンを特徴付け、脳機能との関連を調べることが今後の課題である。

参 考 文 献

- Koyama, S. (2015). On the spike train variability characterized by variance-to-mean power relationship, *Neural Computation*, **27**, 1530–1548.
- Shinomoto, S. et al. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex, *PLoS Computational Biology*, **5**, e1000433.
- Troy, J. B. and Robson, J. G. (1992). Steady discharges of X and Y retinal ganglion cells of cat under photopic illumination. *Visual Neuroscience*, **9**, 535–553.

ノンパラメトリックエントロピー推定とその応用

Non-parametric Entropy Estimation

モデリング研究系 日野 英逸 (Hideitsu Hino)

情報理論において最も基本的な量の一つである情報量は $I_f(x) = -\log f(x)$ で定義される。ここで $f(x)$ はデータ $x \in \mathbb{R}^d$ が従う分布の確率密度関数である。情報量の期待値はエントロピーと呼ばれる: $H(f) = E_f[I_f(X)] = -\int f(x) \log f(x) dx$. 情報量, エントロピーあるいはこれらを用いて導出できる KL ダイバージェンス及び相互情報量は統計学や機械学習など非常に多くの分野で重要な役割を果たしている。

微分エントロピーの推定方法として最も良く利用されている手法の一つが, k -近傍法と確率密度関数の 1 次の展開に基づく推定法である。最近傍法に基づくエントロピー推定量は, Kozachenko and Leonenko (1987) により提案され, 任意の次元の確率変数に対して mean square consistency を持つことが示されている。この結果は一般の k -近傍に基づく推定量に拡張され (Goria et al., 2005), その後も各種の拡張と理論的解析がなされている (Beirlant97, 1997; Paninski, 2003)。確率密度関数 $f(z)$ の検査点 $z \in \mathbb{R}^p$ における値を観測データ集合 $\mathcal{D} = \{x_i\}_{i=1}^n$ を用いて推定する問題を考える。検査点 z を中心とする半径 ε の p 次元超球を $b(z; \varepsilon) = \{x \in \mathbb{R}^p \mid \|z - x\| < \varepsilon\}$ で表す。この超球の体積は $|b(z; \varepsilon)| = c_p \varepsilon^p$ である。ただし, $c_p = \pi^{p/2} / \Gamma(p/2 + 1)$ である。中心 z の ε 球に含まれる確率質量を $q_z(\varepsilon) = \int_{x \in b(z; \varepsilon)} f(x) dx$ で定義する。この定義式の被積分関数を Taylor 展開すると $q_z(\varepsilon) = |b(z; \varepsilon)| (f(z) + O(\varepsilon^2)) = c_p \varepsilon^p f(z) + O(\varepsilon^{p+2})$ を得る。超球の半径 ε を十分小さいと仮定してその 2 次以上の項を無視し, 確率質量を全観測データに占める超球内の点の割合で近似することで, 確率密度関数の推定量 $\hat{f}(z; \varepsilon) = \frac{k_\varepsilon}{nc_p \varepsilon^p}$ を得る。ここで k_ε は観測データ集合 \mathcal{D} の中で半径 ε の超球の中に含まれるものの個数である。一方, ε の代わりに超球に含まれるサンプル数 k を固定した場合, $\hat{f}(z; \varepsilon)$ は $\hat{f}^{nn}(z; k) = k / (nc_p \varepsilon_k^p)$ のようにかける。ここで超球の半径 ε_k は検査点 z からその k 番目に近い点までの距離で決定されることになる。 $\hat{f}_i^{nn}(x_i; k)$ を, データ集合 $\mathcal{D} \setminus \{x_i\}$ を用いて k -近傍法により推定した推定量として, $-\ln \hat{f}_i^{nn}(x_i; k)$ の経験期待値を計算することで, k -近傍エントロピー推定量 $\hat{H}^{nn}(\mathcal{D}; k) = -\sum_{i=1}^n \ln \hat{f}_i^{nn}(x_i; k)$ を得る (Goria et al., 2005)。この方法は確率質量関数の一次展開に基づく方法であるが, 筆者らはより高次の展開に基づく手法を提案した (Hino et al., 2015)。検査点 z を中心とした半径 ε の超球内の確率質量 $q_z(\varepsilon)$ は ε に関する二次の Taylor 展開をすると, $q_z(\varepsilon) = c_p f(z) \varepsilon^p + \frac{n}{4(p/2+1)} c_p \text{Tr} \nabla^2 f(z) \varepsilon^{p+2} + O(\varepsilon^{p+4})$ の形で表わされる。上式左辺の $q_z(\varepsilon)$ を比 k_ε/n で近似し, 両辺を $c_p \varepsilon^p$ で割ることで, $\frac{k_\varepsilon}{nc_p \varepsilon^p} = f(z) + C \varepsilon^2 + O(\varepsilon^4)$ を得る。ここで $C = n \text{Tr} \nabla^2 f(z) / 4(p/2 + 1)$ である。さらに, $Y_\varepsilon = \frac{k_\varepsilon}{nc_p \varepsilon^p}$ と $X_\varepsilon = \varepsilon^2$ を導入し, ε に関する 4 次以上の項を無視することで, 応答変数 Y の説明変数 X に関する一次式 $Y_\varepsilon \simeq f(z) + C X_\varepsilon$ が得られる。この式は説明変数 X による応答変数 Y の線形回帰式とみなせる。複数の半径 $\mathcal{E} = \{\varepsilon_i\}_{i=1}^m$ と置き, \mathcal{E} に含まれる各 ε で定まる $\{(X_\varepsilon, Y_\varepsilon)\}_{\varepsilon \in \mathcal{E}}$ の組を観測データとみなして, 二乗誤差 $R = \frac{1}{m} \sum_{\varepsilon \in \mathcal{E}} (Y_\varepsilon - f(z) - C X_\varepsilon)^2$ を最小化するように $f(z)$ と C を求める。これは単回帰に他ならず, この回帰によって得られた切片が, $f(z)$ の推定量 $\hat{f}(z)$ である。以上より検査点 z における密度の推定量 $\hat{f}^s(z)$ が得られたので, leave-one-out 推定量とし

てエントロピーの推定量 $\hat{H}^s(\mathcal{D}) = -\frac{1}{n} \sum_{i=1}^n \ln \hat{f}_i^s(x_i)$, を得る. ここで $\hat{f}_i^s(x_i)$ は x_i を用いずに求めた密度の推定値である. 筆者らは, この推定量を始めとして, より直接的に一度の回帰問題を解くことでエントロピーを推定する手法, 誤差構造として Poisson 分布を仮定した手法 (Hino et al., 2016), 及び確率質量関数の高次展開に基づく局所フラクタル次元推定量 (Hino et al., 2017) を提案し, ノンパラメトリックなアプローチによるデータ分布の特徴付けの方法論の開発を進めている.

また, 関連して筆者らは一般にデータに重みが与えられている場合に適用可能な情報量推定手法を提案している (Hino and Murata, 2013). 例えば観測されたデータそれぞれに対して, その信頼度が与えられている場合や, あるいは同一の事象が重複して観測されるとして個々の事象を観測した頻度を重みとして表現する場合のように, データ \mathcal{D}_x が与えられた上で, 各データ点 $x_i \in \mathcal{D}_x$ の重要さとして重みが付与されることが考えられる. 過去のデータの重みが小さくなっていくような忘却係数付きのオンライン観測データもこうした重み付きデータの一例である. この重み付き情報量推定量の応用として情報論的クラスタリング (Hino and Murata, 2014), 変化点検知 (Koshijima et al., 2015) などの方法を開発した. 情報論的クラスタリング手法は同種の手法と比較してより正確なクラスタリングを実現し, 変化点検知の手法は従来の手法では捉えられなかったデータの背後にある構造的な変化を抽出することに成功している.

参 考 文 献

- Beirlant, J., Dudewicz, E. J., Györfi, L. and Meulen, E. C. (1997). Nonparametric Entropy Estimation: An Overview, *International Journal of the Mathematical Statistics Sciences*, **6**, 17–39.
- Goria, M. N. and Leonenko, N. N. and Mergel, V. V. and Novi Inverardi, P. L. (2005). A new class of random vector entropy estimators and its applications in testing statistical hypotheses, *Journal of Nonparametric Statistics*, **17**(3), 277–297.
- Hino, H., Fujiki, J., Akaho, S. and Murata, N. (2017). Local Intrinsic Dimension Estimation by Generalized Linear Modeling, *Neural Computation*, **29**(7).
- Hino, H., and Murata, N. (2013). Information estimators for weighted observations, *Neural Networks*, **46**(0), 260–275.
- Hino, H., and Murata, N. (2014). A Non-parametric information theoretic clustering algorithm based on Quantile-based entropy estimator, *Neural Computation*, **26**(9), 2074–2101.
- Hino, H., Koshijima, K. and Murata, N. (2014). Non-parametric entropy estimators based on simple linear regression, *Computational Statistics & Data Analysis*, **89**(0), 72 – 84.
- Hino, H., Akaho, S. and Murata, N. (2016). An Entropy Estimator Based on Polynomial Regression with Poisson Error Structure, *Neural Information Processing - 23rd International Conference, ICONIP 2016, Kyoto, Japan, October 16-21, 2016, Proceedings, Part II*, 11 – 19.
- Koshijima, K., Hino, H. and Murata, N. (2015). Change-Point Detection in a Sequence of Bags-of-Data, *Knowledge and Data Engineering, IEEE Transactions on*, **27**(10), 2632–2644, Oct.
- Kozachenko, L. F. and Leonenko, N. N. (1987). Sample estimate of entropy of a random vector, *Problems of Information Transmission*, **23**, 95–101.
- Paninski, L. (2003). Estimation of entropy and mutual information, *Neural Comput.*, **15**, 1191–1253, June.

組合せ論的系統学における最近の話題

—グラフ理論が拓く系統解析の新展開—

Recent topics in combinatorial phylogenetics

—From graph theory to statistical analysis—

モデリング研究系 早水 桃子 (Momoko Hayamizu)

要 旨

生物の進化の道筋を解明する系統学的なデータ解析において、今日では系統樹を拡張した「系統ネットワーク」というグラフ構造が広く用いられるようになってきているが、記述能力が高い汎用的なモデルが必ずしも優れたモデルであるとは限らないため、系統ネットワークのサブクラスの中で生物学的な妥当さと数学的な性質の良さを兼ね備えたものを見出すことは重要である。特に Francis and Steel (2015) が定義した「系統樹ベースのネットワーク」(tree-based phylogenetic network; TBN) は系統樹に辺を追加する単純な操作で得られる系統ネットワークのサブクラスで、TBN の数学的性質や計算複雑性に関する未解決問題は理論生物学分野のホットトピックスになっている。Hayamizu (2018) は、Francis and Steel が取り上げた決定/探索問題や数え上げ問題だけでなく、TBN に関する列挙問題や最適化問題にもスポットライトを当て、これらの問題を高速に解くアルゴリズムを統一的な視点で生み出す「TBN の構造定理」を証明し、多様な統計学的な応用を可能にした。本稿では、その研究成果の一端を紹介する。

キーワード：系統樹推定，系統ネットワーク，細分系統樹，離散アルゴリズム

1. 研究の背景

生物の進化は古くから系統樹 (phylogenetic tree) を用いて記述されてきたが、例えば植物、菌類、細菌類が進化する過程では異種交雑 (hybridization) や遺伝子の水平伝播 (horizontal gene transfer; HGT) といった木構造で記述しきれない現象が起きうるため、あらゆる種の進化を系統樹だけで描写することはできないといわれている。また、仮にそのような現象を考慮しなくてもよい種を対象にした系統解析を行う場合でも、現実のデータを扱い、その情報を忠実に描写したいなら、木構造よりも融通の利くグラフ構造が欲しいと考えるのは自然であろう。

このようなニーズに動機づけられ、組合せ論的系統学 (combinatorial phylogenetics) という理論生物学の一領域では、系統樹を拡張した系統ネットワーク (phylogenetic network) とその様々なサブクラスに関する研究がこれまでに多数行われている (Huson *et al.* (2010); Steel (2016)). その研究成果は既に実際のデータ解析に応用されており、例えば、SplitTree などのソフトウェアは系統ネットワークを使ってデータを可視化するツールとして広く使われている (Bryant and Moulton (2004)). ただし、このトレンドは系統ネットワークが系統樹に取って代わることを意味しているのではなく、系統樹は依然として進化を記述するファンダメンタルなモデルであることを強調しておく。

2. 系統樹ベースのネットワーク (TBN) と細分系統樹

興味のある現存種の集合 X を葉とする根付き二分系統ネットワーク N が与えられているとき、これらの種が辿った進化の道筋を系統樹モデルで記述するとなれば、 N の中に X を葉とする何らかの系統樹 T を見出したくなる。そこで Francis and Steel (2015) は、系統樹に余分な辺を加えてできる系統樹ベースのネットワーク (*tree-based phylogenetic network*; TBN) を定義した。TBN は細分系統樹 (*subdivision tree*) という全域木を持つ系統ネットワークと定義することもできるため、TBN を論じるうえで細分系統樹という概念は本質的な役割を果たす。

3. TBN の構造定理が導く一連のアルゴリズムと統計学的な意義

Hayamizu (2018) は、根付き二分系統樹 N の細分系統樹の集まり $\{T_1, \dots, T_{\alpha(N)}\}$ を特徴づける構造定理を示し、次の一連の問題を解く高速なアルゴリズムを統一的な視点で記述した。

- (1) 決定／探索問題：系統ネットワーク N が与えられたとき、 N が TBN か (すなわち細分系統樹が存在するか) 否かを決定し、存在するならば一つ見つける問題。Francis and Steel (2015) はこの問題を解く線形時間アルゴリズムを与えたが、次のように数え上げ問題に拡張すると、多項式時間では解けないかもしれないと予想していた。
- (2) 数え上げ問題：系統ネットワーク N が与えられたとき、 N の細分系統樹の個数 $\alpha(N) \in \mathbb{Z}_{\geq 0}$ を求める問題。Hayamizu (2018) は、これを解く線形時間アルゴリズムを与え、 N が TBN のとき、 $\alpha(N)$ は N の複雑さを評価する尺度になるため、モデル選択の文脈に関連する。
- (3) 列挙問題：系統ネットワーク N が与えられたとき、 N の全ての細分系統樹 $T_1, \dots, T_{\alpha(N)}$ を列挙する問題。入力 N のサイズに関する多項式時間でこの問題を解くアルゴリズムが存在しないことはすぐに分かる (列挙したい解の個数 $\alpha(N)$ 自体が N のサイズに関する指数関数で表される場合があるため)。しかし、Hayamizu (2018) は、これを高速に解く線形時間遅延アルゴリズム (列挙アルゴリズムの中で最も効率的なクラスに属するもの) を与えた。全ての解ではなく指定の個数 $k \in \mathbb{N}$ の解のみを列挙するには、 $O(k|V(N)|)$ 時間で十分である。これにより細分系統樹の一樣サンプリングなどの応用が可能になる。
- (4) 最適化問題：系統ネットワーク N と辺の重みづけ関数 $w \geq 0$ が与えられたとき、ある目的関数の値 $f(T)$ を最大化 (または最小化) する細分系統樹 T を求める問題。力まかせ探索では指数時間を要するが、Hayamizu (2018) の構造定理は最適解を線形時間で求めるアルゴリズムを導く。この最適化問題は、 N の各辺が存在するか否かの不確かさに応じた確率 w が与えられ、尤度または対数尤度 $f(T)$ を最大化するベストな細分系統樹 T を求めるという最尤推定の文脈で現れる問題である。

参 考 文 献

- Bryant, D. and Moulton, V. (2004). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks, *Molecular Biology and Evolution*, **21**, 255–265.
- Francis, A. and Steel, M. (2015). Which phylogenetic networks are merely trees with additional arcs?, *Systematic Biology*, **64**, 768–777.
- Hayamizu, M. (2018). A structural theorem for tree-based phylogenetic networks, preprint available at [arXiv:1811.05849](https://arxiv.org/abs/1811.05849) [math.CO].
- Huson, D. H. and Rupp, R. and Scornavacca, C. (2010). *Phylogenetic networks: concepts, algorithms and applications*, Cambridge University Press.
- Steel, M. (2016). *Phylogeny: Discrete and random processes in evolution*, SIAM.

統計数理研究所で開発された R パッケージ R packages developed at the Institute of Statistical Mathematics

モデリング研究系 中野 純司 (Junji Nakano) *

1. はじめに

統計数理研究所は計算機の出現以来、その時代の最新の統計計算環境を維持してきた。そしてその上で種々のソフトウェアが開発・公開されてきている。ただ計算機技術は急速に発展しており、過去のソフトウェアをソースコードでそのまま配布しても使える人は限られてしまう。そのため統計科学技術センターを中心として、それらのソフトウェアをその時代の計算機環境に合うように保守し、さらに利用しやすくすることが試みられてきた。われわれは、統計科学分野でデファクトスタンダードになっている統計解析ソフトウェア R からそれらのソフトウェアを利用できるように、R のパッケージを作成して公開している。本稿ではそれらを紹介するが、多くは CRAN (<https://cran.r-project.org/> またはそのミラーサイト、例えば <https://cran.ism.ac.jp/>) から入手できる。

2. timsac

TIMSAC (TIME Series Analysis and Control program) は、赤池弘次氏を中心として開発された時系列データの解析、予測、制御のための総合的プログラムパッケージである。オリジナル TIMSAC (TIMSAC-72) は 1972 年に発表され、その後、TIMSAC シリーズとして TIMSAC-74, TIMSAC-78, TIMSAC-84 が統計数理研発行の Computer Science Monograph に発表された。工業プロセスの最適制御、経済変動の分析など広い分野で現在でも利用されている。R パッケージ `timsac` は、FORTRAN で書かれているオリジナルプログラムの計算処理機能の多くをサブルーチン化し、R 関数を通して入出力を行い、必要であればその解析結果等を R でグラフィック表示することにより時系列データ解析を容易にしたものである。

3. SAPP

尾形良彦氏を中心として開発された SASeis (Statistical Analysis of Seismicity) は、地震活動などの統計的解析とモデリングのためのプログラムパッケージである。そこでは大森・宇津の公式と点過程 ETAS (Epidemic Type Aftershock Sequence) モデルを扱っているが、それは地震活動の標準化モデルとして世界各国で使用されている。そこで FORTRAN で書かれたこれらの計算処理機能を R パッケージにしたものが SAPP である。

4. TSSS

R パッケージ TSSS は、北川源四郎氏による書籍「FORTRAN 77 時系列解析プログラミング」(岩波書店、1993) に掲載されていたプログラムを基に作成された時系列データ解析のための関数群である。代表的な時系列のモデリングに必要な最小二乗法、最尤法、カルマンフィ

* 中央大学国際経営学部：〒192-0393 東京都八王子市東中野 742-1

ルタによる推定の方法、情報量規準 AIC を用いたモデルの評価・選択の方法が実現されている。TSSS は本書のデータをデータセットとして組み込んでおり、関数のドキュメントにおける例題の一部ではこれらのデータセットを用いている。なお、時変係数 AR モデルの時変分散と時変 AR 係数を推定する関数 (tvvar, tvar) については、OpenMP を使った拡張パッケージ tvvarOMP を作成しており、それにより並列処理も行えるようになっている。このパッケージは <http://jasp.ism.ac.jp/ism/TSSS/> から入手できるが、近々、CRAN で公開予定である。

5. catdap

CATDAP (CATegorical Data Analysis Program) は、坂元慶行氏を中心に開発された最適な分割表 (クロス表) の探索のためのプログラムである。最適な説明変数の選択に AIC (Akaike Information Criterion) が使われている。R パッケージ `catdap` は、FORTRAN で書かれた CATDAP の計算処理機能をサブルーチン化することにより、R からこれら関数として利用できるようにした。最近、主として石黒真木氏によりいくつかの拡張が行われた。例えば現在の関数 `catdap2` では連続値目的変数に適用できたり、目的変数、説明変数に欠測が含まれるデータに適用できたりするようになっている。

なお、関数 `catdap2` の機能をさらに使い易くするため、パッケージ R commander (`Rcmdr`) を使ったメニューインタフェース (`RcmdrPlugin.catdap`) も利用可能になっている。 (<http://jasp.ism.ac.jp/ism/RcmdrPlugin.catdap/>)

6. NScluster

R パッケージ `NScluster` は、ネイマン・スコット型空間クラスターモデルのシミュレーションとパラメータ推定のための関数群である。これらの関数は U.Tanaka, Y. Ogata and K. Katsura, Simulation and estimation of the Neyman-Scott type spatial cluster models (Computer Science Monographs, No.34, 1-44, The Institute of Statistical Mathematics, 2008) の FORTRAN プログラムをもとに開発された。

パラメータ推定のためにシンプレックス法を用いているが、モデルによってはかなりの計算時間がかかる。そのため、この時間のかかる計算処理の部分を OpenMP を使って並列化している。OpenMP が利用可能な環境であれば、実行時間の大幅な短縮を図ることができる。

7. Rhpc

R パッケージ `Rhpc` は基本的な並列計算用パッケージ `snow` の流れをくむ R の並列化のためのパッケージである。その特徴は、並列化のために MPI ライブラリを用いるが 2GB 以上のデータ処理に対応している、多くの部分を C でプログラムして実行速度を上げている、R から MPI 外部プログラムを利用し易くなっている、などである。特に最近のスーパーコンピュータ上での利用を念頭において開発されている。

8. Rmpenv

R パッケージ `Rmpenv` は任意精度による実数と複素数の四則計算および基本的な数学関数、さらに行列積や逆行列を求める関数などを実現するパッケージである。現在機能拡張中であり、まだ公開にはいたっていない。

謝 辞

以上のパッケージの作成は嵯峨優美氏、中間榮治氏との共同研究によるところが大きい。また、オリジナルの FORTRAN プログラム作成者の方々に深く感謝したい。

混合分布モデルとデータ同化

Finite mixture models and data assimilation

モデリング研究系 上野 玄太 (Genta Ueno)

1. 混合分布モデルを用いたプラズマデータ解析

超高層物理学を専攻した大学院時から継続している研究課題である。宇宙空間のプラズマの速度分布は正規分布から離れた形状をとることが多く、複数のピークを持ったり分散の異なる分布の重ね合わせが見られたりする場合が多い。そのような複雑な分布の理解は、地上を優雅に照らすオーロラ現象のメカニズムの理解につながる。プラズマの速度分布を詳細に観測できるようになったのは、1992年に日米共同で打ち上げられた人工衛星ジオテイルが最初であり、現在に至るまで12秒ごとに速度分布データを取得している。

Ueno et al. (2001) は、プラズマ速度分布データに対して正規混合分布モデルを当てはめ、複数の成分に分離を可能としたものである。分離のために推定した各成分のパラメータ (混合比、平均ベクトル、分散共分散行列) は、プラズマの密度が急変する境界層と呼ばれる領域のデータ解析における基礎的な物理量としてそのまま利用できる。このモデルの適用により、25年間余りに取得・蓄積されている大量の速度分布データを対象にした統計解析の道を開いた。この手法により、磁気圏境界層での応用研究を進めた (Lui et al., 2005; Nishino et al., 2007a, d, c, b; Nakai and Ueno, 2011)。かつては高温プラズマに埋もれて取り扱いが困難であった低温プラズマの解析を実施したものである。このモデルはノイズ除去にも応用でき、電子観測データからの光電子成分の除去への応用が Ueno et al. (2001) にある。このモデルはもともとジオテイル衛星に搭載されたプラズマ観測器に対して開発したものであるが、その後、2007年にNASAにより打ち上げられた人工衛星テミスによるプラズマ観測データへ展開している (Chaston et al., 2013)。

応用研究と並行して、新しい方法論の提案を行った。中村 他 (2005) は、プラズマ観測装置に検出不可能な速度があることに動機を得て、データ欠損領域が存在する場合の混合分布モデルを提案したものである。

2. データ同化の方法論への展開

2005年度からは、大気海洋結合モデルを軸としたデータ同化手法の研究を集中的に進めた。Ueno et al. (2007) は大気海洋結合モデルにアンサンブルデータ同化手法を用いた初の研究である。つづいて、推定精度を上げることを目的として、システムノイズ・観測ノイズのパラメータの最適化を行うこととした。状態ベクトルの確率分布をアンサンブルによる近似表現すると、時系列モデルの尤度関数の表式が混合分布モデルのそれと同等の形になるところが面白い。この表式の類似性に注目することで、時系列モデルにおいてもEMアルゴリズムを導出が可能であり、特に観測ノイズの分散共分散行列の推定に有効である (Ueno and Nakamura, 2014)。最尤法によるパラメータ推定を大気海洋結合モデルに対して行った結果、予測精度の向上を実現すると同時に、データ同化による状態推定の限界を明らかにした (Ueno et al., 2010)。限界の一つは、静穏時の大気海洋の状態とエルニーニョなど変動時の状態で、結合モデルによる再現

性に違いが見られたことである。その状況を解決するため、パラメータに時変性を許し、観測状況に適応的にフィルタのゲインを推定するバイズ法を開発した(Ueno and Nakamura, 2016; Nakabayashi and Ueno, 2017)。

参 考 文 献

- Chaston, C. C., Yao, Y., Lin, N., Salem, C. and Ueno, G. (2013). Ion heating by broadband electromagnetic waves in the magnetosheath and across the magnetopause, *Journal of Geophysical Research: Space Physics*, **118**, DOI: <http://dx.doi.org/10.1002/jgra.50506>.
- Lui, A. T. Y., Hori, T., Ueno, G. and Mukai, T. (2005). Plasma transport from multicomponent approach, *Geophysical Research Letters*, **32**, 1, DOI: <http://dx.doi.org/10.1029/2004GL021891>.
- Nakabayashi, A. and Ueno, G. (2017). An extension of the ensemble Kalman Filter for estimating the observation error covariance matrix based on the variational Bayes's Method, *Monthly Weather Review*, **145**, 199–213, DOI: <http://dx.doi.org/10.1175/MWR-D-16-0139.1>.
- Nakai, H. and Ueno, G. (2011). Plasma structures of Kelvin-Helmholtz billows at the dusk-side flank of the magnetotail, *Journal of Geophysical Research*, **116**, DOI: <http://dx.doi.org/10.1029/2010JA016286>.
- Nishino, M. N., Fujimoto, M., Terasawa, T., Ueno, G., Maezawa, K., Mukai, T. and Saito, Y. (2007a). Geotail observations of temperature anisotropy of the two-component protons in the dusk plasma sheet, *Annales Geophysicae*, **25**, 769–777.
- Nishino, M. N., Fujimoto, M., Ueno, G., Maezawa, K., Mukai, T. and Saito, Y. (2007b). Geotail observations of two-component protons in the midnight plasma sheet, *Annales Geophysicae*, **25**, 2229–2245.
- Nishino, M. N., Fujimoto, M., Ueno, G., Mukai, T. and Saito, Y. (2007c). Origin of temperature anisotropies in the cold plasma sheet: Geotail observations around the Kelvin-Helmholtz vortices, *Annales Geophysicae*, **25**, 2069–2086.
- Nishino, M. N., Fujimoto, M., Terasawa, T., Ueno, G., Mukai, T. and Saito, Y. (2007d). Temperature anisotropies of electrons and two-component protons in the dusk plasma sheet, *Annales Geophysicae*, **25**, 1417–1432.
- Ueno, G., Nakamura, N. and Higuchi, T. (2001). Separation of photoelectrons via multivariate Maxwellian mixture model, *Discovery Science, Proceedings of the 4th International Conference, DS 2001* (eds. K. P. Jantke and A. Shinohara), Lecture Notes in Computer Science, **2226**, 470–475, Springer, Washington, D.C. 11.
- Ueno, G. and Nakamura, N. (2014). Iterative algorithm for maximum-likelihood estimation of the observation-error covariance matrix for ensemble-based filters, *Q. J. R. Meteorol. Soc.*, **140**, 295–315, 1, DOI: <http://dx.doi.org/10.1002/qj.2134>.
- Ueno, G. and Nakamura, N. (2016). Bayesian estimation of the observation-error covariance matrix in ensemble-based filters, *Quarterly Journal of the Royal Meteorological Society*, **142**, 2055–2080.
- Ueno, G., Nakamura, N., Higuchi, T., Tsuchiya, T., Machida, S., Araki, T., Saito, Y. and Mukai, T. (2001). Application of multivariate Maxwellian mixture model to plasma velocity distribution function, *Journal of Geophysical Research*, **106**, 25655–25672, 1.
- Ueno, G., Higuchi, T., Kagimoto, T. and Hirose, N. (2007). Application of the ensemble Kalman filter and smoother to a coupled atmosphere-ocean model, *SOLA*, **3** (1), 5–8, 1.
- Ueno, G., Higuchi, T., Kagimoto, T. and Hirose, N. (2010). Maximum likelihood estimation of error covariances in ensemble-based filters and its application to a coupled atmosphere-ocean model, *Q. J. R. Meteorol. Soc.*, **136**, 1316–1343, 7, DOI: <http://dx.doi.org/10.1002/qj.654>.
- 中村永友, 上野玄太, 樋口知之, 小西貞則 (2005). 欠損混合分布モデルとその応用, *応用統計学*, **34**, 57–73.

地球磁気圏の撮像観測とそのデータ同化

Data Assimilation of Imaging Observation in Earth's magnetosphere

モデリング研究系 中野 慎也 (Shin'ya Nakano)

1. はじめに

地球磁気圏は、宇宙空間の中でも地球の持つ磁場の影響が及ぶ範囲を指す。地球磁気圏は、太陽風と呼ばれる太陽からのプラズマ (電荷を持った粒子で構成されるガス) の流れの影響で非対称な形状をしており、太陽側は地上 6 万 km 程度まで、太陽と反対側には地上数百万 km 以上まで広がっている。通常、地球磁気圏の研究では、人工衛星によって直接その場所の環境を観測して得たデータを用いる。しかし、現在運用されている衛星の情報を集めても、広い磁気圏中の数点の情報只得られるに過ぎず、磁気圏の大域的な現象の全体像をつかむのが難しい。

地球磁気圏のプラズマの空間分布を、遠隔から 2 次元的に捉える撮像観測は、衛星による直接観測の欠点を解決する有用な方法である。特に、2000 年から 2005 年に運用されていた人工衛星 IMAGE は、様々な手段による撮像観測を実現した衛星であり、有用なデータが取得されている。しかし撮像観測は、プラズマ密度以外の物理量について情報を得るのが難しいという欠点がある。そこで我々は、データ同化技術を活用することにより、人工衛星 IMAGE による撮像観測データから磁気圏の大域的な現象の全体像を捉える手法の開発を進めてきた。本稿では、これまでに行ってきた撮像観測のデータ同化について紹介する。

2. 高速中性粒子データ同化

地球磁気圏の荷電粒子の中でも、比較的エネルギーの高い 1keV–100keV (eV は荷電粒子のエネルギーを表す単位。電子を 1 ボルトの電圧で加速して得られるエネルギーが 1eV となる) の陽イオン (主に H⁺) は、磁気嵐と呼ばれる地上の全球的な地磁気変動を引き起こす他、オーロラ嵐をはじめとする高緯度域の電離圏現象とも密接に関係している。この 1keV–100keV 程度の陽イオンの空間分布に関する情報を得る手段として、高速中性粒子撮像観測がある。高速中性粒子は、高エネルギーの陽イオンが、磁気圏中に漂う地球起源の中性粒子から電子を受け取ることによって生成される中性の粒子である。磁気圏中の陽イオンは、地球磁場によるローレンツ力の影響で自由に動くことができないが、中性に変化すると力を受けずに高速で直線運動する。これを遠隔で観測することにより、陽イオンの空間分布に関する情報が得られる。我々は、IMAGE 衛星で観測された高速中性粒子のデータを磁気圏荷電粒子モデルに同化し、荷電粒子分布の時間発展を推定する手法の開発を進めてきた Nakano et al. (2008)。ここで問題となるのは、荷電粒子の動きを決める電場と磁場のうち、地球磁気圏においては電場について十分な情報がないということである。そこで、電場についてはデータ同化の過程で推定することにより、全体の感電粒子分布の時間発展を推定することを実現した。高エネルギーの荷電粒子は、運動にエネルギー依存性があるため、磁力線方向の運動を平均化した Boltzmann 方程式で扱っている (Fok et al., 2001)。データ同化には、当初、粒子フィルタを再帰的に適用する手法を用いていたが、開発の過程で、融合粒子フィルタ Nakano et al. (2007)、アンサンブル変換カルマンフィルタ Bishop et al. (2001) に変更され、現在は次に述べる極端紫外光データ同化と

の統合を進めている。

3. 極端紫外光データ同化

IMAGE 衛星では、30.4nm の波長の極端紫外光による撮像観測も行っていた。太陽から来る紫外光のうち、30.4nm の波長のものはヘリウムイオン (He^+) に散乱されるため、これを遠隔から観測することにより、磁気圏の He^+ の分布について情報が得られる。極端紫外光の撮像観測で得られる情報は、磁気圏の中でもエネルギーが低い 1eV–10eV 程度の荷電粒子の情報である。この低エネルギーの荷電粒子は、地球の高度 20000–30000km 以下の領域に高密度で分布しており、プラズマ圏と呼ばれる。プラズマ圏の低エネルギー He^+ は、水素イオン (陽子; H^+) など、他の低エネルギー荷電粒子と同じ方程式にしたがって動くため、 He^+ のデータから、プラズマ圏中のプラズマを構成する荷電粒子全体の動きを推定することができる。我々は、アンサンブル変換カルマンフィルタを用いて、プラズマ圏の物理モデルに、極端紫外光撮像観測データを同化し、プラズマ圏のプラズマ密度分布の時間発展を推定する手法を開発した Nakano et al. (2014)。低エネルギーの荷電粒子の動きも電場と磁場の影響を受けるが、電場について十分な情報がないため、高速中性粒子データ同化と同様にデータ同化の過程で電場の推定を行っている。また、ここで用いているプラズマ圏モデルは、地球磁力線方向の密度構造についてある関数系を仮定した 2 次元モデルであるが、磁力線方向の密度構造に関するパラメータを周辺尤度最大化で推定することにも成功した。

4. おわりに

現在我々は、高速中性粒子データと極端紫外光データの両方を磁気圏統合モデルに同化する新たなデータ同化システム開発を進めている。高速中性粒子の起源となる高エネルギー荷電粒子は、極端紫外光で観測される低エネルギー荷電粒子の分布するプラズマ圏と比較して、磁気圏のやや外側に分布しており、磁気圏内の電場を推定する上で、高速中性粒子データと極端紫外光データは、互いの情報を保管する役割を果たすと考えられる。最初に述べたように、地球磁気圏に関して観測から得られる情報は非常に限られている。我々は、データ同化を活用し、物理法則の知見を活用して直接観測できない物理量を推定しながら、地球磁気圏で起こる様々な現象の全体像を包括的に捉えることを目指している。

参 考 文 献

- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001., Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects, *Mon. Wea. Rev.*, 129, 420–436.
- Fok, M.-C., R. A. Wolf, R. W. Spiro, and T. E. Moore, 2001., Comprehensive computational model of Earth's ring current, *J. Geophys. Res.*, 104, 8417.
- Nakano, S., G. Ueno, and T. Higuchi, 2007., Merging particle filter for sequential data assimilation, *Nonlin. Process. Geophys.*, 14, 395–408.
- Nakano, S., G. Ueno, Y. Ebihara, M.-C. Fok, S. Ohtani, P. C. Brandt, D. G. Mitchell, K. Keika, and T. Higuchi, 2008., A method for estimating the ring current structure and the electric potential distribution using ENA data assimilation, *J. Geophys. Res.*, 113, A05208, doi:10.1029/2006JA011853.
- Nakano, S., M.-C. Fok, P. C. Brandt, and T. Higuchi, 2014., Estimation of temporal evolution of the helium plasmasphere based on a sequence of IMAGE/EUV images, *J. Geophys. Res.*, 119, 3708–3723, doi:10.1002/2013JA019734.

前震識別とその予測可能性

Discrimination of Foreshocks and Its Predictability

モデリング研究系 野村 俊一 (Shunichi Nomura)

要 旨

一連した地震の群れの中の最大地震すなわち本震には、しばしば前震と呼ばれる先行活動が見られ、本震を事前予測する重要な手がかりとなる。本稿では、地震カタログから構成したクラスター（地震群）の時空間的配置および地震規模推移を特徴量として、地震群が前震である確率および一定期間に一定規模の本震が起こる確率を評価する研究について解説する。

キーワード：地震予測、前震、クラスタリング、ロジスティック回帰、スプライン関数

1. はじめに

大地震の後には多数の余震が発生するが、ときに大地震に先駆けて地震が集中して起こることがあり、これを前震活動という。もしも大地震が実際に起こるより前にその前震活動を特定することができれば、大地震を短期的に予測することが可能となる。前震活動とその他の地震活動との完全な事前判別は困難であるが、Ogata et al.(1995)によると前震活動とその他の地震活動では、地震間の時空間的距離やマグニチュード差の傾向に違いがある。以降では、野村・尾形 (2018) を例に前震識別の方法論を 3 つの段階に分けて解説する。

2. 地震群（前震群候補）の構成

まず、地震活動を地震の群れとして捉えるために、地震活動の点群に対してクラスタリング手法を用いて地震群を構成する。ここでは Ogata et al.(1995) に倣い Single-link 法を採用し、日本のマグニチュード 4 以上の地震について、震央間距離 $\Delta d(\text{km})$ と時間距離 $\Delta t(\text{日})$ が $\sqrt{(\Delta d)^2 + (c\Delta t)^2} \leq 33.33$ を満たす地震同士を連結していくことでクラスター（地震群）を構成した。ただし、 $c = 1.11(\text{km/日})$ とおいた。

3. 前震群候補からの特徴量抽出と前震群の定義

次に、構成した地震群内の各地震に対して、その時点までに発生した地震のみからなる部分群を作り、前震群候補とする。さらに、実際の前震群を定義付けた上で、前震群候補である部分群から、前震群の識別に有効な特徴量を抽出する。ここでは、図 1 のように部分群の最後の地震発生時点から 30 日以内に、部分群内の最大マグニチュードを超える地震が起きたときに前震群であると定義し、その事前識別のため各部分群から次の特徴量を抽出した。

- 群内の地震数： $N \geq 2$
- 群内の一番目・二番目に大きいマグニチュード： M_1, M_2
- 群内の期間長： T （日）

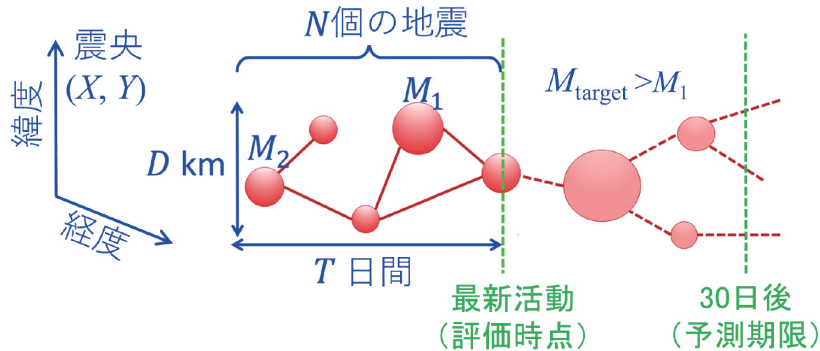


図 1. 前震群候補から抽出する特徴量と前震群の定義

- 群内の平均震央間距離： D (km)
- 群内の中心経度・緯度： (X, Y) (度)

4. 前震群候補の前震確率評価

最後に、前震群候補である部分群について抽出した特徴量に基づき、前震群であるかどうかを確率で評価することにより識別する。ここでは、ロジスティック回帰を用いて前震確率 p を

$$\text{logit} p = \log \frac{p}{1-p} = g(X, Y) + f_1(N, M_1, M_2) + f_2(N, M_1, T) + f_3(N, M_1, D) + \varepsilon_i$$

のように評価した。右辺第 1 項 $g(X, Y)$ は地域による前震確率の変化を表しており、薄板スプライン関数を用いて推定する。 $f_1(N, M_1, M_2)$, $f_2(N, M_1, T)$, $f_3(N, M_1, D)$ は相関の強い特徴量の組合せごとに、特徴量による前震確率への影響を表しており、それぞれ 3 次 B-スプライン関数を用いて推定する。最後の項は、上記には含まれない前震確率に影響する特徴量の効果を、地震群 i ごとの変量効果として取り入れたものである。

1926 年から 1999 年までの気象庁カタログから、前震確率の評価式を学習し、2000 年から 2017 年 10 月までの同カタログに適用して前震確率を評価した結果、最大で 80% 前後の前震確率となった地震群が 2 つあり、そのいずれもが実際にマグニチュード 6 以上の本震を起こした。

5. まとめ

本稿では、進行中の地震群が大地震前の前震活動であるかについて、地震群の特徴量に基づいて確率的に識別する方法を解説した。前震識別に基づく地震予測手法は、短期間での大規模地震の予測について比較的高い発生確率を与える数少ない手法であり、まだ改善の余地は大きいものの、将来的には大地震のリアルタイム予測へと実装されることを期待する。

参 考 文 献

- 野村俊一, 尾形良彦 (2018). マグニチュードと時空間情報に基づく前震確率評価モデルの識別性能, 「地震予知連絡会会報」, **99**, 1-6
- Ogata, Y., Utsu, T. and Katsura, K. (1995). Statistical features of foreshocks in comparison with other earthquake clusters, *Geophysical Journal International*, **121**, 233-254.

「日本人の国民性調査」から「文化多様体解析」へ

From “Japanese National Character Survey” to “Cultural Manifold Analysis (CULMAN)”

データ科学研究系 吉野 諒三 (Ryozo Yoshino) *

「日本人の国民性」調査は昭和 28 年より開始され、戦後民主主義の発展のための「科学的世論調査」の基盤となってきた。これは、世界で希有の長期継続統計の無作為標本抽出調査であり、内閣府「社会意識に関する調査」、NHK「日本人の意識調査」とともに日本の三大意識調査と呼ばれている。その後、これを追従し、米国 GSS、ドイツ ALLBUS、EU の Eurobarometer 等々の一般社会調査が遂行されるようになり、近年ではアジア諸国でも同様の調査が開始されている。

1970 年頃より国民性調査は国際比較に拡大され、筆者は平成元年に入所以来、「日本人の国民性調査」と「意識の国際比較」に携わってきた。「国際比較は意識調査の宝庫である」とは、戦後の科学的世論調査を含む社会調査の発展に大きな足跡を残した林知己夫が到達した認識であった。様々な国を比較する時は、翻訳の問題、各国固有の調査方法の違いなど、そもそも国際比較など可能なかが大問題となる。我々はこの「国際比較可能性」を追求しているのであり、また「データの科学」を計量的文明論のために試行錯誤しているのである。

林知己夫と鈴木達三らにより構築された「連鎖的比較方法論 CLA を、筆者らは「文化多様体解析 CULMAN」へと発展させてきた。特に、この 20 年ほどは一連の大型科研費を獲得して、CULMAN の具現化のため、米国西海岸日系人調査 (1999-2001)、ハワイ日系人調査 (2000-2002)、東アジア価値観調査 (2002-2005)、環太平洋価値観調査 (2004-2009)、アジア太平洋価値観国際比較調査 (2010-2015) 等々を展開してきた。

これらの調査研究が進む中で、共同研究者のハワイ大学黒田安昌名誉教授 (現) の論文はハワイ教育委員会に認められ、初等中等学校の教科書に掲載されるようになった。また、ワシントン大学カシマ・テツデン名誉教授 (現) は、日系人社会への貢献で 2018 年に日本政府から旭日小勲章が授与された。少数者差別問題が世界に渦巻く中で、多様な民族・人種の共存共栄モデルの提示に尽力してきた貢献が認証されたのである。

近年の米大統領選挙やイギリスの EU 離脱国民投票では、事前の世論調査は大きな失敗を見た。欧米では調査方法の質が著しく下がり、信頼性を失っている。また、急速な IT 技術の進展に対し、法律の規制が追いつかぬ中で、Cambridge Analytica 社や Facebook 社の個人情報の扱い等、企業が法律や倫理を踏みこむような事件もある。

世論と選挙に関する Galbraith の次の文章は、今日でも我々に示唆するものは重い。「・・・

* 同志社大学 文化情報学部：〒610-0321 京都府京田辺市多々羅都谷 1-3

経済的に豊かな階層の人々からなる勢力には資金と影響力がある。そして、彼らは投票する。一方、貧しい階層に対する支援者からなる勢力は、人数的には多いが、貧しい人々の多くは残念ながら投票しない。民主主義は存在するが、それは恵まれた人々のための民主主義と言えなくもないのである。・・・」(John Kenneth Galbraith,1996「よい世の中」佐々木直彦・純子訳,p.26, 日本能率協会マネジメントセンター。)]

民衆の感情を掴みながらも、「感情の世論」に流されず、建設的な目標を支える「理性の世論」を掴む調査のために「日本人の国民性調査と意識の国際比較」が貢献し続けていくことを強く望む。

(注) 本研究所の主要な調査は、多数の書籍、研究レポート、WEB サイト上で公開されている。「国際比較データの解析」(吉野・林・山岡,2010, 朝倉書店) 参照。

社会調査方法論の実践的研究

Practical Studies on Social Survey Methodology

データ科学研究系 前田 忠彦 (Tadahiko Maeda)

1. はじめに

社会調査法は社会現象に関するデータ取得のための中心的な手段の一つである。筆者はそのような社会調査を企画・実践するとともに調査データを素材とした調査方法論研究を主なテーマとしている。実践と研究が一体を成すスタイルで研究を続けており、現実問題に合わせた調査設計法自体が研究の重要な一面である。社会調査には様々なプロセスが含まれるが、その全ての段階に調査方法論上の研究課題が潜んでおり、次節で紹介するような具体的な調査を研究材料として研究を進めてきた。実施した社会調査には多くの共同研究も含まれる。

2. 調査プロジェクトの例

2.1 日本人の国民性調査および関連調査

中心的な調査研究の一つが、統計数理研究所が1953年以来5年に一度実施している「日本人の国民性調査」である(最新調査は2018年実施の第14次全国調査)。これは、同じ調査手法(訪問面接法)、同じ調査項目で横断調査を繰り返すことを基本とした継続社会調査で、5年に一度の本調査実施の他に中間年には様々な関連研究を行っている。多くの先輩方の努力により続けられてきたプロジェクトであるから、その資産を活用しながら、近年は自身のアイデアを調査設計等に生かすようにしている。

2.2 共同調査研究

2010年には「2010年格差と社会意識に関する全国調査」(略称SSP-I2010調査)を実施した。吉川徹客員教授と共に多数の研究者を含む共同研究体制を組み、調査の企画・実施から解析まで、大阪大学人間科学研究科との緊密な連携の下で多数の成果を生んだ調査研究である。

2011年には国立国語研究所と「第4回鶴岡市における言語調査」を実施した。1950年に第1回調査が両研究所の協力で実施されて以来、1972年第2回、1991年第3回と約20年間隔で行われてきた。山形県鶴岡市における共通語化の進行を、各回のクロスセクション調査と、パネル調査を組み合わせたデザインで研究する、恐らく世界最長の言語に関する継続調査である。

3. 具体的な研究テーマ例

2節で紹介したプロジェクトでの中心的調査は、調査員が調査対象者を訪ね、面接で回答を取得する「個別訪問面接法」を採用している。最も「正統的」でよく用いられてきた調査手法であるが、そのような調査手法についても未だ様々な研究テーマが残されている。

3.1 調査員効果に関する研究

このような調査員が介在する調査では、調査員の持つ何らかの特徴が調査結果に影響を与える。最近はこの「調査員効果」について二つの面からの研究を行っている。第一は調査員特性

が、回収・非回収に与える影響という面であり、例えば松岡亮二氏（早稲田大学）との共同研究では、「日本人の国民性第 13 次全国調査」について、調査地点の特徴や調査員特性を含めたマルチレベル分析により、対象者・地点・調査員の特性と回収状況の関連を総合的に検討した。第二は調査員属性が回答内容に与える効果についてのもので、特定テーマの調査項目に対して、調査員の属性が与える効果の可能性等を検討している（金沢大学小林大祐氏との共同研究）。

3.2 調査パラデータの解析—訪問記録を例として

調査パラデータとは、質問への回答という最も中心的な調査データに加えて、調査の実施プロセスに付随して得られる様々な情報を指す。最近では、面接調査における調査員の活動（訪問記録）を分析した。この分析を通じて、面接調査での回収・不能がどのような経緯を経て決まるのか、その経緯は対象者の住む住居特性と関係するか、等を分析することによって、調査員の活動に資する知見を得ることを目的とした研究である。

3.3 調査モード間の比較研究

測定プロセスのうち対象者から回答を取得する手段を「調査モード」と呼ぶ。調査員が回答を面接で聴き取る「他記式」に対し、対象者が自ら調査票を読み回答を記入する「自記式」のモードもあり、回答結果にこの調査モードが大きく影響することがある。尾崎幸謙氏（筑波大学）との共同研究では、留置法（自記式）と面接法（他記式）で同時に行われた実験的調査を比較分析した。留置法では、面接法に比べて、暮らし向きの満足度が低いとか、選挙で投票に行く頻度が低いといった違いが明瞭に見られ、この違いは傾向スコアを用いて回答者の属性（共変量）の分布が二つのモードで異なることを調整した上でも消えない。このことから、「社会的望ましさ」への対象者の敏感さが両モード間で異なる反応を引き出す可能性等を推察することができる。モード間の差が生ずるメカニズムも一様ではなく、研究課題が残されている。

3.4 調査不能バイアスの調整

このテーマの背景・動機となっているのは、近年の社会調査特に面接調査全般における回収率の低下傾向である。例えば 2013 年実施の「日本人の国民性第 13 次全国調査」では回収率が 50% と、半数近くの人が調査に協力しない状態での結果が得られている。この協力率の低さで母集団の推定を正確に行えているのかという点（調査不能バイアスの問題）が懸念される事態と言える。伏木忠義氏（新潟大学）との共同研究を進め、例えば 2 節で紹介した SSP-I2010 調査を題材としてこの問題を検討した成果を発表した。

3.5 標本設計・サンプリングの精度とその一貫性に関する検討

サンプリングの設計と、その設計下での標準誤差の大きさ等の調査精度の評価は、社会調査設計上の重要な論点の一つである。各調査プロジェクトでの標本設計を担当し続ける中で、継続社会調査でのそうした精度の一貫性の有無や、市区町村合併のような社会制度の変化が調査設計に与える影響も無視できない論点であることに気づき、こうした点の検討も続けている。

4. 社会調査法研究のこれから

「はじめに」に述べたように、社会調査のプロセスの全てにわたって、調査方法論上の研究課題が含まれており、最も伝統的な調査手法である面接調査法に限っても、様々な研究テーマが残されているのが現状である。他方で、回収率が低下し続ける訪問面接法による調査研究には限界が見え始めていることも感じている。ランダムサンプリングと訪問面接法の組み合わせという伝統的な手段に代わる、現代社会にふさわしい調査方法の研究も必須である。

組織規範継承を可能にする目的指向型成果評価と相互依存性の計量分析

Does Goal-Oriented Managerial Behaviour Applying Performance Management Improve Interpersonal Facilitation Among Public Officials in Japan? A Multilevel SEM Analysis Focusing on Division Level Interdependence

データ科学研究系 朴 堯星 (Yoo Sung Park)

1. 研究の背景と目的

現在、多くの自治体が抱える問題の一つは、行政職員の協力行動をいかに引き出すかである。バブル崩壊以降、日本の行政組織は慢性的な赤字財政を抱えている。そのため、2000年代ごろから成果評価と分権化を軸とする組織運営が取り組まれている。しかし、多くの自治体では、成果評価の導入によって業務遂行の個人化が進み、人材育成、チェック体制が弱まり（村林 2004）、これまで緩やかに維持されてきた組織としての機能が揺らいでいる状況である。元来、成果評価の導入は、官僚制型組織運営の弊害を克服することを狙いとする。すなわち、組織機能の向上のために導入されたものであるにもかかわらず、その実態は、当初の期待とはかけ離れているアイロニックな状況である。総務省(2017)によれば、都道府県・市区町村において 977 団体 (54.4%) がすでに行政評価を導入済みである。導入団体数は毎年、増加していることが知られており、従来の組織運営体制への後戻りはできないだろう。むしろ成果評価をめぐる弊害を払拭するための、組織規範の継承を可能とする職場での協力体制を強化させる組織心理学的要因を計量的に探ることが重要と考える。そこで本研究では、組織の長期的な発展に向けて組織が機能するために必要とされる職員の行動として Borman and Motowidlo(1993) が提唱した task performance (TP) と Contextual performance (CP) に着目して、職場での協力体制を強化させる制度設計の条件としての「相互依存性」がもたらす文脈効果を明らかにする。具体的には、個人-市町-県の多層同時比較調査を遂行し、マルチレベル構造方程式モデリングを用い、目的指向型成果評価のもと、相互依存性が行政職員の協力行動に及ぼす影響を明らかにする。

2. 研究の方法

本研究は以下の手順で進めている。まず、自治体行政に適した課業相互依存尺度を開発するため、自治体行政の業務内容と法律で定まっていない自治体行政の業務内容を整理し、インタビュー調査を行い、自治体行政組織における課業相互依存の実態を把握したうえで、つぎに、「多層的相互依存尺度」の開発を行った。さらに、その後、成果評価と分権化を軸とした行政改革にいち早く取り組んできた三重県庁職員 279 名を対象とした。調査対象者は、多段集落抽出法に基づき、三重県本庁 26 課中、業務内容に応じてより明確に相互依存性が表れると考えられる事業執行部門を対象とし、業務特性に偏りが出ないように 17 課を選出し、これらの課に所属する常勤職員全員 (課によっては、7 名 - 24 名程度が所属) に質問紙調査を実施した。ま

た、三重県庁の職員との比較を試みるため、三重県所在の3市（津市、松阪市、尾鷲市）の職員を対象とした同様の調査を実施している。最後に、調査から得られたデータをもとに、課レベル変数が個人レベル変数に及ぼす影響を確かめるため、マルチレベル分析を行った。調査・分析を終え、相互依存性が、行政職員の協力行動が促進するジョブデザインの条件になりうることについての総括的な考察を行った。

3. 研究成果

まず、新しく開発した「相互依存性」尺度の信頼性および妥当性を検討したうえで、マルチレベルSEMを行い、課の違いがもたらす影響を確かめた。マルチレベルSEMは、withinレベル（個人レベル）におけるrandom slopeやrandom interceptが、1つの値ではなくbetweenレベル（課レベル）で値がバラつく因子として捉える。これが、between levelの変数によって推定されることで、レベル内の関係（within）と各レベル間の関係（between）を同時に検討することができる（Snijders and Bosker, 2012）。本研究では課の主効果の検討を目的としているため、切片のみにランダム効果（random intercept）を仮定し、切片の課レベル残差の程度から推定されたモデル間の比較を行い、切片の課レベルのばらつきを、課レベル変数である業務相互依存性と目標相互依存性で説明するモデルを検討した。その結果、課レベルで相互依存性がうまく取れているほど、個人の対人的促進が高まっていることを確認している（表1）。具体的には、課レベルの業務相互依存性には、直接、個人の対人的促進を高める効果があるが、同時に課レベルの目標

相互依存性を媒介して個人の対人的促進を高める効果もあることが明らかになった。このことは、個人の所属する課の働き方を相互依存的なものに変えることで、職場での協力行動を促すのが可能になることを意味する。

表1. マルチレベルSEMを用いた多層レベル解析結果

	モデル1	モデル2	モデル3	モデル4
固定効果				
個人レベル				
対人的促進 ← 職務充実度	0.498 ***	0.490 ***	0.486 ***	0.490 ***
職務充実度 ← 目的指向型経営管理行動	0.540 ***	0.540 ***	0.540 ***	0.540 ***
課レベル				
目標相互依存性 ← 業務相互依存性	0.544 ***	0.544 ***	0.544 ***	0.544 ***
対人的促進 ← 業務相互依存性		0.715 ***		0.285 †
対人的促進 ← 目標相互依存性			0.973 ***	0.791 ***
変量効果				
切片の課レベル残差	0.026	0.021	0.002	0.002
AIC	1978.80	1962.80	1962.20	1962.80
CFI	0.923	0.940	0.916	0.949

個人レベル: n=414, 課レベル: N=33, *** p < .001, ** p < .01, * p < .05, † p < .10.

これまで多くの自治体組織では、成果評価の導入を期に、職場での知の継承が疎かになってきていることに悩まされていた。これに対し、本研究では、職場での協力体制を構築するための、個人レベルと課レベルでの組織心理学的メカニズムを解明したものである。そもそも仕事を個人で遂行するのではなく、課の全員で相互依存的に行うことによって、課全体の目標に対する認識の共有度合いが高まる。そのような過程で、自然に組織運営にかかわる知の継承が緩やかに促されることになるのではないか。言い換えれば、職場での組織規範の継承には、職員個々が属する部署での仕事のやり方も相互依存的なものへと変えることが重要であると考えられる。今後は、部署内の職員同士のパーソナルネットワークを組み込んだモデリングへの拡張を検討していきたい。

参 考 文 献

- 朴堯星・坂野達郎（2015），“自治体職員の対人的促進に関するマルチレベル分析：課レベルの相互依存性に着目して”，『計画行政』, 38(3), 55-64.
- YOOSUNG PARK “Performance Evaluation System and Interpersonal Facilitation: An Empirical Evidence of Public Officials in Mie Prefectural Governments of Japan”, The spring 2015 conference of The Korea Association for Survey Research, Seoul.2015.6.15.

立川市町丁目別住民意識調査分析追記 —小地域推定モデル活用に向けて—

Addendum to “Hirose et al. (2018)”: Applicability of Small Area Explicit Model to Japanese Survey Data

データ科学研究系 廣瀬 雅代 (Masayo Y. Hirose) *

要 旨

小地域推定モデルに基づく統計的推測法は、わが国の調査データ分析にも大いに貢献し得る。本論文では、そのような手法の有用性を示す資料の説明力を高めるべく、廣瀬ら (2018) の分析結果に補足する形で、手法の有用性の解釈を容易にする資料を提示する。このような資料を通して、わが国の小区別調査データ分析の可能性がさらに広がることを期待する。

キーワード：EBPM, 小区別調査データ分析, 小地域推定

1. わが国の小地域推定モデル活用に向けた資料の追記

Evidence Based Policy Making (EBPM) の重要性が高まっている今日、細かな政策やサービスを計画する際、区分ごとの実態を効率よく把握することは重要な課題のひとつになり得る。しかし、区分数が多くなるにしたがって、慣習的に用いられている区分ごとの推定法は、信頼性を大幅に低下させる懸念があり時には計画遂行にも大きく支障を与えかねない。このような問題に対して、小地域推定モデルに基づく統計的推測法の需要が、理論面及び応用面において急速に高まっている (Rao and Molina, 2015; 久保川, 2016)。わが国の調査データも例外ではない。立川市住民意識調査データ (朴・土屋, 2017) でも、町丁目別に区切ることにより同様の問題が懸念される。廣瀬ら (2018) はその調査データに小地域推定モデルを適用するべく、国勢調査と共通の項目を用いて、国勢調査小地域集計結果と各推定値からの乖離を絶対相対誤差によって測ることで、小地域推定モデルに基づく推測法 (以後 MBA 法と呼ぶことにする) の有用性を評価した。しかし、そのような統計的推測法の活用推進の為には、より解釈が容易となる指標を用いた方が適用手法の有用性を説きやすい。例えば、政策設計やサービス計画を立てる側に MBA 法の有用性を説く場面では、単純であるが解釈が容易になる絶対誤差の指標に基づく資料が大いに役立つと考えられる。

そこで、本論文では、廣瀬ら (2018) の資料に追記する形で、国勢調査の集計結果と各推定法 (慣習的な推定法と MBA 法) に基づく推定値の乖離を絶対誤差 ($AER = |\hat{P}_i - p_i| \times 100$) で測り、その結果を図 1 に示す。ここで、廣瀬ら (2018) の定義と同様に、 \hat{P}_i は、立川市住民意識調査データ (朴・土屋, 2017) から慣習的な推定法と MBA 法によって推定された第 i 町丁目の男性割合に対する各推定値を表しており、平成 27 年度国勢調査小地域集計結果 (<http://www.e-stat.go.jp>)

* 九州大学マス・フォア・インダストリ研究所：〒819-0395 福岡市西区元岡 7447

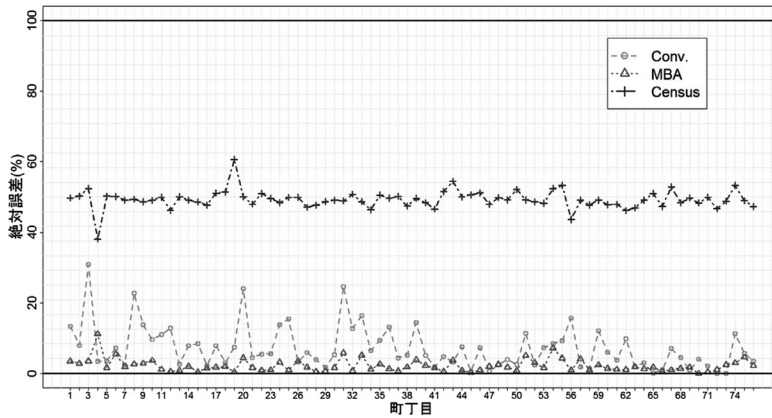


図 1. 立川市 76 町丁目における国勢調査の男性割合からの絶対誤差の比較 (横軸は有効回答サンプルサイズの昇順に町丁目を並べている): 慣習的な推定値の AER(Conv.), MBA 法による推定値の AER(MBA), 国勢調査の各町丁目の男性割合 p_i (Census)

から計算された第 i 町丁目における男性割合 p_i を真としてみなすことにする。この図によって、指標として絶対相対誤差を用いた場合よりも、慣習的な方法と比べて MBA 法がどの程度誤差を抑えているのか解釈しやすくなったように見える。特に、有効回答サンプルサイズが 3 番目に小さい町丁目では、慣習的な推定法によって 31 ポイントもの絶対誤差が引き起こしているのに対し、MBA 法を用いると 3.6 ポイントまで誤差が小さくなっている様子も確認することができる。また、絶対誤差の平均は、慣習的な推定法が 7.3 ポイント、MBA 法は 2.3 ポイントであった。すなわち、国勢調査小地域集計結果 p_i を真として考えると、この結果は、絶対誤差の平均が小地域推定モデルの活用により慣習的な推定法の $1/3$ 以下に抑えられたことを意味している。

小地域推定モデルに基づく推測法は、慣習的な推定法より取り扱いにくい。しかし、それでもこの資料の補足によって、わが国の小区別調査データ分析での小地域推定モデル活用の機会が広がることを期待したい。

謝 辞

横浜市立大学の土屋隆裕氏と統計数理研究所の朴堯星氏には、図 1 を作成する為に、廣瀬ら (2018) の論文で用いた立川市住民意識調査データを引き続き使用する許可をいただいた。また、千葉大学の佐野晋平氏、川久保友超氏、東京大学の菅澤翔之助氏には、今回の資料補足のきっかけとなる、図に関する重要な助言をいただいた。この場をお借りして御礼申し上げたい。

参 考 文 献

- 久保川達也 (2016). 推定における縮小法の展開—高次元解析と小地域推定—日本統計学会誌, 46, 43-67
 朴堯星・土屋隆裕 (2017). 多摩地域 住民意識調査—立川市郵送調査 (2016)—, 統計数理研究所調査研究リポート No.120
 廣瀬雅代・朴堯星・土屋隆裕 (2018). 小地域集計を活用したモデルに基づくアプローチによる防災に関する立川市町丁目別住民意識調査分析, 日本統計学会誌, 48, 49-70.
 Rao, J.N.K. and Molina, I. (2015). Small Area Estimation, 2nd Ed., Wiley, New York

制度的制約下におけるデータベース構造化、モデリング、モデル評価

Database structuring, modeling, and model evaluation under institutional constraints

データ科学研究系 山下智志 (Satoshi Yamashita)

1. はじめに

私個人の今世紀に入って以降の研究活動の特徴は、社会ニーズを吸上げ、データの取得、理論構築、モデリング、モデル評価、社会実装まで一貫通貫的に行い、各ステップの整合性を重視することにあった。特に、秘匿性データや統合データなどのデータベースの構築を長期間にわたって実行してきた。また、社会実装を想定しているため、法律、条約、規制、会計ルールなどのコンプライアンス面を考慮しながらモデリングを行うところに特徴がある。以下、研究活動実績を紹介したい。

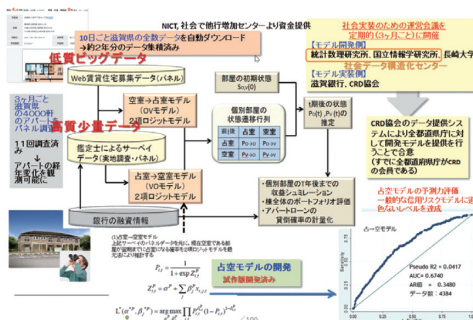
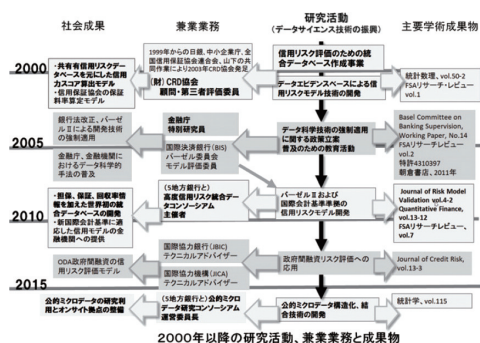
2. 統計モデル、データベース、政策の整合性と効率性

公開された研究業績は、いわゆる数理ファイナンスの分野や企業データ分析の方法論に関する著作が多い [1]。これは数理ファイナンス系方法論の研究が学術論文執筆と親和性が高く、書きやすいもしくは審査に通りやすいと言うことが原因である。実態としては方法論構築に費やしたエフォートは、データベースの作成、高度化やニーズの引き上げや政策化に費やしたエフォートよりも少ないのが実感である。

特に、金融庁特別研究員を 10 年近く併任していたため、全世界の銀行の共通ルールであるバーゼル II (新 BIS 規制) [2] や銀行法改正に関する統計学的根拠の整理を長年続けていた。また、一般的な信用リスクモデルを改良したカントリーリスク計量化モデルは、国際協力機構 (JICA) や国際協力銀行 (JBIC) における政府間金融 (いわゆる ODA) の意思決定に利用されている。

データベース構築に関わる活動としては、日本最大の信用リスクデータベースである CRD 協会を日銀、中小企業庁、全国信用保証協会連合会との連携により 2003 年に設立し、現在も顧問を行っている。そこでは日本の企業の過半数である 160 万社の財務データを蓄積しているだけでなく、そのまた地方銀行とのデータ共有コンソーシアムを立ち上げ、世界で初めて秘匿性の高い与信データの統合を統計数理研究所内で実現した。これにより銀行間で与信構造に差異があることが確認された。現在は総務省との共同で、公的マイクロデータの統合化に取り組んでいる。以下、現在運営管理中のデータベースである。

- 公的マイクロデータ (オンサイト拠点): 政府統計のマイクロデータ (法人統計データ、労働統計データ) を利用するため、統数研内に政府データの窓口である (財) 統計センターと直接つながるオンサイト拠点
- 国際マイクロデータ: アジア 6 カ国の政府マイクロデータ (主として労働統計)



- 高度信用リスク統合データベース： 地方銀行5行（滋賀、群馬、八十二、伊予、北陸）の貸出先法人の全数データの統合データベース。与信情報と回収情報があるのは世界唯一。お互いに秘匿情報のため、分析方法に工夫が必要。
- CRD 協会法人データベース：1996 年より収集を開始した中小企業財務データベース（法人としては2003 年から）。毎年約 160 万社、与信データベースとして日本最大である。
- 民間の購入データ（帝国データバンク、日経 Needs Company）2017 年から開始した政府企業統計ミクロデータと民間信用データとの結合実験として帝国データバンクの企業情報（約3 万件）がある。その他、民間の企業情報データとして日経 Needs と契約を続けている。

3. 賃貸住宅データを用いた、質の異なるデータベースの接続実験

様々なデータベースが増えるとともに、複数のデータベースを用いた統計モデルの構築方法が問題となる。特に質の異なるデータを利用するための方法論が必要とされている。具体的な研究としては、Web データを定時観測（楽天不動産、SUUMO から10日ごとにデータをダウンロード）することによって大規模データベースを作成し、賃貸住宅の入居化要因を分析する。一方、Web データの精度を補完するため、不動産鑑定士による現地パネル調査（賃貸住宅定点観測データ：約 4400 戸）を行った。それぞれのデータを統合することにより、より正確な賃貸住宅の収益予測を行う。2つの質の異なるデータベースを接続することにより、これまでリスク計量化モデルが考案されてこなかったアパートローン与信リスクについて、より正確な賃貸住宅の収益予測を行うことを目指している。現在、モデル精度は個人ローンのデフォルト予測と同等であり、実用化への具体的準備が進んでいる。

参 考 文 献

[1] Satoshi Yamashita, Toshinao Yoshiba, “ A Collateralized Loan’s Loss Under a Quadratic Gaussian Default Intensity Process”, Quantitative Finance, vol.13-12, p.1935-1946, 2013 年 6 月
 [2] Satoshi Yamashita, 他 “Studies on the Validation of Internal Rating Systems”, Basel Committee on Banking Supervision, Working Paper, No.14, 2005 年 5 月

極値分布と指数逆ガウス型分布に関するある一般化について

Generalization for the Extreme Value and Exponential Inverse Gaussian Distributions

データ科学研究系 金藤 浩司 (Koji Kanefuji)

要 旨

本報告では実数上で定義される二つの確率分布の一般化について紹介する。この一般化によって、一般化極値分布と一般化指数逆ガウス型分布を導出している。本報告とは異なり分布関数によって定義された一般化極値分布もある。そこでは母数の値により、三つのタイプの極値分布 (タイプ I: ガンベル型、タイプ II: フレッシュ型、タイプ III: ワイブル型) を表現している。

キーワード: 母変動係数; 寿命分布; 確率素分

1. はじめに

本稿では、タイプ I の極値分布の一般化を検討する。この確率分布の母歪度は零ではなく、非対称の性質を有している。同様の性質を有する実数上の分布として、指数逆ガウス型分布 (Kanefuji and Iwase; 1996) がある。この確率分布は、タイプ I の極値分布が利用される場面においてデータ解析上の別の候補となる一つの確率分布である。指数逆ガウス型分布に関しても同様な一般化を行う。極値分布は指数変換によるガンマ分布との関連性はよく知られている。同様の関連性は、逆ガウス型分布と指数逆ガウス型分布の間にも見られる。この関連性がタイプ I の極値分布と指数逆ガウス型分布の一般化の元となるアイデアである。さらに、3 母数ガンマ分布や 3 母数逆ガウス型分布は、本稿での一般化手法と直接的な関連性を有している。

実際のデータ解析において、タイプ I の極値分布が用いられる場合において、データから計算される標本歪度がその母歪度から大きく外れている場合が多々存在する。これらの変動に対応するため、本報告で定義するような一般化分布が有用となる。

また、Jenkinson(1955) により、一般化極値分布として、三つのタイプの極値分布を包含する分布が提案されている。

2. 二つの確率分布の一般化

タイプ I の極値分布 $EV(\mu, \sigma^2)$ は、次のように定義される。

$$\exp\left(\frac{X - \mu}{\sigma}\right) \sim Ga(1, 1^2),$$

ここで、 $0 < \sigma < \infty$, $-\infty < \mu < \infty$, $-\infty < X < \infty$ であり、記号 $Ga(m, c^2)$ は、母平均 m 、母変動係数 c であるガンマ分布を表している。

定義 1. タイプ I の一般化極値分布 $X \sim GEV(\mu, \sigma^2, \lambda)$ を以下で定義する。

$$\exp\left(\lambda \frac{X - \mu}{\sigma}\right) \sim Ga(1, \lambda^2),$$

ここで、 $0 < \sigma < \infty$, $-\infty < \mu < \infty$, $|\lambda| < \infty$, $-\infty < X < \infty$ であり、 λ は無次元量である。
 $GEV(\mu, \sigma^2, \lambda)$ に従う確率変数の確率素分 $f(x)dx$ は以下である。

$$f(x)dx = \frac{\left(\frac{1}{\lambda^2}\right)^{\frac{1}{\lambda^2} - \frac{1}{2}}}{\Gamma\left(\frac{1}{\lambda^2}\right)} \exp\left(\frac{1}{\lambda} \cdot \frac{x - \mu}{\sigma} - \frac{1}{2\lambda^2} \left\{2 \exp\left(\lambda \frac{x - \mu}{\sigma}\right)\right\}\right) \frac{dx}{\sigma},$$

ここで、 $-\infty < x < \infty$, $-\infty < \mu < \infty$, $0 < \sigma < \infty$, $|\lambda| < \infty$ であり、 $\Gamma(x)$ はガンマ関数である。

指数逆ガウス型分布 $EIG(\mu, \sigma^2)$ は以下の様に定義される。

$$\exp\left(\frac{X - \mu}{\sigma}\right) \sim IG(1, 1^2),$$

ここで、 $0 < \sigma < \infty$, $-\infty < \mu < \infty$, $-\infty < X < \infty$ であり、記号 $IG(m, c^2)$ は、母平均 m 、母変動係数 c である逆ガウス型分布を表している。

定義 2. 一般化指数逆ガウス型分布 $X \sim GEIG(\mu, \sigma^2, \lambda)$ を以下で定義する。

$$\exp\left(\lambda \frac{X - \mu}{\sigma}\right) \sim IG(1, \lambda^2),$$

ここで、 $0 < \sigma < \infty$, $-\infty < \mu < \infty$, $\lambda < \infty$, $-\infty < X < \infty$ であり、 λ は無次元量である。
 $GEIG(\mu, \sigma^2, \lambda)$ に従う確率変数の確率素分 $f(x)dx$ は以下である。

$$f(x)dx = \frac{1}{\sqrt{2\pi}} \left[\exp\left(\lambda \frac{x - \mu}{\sigma}\right) \right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\lambda^2} \left(2 \sinh\left(\lambda \frac{x - \mu}{2\sigma}\right)\right)^2\right\} \frac{dx}{\sigma},$$

ここで、 $-\infty < x < \infty$, $-\infty < \mu < \infty$, $0 < \sigma < \infty$, $|\lambda| < \infty$ であり、 $\sinh(x)$ は双曲線正弦関数である。

謝 辞

本稿の内容は、岩瀬晃盛広島大学名誉教授との共同研究の成果である。

参 考 文 献

- Jenkinson, A. F. (1955). The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements, *Quarterly Journal of the Royal Meteorological Society*, **81** (348), 158–171.
- Kanefuji K. and Iwase K. (1996). Exponential Inverse Gaussian Distribution, *Computational Statistics*, **11** (*), 315–326.

医師主導治験による医療機器開発の経験

An experience of developing medical device by investigator-initiated clinical trials

データ科学研究系 伊藤 陽一 (Yoichi M. Ito)

要 旨

本稿では、アカデミア発の医療機器開発において、規制当局のどのような話し合いの上で、臨床試験(治験)が立案され、承認に至ったということに関して記述を行う。医療機器に関する承認審査の規制が改正され、開発中断となっていた医療機器に関して、医師主導治験を実施することによって、承認申請を行うことができた。医師主導治験の実施にあたっては、主に有効性の判断基準、対象施設と症例数について、規制当局(PMDA)との間で話し合いが行われ、当初の計画が変更された。治験の結果として、対象者 20 名中 17 名で有効性が確認され、新規の医療機器として承認された。

1. 開発の経緯

平成 14 年(2002 年)に薬事法が改正され(平成 17 年(2005 年)施行)、医療機器に係る安全対策の見直しが行われた。本改正によって、リスクに応じたクラス分類制度が導入され、体内に留置して不具合が生じた場合に生命に危険を及ぼす可能性が高いものをクラス IV(高度管理医療機器)とし、医薬品と同様の Good Clinical Practice(GCP)基準に基づく治験が必要とされた。一方、平成 15 年(2003)に「医薬品の臨床試験の実施の基準に関する省令」(GCP)の改正により、従来、医師自ら実施することはできなかった薬事法上未承認・適応外の医薬品を用いた治験を医師主導治験として実施することができるようになった。このような流れを受けて、アカデミア発の医薬品・医療機器開発を推進するために、平成 19 年(2007)年に、文科省の橋渡し研究支援推進プログラムにより、北海道大学、札幌医科大学、旭川医科大学の 3 大学共同で、北海道臨床開発機構という組織が作られた。

北海道大学病院整形外科三浪明男教授は、平成 16~18 年度厚生労働省科学研究費補助金/免疫とアレルギー疾患予防・治療研究事業「関節リウマチ上肢人工関節に関する研究」によって、新規の人工関節の開発を行った。当該研究を通して、人工関節の有効性および安全性を確保するためには、インプラントの設置後のアライメント、コンポーネントの固定、軟部組織とのバランスが重要であり、また、正常手関節の運動解析によって、矢状面での伸展、屈曲より Dart thrower(投げ矢)面での伸展、屈曲運動がより生理的であり、手関節周囲軟部組織に加わるストレスが少ないことが確認された。これらの条件を満たすべく、摺動面が卵円形をした橈尺屈半拘束型のモデルが採用され、原材料に関してはシリコンではなく、他関節において最も多くの実績のあるチタン合金および超高分子量ポリエチレンを使用することとされた。本製品を製造したナカシマプロペラ株式会社(現ナカシマメディカル株式会社)が、2007 年 3 月に医療機器製造販売承認申請した。しかし、承認審査の過程において、審査側から、本開発品目に用いら

れている材料は人工膝関節や人工股関節において多くの使用実績があるものの適用部位が手関節の同等製品が存在しないことから、安全性および有効性を確認した臨床データが必要との見解が示された。しかし、ナカシマプロペラ株式会社は人を対象とした治験の経験を有していなかったことから、申請を取り下げ、開発は中断されていた。このような経緯から、北海道臨床開発機構が医師主導治験を支援することにより、開発が継続されることとなった。

2. 有効性指標の選択と有効性の判断基準

ICH-E9 統計ガイドラインでは、「主要変数の選択には、関連した研究領域で一般に認められている規範と標準を反映させるべきである。先行研究や公表論文で使用された実績のある、信頼性及び妥当性の確立した変数を使用することが薦められる」と述べられている。そこで、本治験では、Figgie らによって用いられた人工手関節に関する評価尺度 (Wrist Scoring System by Figgie) を用いることとした。Wrist Scoring System by Figgie では、除痛 (Pain relief) の程度について 50 点、関節を何度動かすことができるかという関節の可動性 (Range of motion) について 20 点、関節を様々な角度で止めることができるかどうかという関節の安定性 (Function) について 30 点の計 100 点満点で評価される。(Figgie et al. 1990).

当初、観察期間は人工手関節の埋植後 12 ヶ月、Wrist Scoring System by Figgie 70 点以上を有効性の判断基準として提案したが、PMDA との相談の結果、観察期間は 12 ヶ月から 18 ヶ月に延長され、実際に関節が動くことを評価するため、Range of Motion のスコアが 10 点以上という条件が追加された。

3. 対象施設と症例数

有効性指標に基づき、統計学的に有効性を証明できる必要症例数は 8 例と推定されたため、北海道大学病院において、目標症例数 10 例での治験実施を提案した。しかし、PMDA との相談の結果、どこの施設でも手術できることを確認するために、実施施設は 2 施設以上、また、有害事象の発現割合の推定の観点から、20 例以上の実施が要求され、2 施設、20 例で治験を実施することとなった。

4. 対象施設と症例数治験成績および考察

有効性については、主要評価項目である埋植後 18 か月時点における有効性の判断基準を達成した被験者の割合は 85 % (17 例)、安全性については、有害事象・不具合等が評価され、本品との因果関係が否定できない有害事象が 1 例、2 事象、不具合が 2 件、X 線学的評価で 18 か月に緩みありと判断された被験者が 4 例であった。本成績を持って、新規の医療機器として承認された。治験デザイン上の PMDA との相談のポイントとしては、有効性指標の明確化、安全性については長期の安全性を重視したものであった。レギュラトリーサイエンスの観点から、有効性と安全性のバランスが考慮された結果であると思われる。

参 考 文 献

- Figgie, M. P., Ranawat, C. S., Inglis, A. E., Sobel, M., & Figgie, H. E., 3rd. (1990). Trispherical total wrist arthroplasty in rheumatoid arthritis. *J Hand Surg Am*, 15(2), 217-223.
- 医薬品医療機器総合機構 (2016). 審査報告書. http://www.pmda.go.jp/medical_devices/2016/M20161117001/510462000_22800BZX00385_A100_2.pdf

経時データ解析と健康指標の長期推移

Longitudinal Data Analysis and Long Term Trends of Health Related Measures

データ科学研究系 船渡川 伊久子 (Ikuko Funatogawa)

1. 経時データ解析のための自己回帰線形混合効果モデル

生物統計学分野において、経時データ解析の手法は線形混合効果モデルの発表から、大きく発展しましたが、その多くは静学的内容です。線形混合効果モデルと時系列解析で用いられる自己回帰モデルを拡張した自己回帰線形混合効果モデルを提案しました (Funatogawa et al., 2007)。このモデルは、反応を直前の反応と固定効果および変量効果の共変量に回帰し、誤差の分散共分散構造を拡張しています。反応は漸近値に向かい推移し、変量効果は漸近値の個体間差を表し、非線形混合効果モデルでの monomolecular 成長曲線に対応します。周辺モデルとして表現することで、従来提案されていなかった儉約的で汎用性の高い分散共分散構造を導出し、特に欠測のあるデータで活用できます (Funatogawa et al., 2008b)。複数の反応変数への拡張も容易です (Funatogawa et al., 2008a)。反応を一時点前の反応に回帰すると、時間依存性共変量の影響を線形および非線形混合効果モデルとは異なる方法で扱い、過去の共変量の影響を取り入れるダイナミックなモデルとなります。投与量が時間依存性共変量で、観測間隔が一定でない場合に、状態空間表現を用いて最尤推定を行う方法を提案しました (Funatogawa and Funatogawa, 2012a)。また、投与量が時間依存性共変量で、観測された反応の値によって変更される場合、モデルが正しければ最尤推定値に偏りはありません (Funatogawa and Funatogawa, 2012b)。モデルのメカニスティックな側面に着目し、和書 (船渡川・船渡川, 2015) および英文書籍 (Funatogawa and Funatogawa, in press) を出版しました。一方、経済学の分野では、動学的パネルデータ分析と呼ばれ、反応を以前の反応に回帰するモデルが個体間差を考慮する形で観察研究に用いられています。

2. 健康指標の長期推移

喫煙は、喫煙開始から死亡までが非常に長いことや出生コホート間の喫煙習慣の複雑な違いが、その影響の大きさを分かりにくくしています。喫煙開始年齢は喫煙期間も表す重要な指標ですが、各国の長期推移の報告は限られます。喫煙開始割合、喫煙率、肺癌死亡率の加齢変化の出生年による長期推移を、日本および英国についてそれぞれ WHO Bulletin と BMJ Open に発表しました (Funatogawa et al., 2013; Funatogawa et al., 2012)。喫煙と肺癌の研究は、関心の高い分野で、しばしば医学の主要ジャーナルに掲載されますが、喫煙の早期中止でリスクが減少するという主張が目立ち、若年期の早期喫煙開始の危険性が十分に伝えられていません。背景には、統計学や疫学の方法論上の問題があると考えています。米国や英国女性の喫煙開始、喫煙率、肺癌死亡率の出生年による変化と関連した内容を、N Engl J Med (レター) や Lancet (レター) で報告しました (Funatogawa, 2018; Funatogawa, 2013)。

Body Mass Index (BMI) は重要な健康関連指標ですが、数十年単位の経年的加齢変化を計量的に評価した報告はなく、横断調査から出生コホートを考慮せずに求めた加齢変化を用いて

いました。そこで、日本の代表的な繰り返し横断調査で、無作為抽出で行われ、60 年以上の記録が存在する国民健康・栄養調査のデータを利用し、出生コホートを考慮した BMI の加齢変化を、0~25 歳女性の結果を BMJ, 20~60 歳代成人男女の結果を Int J Epidemiol に発表しました (Funatogawa et al., 2008c; Funatogawa et al., 2009)。日本人女性は、より最近の出生コホートほど、子供の頃はより過体重ですが、成人するとより痩せていること、横断調査と出生コホート別では加齢変化パターン自体が異なること等を示しました。喫煙や肥満に関する日本語での解説を公表しています (船渡川, 2014a; 船渡川, 2014b; 船渡川・船渡川, 2015)。

参 考 文 献

- Funatogawa, I. (2013). The first generation in which many women began smoking, *Lancet*, **381**(9876), 1455.
- Funatogawa, I. (2018). Incidence of lung cancer among young women, *The New England Journal of Medicine*, **379**(10), 988.
- Funatogawa, I. and Funatogawa, T. (2012a). An autoregressive linear mixed effects model for the analysis of unequally spaced longitudinal data with dose-modification, *Statistics in Medicine*, **31**(6), 589–599.
- Funatogawa, I. and Funatogawa, T. (2012b). Dose-response relationship from longitudinal data with response-dependent dose modification using likelihood methods, *Biometrical Journal*, **54**(4), 494–506.
- Funatogawa, I. and Funatogawa, T. (in press). *Longitudinal Data Analysis: Autoregressive Linear Mixed Effects Models*, Springer, Singapore.
- Funatogawa, I., Funatogawa, T., Nakao, M., Karita, K. and Yano, E. (2009). Changes in body mass index by birth cohort in Japanese adults: results from the National Nutrition Survey of Japan 1956–2005, *International Journal of Epidemiology*, **38**(1), 83–92.
- Funatogawa, I., Funatogawa, T. and Ohashi, Y. (2007). An autoregressive linear mixed effects model for the analysis of longitudinal data which show profiles approaching asymptotes. *Statistics in Medicine*, **26**(2113–30), 2113–2130.
- Funatogawa, I., Funatogawa, T. and Ohashi, Y. (2008a). A bivariate autoregressive linear mixed effects model for the analysis of longitudinal data, *Statistics in Medicine*, **27**(6367–78), 6367–6378.
- Funatogawa, T., Funatogawa, I. and Takeuchi, M. (2008b). An autoregressive linear mixed effects model for the analysis of longitudinal data which include dropouts and show profiles approaching asymptotes. *Statistics in Medicine*, **27**, 6351–6366.
- Funatogawa, I., Funatogawa, T. and Yano, E. (2008c). Do overweight children necessarily make overweight adults? Repeated cross sectional annual nationwide survey of Japanese girls and women over nearly six decades, *British Medical Journal*, **337**(a802).
- Funatogawa, I., Funatogawa, T. and Yano, E. (2012). Impacts of early smoking initiation: long-term trends of lung cancer mortality and smoking initiation from repeated cross-sectional surveys in Great Britain, *BMJ Open*, **2**(5).
- Funatogawa, I., Funatogawa, T. and Yano, E. (2013). Trends in smoking and lung cancer mortality in Japan, by birth cohort, 1949–2010, *Bulletin of the World Health Organization*, **91**(5), 332–340.
- 船渡川伊久子 (2014a). 肺の健康とタバコ 近年の日本における肺癌発生の推移と関連因子, *健康管理*, **61**(8), 19–25.
- 船渡川伊久子 (2014b). 思春期の栄養と運動を考える 小児・思春期の発育についての疫学的検討, *思春期学*, **32**(1), 145–149.
- 船渡川伊久子, 船渡川隆 (2015). 『経時データ解析』, 朝倉書店, 東京.

先端医学研究の発展を支える統計数理とデータサイエンス

Statistical Mathematics and Data Science for the Developments of Advanced Medical Researches

データ科学研究系 野間久史 (Hisashi Noma)

1. はじめに

世界規模で進む社会の高齢化により、医療費・医療資源の効率的な配分は、高水準の医療・福祉を維持するべく、先進諸国において重要な問題となっている。特に日本では、既に 65 歳以上の高齢者の割合が人口の 4 分の 1 を超えており、WHO が定める「超高齢社会」となっている。年間の医療費も 40 兆円を超えており、深刻な状況にある。このような中で、将来に向けて高水準の医療・福祉の持続、および、医療技術のさらなる発展を図るためには、その基盤となる信頼できる科学的根拠が不可欠であり、そのために、統計学・データサイエンスは極めて重要な役割を果たしている。統計数理研究所においても、2018 年 4 月に、医療・健康科学領域におけるデータサイエンスの最先端の研究および人材育成の拠点として、医療健康データ科学研究センターが設立されている。本稿では、著者らが取り組む研究プロジェクトの一部を紹介する

2. ネットワークメタアナリシス

医療費・医療資源の効率的な配分のために、医療政策・診療ガイドラインの策定において重要になるのが、既に多く存在する医薬品・医療技術のいずれが最も高い有効性・安全性を持ち、経済的であるか、ということである。しかしながら、多くの治療法をすべて比較して、それらの優劣を比べるための十分な検出力を達成する臨床試験を行うためには、一般的に、数万人～数十万人以上の規模の試験を行う必要があり、現実的には不可能である。この問題を解決するために、近年の医療統計学の研究から、ネットワークメタアナリシスという新しい方法論が開発された。ネットワークメタアナリシスは、過去に行われた臨床試験の結果を統合し、対象となる治療法間の比較評価を行ったエビデンスを提供してくれる新しい方法として、近年、先端的な臨床医学・医療技術評価で急速に普及している。本邦からも、優れた先進的な研究成果がいくつか報告されており、例えば、著者の野間も参加した、双極性障害の 17 種類の薬物療法のネットワークメタアナリシス (Miura et al., 2014) などがある。

ネットワークメタアナリシスにおいて、統計科学の方法論は、そのエビデンスの科学的な妥当性および精確性の根幹を支えるための中心的な役割を果たしている。ネットワークメタアナリシスは、複数の異なる情報源から得られるエビデンスを統合するため、その異質性を適切に考慮した複雑な構造を持つマルチレベルモデルを用いる必要がある。これらの統計的推測手法には、最尤法やベイズ法を基礎とした方法が一般的に用いられるが、Noma et al. (2018) などによって、一般的なネットワークメタアナリシスが行われる条件下におけるこれらの推測手法の不正確性が明らかにされ、それを解決するための方法論の開発・整備が活発に進められてい

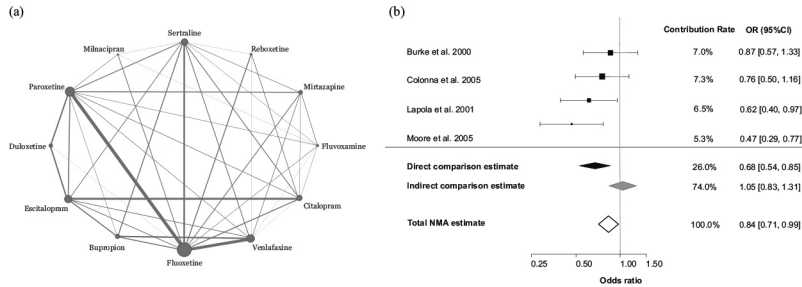


図 1. (a) 新世代抗うつ薬の臨床試験のエビデンスのネットワーク。(b) Escitalopram vs. Citalopram の比較における複合尤度法による推定量の分解の結果 (Noma et al. (2017) より転載)。

る。Noma et al. (2018) では、リサンプリング法を用いて、この不正確性の問題を解決した推定・検定の手法を提案している。また、特に、近年、試験間の異質性を考慮した治療効果の評価方法として、将来の集団において期待される治療効果の「予測」も重要な問題となっているが、Nagashima et al. (2018) は、現在、国際的なスタンダードとなっている予測区間の構成方法の不正確性を明らかにし、基礎的な 2 治療の比較試験の統合解析において、これより大幅に優れた性能を持つ正確な予測区間の構成方法を提案した。この正確な予測手法のネットワークメタアナリシスへの拡張に関する研究も現在進行中である。加えて、治療効果の評価におけるバイアスの評価において重要となる、ネットワーク上でのエビデンスの不整合性を評価するための方法の開発研究も活発に行われており、Noma et al. (2017) は、複合尤度法を用いた新しい不整合性の評価方法を提案している。Oxford 大学、京都大学の研究グループと協同し、精神医学領域のネットワークメタアナリシスで史上最大規模の研究であった新世代抗うつ薬の臨床試験のネットワークにこれを応用し、興味深いスポンサーシップバイアスの可能性を示唆する結果を報告している (図 1)。

参 考 文 献

- Miura, T., Noma, H., Furukawa, T. A., et al. (2014). Comparative efficacy and tolerability of pharmacological treatments in the maintenance treatment of bipolar disorder: a systematic review and network meta-analysis. *Lancet Psychiatry* 1, 351-359.
- Nagashima, K., Noma, H., and Furukawa, T. A. (2018). Prediction intervals for random-effects meta-analysis: A confidence distribution approach. *Statistical Methods in Medical Research*, doi: 10.1177/0962280218773520.
- Noma, H., Nagashima, K., Maruo, K., Goshu, M., and Furukawa, T. A. (2018). Bartlett-type corrections and bootstrap adjustments of likelihood-based inference methods for network meta-analysis. *Statistics in Medicine* 37, 1178-1190.
- Noma, H., Tanaka, S., Matsui, S., Cipriani, A., and Furukawa, T. A. (2017). Quantifying indirect evidence in network meta-analysis. *Statistics in Medicine* 36, 917-927.

集約的シンボリックデータ解析

Aggregated Symbolic Data Analysis

データ科学研究系 清水信夫 (Nobuo Shimizu)

近年、様々な分野において、Web システムを用いたデータの収集が多用されており、各分野における活動の詳細なデータが計算機上に連続的に蓄積されるようになってきている。それらのデータは連続変数とカテゴリ変数が混在した多次元データであることが多く、データの個体数についても非常に大きな場合が多々存在する。

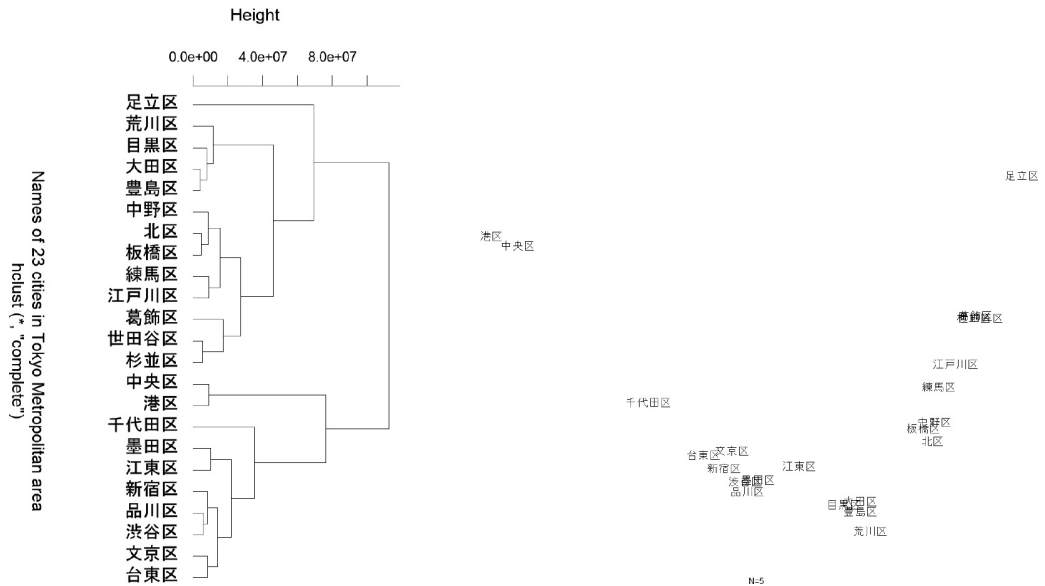
このようなデータは、全体像を見たり詳しい解析を行ったりするための、一般的なデータ管理・処理ソフトウェアによる取り扱いが困難なほどの巨大かつ複雑なデータ、いわゆる“ビッグデータ”の代表例と考えられ、従来多用されていた個体データに着目する方法以外の解析手法の開発が求められる。ただし、そのようなときは、個体データが意味のある自然な比較的小数のグループに分かれる場合が少なからず存在する。したがって、個体データそのものではなく、個体がまとめられたグループに関心を向けた手法が必要である。その方法の一つとして、Diday(1988)によりシンボリックデータ (Symbolic Data, SD) という概念が提案され、データの中で個体がまとめられたグループを SD として解析するシンボリックデータ解析 (Symbolic Data Analysis, SDA) が提唱されている。

SDA においては、データとして各連続変数ごとに 1 つの値ではなく、ある値を中心としてばらつきをもつデータ (区間データや分布値データ) などで表されるものが考えられ、それらを SD と考えた場合の解析として従来の各種多変量解析手法を拡張する研究が、Bock and Diday(2000), Billard and Diday(2005), Diday and Noirhomme-Fraiture(2008) などにまとめられている。それら以外にも、SD に関するクラスタリングに関しては Verde(2004) や Irpino and Verde(2006) など、多次元尺度構成法に関しては Denœux and Masson(2000) や Groenen et al.(2006) などの研究がある。これまでの SDA においては、データは最初から区間や集合のような形で与えられている場合が多く、そこではグループ内の複数の変数間の関係は無視される。例えば、2 つの連続変数間の相関関係は考慮されない。

しかしながら、現代のビッグデータにおいては、元の個体データは保持されている。個体数や変数が極めて多いデータの場合は移動や計算に困難を伴うが、どうしても必要ならばグループに関するいかなる記述統計量も計算することは可能である。そこで、グループにおける多次元データの情報を可能な限り簡潔な形で持つために、複数の記述統計量を考えることにし、それを集約的シンボリックデータ (Aggregated symbolic data, ASD) と呼ぶこととする。

ASD は、グループ内の個体データのそれぞれの変数および複数の変数の組み合わせに関し、情報の脱落を可能な限り抑えつつ保持すべき値を可能な限り少なくして取り扱いを容易にするために、2 次までのモーメントについて求められる統計量の集合として表される。ASD に含まれる統計量の例としては連続変数における平均および分散共分散行列、カテゴリ変数における分割表などがある。これらを用いて、ASD 間の非類似度を尤度比検定統計量やカイ 2 乗統計量などを用いて定義し、それらを用いてクラスタリングや多次元尺度構成法を行う方法を提案した。この方法を東京都区部の不動産情報データに適用し、23 の特別区をそれぞれグループとして考え、データの各変数から各グループ間の非類似度を ASD を用いて求めた上で階層的ク

ラスティングおよび多次元尺度構成法を行った例を図 1 に示す。



(a) 階層的クラスタリング

(b) 多次元尺度構成法

図 1. 集約的シンボリックデータを用いた不動産情報データの解析例
(清水・中野・山本 (2018))

参 考 文 献

- Billard, L. and Diday, E. (2006). *Symbolic data analysis: Conceptual statistics and data mining*, John Wiley & Sons Ltd, Chichester, UK.
- Bock, H.-H. and Diday, E. (2000). *Analysis of symbolic data: exploratory methods for extracting statistical information from complex data*, Springer-Verlag, Berlin.
- Deneoux, T., Masson, M. (2000). Multidimensional scaling of interval-valued dissimilarity data, *Pattern Recognition Letters*, **21**, (1), 83–92.
- Diday, E. (1988). The symbolic approach in clustering and related methods of data analysis, *Classification and Related Methods of Data Analysis*, 673–684.
- Diday, E. and Noirhomme-Fraiture, M. (2008). *Symbolic Data Analysis and the SODAS Software*, John Wiley & Sons Ltd, Chichester, UK.
- Groenen, P. J. F., Winsberg, S., Rodriguez, O. and Diday, E. (2006). I-Scal: Multidimensional scaling of interval dissimilarities, *Computational Statistics and Data Analysis*, Elsevier, **51**, (1), 360–378.
- Irpino, A. and Verde, R. (2006). A new Wasserstein based distance for the hierarchical clustering of histogram symbolic data, *Data Science and Classification*, Springer, Berlin, 185–192.
- Verde, R. (2004). Clustering methods in symbolic data analysis, *Data Science and Classification*, Springer, Berlin, 299–317.
- 清水信夫, 中野純司, 山本由和 (2018). 集約的シンボリックデータのカイ 2 乗統計量を用いた非類似度とその不動産情報データへの適用, *統計数理*, **66** (2), 279–294.

ツイートを利用した都市気温の解析 —都市インテリジェンス向上をめざして— Urban Temperature Analysis Using Participatory Sensing Data “Tweets” —Toward Improvement of Urban Intelligence—

モデリング研究系 松井 知子 (Tomoko Matsui)

1. はじめに

近年、地球温暖化によって異常気象リスクが顕在化している。特に都市での熱波リスクは深刻である。このリスクに対処するには各地点、各時刻での気温状況を正確に把握することが重要である。気象庁などは都市内にいくつかの観測拠点を設けて、定期的に気温を測定している。しかし、観測拠点数や時間間隔はスパースである。そこで本研究では、都市の各所で絶え間なくつぶやかれる「暑い」「だるい」などの暑さに関係すると想定するツイートデータ（ヒートツイート）を利用して、上記のスパースな気温観測データを空間的、時間的に補間する統計・機械学習の方法について研究を行った。将来的には本方法を熱波対策に役立て、都市インテリジェンス向上を目指したいと考えている。

2. ヒートツイートと気温観測データとの関係

気温補間においてヒートツイートが有用できることを確かめるために、コンピュータモデルを用いて各観測拠点におけるヒートツイートと気温観測データとの関係を調べた。ここでヒートツイートは、一般化加法モデルとロジスティックモデルにより、主要駅／公園／河川からの距離、日中・夜間人口密度、土地利用の種類などを考慮した確率的な強度で表した。表 1 に気温、気温変化のそれぞれとヒートツイートの lower tail（気温は低い／気温変化は小さい、ヒートツイートは少ない時に相当）と upper tail（気温は高い／気温変化は大きい、ヒートツイートは多い時に相当）における依存性を示す。一般に気温が高い時にヒートツイートは多くなると考えられてきたが、熱波で問題となる upper tail では特に気温変化がヒートツイートに関係していることを新たに見出した。

表 1. 気温／気温変化とヒートツイートの依存性

	Lower tail	Upper tail
気温	0.40	0.00
気温変化	0.20	0.23

3. S-BLUE 法による気温補間

S-BLUE (spatial best linear unbiased estimation) 法 (Peters, G. W., Nevat, I. and Matsui, T. (2015)) の枠組みを用いて、各観測拠点において正確ではあるがスパースに観測される気温データとヒートツイートを組み合わせ、都市気温の時空間モデルを表す線形汎関数を構成す

る。S-BLUE 法ではガウス過程を内包しており、連続／離散量の異種混合データの複雑な非線形な関係性をうまく表現することができる。また、平均二乗誤差最小化に基づいてパラメータを効率的に推定することができる。図 1 に、都内について上記モデルを用いて、朝 6 時、昼 12 時、夕方 18 時の都内の気温を補間した結果を示す。ヒートツイートを利用することにより、昼 12 時には東京の中心部（山手線内）がヒートアイランド現象によって気温が上昇する様子を捉えることができた。

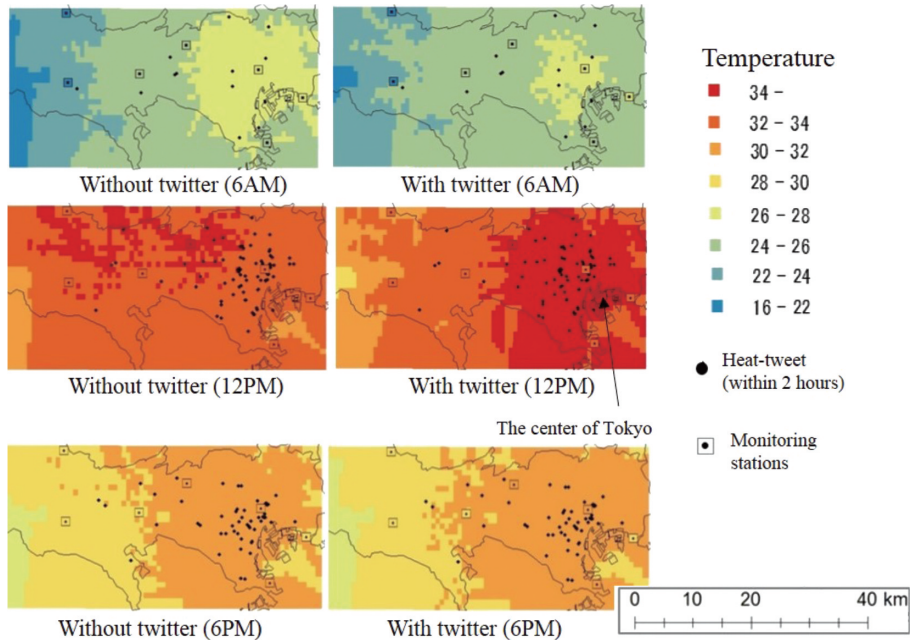


図 1. 朝 6 時、昼 12 時、夕方 18 時の都内の気温を補間
(左側：ヒートツイートを利用しない場合、右側：ヒートツイートを利用する場合)

4. おわりに

本研究は統計的機械学習研究センター「都市インテリジェンス」プロジェクトにおける、村上大輔助教、Gareth W. Peters 教授 (Heriot-Watt 大)、山形与志樹博士 (国立環境研究所) との共同研究の成果 (Murakami, D., Peters, G. W., Yamagata, Y. and Matsui, T. (2016)) である。今後もこのプロジェクトを基盤として、都市レジリエンス向上を目標とした、統計数理／機械学習に基づく技術とその理論の研究開発を進めていきたい。

参 考 文 献

- Murakami, D., Peters, G. W., Yamagata, Y. and Matsui, T. (2016). Participatory Sensing Data Tweets for Micro-Urban Real-Time Resiliency Monitoring and Risk Management, *IEEE Access*, 4, 347-372, doi:10.1109/ACCESS.2016.2516918.
- Peters, G. W., Nevat, I. and Matsui, T. (2015). How to Utilize Sensor Network Data to Efficiently Perform Model Calibration and Spatial Field Reconstruction, *Modern Methodology and Applications in Spatial-Temporal Modeling* (eds. G. W. Peters and T. Matsui), Chap. 2, 25-62, Springer.

機械学習による新物質の発掘

Machine learning for accelerated materials discovery

データ科学研究系 吉田 亮 (Ryo Yoshida)

キーワード：マテリアルズインフォマティクス，機械学習，バイズ推論，転移学習

1. マテリアルズインフォマティクス

MI の問題の多くは，順問題と逆問題の形式に帰着する．順問題の目的は，系の入力 S に対する出力 Y の予測である．物性予測の文脈では，入力は物質（分子，組成，結晶等），出力は物性値（エネルギー，電子状態等）に相当する．これまでの材料研究では，第一原理計算や分子動力学計算等の理論計算が順方向の予測を担ってきた．このタスクをデータ科学のモデルに代替させることが，MI の中心的課題のひとつである．これに対し，逆問題では文字通り逆方向の予測を行う．すなわち，出力 Y の値（例えば目標物性）を設定した上で，それを達成する入力 S の状態（構造）を予測する．データ科学の観点からみると，これらの計算は，物質構造の“表現・学習・生成”を行うことに相当する．記述子と呼ばれる特徴ベクトルを通して物質の構造を“表現”し，データのパターンから構造から物性の数学的写像を“学習”する．さらに，計算機を用いて所望の物性値を有する物質を“生成”し，有望な候補物質を炙り出す．対象となる入力 S は，分子，組成，結晶，混合物，プロセス，合成経路等，問題に応じて多様な形式をとりうる．

2. 機械学習によるハイスループットスクリーニング

構造と物性の関係を表す実験や理論計算のデータから，物質 S の物性 Y の予測モデル $f(S)$ を導くことが目的である．記述子と呼ばれる特徴ベクトルを通して物質の構造を“表現”し，データのパターンから構造から物性の数学的写像を“学習”する．記述子は MI における最も基本的な要素技術である．入力 S の形式が多様であることから，対象領域ごとに独自に研究が展開している．膨大な数の候補物質のライブラリを作製した上で，訓練済みモデルを用いてスクリーニング実験を実施する．実験や理論計算に比べて機械学習のモデルは圧倒的に計算コストが低いいため，膨大な数の候補物質を対象とする物性評価を行うことができる．

3. 転移学習

機械学習の他の応用領域に比べて，材料研究のデータ数は圧倒的に少ない．データ科学が本格的に導入されて間もないこともあり，データベースの整備は発展途上の段階にある．とりわけ，研究対象が最先端に近づくにつれて，スモールデータの傾向はより顕著になる．スモールデータに対する解決策として，転移学習と呼ばれるアプローチが有望視されている．転移学習では，あるタスクの訓練済みモデルを別のタスクに再利用する．我々は，XenonPy.MDL という訓練済みモデルライブラリを開発している (Yamada et al. (2019))．低分子化合物，高分子，無機結晶等，様々な物質に対する >105 の物性推算モデルが収録されている．図 1 は，ニューラルネットワークの転移学習に基づくポリマー定圧熱容量 (C_p) の予測を例示したものであ

る。第一原理計算で低分子化合物の化学構造と定容熱容量 (C_v) の関係を表す 133,805 個のデータを取得し (Ramakrishnan et al. (2014)), ソースモデルを導いた。ソースモデルの部分ネットワークを用いて特徴量の計算を行い、高分子材料データベース PoLyInfo (Otsuka et al. (2011)) に登録されている 58 個の C_p のデータを用いて予測モデルを構築した。ソースタスクの学習過程で C_v と C_p に共通する縮約特徴量を獲得し、これを用いることでたった 58 個のデータからポリマー C_p の予測モデルを導くことに成功した。

4. ベイズ推論による物質構造の設計

我々は、ベイズ推論や機械学習を方法論の基軸とし、新物質の創製を目的に設計と合成を対象とする機械学習の手法とソフトウェアを開発してきた (Ikebata et al. (2017)). 実験やシミュレーションから得られるデータを用いて、物質の構造から物性の順方向の予測モデルを構築する。これに条件付き確率のベイズ則を適用し、物性から構造の逆方向のモデルを導き、このモデルから仮説物質を発生させることで、所望の物性を有する埋蔵物質を炙り出す。確率的言語モデルに基づく構造生成器や機械学習の様々な技術を結集させて構築した確率推論のアルゴリズムである。現在、この手法を用いて様々な材料系を対象に実証研究を進めている (Wu et al. (2019) など)。機械学習で物質構造の設計図が描き、大量の埋蔵物質を発掘する。これが本研究のグランドチャレンジである。

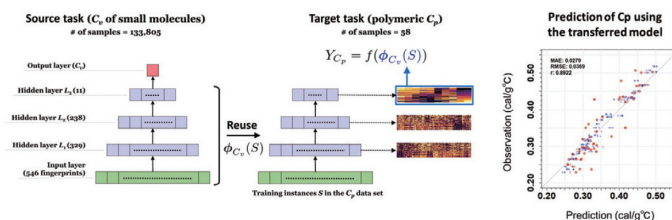


図 1. 転移学習によるポリマー定圧熱容量 (C_p) の予測

参 考 文 献

- Ramakrishnan, R., Dral, P. O., Rupp, M. and von Lilienfeld, O. A. (2014). Quantum chemistry structures and properties of 134 kilo molecules, *Scientific Data*, 1, 140022.
- Otsuka, S., Kuwajima, I., Hosoya, J., Xu, Y. and Yamazaki, M. (2011). PolyInfo: Polymer database for Polymeric Materials Design, 2011 International Conference on Emerging Intelligent Data and Web Technologies, 22-29.
- Yamada, H., Liu, C., Wu, S., Koyama, Y., Ju, S., Shiomi, J., Morikawa, J. and Yoshida, R. (2019). Transfer learning: accelerated materials discovery with small data.
- Ikebata, H., Hongo, K., Isomura, T., Maezono, R. and Yoshida, R. (2017). Bayesian molecular design with a chemical language model, *Journal of Computer-Aided Molecular Design*, 31(4), 379-391.
- Wu, S., Kondo, Y., Kakimoto, M., Yang, B., Yamada, H., Kuwajima, I., Lambard, G., Hongo, K., Xu, Y., Shiomi, J., Schick, C., Morikawa, J. and Yoshida, R. (2019). Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm.

方向統計学と動物移動生態学

Circular statistics and movement ecology

データ科学研究系 島谷 健一郎 (Kenichiro Shimatani)

動物の移動軌跡データ

動物装着型 GPS を用いる研究は、水中・空中・陸上を問わず、様々な動物について世界各地で盛んに行われている。GPS データには、動物がいつどこにいたのかという情報に加え、各時点における速さと方向を求めることで、どのように動いていたかという情報も含まれる。それは、いつどこで何をしていたかという推定につながる。

円周自己回帰モデル

速さに特徴的なパターンが見られなくても、方向には見られる場合がある。そこでは、方向とその変化に関する時系列モデルが基本的なデータ解析の道具となる。そこで基本となるのは、各時点で方向はランダムに変動するランダムウォークである。しかし、実際の動物の動きで多く見られるのは、巣や餌場などへ向かっていく動きや、ある場所の周辺をうろついているようなパターンである。こうした移動軌跡を表現するには、360 度で元の 0 度に戻るという方向データの特性を鑑みた、円周上の統計モデルが適している。そこで、Shimatani et al. (2012) では、Kato (2010) の円周自己回帰モデル (図 1a) の適用を提唱した。

この自己回帰モデルでは、目的変数は特定の方向 α に集中する傾向を示す。シミュレーションで順に方向を生成し、速さを一定にして軌跡データを作ると、ある方向に向う軌跡を描く。ランダムな変動には、von-Mises 分布などの円周上の確率分布を用いる。確率密度関数のグラフが尖がっていてランダムな変動が小さいと、目標方向へ向かって進む軌跡が得られる (図 1b ㉑)。変動の大きい確率分布を用いると、軌跡は震えながらゆっくり目標方向に進んでいく (図 1b ㉒)。言い換えると、確率分布の尖り具合を決めるパラメータは、行くべき方向へ向かう調整力という解釈を伴う。一方、自己回帰モデルの回帰係数 w の値が 1 に近いと、目標方向への集中が弱いため方向修正する効果が弱く、何かの拍子にあらぬ方向に向かうとしばらく進んでから元の方向に戻る (図 1b ㉓)。つまり、回帰係数 w は、その方向に向かおうという動物の意志の強さという解釈を伴う。なお、 $w = 1$ とするとランダムウォーク (動物行動では correlated random walk と呼ばれる) となる。

実際のデータが与えられたときは、最尤法などでパラメータを最適化することで、動物の意志の強さ、目標方向、方向の調整力に関する定量的評価を与えられる。また、 $w = 1$ としたランダムウォークと AIC などの情報量規準で相対評価することにより、動物が進みたい方向を持っていたのかどうかを検証できる。

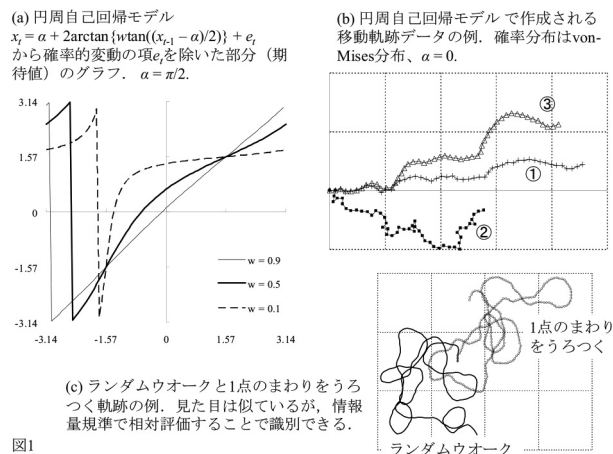
基本統計モデルが満たすべき条件

一般に、ある現象についてのある型のデータに関する統計モデルは、次の 3 つの条件を満た

しているべきである (Shimatani et al. 2012)。

1. 未知パラメータにその現象に関する定量的解釈が伴う。
2. 数学的あるいは数値計算で扱いやすい数式からなる。
3. 広範な現象に適用でき、状況に応じ適宜拡張できる。

上述したように、Kato (2010) の円周自己回帰モデルを動物軌跡に適用すると動物の行動に関する解釈を伴っており、1 を満たす。2 もよく知られた関数で尤度が書けるなど、数学的扱いも容易である。説明変数を、1 時点だけでなく 2 時点や 3 時点前まで含め、それらの線形和 (方向なので 1 度複素平面の単位ベクトルに直して線形和を取り、再び argument で方向に戻す) を使う式に拡張できる。すると、左に回り始めたらしばらく左に回るなどの、滑らかな軌跡を描く。さらに、集中させる方向を、特定の地点と各時点における地点を結ぶ方向にする (α は時間とともに変わる) と、特定地点の周辺をうろつきまわる軌跡を描ける (図 1c)。ここでは回帰係数 w は、その地点に留まりたい意志の強さといった解釈ができる。さらに、3 次元空間における軌跡へも、若干の修正で可能である (島谷 2015) など、Kato (2010) の円周自己回帰モデルは豊かな拡張性を有する。すなわち、3 も満たし、動物行動軌跡の方向に関する基本統計モデルとしての役割を担える。このように、方向統計学は動物の移動生態学 (movement ecology) に関する基本モデルを与え、様々な発展をもたらす可能性を有する。逆に、動物行動に関するデータは、方向統計学の発展を促す。両者は表裏一体となって相性よく発展していくと期待できる。



参 考 文 献

- Kato, S. 2010. A Markov process for circular data. *Journal of the Royal Statistical Society B* 72:655-672.
- Shimatani I.K., Yoda, K., Katsumata, N., Sato, K. (2012). Toward the Quantification of a Conceptual Framework for Movement Ecology Using Circular Statistical Modeling, *PLoS One*, doi:10.1371/journal.pone.0050309.
- 島谷健一郎 (2015) 3 次元軌跡データの基本モデルとその限界. *数理解析研究所講究録* 1940: 95-100.

機械学習に基づく高熱伝導率ポリマーの設計事例

A case study on high thermal conductivity polymer design based on machine learning

データ科学研究系 ウ ステファン (Stephen Wu)

キーワード：機械学習，マテリアルズ・インフォマティクス，ポリマー

1. はじめに

新しい材料を開発する従来の方法は，専門家の知識を使用して特定の物性要件を持つ少数の候補者を提案することです。このような方法は，膨大な化学空間（およそ 10^{60} 候補者 (Bohacek et al., 1996)）のために非効率的であることが知られている。人工知能技術の急速な発展のおかげで，機械学習に基づくより効率的な候補生成方法は，化学フロンティアの拡大ペースを加速すると期待されている。しかし，機械学習はどれほど強力であっても，単独で材料開発の挑戦を克服することはできません。計算分子設計法の開発の長い歴史にもかかわらず，合成までたどり着いたアルゴリズム的に設計された新規分子の例は限られている。これは，合成の難しさ，専門家の知識と学習された機械の知能の不一致，実用的なアプリケーションで厳しい要件などによるものであると推測しています。本研究では，Bayesian 分子設計という機械学習技術によって設計された高熱伝導率の新規ポリマーの発見を成功した。

2. 手法

ポリマーの熱伝導率 (λ) は実用的重要性のために多くの注目を集めている (Anderson, 1966)。しかし，設計のためにポリマーのメカニズムを完全には理解していない。本研究はこれまでに提案されているいくつかの計算設計法 (Zhong et al., 2001) と異なり，iqspr と呼ばれるベイジアン分子設計アルゴリズムに基づくデータ駆動アプローチを使用している (Ikebata et al., 2017)。データのソースは PolyInfo データベース (Otsuka et al., 2011) である。

R パッケージの iqspr は，ユーザーが指定した材料特性が与えられた探索空間を表す事後分布に対してサンプリングを実行します。ベイズの定理により，事後分布は尤度と事前分布との積として計算することができる。事前分布は，材料の物性に何の制約もなく化合物らしいものを表し，尤度は材料が与えられた場合に所望の材料物性を得る確率である。この問題設定の大きな課題は，信頼性の高い尤度モデルを訓練するためのデータが不十分であり，事後分布サンプリング・ステップ中に欠陥が生じることです。小さなデータの問題を克服するために，iqspr で元の目標の λ ではなく， λ に関連する物性を選択する。図 1 は，PoLyInfo におけるいくつかの関連するポリマー物性のデータを示す。最終的に，高い λ の元のターゲットに代えて，高いガラス転移温度 (T_g) および熔融温度 (T_m) を目標とした。その後，iqspr から生成された提案候補は転移学習と専門知識を備えたポストスクリーニングを経た。我々の問題における転移学習は， λ に相関し，十分に大きなデータセットを有する物性のモデルを構築して，これらの

モデルの一部は、 λ のモデル構築を助けるために利用される。そして、専門知識による産業応用の観点から設計の要素を追加する。

3. 結果

iqspr パッケージでは、 T_g が > 250 °C、 T_m が > 350 °C になるようにターゲット領域を設定する。回帰モデルにはデフォルト値が使用され、 T_g 値と T_m 値が記録されたすべてのホモポリマーを含む訓練データが使用されます。ホモポリマーらしい分子を生成するために使用された従来のモデルでは、PoLyInfo データベースで利用可能な 14,424 ホモポリマーのすべてを使用した。その後、生成された分子はすべてポストスクリーニングを経て、三つの候補が専門家によって選択され、うまく合成された。図 2 は、我々の転移学習モデルからの予測値と比較した、3 つの候補の λ を示す。

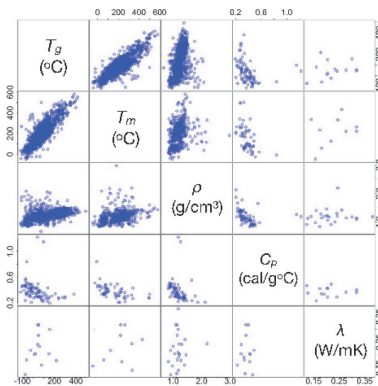


図 1. PoLyInfo 材料物性データの散布図
(ρ : 密度, C_p : 定圧比熱)

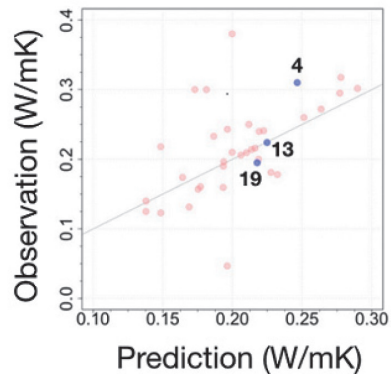


図 2. PoLyInfo の λ 観測値と転移学習モデルからの予測値との比較 (赤い点: 訓練データ, 青い点: 新しく合成したポリマー)

参 考 文 献

- Anderson, D.R. (1966). Thermal Conductivity of Polymers, *Chemical Reviews*, 66(6), 677-690.
- Bohacek, R.S., McMartin, C. and Guida, W.C. (1996). The art and practice of structure-based drug design: A molecular modeling perspective, *Medicinal Research Reviews*, 16(1), 3-50.
- Ikebata, H., Hongo, K., Isomura, T., Maezono, R. and Yoshida, R. (2017). Bayesian molecular design with a chemical language model, *Journal of computer-aided molecular design*, 31(4), 379-391.
- Otsuka, S., Kuwajima, I., Hosoya, J., Xu, Y. and Yamazaki, M. (2011). PolyInfo: Polymer database for Polymeric Materials Design, 2011 International Conference on Emerging Intelligent Data and Web Technologies, 22-29.
- Zhong, C., Yang, Q. and Wang, W. (2001). A group contribution model for the prediction of the thermal conductivity of polymer melts, *Industrial & Engineering Chemistry Research*, 40, 4151-4153.

大規模空間データのための空間可変係数モデリング

Spatially varying coefficient modeling for large dataset

データ科学研究系 村上 大輔 (Daisuke Murakami)

1. 背景

センサ技術の発展に伴い、地理空間データが急速に大規模化してきている。例えば土地被覆や気象関連の情報は高解像度の衛星画像として提供されており、世帯構成や産業構成のような社会経済統計は街区のような空間詳細な単位毎に整備されつつある。それら標本数 N の大きな地理空間データを柔軟にモデル化する方法が研究・実務で求められている。

Gaussian process (GP) は、地理空間データの背後にある空間過程をモデル化するために幅広く用いられてきた。GP の共分散が距離減衰関数に従うことを仮定することで、地点間の空間的な従属関係を柔軟にモデル化することができる。一方で、GP を推定するためには $N \times N$ 次元の共分散行列の逆行列を評価する必要があり、標本数 N が数百万あるいはそれ以上になりうる昨今、その計算負荷は実用上の課題となってきた。そのような中、GP を高速に推定するための近似手法が近年活発に議論されてきた。

本節では、GP に基づいた空間モデルの高速化の一環として取り組んでいる Spatially varying coefficient (SVC) モデルの高速化に関する研究を紹介する。

2. SVC モデルとその高速化

SVC モデルとは、各回帰係数の背後に GP を仮定することで回帰係数を場所毎に推定しようというモデルである。例えば都市部と郊外部の違いなどを柔軟に捉えることができるなど、SVC モデルは実用上便利である。しかしながら、回帰係数毎の GP を推定しようという同モデルの計算量は極めて大きいことが知られている。そこで本研究では、大規模データへの適用を見据え、SVC モデルの推定を次の手順で高速化した：(i) 各 GP を主成分のみを残して低ランク近似する；(ii) サイズが N に依存する行列・ベクトルを予め処理する（内積をとる）ことで、 N に依存しない計算量で評価できるように事後確率を書き直す；(iii) 同事後確率を逐次的に最大化していくことで回帰係数の空間分布を決める各 GP (低ランク) を推定する。手順 (i) は上述の逆行列の計算量を削減するための近似、手順 (ii) と (iii) は各 GP を特徴づけるパラメータの推定を高速化するための処理である。

以上で高速化した SVC モデルの計算時間を、ベイズ SVC モデル（提案手法はこれを近似）および地理的加重回帰モデル（従来手法）と比較した。計算はすべて統計ソフトウェア R 上で行った。計算時間の比較結果は図 1 に示すとおりである。同図より、ベイズ SVC モデルは計算負荷が極めて大きく、実用上難があることを確認した。従来手法の計算時間もまた標本数 N が大きくなるにしたがって急激に増加しており、大規模データのモデリングの観点からは課題が残されているとの示唆を得た。

対照的に、提案手法の計算時間は N に関して線形にしか増加しておらず、例え SVC 数 K が 8 の（8 個の GP を同時推定する）場合であってもその増加は極めて遅い。例えば $N = 5,000$

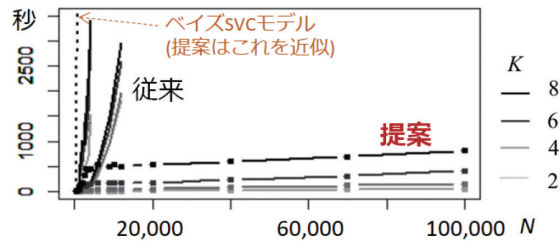


図 1. 計算時間の比較 (N は標本数, K は SVC の数).

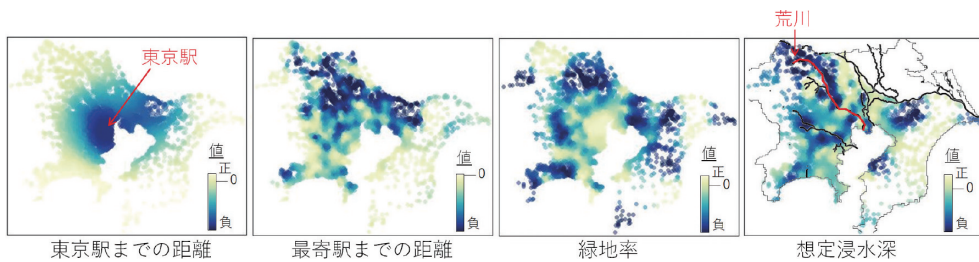


図 2. 推定された場所毎の回帰係数

の場合の提案モデルの推定時間は平均で 454 秒となり、従来モデルが適用困難な $N = 100,000$ の場合でも、その推定時間はわずか 836 秒 (平均値) となった。提案手法を用いることで、計算時間が大幅に短縮されることを確認した。なお、多くの場合に提案手法の SVC 推定精度が従来手法を上回ることもまたモンテカルロシミュレーションにより確認している。

3. 住宅地価分析への応用

提案モデルを東京都市圏の住宅地公示地価 (2010) 年の分析に適用した。説明変数は東京駅までの距離、最寄り駅までの距離、1km グリッド内の緑地率、想定浸水深である。各データは国土数値情報ダウンロードサービス (<http://nlftp.mlit.go.jp/ksj/>) で公開されている。

推定された回帰係数 (SVC) を図 2 にプロットした。この図から東京駅までの距離は都心を中心とした広域的な影響パターンを持つことが確認された。対照的に、最寄り駅までの距離は郊外部での影響力が大きく、特に鉄道網が比較的疎な北部ではその影響力が強まっていることが確認できる。緑地率もまた郊外部で強い効果を持つが、その影響は負であり、緑地の多さは地価を低下させるという結果が得られた。これは緑地よりも都市施設の多い土地のほうが好まれる傾向があるためである可能性がある。なお、例外的に都心部と横浜市中心部では緑地率は正に有意となっており、それら地域では緑地が好まれているの示唆を得た。想定浸水深は荒川沿岸の地価をより強く低下させていると推定された。この結果は、荒川沿岸の地価が水害リスクが適切に反映された値付けになっており、水害リスクにより適応した都市パターンであることを意味するものである。以上の結果は直感に整合する。

オイラー標数法によるウィシャート行列の 最大固有値分布の近似

The Euler characteristic method for approximating the distributions of the largest eigenvalues of Wishart matrices

数理・推論研究系 栗木 哲 (Satoshi Kuriki)

要 旨

実対称ランダム行列の最大固有値は、行列の 2 次形式として定義される確率場の最大値であるため、オイラー標数法 (チューブ法) によってその近似分布を与えることができる。本稿では、ウィシャート行列の行列サイズと自由度パラメータが無限大に発散する状況でも、オイラー標数法近似が真の極限分布の裾確率を精確に近似することを確認する。

キーワード：チューブ法, オイラー標数法, ランダム行列, Tracy-Widom 分布

1. 最大固有値とエクスカーション集合のオイラー標数

$n \times n$ 実対称行列 A が自由度 N のウィシャート分布 $W_n(N, I_n)$ に従うとする。その最大固有値は、 M をランク 1 の $n \times n$ 直交射影行列の全体とすると

$$\lambda_1(A) = \max_{\|h\|=1} h^\top A h = \max_{U \in M} \text{tr}(UA),$$

すなわち M を添字集合とする確率場 $\{\text{tr}(UA)\}_{U \in M}$ の最大値である。これより、エクスカーション集合 $M_x = \{U \in M \mid \text{tr}(UA) \geq x\}$ が定義される。このランダム集合のオイラー標数 $\chi(M_x)$ の期待値が、最大固有値のオイラー標数法近似である (栗木, 2019) :

$$\Pr(\lambda_1(A) \geq x) \approx E[\chi(M_x)] \quad (x \text{ が大きいとき}).$$

\tilde{A} が自由度 N の $n \times n$ 複素ウィシャート分布に従う場合も、添字集合 \tilde{M} を適当にとることにより同じ議論ができる。ラゲール多項式を $L_n^{(\alpha)}(x) = x^{-\alpha} e^x (d/dx)^n (x^{n+\alpha} e^{-x}) / n!$ とおく。

定理 1. (i) $A \sim W_n(N, I_n)$, $\alpha = N - n$.

$$E[\chi(M_x)] = \frac{\sqrt{\pi}(-1)^{n-1}(n-1)!}{2^{\frac{N+n-1}{2}} \Gamma(\frac{N}{2}) \Gamma(\frac{n}{2})} \int_x^\infty \lambda^{\frac{N-n-1}{2}} e^{-\frac{\lambda}{2}} L_{n-1}^{(\alpha)}(\lambda) d\lambda.$$

(ii) $\tilde{A} \sim CW_n(N, I_n)$, $\alpha = N - n$.

$$E[\chi(\tilde{M}_x)] = \frac{n!}{\Gamma(N)} \int_x^\infty \lambda^{N-n} e^{-\lambda} \{L_{n-1}^{(\alpha)}(\lambda) L_{n-1}^{(\alpha+1)}(\lambda) - L_n^{(\alpha)}(\lambda) L_{n-2}^{(\alpha+1)}(\lambda)\} d\lambda.$$

(i) と同等な表現は、Kuriki and Takemura (2001, 2008) で与えられている。

2. エッジ極限

Marchenko-Pastur 則より, ウィンシャート行列の固有値の極限分布は適当なスケーリングのもとで有限サポートをもつ. サポートの上界 μ_+ を拡大するスケーリング (エッジ極限) を行う.

$$x \mapsto s = \frac{x - \mu_+}{\sigma} \quad \text{ただし} \quad \frac{N}{n} \rightarrow \gamma, \quad \mu_+ = (1 + \sqrt{\gamma})^2 n, \quad \sigma = (1 + \sqrt{\gamma}) \left(1 + \frac{1}{\sqrt{\gamma}}\right)^{\frac{1}{3}} n^{\frac{1}{3}}.$$

ラグール多項式の極限定理 (Johnstone, 2001) によって次を得る.

定理 2. (i) $A \sim W_n(N, I_n)$, $N, n \rightarrow \infty$ s.t. $N/n \rightarrow \gamma$ のとき

$$E[\chi(M_x)] \Big|_{x=\mu_+ + \sigma s} \rightarrow \frac{1}{2} \int_s^\infty \text{Ai}(x) dx.$$

ここで Ai は第一種エアリー関数.

(ii) $\tilde{A} \sim CW_n(N, I_n)$, $N, n \rightarrow \infty$ s.t. $N/n \rightarrow \gamma$ のとき

$$E[\chi(\tilde{M}_x)] \Big|_{x=\mu_+ + \sigma s} \rightarrow \int_s^\infty \{\text{Ai}'(x)^2 - \text{Ai}(x)^2\} dx.$$

図 1 より, オイラー標数法近似は, 真の極限分布である Tracy-Widom 分布の上側裾確率を精確に近似することがわかる. 実際それらの相対誤差は, $s \rightarrow \infty$ のとき

$$\Delta(s) \sim -2^{-5} \pi^{-\frac{1}{2}} s^{-\frac{9}{4}} e^{-\frac{2}{3} s^{3/2}} \quad (\text{実の場合}), \quad 2^{-10} \pi^{-1} s^{-\frac{9}{2}} e^{-\frac{4}{3} s^{3/2}} \quad (\text{複素の場合}).$$

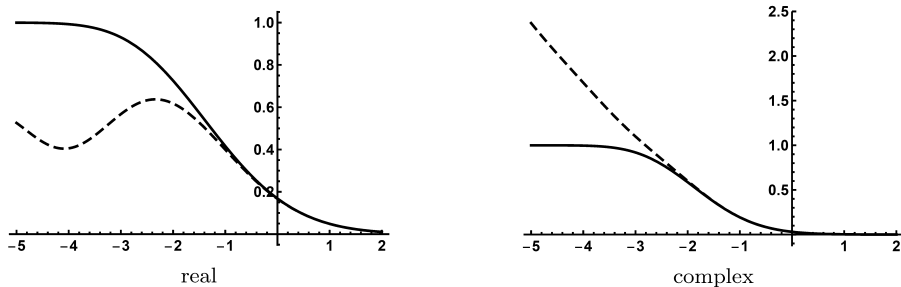


図 1. 極限分布 (Tracy-Widom 分布) の上側確率 (実線) とそのオイラー標数法近似 (破線).
左: 実ウィンシャートの場合, 右: 複素ウィンシャートの場合.

定理 2 で扱う極限では, オイラー標数法の確率場の添字集合の次元と体積が無限に発散するため, オイラー標数法が有効に働くかどうかは自明ではない. 無限次元の添字集合を直接扱うことのできるオイラー標数法の開発は, これからの大きな研究課題である.

参 考 文 献

- Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal components analysis, *The Annals of Statistics*, **29** (2), 295–327.
- 栗木哲 (2019). チューブ法の理論・応用とその周辺, 「統計数理」, 投稿中.
- Kuriki, S. and Takemura, A. (2001). Tail probabilities of the maxima of multilinear forms and their applications, *The Annals of Statistics*, **29** (2), 328–371.
- Kuriki, S. and Takemura, A. (2008). Euler characteristic heuristic for approximating the distribution of the largest eigenvalue of an orthogonally invariant random matrix, *Journal of Statistical Planning and Inference*, **138** (11), 3357–3378.

変化点モデルのための AIC

AIC for change-point models

数理・推論研究系 二宮 嘉行 (Yoshiyuki Ninomiya)

要 旨

変化点問題は長い間議論されているが、その理由の一つは、変化点モデルに通常の統計的漸近理論を満たさせない非正則性が存在することにある。本研究の目的は、そのような変化点モデルの AIC を導出することである。AIC の罰則項は最大対数尤度の漸近バイアスの二倍であり、正則条件を満たすモデルであればそれはパラメータの数の二倍となる。一方変化点モデルでは、その非正則性がゆえにそれは $2m + 2p_m$ とならない。ここで m と p_m はそれぞれ変化点の数と変化点以外のパラメータの数である。本稿では、その漸近バイアスが $6m + 2p_m$ で評価できることを示した Ninomiya (2015) の結果を紹介する。

キーワード：構造変化，情報量規準，非正則性，Brown 運動，ランダムウォーク

1. モデル

独立観測系列 $\{x_i, 1 \leq i \leq n\}$ に対し、 m 個の変化点 $k^{(1)}, \dots, k^{(m)}$ をもつモデルを考える。簡単のため、 x_i の分布は指数型分布族に属する、つまり確率関数は

$$(1.1) \quad k^{(j-1)} + 1 \leq i \leq k^{(j)} \quad \text{のとき} \quad \exp\{\theta^{(j)T} \mathbf{T}(\cdot) + S(\cdot) - A(\theta^{(j)})\}$$

と書けるものとする ($1 \leq j \leq m+1$)。ここで $k^{(0)} = 0$ かつ $k^{(m+1)} = n$ としている。また、 $\theta^* = (\theta^{*(1)T}, \dots, \theta^{*(m+1)T})^T$ と $\mathbf{k}^* = (k^{*(1)}, \dots, k^{*(m)})^T$ を、 $\theta = (\theta^{(1)T}, \dots, \theta^{(m+1)T})^T$ と $\mathbf{k} = (k^{(1)}, \dots, k^{(m)})^T$ の真値とする。いま、

$$(1.2) \quad \theta^{*(1)} \neq \theta^{*(2)} \neq \dots \neq \theta^{*(m+1)},$$

を仮定し、 θ^* と \mathbf{k}^* は未知であるとする。加えて、 $\theta^{*(1)}, \dots, \theta^{*(m+1)}$ は自然パラメータ空間に含まれるパラメータ集合の内点であり、そのパラメータ集合上で $\partial^2 A(\theta) / \partial \theta \partial \theta^T$ は正定値であるとする。さらに、後の漸近論のため、 $1 \leq j \leq m$ に対して $\lim_{n \rightarrow \infty} k^{*(j)} / n = \kappa^{(j)}$ であることを仮定する。ただし $0 < \kappa^{(1)} < \dots < \kappa^{(m)} < 1$ としている。

2. 結果

$\hat{\mathbf{k}}_x$ と $\hat{\theta}_x$ を \mathbf{k}^* と θ^* の $x = (x_1^T, \dots, x_n^T)^T$ に基づく最尤推定量とし、 $f(x|\mathbf{k}^*, \theta^*)$ を x の同時確率関数とする。モデル選択は $f(\mathbf{y}|\mathbf{k}^*, \theta^*)$ と $f(\mathbf{y}|\hat{\mathbf{k}}_x, \hat{\theta}_x)$ の間の Kullback-Leibler ダイバージェンスの二倍

$$2\text{KL}\{f(\mathbf{y}|\mathbf{k}^*, \theta^*), f(\mathbf{y}|\hat{\mathbf{k}}_x, \hat{\theta}_x)\} = 2\text{E}_y\{\log f(\mathbf{y}|\mathbf{k}^*, \theta^*)\} - 2\text{E}_y\{\log f(\mathbf{y}|\hat{\mathbf{k}}_x, \hat{\theta}_x)\},$$

を小さくしようとするところに行うことができる。ここで、 \mathbf{y} は x のコピーであり、 E_y はその y

に関する期待値を表すとする。右辺の第一項はモデルに依らないため、第二項を考えさえすればよい。そのシンプルな推定量は $-2\log f(\mathbf{x}|\hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x)$ であるが³、これは過小評価する。そこで、AIC 型の情報量規準では、それをバイアス補正した

$$(2.1) \quad \begin{aligned} & -2\log f(\mathbf{x}|\hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x) + 2E_x[\log f(\mathbf{x}|\hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x) - E_y\{\log f(\mathbf{y}|\hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x)\}] \\ & = -2\log f(\mathbf{x}|\hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x) + 2E \left[\sup_{(\mathbf{k}, \boldsymbol{\theta})} L_x(\mathbf{k}, \boldsymbol{\theta}) - L_x \left\{ \operatorname{argsup}_{(\mathbf{k}, \boldsymbol{\theta})} L_y(\mathbf{k}, \boldsymbol{\theta}) \right\} \right], \end{aligned}$$

を考える。ここで、 E は \mathbf{x} と \mathbf{y} の両方での期待値を表し、また $L_x(\mathbf{k}, \boldsymbol{\theta}) = \log f(\mathbf{x}|\mathbf{k}, \boldsymbol{\theta}) - \log f(\mathbf{x}|\mathbf{k}^*, \boldsymbol{\theta}^*)$ である。しかし、(2.1) における期待値は陽に求められないため、通常の AIC と同じようにその漸近評価を用いることを考える。つまり、(2.1) の代わりに

$$(2.2) \quad -2\log f(\mathbf{x}|\hat{\mathbf{k}}_x, \hat{\boldsymbol{\theta}}_x) + 2E\{b(\mathbf{k}^*, \boldsymbol{\theta}^*)\}$$

を考える。ここで、 $b(\mathbf{k}^*, \boldsymbol{\theta}^*)$ は $\sup_{(\mathbf{k}, \boldsymbol{\theta})} L_x(\mathbf{k}, \boldsymbol{\theta}) - L_x\{\operatorname{argsup}_{(\mathbf{k}, \boldsymbol{\theta})} L_y(\mathbf{k}, \boldsymbol{\theta})\}$ の弱極限であるとする。ただし、 $\sup_{(\mathbf{k}, \boldsymbol{\theta})}$ と $\operatorname{argsup}_{(\mathbf{k}, \boldsymbol{\theta})}$ は $L_x(\mathbf{k}, \boldsymbol{\theta})$ が $O_P(1)$ あるいは正の値となるような $(\mathbf{k}, \boldsymbol{\theta})$ の集合上でとるものとする。

$A'(\boldsymbol{\theta}^{*(j)})$ を $\partial A(\boldsymbol{\theta}^{(j)})/\partial \boldsymbol{\theta}^{(j)}|_{\boldsymbol{\theta}^{(j)}=\boldsymbol{\theta}^{*(j)}}$, $B_1^{(j)}(\boldsymbol{\theta}^*) = A(\boldsymbol{\theta}^{*(j+1)}) - A(\boldsymbol{\theta}^{*(j)}) - (\boldsymbol{\theta}^{*(j+1)} - \boldsymbol{\theta}^{*(j)})^T A'(\boldsymbol{\theta}^{*(j)})$, $B_2^{(j)}(\boldsymbol{\theta}^*) = A(\boldsymbol{\theta}^{*(j)}) - A(\boldsymbol{\theta}^{*(j+1)}) - (\boldsymbol{\theta}^{*(j)} - \boldsymbol{\theta}^{*(j+1)})^T A'(\boldsymbol{\theta}^{*(j+1)})$ とし、 $Q_{k, \mathbf{x}}^{(j)}$ を

$$\begin{aligned} & I_{\{k < k^{*(j)}\}} \sum_{i=k+1}^{k^{*(j)}} [(\boldsymbol{\theta}^{*(j+1)} - \boldsymbol{\theta}^{*(j)})^T \{\mathbf{T}(\mathbf{x}_i) - A'(\boldsymbol{\theta}^{*(j)})\} - B_1^{(j)}(\boldsymbol{\theta}^*)] \\ & + I_{\{k > k^{*(j)}\}} \sum_{i=k^{*(j)+1}^k} [(\boldsymbol{\theta}^{*(j)} - \boldsymbol{\theta}^{*(j+1)})^T \{\mathbf{T}(\mathbf{x}_i) - A'(\boldsymbol{\theta}^{*(j+1)})\} - B_2^{(j)}(\boldsymbol{\theta}^*)]. \end{aligned}$$

とする。すると次の定理が得られる。

定理 1. \mathbf{x} が (1.1) にしたがって、また (1.2) が成立すると、(2.2) における漸近バイアスは

$$(2.3) \quad E\{b(\mathbf{k}^*, \boldsymbol{\theta}^*)\} = \sum_{j=1}^m E \left(\sup_k Q_{k, \mathbf{x}}^{(j)} + Q_{\operatorname{argsup}_k Q_{k, \mathbf{y}}^{(j)}, \mathbf{x}}^{(j)} \right) + p_m,$$

で与えられる。ここで p_m は $\boldsymbol{\theta}$ における異なるパラメータの数である。

情報量規準を使うメリットのの一つはその使いやすさであることを鑑み、 $1 \leq j \leq m$ に対して

$$(2.4) \quad \boldsymbol{\theta}^{*(j+1)} - \boldsymbol{\theta}^{*(j)} = \alpha_n^{-1/2} \Delta_{\boldsymbol{\theta}^*}^{(j)} \quad \text{かつ} \quad O(1) \neq \alpha_n = o(n)$$

なる近接条件を仮定する。ここで $\Delta_{\boldsymbol{\theta}^*}^{(j)}$ は定数ベクトルである。すると変化点推定量の漸近挙動が変わり、定理 1 の代わりに以下が得られる。

定理 2. (2.4) のもとで (2.2) における漸近バイアスは

$$E\{b(\mathbf{k}^*, \boldsymbol{\theta}^*)\} = 3m + p_m$$

で与えられる。

参 考 文 献

Ninomiya (2015). Change-point model selection via AIC. *Annals of the Institute of Statistical Mathematics* 67, 943–961.

抽出法と計算代数

Samplers and Computational Algebra

数理・推論研究系 間野 修平 (Shuheï Mano)

要 旨

Diaconis と Sturmfels は, Markov chain Monte Carlo について, Markov 連鎖の推移の基底である Markov 基底を導入し, それが十分統計量が定める多項式環のトーリックイデアルの Gröbner 基底により与えられることを示した. 筆者は, 微分作用素環の考察により, 標本経路が目的の分布に従う Markov 連鎖を構成することで, 直接抽出が可能であることを示した.

キーワード: 代数統計, 抽出法, 計算代数, Gröbner 基底, 超幾何系

1. はじめに

近年の代数統計の研究は, Pistone と Wynn (1996) による実験計画法における母数の識別性に関する研究と, Diaconis と Sturmfels (1998) による Markov chain Monte Carlo (MCMC) における定常分布を目的の分布とする Markov 連鎖の推移の基底である Markov 基底の導入が源流とされる. 後者の周辺の発展については Aoki et al. (2012) を参照. MCMC の長所は正規化定数の計算を要しないこと, 短所は定常分布からの抽出を保証し難いことで, 正規化定数を効率的に計算できるなら MCMC を使う必要はない. 筆者は, 微分作用素環の考察により, 標本経路が目的とする分布に従う Markov 連鎖を構成し, 推移確率に現れる正規化定数を計算すれば, 目的の分布からの直接抽出が可能であることを示した (Mano 2017). 本稿では, その概略を紹介する. 詳細と参考文献については Mano (2018) を参照されたい.

2. Markov 基底

カウントベクトル $c \in \mathbb{N}_0^m$ が従う分布からの MCMC における Markov 連鎖の状態空間は, 十分統計量 $b \in \mathbb{C}^d$ について $\mathcal{F}_b(A) := \{c; Ac = b, c \in \mathbb{N}_0^m\}$ と表せる. A は階数 d の非負整数値 $d \times m$ 配置行列 (行空間に $(1, \dots, 1)$ を含む) とする. 集合 $\mathcal{M}(A) = \text{Ker}A \cap \mathbb{Z}^m$ を A に関する移動という. Markov 基底 $\mathcal{B} \subset \mathcal{M}(A)$ は $\mathcal{F}_b(A)$ に既約な Markov 連鎖を与える移動の集合である. 任意の移動 z は $z = z^+ - z^-$, $z_i^+ := \max\{z_i, 0\}$, $z_i^- := \max\{-z_i, 0\}$ と表され, 多項式環の二項式と 1 対 1 に対応づけられる. 特に, $I_A := \{x^{z^+} - x^{z^-}; z \in \mathcal{M}(A)\}$ はトーリックイデアルである. ここで, $x^z := \prod_{i=1}^m x_i^{z_i}$ とした.

定理 1 (Diaconis, Sturmfels 1998). $\mathcal{B} = \{z_i; i \in \{1, \dots, s\}\} \subset \mathcal{M}(A)$ が Markov 基底であることは, $\{x^{z_j^+} - x^{z_j^-}; j \in \{1, \dots, s\}\}$ が I_A の生成系であることの必要十分条件である.

I_A の Gröbner 基底は生成系だから Markov 基底である. Gröbner 基底を求める一般的なアルゴリズム (Buchberger 1976) があるので, 原理的には任意の A に対し Markov 基底が得られる.

3. A 超幾何系

定義 1 (Gelfand et al. 1990). 2 節で導入した行列 A , 十分統計量 b に対し, 次の消去作用素が定める線形偏微分方程式系を A 超幾何系 $H_A(b)$ とよぶ.

$$\sum_{j=1}^m a_{ij} x_j \frac{\partial}{\partial x_j} - b_i, \quad i \in \{1, \dots, d\}, \quad \partial^{c^+} - \partial^{c^-}, \quad c \in \text{Ker} A \cap \mathbb{Z}^m.$$

$H_A(b)$ は微分作用素環の左イデアルで, A 超幾何イデアルとよばれる.

A 超幾何系 $H_A(b)$ の原点周りの級数解

$$Z_A(b; x) := \sum_{c \in \mathcal{F}_b(A)} \frac{x^c}{c!}, \quad c! := \prod_{i=1}^m c_i!$$

を A 超幾何級数とよぶ. ただし, $b \notin \text{AN}_0^m$ のときは $Z_A(b; x) = 0$ と規約する. m 個の対数アフィンモデルからの長さ n の多項抽出によるカウントベクトル $c \in \mathbb{N}_0^m$ の分布は離散指数型分布族であり, 十分統計量 $b \in \mathbb{N}_0^d$ で $Ac = b$ と条件づけた分布の確率関数は $x^c/c!$ に比例する. 正規化定数が A 超幾何多項式になるので, Takayama et al. (2018) は A 超幾何分布とよんだ. 周辺度数が所与の分割表の分布など, カウントデータ解析における典型的な統計モデルを含む.

4. 直接抽出法

A 超幾何多項式の性質から, A の第 i 列ベクトルを a_i として,

$$p_A(b; i) := \frac{\mathbb{E}(C_i | AC = b)}{n} = \frac{Z_A(b - a_i; x) x_i}{Z_A(b; x) n}, \quad \sum_{i=1}^m p_A(b; i) = 1$$

が従う. $p_A(b; i)$ を, A 超幾何多項式を状態空間とし, 推移ごとに次数が 1 下がる Markov 連鎖において, $Z_A(b; x)$ から $Z_A(b - a_i; x)$ への推移確率とみなすと, 次のアルゴリズムが得られる.

アルゴリズム 1 (Mano 2017). A 超幾何分布からの逐次直接抽出.

1. $t_1 = j$ を確率 $p_A(b; j)$ で抽出.
2. $i = 2, \dots, n$ について, $t_i = j$ を確率 $p_A(b - (a_{t_1} + \dots + a_{t_{i-1}}); j)$ で抽出.

A 超幾何多項式は $H_A(b)$ の標準単項式のベクトル Q が従う Pfaffian 系 $\partial_i Q = P_i Q$ を用いて求める. 有理関数の行列 $P_i, i \in \{1, \dots, m\}$ は標準単項式の Gröbner 基底による標準形として原理的には任意の A に対して得られる (Saito et al. 2010; 日比ら 2011).

5. おわりに

様々な統計モデルについて MCMC による抽出を直接抽出に置き換えていくことは, 代数統計の新しい挑戦のひとつになると期待される.

参 考 文 献

- Mano, S.: Partition structure and the A -hypergeometric distribution associated with the rational normal curve. *Electron. J. Statist.* **11**, 4452–4487 (2017)
- Mano, S.: *Partitions, Hypergeometric Systems, and Dirichlet Processes in Statistics*. JSS Research Series in Statistics, SpringerBriefs in Statistics (2018)

角度の観測を含むデータのための統計モデル

Statistical Models for Data Which Include Angular Ones

数理・推論研究系 加藤 昇吾 (Shogo Kato)

1. はじめに

様々な学問分野において、角度として表される観測値が得られることがある。例えば、気象学における風向の観測はその一例である。風向は、西を $-\pi$ とし、反時計回りを正の向きとすれば、南を $-\pi/2$ 、東を 0 、北を $\pi/2$ のように角度で表すことができる。つまり、任意の風向は $-\pi$ 以上 π 未満の角度 θ 、もしくは円周上の点 $(\cos \theta, \sin \theta)$ 、として表現できる。他にも、医学・地震学・動物行動学・生命情報学など、多くの分野で角度の観測が存在している。

角度の観測を含むデータには、統計解析をする上で大きな問題がある。それは、このようなデータを解析する上では、統計学が主に対象としている実数値データのための解析手法をそのまま使うことができないという問題である。例えば、平均や分散などの要約統計量や、実数値データのための確率分布・回帰モデル・時系列モデルなどの統計モデルは、角度データに直接応用すると、しばしば不自然な解析結果を与えることにつながってしまう。この問題を解決するため、角度の観測を含むデータの統計的手法を考えることが統計学における重要な研究テーマとなっている。本報告では、このテーマにおける研究の背景と著者の研究結果を概観する。

2. 角度データのための確率分布

方向統計学における中心的なアプローチは、データに何らかの確率分布を仮定したパラメトリックな統計解析法である。著者は、角度データのための確率分布（円周上の確率分布）の 1 つとして知られる「円周上のコーシー分布」に着目し、これに関連したパラメトリックな統計モデルを研究してきた。円周上のコーシー分布は確率密度関数

$$f(\theta) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}, \quad -\pi \leq \theta < \pi; \quad -\pi \leq \mu < \pi, \quad 0 \leq \rho < 1,$$

で定義される確率分布である。ここで、 μ は分布の位置を定めるパラメータ、 ρ は分布の集中度を調節するパラメータである。

方向統計学においては、円周上のコーシー分布は古くから知られていたもののほとんど注目されていない分布であった。この分野で長きにわたり研究の中心だったのは、「フォン・ミーゼス分布」とよばれる円周上の確率分布およびそれを応用したパラメトリックな統計手法である。フォン・ミーゼス分布は、実数上の正規分布からの自然なアナロジーにより導かれることから、「円周上の正規分布」とよばれることもある。一方、正規分布が持ついくつかの扱いやすい性質が成立しない問題点も指摘されており、そのことがフォン・ミーゼス分布およびそれを応用した統計手法の理論的性質を導くことを困難にしている一面もあった。そのような中、McCullagh (1996) などによって開拓された円周上のコーシー分布に著者は興味を持ち、この分布に関連した統計モデルの研究を行うようになった。

3. 角度の観測を含むデータのための統計モデル

著者は、円周上のコーシー分布に関連した統計モデルの論文を今までに複数篇執筆してきたが、ここでは、その中から 2 篇についてある程度詳しく説明し、4 篇について簡潔に紹介する。

Kato (2010) では、円周上のコーシー分布を誤差分布として用いた新たな円周上のマルコフ過程を提案した。この研究は Fisher and Lee (1994) によるマルコフ過程と関連がある。彼らは、回帰曲線としてメビウス変換、誤差分布としてフォン・ミーゼス分布を仮定した円周上のマルコフ過程を提案した。それに対し、Kato (2010) では、誤差分布としてフォン・ミーゼス分布の代わりに円周上のコーシー分布を用いることにより、Fisher and Lee (1994) のモデルでは得られなかった多くの挙動に関する性質を得ることに成功した。例えば、任意時点 t の角度を所与としたときの $t+h$ 時点における角度の条件付分布が円周上のコーシー分布となることや、その条件付分布のパラメータも複素数を用いれば簡潔に表現することができることを示した (t と h は自然数をあらわす)。

Kato and Pewsey (2015) では、円周上のコーシー分布の拡張として、2次元トーラス $[-\pi, \pi]^2$ 上の分布を提案した。2次元トーラス上の分布の既存研究としては、フォン・ミーゼス分布を拡張した 2 変量フォン・ミーゼス分布がよく知られている。この分布は、条件付分布がフォン・ミーゼス分布となる利点があるが、正規化定数が特殊関数の無限和で表されることや、パラメータの解釈が困難であること、周辺分布がよく知れていない分布になること、などの問題点がある。それに対して、Kato and Pewsey (2015) で提案した 2次元トーラス上の分布ではこれらの問題がすべて解決されたモデルとなっている。

その他の研究結果について簡潔に紹介する。Kato *et al.* (2008) では説明変数・被説明変数が共に角度となる回帰モデルを提案し、誤差分布として円周上のコーシー分布を用いることで回帰モデルのいくつかの扱いやすい性質を得られることを示した。Kato and Jones (2010, 2015) では、円周上のコーシー分布を特別な場合として含む柔軟な円周上の確率分布を提案した。Kato and McCullagh (2018) では、円周上のコーシー分布を球面上の分布へと拡張し、メビウス変換に関して閉じていることやパラメータ推定が容易にできることを示した。

参 考 文 献

- Fisher, N. I. and Lee, A. J. (1994). Time series analysis of directional data, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **56**, 327–339.
- Kato, S. (2010). A Markov process for circular data, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72**, 655–672.
- Kato, S. and Jones, M. C. (2010). A family of distributions on the circle with links to, and applications arising from, Möbius transformation, *Journal of the American Statistical Association*, **102**, 249–262.
- Kato, S. and Jones, M. C. (2015). A tractable and interpretable four-parameter family of unimodal distributions on the circle, *Biometrika*, **102**, 181–190.
- Kato, S. and McCullagh, P. (2018). Möbius transformation and a Cauchy family on the sphere, *arXiv:1510.07679v2*, 1–30.
- Kato, S. and Pewsey, A. (2015). A Möbius transformation-induced distribution on the torus, *Biometrika*, **102**, 359–370.
- Kato, S., Shimizu, K. and Shieh, G. S. (2008). A circular–circular regression model, *Statistica Sinica*, **18**, 633–645.
- McCullagh, P. (1996). Möbius transformation and Cauchy parameter estimation, *The Annals of Statistics*, **24**, 787–808.

極値分布の吸引領域と離散性

Maximum domain of attraction and Discreteness

数理・推論研究系 志村 隆彰 (Takaaki Shimura)

キーワード：極値理論, 極値分布, 確率分布の裾, 最大値吸引領域, 離散化.

近年, 東日本大震災のような巨大地震や西日本豪雨のような豪雨など, 極端に大きな被害をもたらす自然災害 (激甚災害) が続き, 大きな社会問題になっている. このような災害をランダムな現象とみなせば, “起こる頻度 (確率) は少ないが, 一旦起こったときの被害 (量) が極端に大きい” ことが特徴である. このような “稀に起こる極端な事象とその起こり方” を研究するのが極値理論 (Extreme Value Theory) である. 極値理論では, 極端な事象の起こり方を表す確率分布 F の裾 (確率) $\bar{F}(x) \equiv 1 - F(x)$ (x より大きい値が起こる確率を表す) の挙動が重要で, x が F の上限 $x_F \equiv \sup\{x : F(x) < 1\}$ に近づくときの 0 への収束が, 正規分布のように収束が速いとき裾が軽い, コーシー分布のように遅いとき裾が重いという (heavy tail は最近広く知られるようになってきている).

確率統計でもっとも重要な定理は中心極限定理であるが, 極値理論でこれに当たるのが多数の中の最大値の振る舞いを表す次の定理である.

定理 0. X_1, X_2, \dots を共通の確率分布 F に従う実数値独立確率変数列とし, X_n までの最大値を $M_n \equiv \max\{X_1, \dots, X_n\}$ とする. このとき, 共通分布 F が適当な条件を満たせば, 非退化分布 G が存在して, 適当な定数列 $a_n > 0$ と $b_n \in \mathbf{R}$ により,

$$\mathcal{L}\left(\frac{M_n - b_n}{a_n}\right) \rightarrow G \quad (n \rightarrow \infty)$$

となる ($\mathcal{L}()$ はカッコ内の確率変数の分布を, \rightarrow は分布の収束を表す). このときの極限分布 G は極値分布と呼ばれ, 3 種類ある (フレシェ分布, グンベル分布, (極値) ワイブル分布).

各極値分布に対して, 上記の収束が成り立つような確率変数列の共通分布 F の集合をその極値分布の (最大値) 吸引領域といい, 分布の裾 $\bar{F}(x)$ が分布の上限に近づくときの漸近挙動で特徴づけられる. 本稿では, F として上限が無限の分布を扱うので, 極値分布はフレシェ分布あるいはグンベル分布になる. 裾の重さに注目すると, パレート分布のような裾の挙動がべきオーダーで重い (正確には裾が正則変動する) ものはフレシェ分布の吸引領域に, 指数分布などの裾がより軽いものはグンベル分布の吸引領域に属する. また, 分布の連続性の観点からは, 正規分布, 指数分布, コーシー分布など様々な裾挙動を持つ連続分布がいずれかの極値分布の吸引領域に属する半面, 幾何分布やポアソン分布などの裾が重くない離散分布はどの極値分布の吸引領域にもはまらない (指数分布と幾何分布は連続分布か離散分布かという違いであることに注意). この事実は, 吸引領域への属性と離散性との相性が悪いことを示唆している. 分布が吸引領域に入らないということは, 理論的に, それに従う独立同分布列の最大値の挙動がとらえきれないだけでなく, 応用上も, 実データが必然的に離散であるから, 統計解析の精度に影響することを意味している.

この背景の元, 最大値吸引領域と分布の離散性の関係を考察した結果を述べる. 実数上の分

布に対して、区間 $(n-1, n]$ の確率を一点 $\{n\}$ に集中させて離散分布（整数値分布）を作る操作を分布の”離散化”と呼び、離散分布に対して、連続分布でその離散化が元の離散分布と一致するものを離散分布の”連続化”と呼ぶ（連続化は一意ではない）。幾何分布は指数分布の離散化、幾何分布の連続化のひとつが指数分布である。このような場合、指数分布は離散化で吸引領域への属性を失い、幾何分布は連続化で属性が回復可能ということにする。

問題は、第一に、吸引領域に属する分布で、離散化によりその属性を保つか、失うかの判別条件を与えること。第二に、連続化で属性が回復可能な分布の特徴付けである。

定理 A. 吸引領域に属する分布が、離散化により、吸引領域に留まるための必要十分条件は、長い裾を持つことである： $\lim_{x \rightarrow \infty} \bar{F}(x+1)/\bar{F}(x) = 1$ (long-tailed distribution).

定理 B. 分布 F が吸引領域に回復可能であるための必要十分条件は、そのフォン・ミーゼス関数 \bar{F}_0 (\bar{F} を滑らかにしたもの) が次を満たすことである。

$$\lim_{n \rightarrow \infty} \left(\log \frac{\bar{F}_0(n+1)}{\bar{F}_0(n+2)} \right)^{-1} - \left(\log \frac{\bar{F}_0(n)}{\bar{F}_0(n+1)} \right)^{-1} = 0.$$

定理 A から、パレート分布や対数正規分布は離散化で属性を失わないこと、定理 B から幾何分布の他に、ポアソン分布も回復可能であることがわかる（定理 B の吸引領域は実はグンベル分布の吸引領域）。

次に、幾何的分布と呼ばれる離散分布に対して、吸引領域に属する連続化を与える具体的方法を与える。離散分布 F_1 が幾何的分布であるとは、ある正の指数 γ が存在して、 $\lim_{n \rightarrow \infty} \bar{F}_1(n+1)/\bar{F}_1(n) = e^{-\gamma}$ となるときをいう。また、 F が指数分布であるとは、ある正の指数 γ が存在して、 $\lim_{x \rightarrow \infty} \bar{F}(x+k)/\bar{F}(x) = e^{-\gamma k}$ となるときをいい、その全体を $\mathcal{L}(\gamma)$ であらわす。 $\mathcal{L}(\gamma)$ はグンベル分布の吸引領域に含まれることが知られている。

定理 C. F_1 を指数 $\gamma (> 0)$ の幾何的分布、 G を $[0, 1]$ 上の分布とする。このとき、 $F_1 * G \in \mathcal{L}(\gamma)$ となるための必要十分条件は、

$$G(x) = \frac{1 - e^{-\gamma x}}{1 - e^{-\gamma}} \quad (0 \leq x \leq 1)$$

である。

幾何的分布、指数分布はそれぞれ幾何分布、指数分布の一般化であり、ガンマ分布などが指数分布である。この定理は、幾何的分布は指数が共通である限り、同じ $[0, 1]$ 上の分布を（確率変数の意味で独立に）足すことで、吸引領域へ回復可能であることを主張している。たとえば、幾何分布に従う実データは定理 0 の極限定理は成り立たないが、データにランダムな補正をすることで成り立つようにできるのである。

参 考 文 献

- Shimura, T. (2012). Discretization of distributions in the maximum domain of attraction, *Extremes* **15**, 299-317.
- Shimura, T. (2012). Limit distribution of a roundoff error, *Statistics and Probability Letters*, **82**, 713-719.
- 高橋 倫也・志村 隆彰. (2016). 「極値統計学」(ISM シリーズ 進化する統計数理 5), 近代科学社.

拡散過程の非同期・ノイズ付観測データに対する 最尤型推定法

Maximum likelihood type estimation for diffusion processes with noisy, nonsynchronous observations

数理・推論研究系 荻原 哲平 (Teppei Ogihara) *

近年株式市場における一日内の全取引の情報を記録した「高頻度データ」の利用可能性が高まり、従来のデータに比べて膨大な情報量をもつため、金融市場のミクロ構造のさらなる解明が期待され、このようなデータを用いたデータ解析手法が活発に研究されている。通常、日内において株価が観測されるのは株価の約定時であるため、異なる株式に対して観測時刻が一致していないという「非同期観測」の問題が必然的に生じ、特にデータの共変動（共分散）の推定が困難になる。データの線形補完や直前データを用いた補完などによるシンプルな「同期化」を行ったデータに対する共変動推定量には深刻なバイアスが存在することが知られている。

確率空間 (Ω, \mathcal{F}, P) 上の二次元確率過程 $X = \{X_t\}_{0 \leq t \leq T}$ が以下を満たすとする：

$$dX_t = \mu(t, X_t, \sigma_*)dt + b(t, X_t, \sigma_*)dW_t, \quad t \in [0, T].$$

ここで、 σ_* はモデルの d 次元パラメータ、 $\{W_t\}_{0 \leq t \leq T}$: 二次元標準ブラウン運動、 μ は \mathbb{R}^2 値、 b は $\mathbb{R}^2 \otimes \mathbb{R}^2$ 値既知関数とする。 X の成分を X^1, X^2 と書き、 X^1, X^2 の観測時刻をそれぞれランダム時刻 $\{S_i^n\}_{i=0}^{\ell_1, n}, \{T_j^n\}_{j=0}^{\ell_2, n} \subset [0, T]$ で表す。データが高頻度観測になる極限を扱うため、 $n \rightarrow \infty$ の時 $\max_{i,j} |S_i^n - S_{i-1}^n| \vee |T_j^n - T_{j-1}^n| \rightarrow^p 0$ を仮定する。この時観測データ $\{S_i^n\}_i, \{T_j^n\}_j, \{X_{S_i^n}^1\}_i, \{X_{T_j^n}^2\}_j$ からモデル・パラメータ σ_* を推定する問題を考える。

非同期観測モデルに対する最尤型推定法

Ogihara and Yoshida (2014) において、 σ_* に対する疑似対数尤度関数を用いた最尤型推定法が提案されている。 $\mu \equiv 0$ かつ b, X_0 と $\{S_i^n\}_i, \{T_j^n\}_j$ が非ランダムの時、 $\{X_{S_i^n}^1\}_i, \{X_{T_j^n}^2\}_j$ は多変量正規分布に従い、共分散関数を計算することが可能であるため、対数尤度関数を二次形式で与えることができる。一般の場合でも同様に疑似対数尤度関数 $H_n(\sigma)$ を近似的に構成可能であり、 $H_n(\sigma)$ を最大にするパラメータの値として最尤型推定量 $\hat{\sigma}_n$ が定義される。このように定義された $\hat{\sigma}_n$ に対して、 μ, b の微分可能性や非退化性等の条件と、観測時刻列 $\{S_i^n\}_i, \{T_j^n\}_j$ の漸近挙動に関する条件の下、漸近混合正規性：

$$(0.1) \quad \sqrt{n}(\hat{\sigma}_n - \sigma_*) \rightarrow^{s-\mathcal{L}} \Gamma^{-1/2} \mathcal{N} \quad \text{as } n \rightarrow \infty$$

が示される。ただし、 $\rightarrow^{s-\mathcal{L}}$ は stable convergence を表し、 ∂_σ を σ に関する微分、 $\Gamma = P\text{-}\lim_{n \rightarrow \infty} (-\partial_\sigma^2 H_n(\sigma_*)/n)$, \mathcal{N} を (Ω, \mathcal{F}, P) のある拡張上で定義された、 \mathcal{F} と独立に $N(0, I_d)$ に従う確率変数とする。ここで I_d は d 次元単位行列である。

* 東京大学 数理・情報教育研究センター：〒113-8656 文京区弥生 2-11-16

非同期観測モデルにおける局所漸近混合正規性

Jeganathan (1983) では, 統計モデル $\{P_{\theta,n}\}$ の局所漸近混合正規性の下, 任意の推定量の漸近分散の下界を与える minimax 不等式が示された. この下界を達成する推定量は漸近有効推定量と呼ばれる. 拡散過程の高頻度観測モデルに対しては, 局所漸近混合正規性を示すには拡散過程の推移確率密度関数の漸近挙動を解析する必要がある, 非同期性のない規則的な観測モデルに対して Gobet (2001) では Malliavin 解析の技術を用いて局所漸近混合正規性を示した. このモデルにおいては Genon-Catalot and Jacod (1993) の最尤型推定量が漸近有効となる. Ogihara (2015) では, Gobet (2001) の Malliavin 解析を用いた手法を応用し, 拡散過程の非同期観測モデルに対する局所漸近混合正規性を証明し, 最尤型推定量 $\hat{\sigma}_n$ が漸近有効であることを示した.

非同期・ノイズ付観測モデルに対する最尤型推定法

高頻度データの解析上の問題点として, 非同期観測に加えて, 拡散過程によるモデリングにおける仮想的な観測ノイズの存在が実証研究から示唆されている. このような観測ノイズは「マーケット・マイクロストラクチャー・ノイズ」と呼ばれる.

観測ノイズ $\{\epsilon_i^k\}_{k=1,2,i \in \mathbb{Z}_+}$ を $(X_t, W_t)_{0 \leq t \leq T}$ と独立で, ある正定数 v_1, v_2 に対して, $E[\epsilon_i^k] = 0$, $E[\epsilon_i^k \epsilon_j^l] = v_k \delta_{kl} \delta_{ij}$ を満たす確率変数として, 観測が $\{X_{S_i^n}^1 + \epsilon_i^1\}_{i=0}^{\ell_1, n}, \{X_{T_j^n}^2 + \epsilon_j^2\}_{j=0}^{\ell_2, n}$ で与えられるような統計モデルとして非同期・ノイズ付観測がモデル化される.

この時 Ogihara (2018) において, モデル・パラメータ σ_* の最尤型推定量が以下のように構築されている. まず L_n を正整数列で, $L_n \rightarrow \infty, (L_n/\sqrt{n}) \vee (n^{1/4}/L_n) \rightarrow 0$ を満たすものとし, 観測区間全体 $[0, T]$ を L_n 個の区間に等分する. 等分された各区間において, ϵ_i^k が正規分布に従うと仮定すれば拡散過程 X が局所的に条件付正規分布で近似されることを用いて対数尤度関数を正規近似し, 各区間の近似対数尤度関数を足し合わせることで疑似対数尤度関数 $\mathcal{H}_n(\sigma)$ が構築される. 最尤型推定量は $\hat{\sigma}_n = \operatorname{argmax}_{\sigma} \mathcal{H}_n(\sigma)$ と計算される.

Ogihara (2018) において, μ, b の微分可能性や非退化性等の条件と, 観測時刻列 $\{S_i^n\}_i, \{T_j^n\}_j$ の漸近挙動に関する条件の下, 漸近混合正規性:

$$(0.2) \quad n^{1/4}(\hat{\sigma}_n - \sigma_*) \rightarrow^{s-\mathcal{L}} \tilde{\Gamma}^{-1/2} \mathcal{N}$$

が示された. ただし, $\tilde{\Gamma} = \text{P-lim}_{n \rightarrow \infty} (-\partial_{\sigma}^2 \mathcal{H}_n(\sigma_*)/\sqrt{n})$ である. 特にこの結果は ϵ_i^k が正規分布に従わない場合でも成立する.

参 考 文 献

- Genon-Catalot, V., Jacod, J. (1993): On the estimation of the diffusion coefficient for multi-dimensional diffusion processes, *Annals of Institute of Henri Poincare*, 29, 119-151.
- Gobet, E. (2001) Local asymptotic mixed normality property for elliptic diffusion : a Malliavin calculus approach. *Bernoulli*, 7, 899-912.
- Jeganathan, P. (1983) Some asymptotic properties of risk functions when the limit of the experiment is mixed normal. *Sankhya Ser. A*, 45, 66-87.
- Ogihara, T. (2015): Local asymptotic mixed normality property for nonsynchronously observed diffusion processes, *Bernoulli*, 21, 2024-2072.
- Ogihara, T. (2018): Parametric Inference for Nonsynchronously Observed Diffusion Processes in the Presence of Market Microstructure Noise, *Bernoulli*, 24, 3318-3383.
- Ogihara, T. and Yoshida, N. (2014): Quasi-likelihood analysis for nonsynchronously observed diffusion processes, *Stochastic Processes and their Applications*, 124, 2954-3008.

一般化平均による統計モデル

A statistical model via generalized mean

数理・推論研究系 江口 真透 (Shinto Eguchi),

1. 一般化平均と情報幾何

1930 年に, コルモゴロフと南雲によって独立に発表された一般化平均について考察する. 正の数 x, y に対する一般化平均は

$$(1.1) \quad \text{GM}_\phi(x, y) = \phi((1 - \pi)\phi^{-1}(x) + \pi\phi^{-1}(y))$$

と定まる. ここで $\phi: \mathbb{R} \rightarrow (0, \infty)$ は単調関数とする. 同様に x, y, z の一般化平均を考えると, 定義から $\text{GM}_\phi(x, \text{GM}_\phi(y, z)) = \text{GM}_\phi(\text{GM}_\phi(x, y), z)$ がいえる. 典型例の算術平均, 幾何平均, 調和平均などが含まれるが, ‘平均’ の持つべき公理から特徴付けられた一般化平均は生成関数 ϕ の関数自由度を持つので多様な平均の考えが展開できる.

統計学のために一般化平均を積極的に援用した方法を提案した. はじめに情報幾何との関連について考察し, つぎに統計モデリングのために幾つかの応用を紹介する. ロジスティック回帰や比例ハザードモデルにおける予測関数を一般化平均を使った準線形モデルを提案した. さらにクラスタリングのエネルギー関数やロス関数の混合のために一般化平均がキーになることが示された.

情報幾何は確率密度の関数空間の上に双対リーマン幾何をベースに豊かな直観を与え, 確率に関連する全ての分野へ幾何的考えを構築している. その中の基本定理として, この関数空間の上でのピタゴラス定理が挙げられる (Amari-Nagaoka, 2007). 確率密度関数 $p(x)$ と $q(x)$ を混合測地線でつなぎ, 一方で, $r(x)$ と $q(x)$ を指数測地線でつないだとき, この 2 つの測地線が $q(x)$ で直交するならば, またその時に限り,

$$(1.2) \quad D_0(p, r) = D_0(p, q) + D_0(q, r)$$

が成立する. ここで D_0 は KL ダンバージェンスとする. この性質から最尤推定と十分統計量の関係, 赤池情報量規準の妥当性などが導かれる. この考察において $r(x)$ と $q(x)$ をつなぐ指数測地線とは

$$(1.3) \quad \text{EG}(q, r) = \exp((1 - \pi) \log q(x) + \pi \log r(x) - \kappa(\pi))$$

と定められる. ここで $\kappa(\pi)$ は正規化定数とする. このように $\text{EG}(q, r)$ は生成関数 $\phi = \exp$ を使って正の数の代わりに密度関数に対する一般化平均と見れる. したがって, 一般の ϕ に

$$(1.4) \quad \text{EG}_\phi(q, r) = \phi((1 - \pi)\phi^{-1}q(x) + \pi\phi^{-1}r(x) - \kappa_\phi(\pi))$$

が定まり, これを一般化指数測地線と呼ぶ (Eguchi-Komori, 2015). 同様な考えから一般化 KL ダイバージェンスを導出すると, ピタゴラス定理が示される. 次の節では一般化平均を直接に統計モデリングに応用することを考察する.

2. 準線形ロジスティック回帰モデル

一般化平均 (1.1) は正の数の平均で在ったが, 実数 x と y に対しては

$$(2.1) \quad \text{RGM}_\phi(x, y) = \phi^{-1}((1 - \pi)\phi(x) + \pi\phi(y))$$

と定める. (1.1) の生成関数 ϕ の代わりに ϕ^{-1} を取られていることに注意する. 統計の応用としては実数の代わりに, 回帰関数, 予測関数, エネルギー関数, ロス関数などの実数値関数の一般化平均を考えることができる. 例えば, ロジスティック回帰において p 変数の説明変数 $X = x$ を与えたとき 2 値反応変数 y の条件付き確率関数を

$$(2.2) \quad p(y|x) = \frac{\exp\{yf(x)\}}{1 + \exp\{f(x)\}} \quad (y = 0, 1)$$

とする. 予測関数を一般化平均によって

$$(2.3) \quad f_\tau(x, \beta, \pi) = \frac{1}{\tau} \log \left(\sum_{k=1}^K \pi_k \exp(\tau \beta_k^\top x_k) \right)$$

とし, 準線形予測関数と呼ぶ. ここで τ は逆温度パラメータ, $x = (x_1, \dots, x_K)$ と $\beta = (\beta_1, \dots, \beta_K)$ は p 変数の同じ K 分割とする. Cf. Omae et al. (2017). 逆温度パラメータ τ を極限 ∞ を取ると $f_\tau(x, \beta, \pi) = \max_{1 \leq k \leq K} \beta_k^\top x_k$ となり, 極限 $-\infty$ を取ると $f_\tau(x, \beta, \pi) = \min_{1 \leq k \leq K} \beta_k^\top x_k$ となる. 極限 0 を取ると線形予測関数に帰着される.

$Y = 0$ のときの X の条件分布 $p(x|Y = 0)$ は正規分布 $N(\mu_0, \Sigma)$ に従っているが, $Y = 1$ のときの X の条件分布 $p(x|Y = 1)$ は混合正規分布 $\sum_{k=1}^K \pi_k^* N(\mu_k, \Sigma)$ に従っていると仮定する. これは $Y = 0$ サンプルは均一な母集団から得られたが $Y = 1$ サンプルは非均一な異質な母集団の混合から得られた状況を考えている. このとき, $\tau = 1$ のとき

$$(2.4) \quad f_\tau(x, \beta, \pi) = \log \frac{p(x|Y = 1)}{p(x|Y = 0)}$$

が成立する. このような仮定のもとでは, 準線形予測関数の判別の最適性が示される.

準線形予測関数の定義において説明変数 x の K 分割が必要であるが, これは教師なし学習によって構成できる. 典型的にはクラスター分析によって K 分割が求まる. またスパース学習を準線形モデル (2.3) をロジスティック回帰 (2.2) に代入したモデルの対数尤度関数に (2.3) のパラメータ β_k の積の L_1 ペナルティを入れたものを考える方法も有力である.

比例ハザードモデル, 分割クラスタリング, 混合ロス関数, モデル平均, メタアナリシス, 半教師学習などについても一般化平均を使うと興味深い展開ができると思われる. 線形モデルから準線形モデルへの拡張や, エネルギー関数やロス関数の混合について, 幾つかの新しい知見が得られた. 広い実用のためには未だ多くの解決すべき問題があるが, 線形モデルを柔軟に結合する統計方法として着実な完成に近い将来になされることが望まれる.

参考文献

Amari, S. I., & Nagaoka, H. (2007). *Methods of information geometry* (Vol. 191). American Mathematical Soc.

Eguchi, S., & Komori, O. (2015). Path connectedness on a space of probability density functions. In *International Conference on Networked Geometric Science of Information* (pp. 615-624). Springer, Cham.

Omae, K., Komori, O., & Eguchi, S. (2017). Quasi-linear score for capturing heterogeneous structure in biomarkers. *BMC bioinformatics*, 18(1), 308.

ガンマ・ダイバージェンスに基づいた ロバスト統計

Robust Statistics via Gamma-Divergence

数理・推論研究系 藤澤 洋徳 (Hironori Fujisawa)

1. はじめに

外れ値が存在するとき推定値にはバイアスが生じる。たとえば標本平均を考えよう。データの一つが非常に大きな値を取るとき、標本平均は非常に大きな値となってしまう。このバイアス問題を克服する単純な手法は中央値である。外れ値が一つであれば、中央値は大きく変わらない。ただし、外れ値の割合が大きい時には、中央値は大きくずれる。実は、外れ値の割合が大きい場合にも、バイアスを十分に小さくできる手法、というのは、長らくきちんと議論されていなかった。この問題はガンマ・ダイバージェンスによって解決される。ロバスト統計で 25 年以上未解決だった重要問題が解決されたと考えている (Fujisawa and Eguchi, 2008)。

2. 本研究の特徴

過去の研究との大きな違いは二つある。一つはバイアスの議論の仕方であり、もう一つは外れ値への意識の仕方である。

上述したバイアスと言うのは、厳密には、統計科学で通常使われているバイアスではない。厳密に言えば「潜在バイアス」と呼ばれるものである。この潜在バイアスを小さくすることはロバスト統計の最大の目的の一つである。しかしながら、潜在バイアスを直接に議論することが難しかったため、理論的には、影響関数のような代替指標を使って議論することが主流であった。これまで使われてきた代替指標を使わずに、潜在バイアスを直接に議論した点が、本研究の独創的な点である。

もう一つは潜在バイアスを議論するときの前提条件に対する意識の違いである。過去のロバスト統計では、前述した影響関数という、強力で便利な代替指標を使って議論することが主流であった。しかし、この道具を使うためには、外れ値の割合が小さいということを暗に想定している。影響関数を使いながら、外れ値の外れ値らしさが軽く登場してくる。本研究では意識レベルを逆にしている。外れ値の外れ値らしさの方を重視する。外れ値の割合が小さいという仮定は一切おいていない。この点も本研究の独創的な点である。

結果的に、外れ値の割合が大きくても潜在バイアスが小さいロバスト推定、が可能になった。それはガンマ・ダイバージェンス $D_\gamma(g, f)$ に基づく手法である：

3. ガンマ・ダイバージェンス

ガンマ・ダイバージェンス $D_\gamma(g, f)$ は以下で与えられる：

$$d_\gamma(g, f) = -\frac{1}{\gamma} \log \int g(x)f(x)^\gamma dx + \frac{1}{1+\gamma} \log \int f(x)^{1+\gamma} dx,$$

$$D_\gamma(g, f) = d_\gamma(g, f) - d_\gamma(g, g).$$

ここで、 g と f は密度関数であり、 $\gamma > 0$ はロバスト性を調整するパラメータである。実際のパラメータ推定は、 g に基づく期待値を経験密度関数 \bar{g} で置き換え、 f をパラメトリックモデル f_θ に置き換えて、 $d_\gamma(\bar{g}, f_\theta)$ の最小化で行う。

図 1 はガンマ・ダイバージェンスに対して成立する近似的なピタゴリアン構造である。詳細は省くが、外れ値が外れ値らしいという仮定を置いている。このピタゴリアン構造は、ガンマ・ダイバージェンスに基づくロバスト推定が、なぜ上手く働くかの直感的な説明を与える。データ発生分布 g とパラメトリック分布 $h = f_\theta$ のダイバージェンス $D_\gamma(g, h)$ を小さくするのが普通の方法である。ここで次の二点に注意する：(i) 直交性が近似的に成り立っている。(ii) g と f は固定されていて自由に動けるのはパラメトリック分布 h だけである。そうすると、 $D_\gamma(g, h)$ を小さくしようとする、 $D_\gamma(f, h)$ が小さくなり、結果的に、パラメトリック分布 h はターゲット分布 f に近づく。これがガンマ・ダイバージェンスに基づいたロバスト推定が上手く働く理由である。

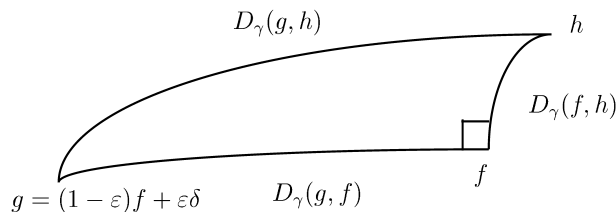


図 1. ピタゴリアン構造。 g はデータ発生分布であり、ロバスト統計で慣習的に用いられている汚染分布 $g(x) = (1 - \epsilon)f(x) + \epsilon\delta(x)$ を想定。 ϵ は外れ値の割合。 f はターゲット分布（外れ値に汚染されなかった場合の分布）。 δ は外れ値の分布。 h はパラメトリック分布 f_θ 。

また、ある種の仮定の下では、外れ値の割合が大きくても潜在バイアスが小さくなる手法は、本質的に、ガンマ・ダイバージェンスに基づく手法だけであるという、ある種の唯一性が証明できる。加えて、パラメトリックモデルが指数型分布族に入っている場合には、ピタゴリアン構造を利用して、ロス関数の単調減少性をもつきれいなパラメータ推定アルゴリズムも提案できる。

4. 発展

唯一性の定理の仮定は弱く、ガンマ・ダイバージェンスはある種の決定打である。そのため、研究の発展は難しい状態が続いたが、最近になって研究が進むようになった。パラメトリック分布を、面積を変化させるパラメータ λ をわざと組み込んだ拡張モデル $\lambda f(x; \theta)$ に代えることで、ガンマ・ダイバージェンスでなくても、外れ値の割合が大きくても潜在バイアスが小さくなる手法を構築することができた (Kanamori and Fujisawa, 2015)。この研究と関連研究で、日本統計学会研究業績賞を頂いた。当初に提案したパラメータ推定アルゴリズムは、ピタゴリアン構造を利用してきれいであったが、スパース罰則と組み合わせると、使いやすしいパラメータ推定アルゴリズムの構築が容易でなく、ロバスト性とスパース性を同時に併せもつ手法の提案が難しかった。この問題は、Majorization-Minimization アルゴリズムの適用で克服されて、ガウシアン・グラフィカル・モデリングや回帰モデリングに適用されている (Hirose, Fujisawa and Sese, 2015; Kawashima and Fujisawa, 2017)。Google Scholar などで検索することで、その他にも様々な発展をして注目を浴びていることを見取ることができる。

統計的自然言語処理と統計学

Natural language processing and Statistics

数理・推論研究系 持橋 大地 (Daichi Mochihashi)

1. 「静的な統計」から「動的な統計」へ

2011年に統数研に准教授として着任後、2016年の『統計数理』64巻2号において、特集「統計的自然言語処理の現在」を企画・担当した。言語に関しては『統計数理』では2000年の48巻2号で特集「“ことば”新研究」が組まれており、二者を比較すると、16年の間に大きな変化があったことに気づく。2000年の特集では文章のジャンルの分類や編集距離による類似和歌の発見、多変量解析による文書の因子分析などが主な内容であったが、2016年では構文構造の教師なし学習、言語変化への統計的アプローチ、CRF(条件付確率場)の詳細な解説など、内容が大きく様変わりしている。統計的な手法としても、前者が多変量解析をベースとしているのに対し、後者では系統樹の統計モデル、ポアソン過程、条件付確率場(ロジスティック回帰の隠れマルコフモデル化)、階層ベイズモデルのように最新の幅広い統計あるいは機械学習の手法が取り入れられるようになった。

手法以外にも、研究の哲学ともよぶべきものが、「静的な統計」から「動的な統計」へと変わったように感じられる。前者では、文や文章は確定した「モノ」であり、それをどう扱うかがテーマとなっていたが、後者では言語自体も変化するものであり、内部に多くの文脈依存性を持っていることがフォーカスされている。例えばCRFは文の構文解析や、各単語の品詞が互いに依存するマルコフ確率場のモデルであり、特集に含まれている文の読み時間の動的な推定やTwitterのツイートのモデルも、時間依存性や空間依存性を含んでいる。

現在、「動的」とは主に、ある文や文章内の現象を動的にモデル化しているが、上の言語系統樹の話にあるように、世代を超えて言語自体が時間的・空間的に変わりうる様相を統計的に明らかにすることにも今後取り組みたく、国立国語研究所や国立民族学博物館との共同研究を現在行っている。

2. 単語と分節化の理論

英語のように単語に分かれていない¹日本語や中国語、タイ語のような言語にとって、文を「単語」に分けることは最も基礎となる重要な課題である。従来はこのために、人手で準備した大量の「正解」の単語列をもとにCRFなどを学習することで単語分割や品詞推定がなされてきたが、「正解」が真に正解であるという保証はなく、また日々無数に生まれる新語には対応できないという限界がある。

一の皇子は、右大臣の女御の御腹にて、寄せ重く、疑ひなきまうけの君と、世にもてかしづききこゆれど、この御にほひには並びたまふべくもあらざりければ、おほかたのやむごとなき御思ひにて、この君をば、私ものに思ほしかしづきたまふこと限りなし。はじめよりおしなべての上宮仕したまふべき際にはあらざりき。

図1.『源氏物語』の教師なし形態素解析の結果の一例。辞書や文法は一切用いていない。

¹ ラテン語や英語も、もともとは単語を分けて書かず一続きに書くのが普通であった。

これに対し、ノンパラメトリックベイズ法による文字-単語の階層ベイズモデルを仮定し、出力された生の文字列のみから「単語」を逆に推定する統計モデルを 2009 年頃に発表した (Mochihashi et al., 2009). これにより、例えば「源氏物語」の文字列のみから単語を図 1 のように推定できる. その後の統数研での共同研究で、さらに品詞を同時に教師なし推定したり (Uchiumi et al., 2015), 教師データも利用した半教師あり学習も高精度で行えるようになった (Fujii et al., 2017).

統計的には、これはセミマルコフモデルの一種による分節化であると考えられる. したがって、同様の統計モデルを言語だけでなく、他の時系列データにも適用することができる. 言語において文字列にあたる出力をガウス過程からの波形とすれば、ロボティクスにおいてロボットの関節角の時系列を分節化して「動作」を取り出すことも可能になった (Nagano et al., 2018). 音声認識においても、音声から単語を自動認識することが可能になるため、共同研究を行っている.

3. 離散データと「科学の科学」

離散データを扱う統計的自然言語処理は、他の多くの分野と繋がりを持っている. 2015 年から現在まで研究員を務めている日本学術振興会の学術情報分析センターでは、科研費の申請に査読者が付した“5”、“2”などの審査点から、項目反応理論により図 2 のように正規分布に従う各申請のスコアを統計的に算出する研究を行った. 多次元で行えば、矢印で示した各審査員の「評価軸」も客観的なデータから数学的に明らかになり、公正な審査に貢献する. また、大規模な確率的潜在意味解析 (LDA) を用いて、学振に登録されている研究者の専門性を計算し、審査の際に適切な審査員を推薦する試みも行っている.

科学的な発見は論文、つまり言語の形で発表される. 論文の生産自体も離散的なイベントであり、点過程として捉えることができる. また、そこには複雑な相互依存関係があると考えられる. 自然言語処理の背景を生かし、こうした「科学の科学」へも今後取り組みたいと考えている.

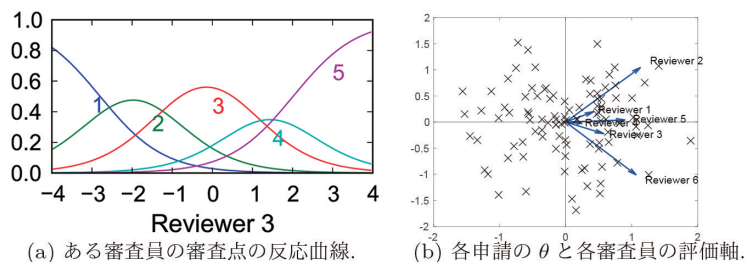


図 2. 項目反応理論による潜在的な得点と評価軸の計算

参 考 文 献

- Fujii, R., Domoto, R. and Mochihashi, D. (2017). Nonparametric Bayesian Semi-supervised Word Segmentation, *Transactions of ACL*, 5, 179–189.
- Mochihashi, D., Yamada, T. and Ueda, N. (2009). Bayesian Unsupervised Word Segmentation with Nested Pitman-Yor Language Modeling, *Proceedings of ACL-IJCNLP 2009*, 100–108.
- Nagano, M., Nakamura, T., Nagai, T., Mochihashi, D., Kobayashi, I. and Kaneko, M. (2018). Sequence Pattern Extraction by Segmenting Time Series Data Using GP-HSMM with Hierarchical Dirichlet Process, *IROS 2018*, 4067–4074.
- Uchiumi, K., Tsukahara, H. and Mochihashi, D. (2015). Inducing Word and Part-of-speech with Pitman-Yor Hidden Semi-Markov Models, *ACL-IJCNLP 2015*, 1774–1782.

メタアナリシスにおける公表バイアスの最悪評価に基づく感度解析

Worst-case sensitivity analysis for publication bias in meta-analysis

数理・推論研究系 逸見 昌之 (Masayuki Henmi)

1. 医学研究におけるメタアナリシス

同じ目的あるいは何らかの共通点を持った複数の研究から得られた統計解析の結果を統合して、よりエビデンスの高い結果を得るための統計解析のことをメタアナリシスと呼ぶ。医学研究における最も典型的な応用例は、医薬品の効果を検証するための臨床試験や疾患等の原因を探る疫学研究に対するメタアナリシスであり、例えば臨床試験であれば、ある疾患に対する新薬 A と既存薬 B とを比較するランダム化臨床試験の結果が複数ある際に、個々の研究では新薬 A の効果を支持する十分なエビデンスが得られなかったとしても、複数個の結果を統合することによって十分なエビデンスが得られることがある。

2. 公表バイアスの問題

一般に統計解析においては、観測データが関心のある母集団から偏りなく得られていることが重要であるが、メタアナリシスの場合は、観測データとなる複数の研究結果は公表されている論文等からしか得られないので、データに偏りが生じやすくなる。例えば、臨床試験や疫学研究などにおける二群比較では、統計的検定で有意な結果が公表されやすく、それらの結果だけを集めてメタアナリシスを行えば、当然その結果も有意となる。このように、公表データの偏りによりメタアナリシスの結果に生じるバイアスのことを公表バイアス (publication bias) と呼ぶ。図 1 は、ある薬剤を妊婦に投与した場合に早産を予防できるかどうかを検証するために行われた、複数のランダム化臨床試験の結果をプロットしたものである。横軸は各試験での対数オッズ比の推定量、縦軸はその標準誤差の逆数を表している。図の下の方ほど標準誤差が大きくなり、検定結果が有意になりにくくなるが、その部分で片側が欠けていることから、公表バイアスの存在が強く疑われる。このような図は一般にファンネルプロット (funnel plot) と呼ばれ、公表バイアスの存在の可能性を視覚的に検討するためによく用いられる。

3. 最悪評価による感度解析

公表バイアスを調整して妥当な統合結果を得るためには、個々の研究結果の公表のプロセスに関する強い仮定が必要であるが、それを観測される研究結果から検証することは不可能である。そこで、その仮定を有り得る範囲内で変化させて統合結果がどのように変わるかを調べるといった感度解析を行うことが推奨されているが、その仮定をどのように変化させるかというは難しい問題である。この問題に対し我々は、標準誤差が大きい推定量を有する研究ほど結果が公表されにくいという、定性的な弱い仮定のみを用いて、統合結果として得られる信頼区間

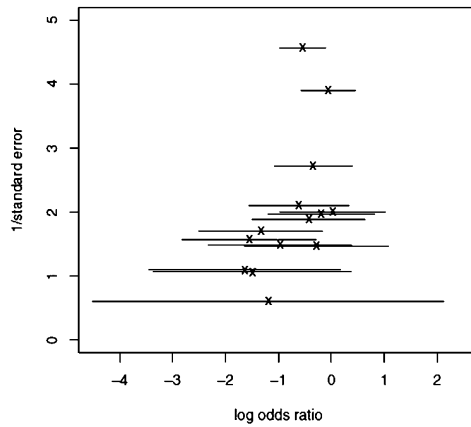


図 1. 臨床試験のファンネルプロット

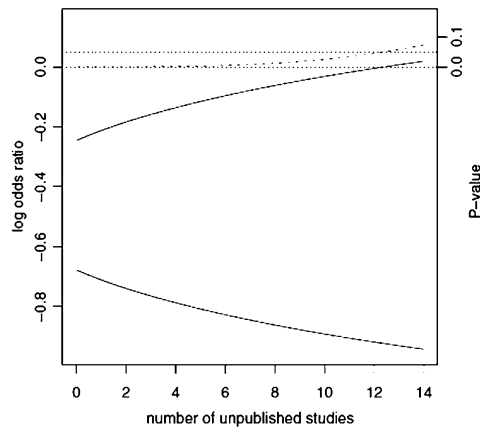


図 2. 信頼区間の存在範囲 (実線) と P-値の上限 (破線)

と P-値の最悪評価を行った。図 2 は、図 1 のデータを用いたメタアナリシスに対して、未公表の研究数ごとに、統合結果から得られる信頼区間の存在範囲 (実線) と P-値の上限 (破線) を表しており、この図から未公表の研究数が 13 を超えると統合結果としての検定の有意性が逆転する可能性があることが分かる。

参 考 文 献

Henmi, M., Copas, J.B. and Eguchi, S. (2007). Confidence Intervals and P-values for Meta-analysis with Publication Bias. *Biometrics* **63**, 475-482.

スピングラス理論による制限等長定数評価

Evaluation of restricted isometry constants using spin-glass theory

数理・推論研究系 坂田 綾香 (Ayaka Sakata)

要 旨

制限等長定数は、圧縮センシングにおける完全再構成条件を与えるが、その厳密評価は困難である。我々は統計物理学におけるスピングラス理論を用いて、制限等長定数を制度よく見積もる方法を提案した。

キーワード：圧縮センシング、制限等長性

原信号がゼロ成分を多く持つ (スパース性) という事前知識の下で、次元より少ない観測から信号を復元する枠組みである圧縮センシング (CS) は、様々な分野において利用されている。 $\mathbf{A} \in \mathbb{R}^{M \times N}$ を観測行列とすると、線形観測 $\mathbf{y} = \mathbf{A}\mathbf{x}$ から $S (< N)$ 個の非ゼロ要素を持つ信号 \mathbf{x} (S スパースベクトル) を再構成する問題として CS は定式化される。 ℓ_0, ℓ_1 再構成法は代表的手法であり、 S スパースベクトル \mathbf{x} が完全復元される十分条件は制限等長定数により与えられる [Candés and Tao (2005)]。ここでは、観測行列 $\forall \mathbf{A} \in \mathbb{R}^{M \times N}$ はコラムが典型的に $(\mathbf{A}^T \mathbf{A})_{ii} = 1$ ($i \in \{1, \dots, N\}$) を満たすように規格化されているとする。全ての S スパースベクトル \mathbf{x} に対して次の不等式が成立するとき、行列 \mathbf{A} は制限等長性を満たす。

$$(0.1) \quad (1 - \delta_S^{\min}) \|\mathbf{x}\|_F^2 \leq \|\mathbf{A}\mathbf{x}\|_F^2 \leq (1 + \delta_S^{\max}) \|\mathbf{x}\|_F^2$$

また $0 < \delta_S^{\min} \leq \delta_S^{\max}$ を制限等長定数 (RIC) と呼ぶ。RIC が与える ℓ_0, ℓ_1 再構成による完全復元条件は [Candés et al. (2006)] 等に示されている。

RIC は \mathbf{A} のグラム行列の固有値と関係づけられる。 S スパースベクトル \mathbf{x} の非ゼロ要素の位置を $T \subseteq V = \{1, \dots, N\}$, $|T| = S$ として表現し、 \mathbf{A} の $i \in T$ コラムからなる行列を \mathbf{A}_T 、また $\mathbf{x}_T = \{x_i | i \in T\}$ とすると、 $\mathbf{A}\mathbf{x} = \mathbf{A}_T \mathbf{x}_T$ である。そして全ての T について、次の不等式が成立する。

$$\lambda_{\min}(\mathbf{A}_T^T \mathbf{A}_T) \|\mathbf{x}_T\|_F^2 \leq \|\mathbf{A}_T \mathbf{x}_T\|_F^2 \leq \lambda_{\max}(\mathbf{A}_T^T \mathbf{A}_T) \|\mathbf{x}_T\|_F^2$$

$\lambda_{\min}(\mathbf{B})$, $\lambda_{\max}(\mathbf{B})$ は \mathbf{B} の最小・最大固有値を表し、上付きの T は転置を表す。

(0.1) との比較から、 $\lambda_{\min}^*(\mathbf{A}; S) = \min_{T: T \subseteq V, |T|=S} \lambda_{\min}(\mathbf{A}_T^T \mathbf{A}_T)$ および $\lambda_{\max}^*(\mathbf{A}; S) = \max_{T: T \subseteq V, |T|=S} \lambda_{\max}(\mathbf{A}_T^T \mathbf{A}_T)$ を用いて、RIC は次のように表現される。

$$(0.2) \quad \delta_S^{\min} = 1 - \lambda_{\min}^*(\mathbf{A}; S), \quad \delta_S^{\max} = \lambda_{\max}^*(\mathbf{A}; S) - 1$$

(0.2) を厳密に評価するには、あらゆる T について固有値を評価しなくてはならない。これは計算量的に困難であるため、様々な近似方法が考えられてきた [Bah and Tanner (2010)]。我々は統計物理学におけるスピングラス理論を用いて、RIC の評価を改善することに成功した。そ

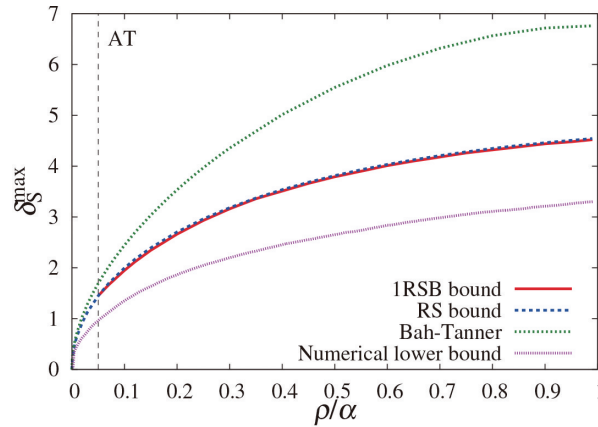


図 1. $\alpha = 0.5$ での RIC の ρ/α 依存性。

の際、 $M = N\alpha$, $S = N\rho$ ($\alpha, \rho \sim O(1)$) として $N \rightarrow \infty$ 極限を考えた。

図 1 は、観測行列がガウシアンランダム行列の場合の δ_S^{\max} について、先行研究と我々の評価法を比較したものである。垂直点線の ρ/α において、レプリカ対称性の破れに伴った相転移現象が起きる。"Numerical lower bound"とは数値的に評価された RIC の下界であり、RIC がこれより大きいことを保証している。"Bah-Tanner"は先行研究による RIC の上界評価である。我々の"1RSB bound", "RS bound"も上界評価であるが、先行研究よりも精度の高い上界評価となっている。1RSB bound とは、レプリカ対称性の破れを 1 段階考慮した評価となっており、逐次的に対称性の破れを考慮していくことで上界評価が正しい値に近づいていくことが数学的に保証される。また δ_S^{\min} についても精度の良い評価を達成し、更にランダム一次転移や Gardner 転移と呼ばれる相転移現象が見られることも発見した。具体的評価法は [Sakata and Kabashima (2015)] に示されている。

提案手法は、ランダム直交行列から構成された観測行列に対しても適用可能である。様々な行列に対して RIC を評価することで、効率的観測方法などが提案できると考えられる。

謝 辞

本研究は樺島祥介氏 (東京工業大学) との共同研究である。

参 考 文 献

- Bah, B. and Tanner, J. (2010). Improved Bounds on Restricted Isometry Constants for Gaussian Matrices, *SIAM Journal on Matrix Analysis and Applications*, **31**, 2882–2898.
- Candès, E. J. and Tao, T. (2005). Decoding by linear programming, *IEEE Transactions on Information Theory*, **51**, 4203–4215.
- Candès, E. J., Romberg, J. and Tao, T. (2006). Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information, *IEEE Transactions on Information Theory*, **52**, 489–509.
- Sakata, A. and Kabashima, Y. (2015). Replica Symmetric Bound for Restricted Isometry Constant, *Proceedings of IEEE International Symposium on Information Theory*, **2015**, 2006–2010.

多層整数計画に基づくクリンチ／エリミネーション数の計算

Calculation of Clinch and Elimination Numbers Based on Multilayered Integer Programming

数理・推論研究系 伊藤 聡 (Satoshi Ito)

1. はじめに

リーグスポーツのシーズン中のどの時点においても、最終的にリーグ優勝やプレーオフ出場権など特定の状況（指標）が達成されることが確定する最小の勝ち試合数（クリンチ数），もしくは逆にその状況（指標）に届かないことが確定する最小の負け試合数（エリミネーション数）が存在する。本研究は，順位決定に係る複数の判定基準が存在する場合のクリンチおよびエリミネーション数の計算を，多層の整数計画問題を解くことにより高速に行う汎用的な枠組みを開発することを目的としている。

2. 整数計画モデルと上下界の多層構造

リーグに属するチームの集合を L とし，そのチーム数を n とする。リーグ L の全チーム間のこれまでの勝敗記録と残り試合数が与えられているとし， w_{ij} をチーム $i \in L$ のチーム $j \in L$ に対する現時点での勝数， g_{ij} をチーム i, j 間の残り試合数とする。チーム $i \in L$ のチーム $j \in L$ に対する今後の勝数を x_{ij} と表すことにすると，

$$X := \{ x = (x_{ij}) \in \mathbb{Z}^{n \times n} \mid x_{ij} + x_{ji} \leq g_{ij}, x_{ii} = 0, x_{ij} \geq 0 \ (\forall i, j \in L) \}$$

は今後起こり得る勝敗に関するシナリオを過不足無く与える（ \mathbb{Z} は整数の集合であり，引分がない場合は第 1 式の不等号を等号に変えるものとする）。

順位判定基準が複数あり，そのうち最初の m 個が勝敗数のみに基づく基準（すなわち $m+1$ 番目の基準は得失点差など勝敗数以外に基づくもの）であるとする。このときチーム $a \in L$ の例えば（簡単な例として）第 k 位クリンチ数は，チーム a が今後引き分けることなく，かつ m 個のうちのいずれかの基準でチーム a より上位の成績を持つチームが k 個存在するという制約条件のもとで，チーム a の今後の勝数 $\sum_{j \in L} x_{aj}$ の最大値 \bar{z} を求めることにより得られる（この最大化問題に許容解がなければ既に第 k 位以上が確定していることになるし， \bar{z} が残り試合数と等しければ今後全勝しても第 $k+1$ 位以下の可能性があることになり，そうでなければ第 k 位クリンチ数は $\bar{z}+1$ となる）。ここで注意すべきことは， m 個の順位判定のうち最初の $m-1$ 個は等号なし（ $<$ ）であり，最後の m 番目のみが等号つき（ \leq ）の順位判定となることである。例えば 2016 年に始まった男子プロバスケットボールの B.LEAGUE では，①勝率（勝数）→ ②当該クラブ間勝率（当該クラブ間 1 試合平均勝数）→ ③当該クラブ間得失点差 → ④当該クラブ間 1 試合平均得点 → ⑤得失点差 → ⑥ 1 試合平均得点（総得点）→ ⑦理事会が必要と判断した場合に抽選と判定基準が続くが，この場合 $m=2$ であり，判定基準②に関して同成績であっても③以降の基準により順位が変わる可能性が残っているため，チーム a にとってより都合の悪い状況を想定するという意味で②に対する順位判定は等号つきである必要がある。

上記の例のような整数計画モデルは当該チーム間成績が絡む場合などには非線形となり、汎用最適化パッケージを用いて直接解くことは必ずしも容易でない。そこで、最大目的関数値の上界と下界をうまく利用することを考える。図 1 のように、 $m - 1$ 番目までの順位判定のうちいずれか一つの不等号 ($<$) を等号つき (\leq) に変えることにより $m - 1$ 個のレベルの上界が得られ、また、1 番目を除く順位判定をそれ以降を含めて取り去る（すなわち順位判定を途中で打ち切る）ことにより $m - 1$ 個のレベルの下界が得られる。さらに、 m 番目の順位判定を等号なし ($<$) に変えて得られる、よりタイトな下界を用いることにより、勝敗数以外の要素に基づく $m + 1$ 番目以降の判定基準を考慮する必要性の有無も判定することができる。

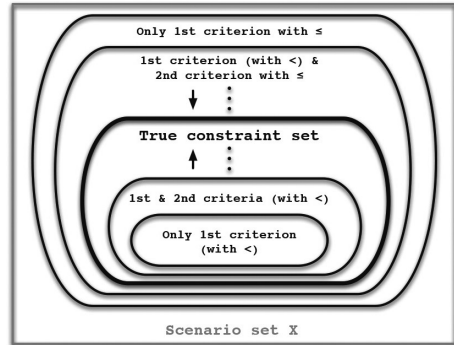


図 1: 上下界の多層構造

3. 数値実験

B.LEAGUE において考えられる指標のうち、B1 チャンピオンシップ・トーナメント進出のエリミネーション数について、2016-2017 シーズン第 15 節以降の 50 日分の勝敗記録に対して数値実験を行った結果を図 2 に示す。全 18 クラブのエリミネーション数を計算するのに要した個々の時間を一日単位で箱ひげ図にプロット（四分位点から 1.5 IQR を超える外れ値を点で表示）しているが、上下界の情報を使わずに解いた (a) に対して、これを活用した (b) では、極めて時間がかかるケースを抑えることにより平均計算時間を短縮することに成功している。

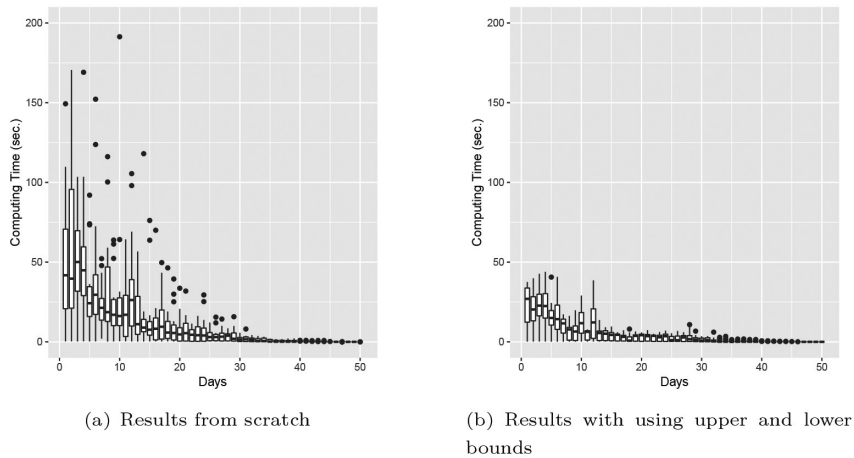


図 2: B1 チャンピオンシップ・トーナメント進出エリミネーション数の計算時間

参 考 文 献

Ito, S. and Shinano, Y. (2018). Calculation of clinch and elimination numbers for sports leagues with multiple tiebreaking criteria, ZIB-Report 18-51, urn:nbn:de:0297-zib-70591.

天文学とデータ科学

Astronomy and Data Science

数理・推論研究系 池田 思朗 (Shiro Ikeda)

要 旨

深層学習の成功をきっかけとして、データ科学分野へ大きな注目が集まっている。新たなデータ処理の方法は産業を通じた社会への貢献だけでなく、自然科学の分野にも影響を及ぼすはずである。生物学ではこうした流れからバイオインフォマティクスが生まれた。天文学ではまだそうした大きな流れは生まれていないが、今後はデータ科学によって大きく変わっていくはずである。

キーワード：天文学, データ科学, Event Horizon Telescope, すばる望遠鏡.

1. はじめに

天文学では、観測結果がデジタルデータとして保存されるようになって以来、計算機上で統計的な解析が行われてきた。しかし、これからの天文観測では爆発的にデータが増えていくため、新たなデータ解析手法の導入は必要不可欠である。今後の天文学は、データ科学との有機的な共同研究が必要となる。

この数年、天文学者と議論を重ねてきた経験から、データ科学から天文学への貢献の可能性には大きくふたつの方向性があると考えている。ひとつは新たな方法による解析精度の向上だろう。限られたデータからデータ科学の方法によってより多くの情報を得ようという方向性である。もうひとつはビッグデータへの対応である。今後、観測機器の高性能化によって高精度なデータが非常に多く得られるようになると、これまで天文学者が解析に用いてきた方法では対応しきれない。ここにも新しいデータ科学の方法が求められている。

2. 電波干渉計とスパース推定

電波干渉計は天体から発せられる電波を複数のアンテナで受信し、相関処理をしたのちに天体のイメージを得ようというものである。離れた位置にあるアンテナの信号に相関処理を行えば、光の干渉と同じように画像のフーリエ変換に対応する情報を得られる。したがって、この逆問題は理想的には逆フーリエ変換によってイメージングできるはずである。しかし、アンテナ数は限られており、一般にフーリエ空間上の観測点の数はイメージのピクセル数に比べて少ない。このため、電波干渉計のイメージングの問題は不良設定問題である。

条件が足りない不良設定問題に対して、データ科学ではこの 20 年、LASSO や圧縮センシングといったスパース推定の手法が開発された。例えば画像上で局所的に光源が分布しているコンパクトな天体の場合、こうした方法は有効である。我々はブラックホールシャドウの撮像のためのプロジェクト Event Horizon Telescope (EHT) に参加していて (Honma et al., 2014;

Ikeda et al., 2016; Akiyama et al., 2017), スパース推定は EHT に欠かせない方法のひとつになっている。こうした貢献はまさに、新たな方法による解析精度の向上であろう。

3. すばる望遠鏡の観測データからの超新星の選別

もうひとつ、現在共同研究を行っているのは、すばる望遠鏡の超広視野主焦点カメラ (HSC: Hyper Suprime-Cam) を用いたサーベイ観測「すばる戦略枠プログラム」に関するプロジェクトである。このプログラムは 2014 年に開始され、5 年間で 300 晩の観測を行う。最終的に観測データはペタバイトのオーダーになると見積られている。そこで期待されている成果のひとつは、遠方で発生する Ia 型超新星を数多く検出することである。これにより、宇宙論パラメータの推定精度が向上すると考えられる。

突然現れる超新星探索の発見は、ある日の画像から以前に撮った画像を差し引き、差分イメージに引き残された超新星を見つけ出すことになる。これまでは差分から超新星の候補となる画像パッチを自動的に取り出し、それを人間が本物か偽物か判定を行っていた。しかし、HSC の差分画像では、一晩に観測される超新星候補の数は数万以上に登り、その中にある目的の超新星は高々数十個しかないと見積もられている。このため、人間の目に頼るのではなく、本物・偽物の判定結果を返す関数を機械学習の方法によって作成し、実際の観測に用いている (Morii et al., 2016)。この研究は、これまでの天文学の方法をビッグデータでも引き続き行うための対応である。

4. まとめ

天文学は、歴史の初期から存在する最も古い学問のひとつである。歴史的にみれば、常に最先端の技術が投入されてきた。現在大きく発展しているデータ科学の方法が取り入れられるのも当然の流れである。

ここに挙げたように、天文学からデータ科学への期待は、より多くの情報を引き出し、来たるビッグデータに対応するという 2 点になるだろう。その先、データ科学が主導して天文学へ新たな提案をすることができれば、天文データ科学と呼べる分野が確立するだろう。

これまで天文学に関するさまざまなデータ解析の相談を受けてきた。今後 10 年の間に起きる天文学の変化の力になりたいと考えている。

参 考 文 献

- Akiyama, K., Ikeda, S., Pleau, M., Fish, V. L., Tazaki, F., Kuramochi, K., Broderick, A. E., Dexter, J., Mościbrodzka, M., Gowanlock, M., Honma, M. and Doeleman, S. S. (2017). Superresolution Full-polarimetric Imaging for Radio Interferometry with Sparse Modeling, *The Astronomical Journal*, **153** (4), 159(12pages).
- Honma, M., Akiyama, K., Uemura, M. and Ikeda, S. (2014). Super-resolution imaging with radio interferometry using sparse modeling, *Publications of Astronomical Society of Japan*, **66** (5), 95(14pages).
- Ikeda, S., Tazaki, F., Akiyama, K., Hada, K. and Honma, M. (2016). PRECL: A new method for interferometry imaging from closure phase, *Publications of Astronomical Society of Japan*, **68** (3), 45(9pages).
- Morii, M., Ikeda, S., Tominaga, N., Tanaka, M., Morokuma, T., Ishiguro, K., Yamato, J., Ueda, N., Suzuki, N., Yasuda, N. and Yoshida, N. (2016). Machine-learning Selection of Optical Transients in Subaru/Hyper Suprime-Cam Survey, *Publications of Astronomical Society of Japan*, **68** (6), p. 104(8pages).

制約付き非凸スパース最適化問題に対する DC アルゴリズム

DC Algorithm for Constrained Nonconvex Sparse Optimization

数理・推論研究系 田中 未来 (Mirai Tanaka)

1. 制約付き非凸スパース問題

近年, 統計数理のさまざまな分野において Lasso をはじめとするスパース最適化の研究が進められている. 本稿では制約領域 $S \subseteq \mathbb{R}^n$ 上における損失関数 $l(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ とスパース正則化関数 $r(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ の和を最小化する問題

$$(1.1) \quad \text{最小化 } l(\mathbf{x}) + r(\mathbf{x}) \quad \text{制約条件 } \mathbf{x} \in S$$

に対する効率のよいアルゴリズムを提案した著者らの論文 (Tanaka and Takeda, 2018) を紹介する.

以下では l, r, S について以下のような仮定をおく. まず l については連続的微分可能であることと, ある $L \in \mathbb{R}_+$ が存在して $(L/2)\|\mathbf{x}\|_2^2 - l(\mathbf{x})$ が凸関数となることを仮定する. 次に r については連続であることと, ある $\lambda \in \mathbb{R}_+$ が存在して $\phi(\mathbf{x}) := \lambda\|\mathbf{x}\|_1 - r(\mathbf{x})$ が凸関数となることを仮定する. 後者の仮定は多くのスパース正則化関数が満たすものである. さらに $l+r$ が下に有界であることと S が空でない凸集合であることを仮定する. 以上の仮定をおいたとしても問題 (1.1) は符号制約付き回帰, 主成分分析, 標準単体上の最小 2 乗問題, C -SVM の双対問題などを含むさまざまな問題のスパースな解を求める問題に対応する. しかしながら問題 (1.1) は制約付き非凸最適化問題であり, 一般に大域的最適解を求めることは難しい. 以下では問題 (1.1) の停留点を効率よく求める DC アルゴリズムについて述べる.

2. DC アルゴリズムとその各反復で解く子問題

2 つの凸関数 $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}$ について, これらの差を最小化する問題

$$(2.1) \quad \text{最小化 } g(\mathbf{x}) - h(\mathbf{x})$$

を DC 最適化問題と呼ぶ. この問題の停留点を求めるために DC アルゴリズム (アルゴリズム 1) がよく用いられる. このアルゴリズムは緩やかな仮定の下で DC 最適化問題の停留点に収束する点列を生成する.

アルゴリズム 1 問題 (2.1) に対する DC アルゴリズム

- 1: 適当な初期点 $\mathbf{x}^{(0)} \in \mathbb{R}^n$ をとる.
 - 2: **for** $t = 0, 1, \dots$ (収束するまで)
 - 3: h の劣勾配 $\mathbf{s}^{(t)} \in \partial h(\mathbf{x}^{(t)})$ を求める.
 - 4: $\mathbf{x}^{(t+1)} \in \operatorname{argmin}_{\mathbf{x}} \{g(\mathbf{x}) - (\mathbf{s}^{(t)})^\top \mathbf{x}\}$ と更新する.
-

問題 (1.1) を DC 最適化の枠組みで解くために, 次のように凸関数 g, h を定める:

$$g(\mathbf{x}) = \frac{L}{2} \|\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 + \delta(\mathbf{x} | S), \quad h(\mathbf{x}) = \frac{L}{2} \|\mathbf{x}\|_2^2 - l(\mathbf{x}) + \phi(\mathbf{x}).$$

ここで $\delta(\mathbf{x} | S)$ は $\mathbf{x} \in S$ のときに 0, そうでないときに $+\infty$ をとる関数である. このように g, h を定めると問題 (1.1) を DC 最適化問題 (2.1) に帰着できる.

問題 (1.1) に対する DC アルゴリズムの各反復では次の子問題を解く:

$$(2.2) \quad \text{最小化} \quad \frac{1}{2} \|\mathbf{x} - \mathbf{v}\|_2^2 + \gamma \|\mathbf{x}\|_1 \quad \text{制約条件} \quad \mathbf{x} \in S.$$

この子問題を高速に解くことができるとき, DC アルゴリズムは全体として高速なものとなる.

著者らは次の 4 つの場合について子問題 (2.2) の最適解ないしその近似が効率よく計算できることを示した. 1 つ目は S が超直方体 $\{\mathbf{x} : \mathbf{x} \leq \mathbf{x} \leq \bar{\mathbf{x}}\}$ の場合である. この場合は簡単な解析により子問題 (2.2) の最適解を陽に書き下すことができる. また, S が ℓ_2 ノルム球 $\{\mathbf{x} : \|\mathbf{x}\|_2 \leq 1\}$ の場合についても最適解を陽に書き下すことができる. S が標準単体 $\{\mathbf{x} : \mathbf{1}^\top \mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\}$ の場合, 子問題 (2.2) は \mathbf{v} の標準単体への射影の計算となるので, 既存の $O(n)$ 時間アルゴリズムを用いることができる. 著者らが解析した S の中で最も興味深いものは箱型制約に単一線形制約が付加した $\{\mathbf{x} : \mathbf{a}^\top \mathbf{x} = b, \mathbf{x} \leq \mathbf{x} \leq \bar{\mathbf{x}}\}$ である. 著者ら是对応する子問題に対する 2 分探索に基づく多項式時間アルゴリズムを構築した.

3. 計算機実験

制約付き非凸スパース最適化問題に対する既存の DC アプローチ (Tono et al., 2017) との比較実験の結果の一部として, ある問題例をそれぞれの手法で解いたときの解パスの比較を図 1 に示す. 既存手法はパラメータ λ の値を大きくとつても厳密にスパース解を出力したとは言えないのに対し, 著者らによる提案手法は $\lambda > 2$ で厳密なスパース解を出力した. また, 詳細は割愛するが, 著者らによる提案手法は既存手法や汎用ソルバと比べて高速に優れた解を出力した.

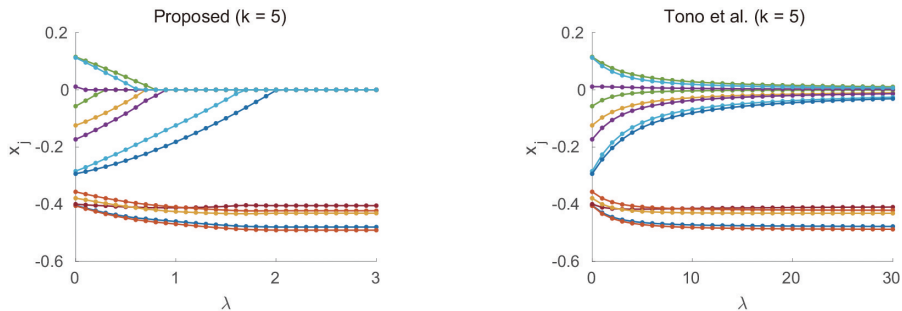


図 1: 解パスの比較 (左: 著者らによる提案手法によるもの, 右: Tono et al. (2017) による既存手法によるもの)

参 考 文 献

- Tanaka, M. and Takeda, A. (2018). Efficient iterative algorithm for constrained nonconvex sparse optimization, submitted.
- Tono, K., Takeda, A. and Gotoh, J. (2017). Efficient DC algorithm for constrained sparse optimization, arXiv:1701.08498v1.

関数推定の理論に基づく深層学習の原理解析

Analysis for Deep Learning by Function Estimation Theory

数理・推論研究系 今泉 允聡 (Masaaki Imaizumi)

1. はじめに

本稿では、深層ニューラルネットワーク (DNN) が他手法より良い性能を発揮する原理を、統計理論を用いて解析した。DNN は既存手法よりも高い性能を発揮することが経験的に知られているが、なぜその性能が発揮されるのかという原理は充分には解明されていない。既存の統計理論では、データが滑らかな関数から生成されている場合、多くの既存の統計・機械学習の手法が理論上の最適精度を達成することが示されており、DNN の相対的優位を説明することは難しい。本稿はその困難さを解決するため、データが非滑らかな関数から生成されている状況で各手法の汎化誤差評価を行った。具体的には、DNN による推定量の汎化誤差の収束レートを導出し、そのレートがミニマックスの意味での最適性を満たすことを示した。加えて、いくつかの既存手法がその収束レートを達成しないことを示し、DNN が他手法に理論的な優越する状況を明らかにした。

2. 問題設定

非滑らかな関数による回帰問題を考える。 $I = [0, 1]$ とし、独立同一分布より生成された観測値の集合 $\{(X_i, Y_i) \in I^D \times \mathbb{R}\}_{i \in [n]}$ が与えられ、またそれらのデータ生成過程は以下の関係を満たしているとする：

$$Y_i = f^*(X_i) + \xi_i.$$

ここで、 $f^* : I^D \rightarrow \mathbb{R}$ はデータ生成過程を特徴付ける真の関数 (未知) であり、また ξ_i は平均 0 で分散 $\sigma^2 > 0$ のガウスノイズであるとする。また、 f^* は区分以上でのみ滑らかな関数であるとする。即ち、 f^* の定義域 I^D が α -Smooth な境界を持つ複数の区分に分割され、その区分の内部で f^* は β -Smooth であるとする。区分の境界線上では、 f^* は非連続になりうる。

観測の集合 $\mathcal{D}_n := \{(X_i, Y_i)\}_{i \in [n]}$ による f^* の推定量を考える。DNN によるモデル Ξ_{NN} を用いて、経験リスクを最小化する最小二乗推定量を

$$\hat{f} \in \operatorname{argmin}_{f \in \Xi_{NN}} \frac{1}{n} \sum_{i \in [n]} (Y_i - f(X_i))^2,$$

と定義し、この関数 \hat{f} を f^* の推定量として用いる。

3. 結果

3.1 DNN による汎化誤差の評価

\hat{f} による汎化誤差は以下のように評価される。

Theorem 1. (\hat{f} による汎化誤差)

ある定数 $c_1, C_L > 0$ と DNN のあるネットワークのもとでの推定量 \hat{f} が

$$\|\hat{f} - f^*\|_{L^2(P_X)}^2 \leq C_L \max\{n^{-2\beta/(2\beta+D)}, n^{-\alpha/(\alpha+D-1)}\} (\log n)^2,$$

を確率 $1 - c_1 n^{-2}$ 以上で満たす。

収束レートのうち、一つ目の項 $n^{-2\beta/(2\beta+D)}$ は、各区分内部の f^* の滑らかさな部分を推定する影響、二つ目の項 $n^{-\alpha/(\alpha+D-1)}$ は各区分そのものを推定する影響を表現している。

3.2 DNN の最適性

定理 1 で得られた結果の最適性を議論するため、区分上でのみ滑らかな関数 f^* を推定する際のミニマックスな収束レートを導出する。

Theorem 2. (区分上でのみ滑らかな関数推定のミニマックスレート)

\bar{f} を \mathcal{D}_n に依存する任意の推定量とする。この時、ある定数 $C_{mm} > 0$ のもとで以下が成立：

$$\inf_{\bar{f}} \sup_{f^*} \mathbb{E} [\|\bar{f} - f^*\|_{L^2(P_X)}^2] \geq C_{mm} \max\{n^{-2\beta/(2\beta+D)}, n^{-\alpha/(\alpha+D-1)}\}.$$

定理 2 の結果より、定理 1 で得られた汎化誤差の収束レートは、ミニマックスな汎化誤差の収束レートに対数項の影響を除いて一致している。すなわち、区分上でのみ滑らかな関数の推定問題において、DNN による推定量は理論的な最適性を達成していると言える。

3.3 DNN と他手法の比較

区分上でのみ滑らかな関数を推定する際の、他手法の非最適性について議論する。本稿では、以下の形式で書かれる線形推定量と呼ばれる推定量のクラスを考える：

$$(3.1) \quad \hat{f}^{\text{lin}}(x) = \sum_{i \in [n]} \Upsilon_i(x; X_1, \dots, X_n) Y_i.$$

なお、 Υ_i は X_1, \dots, X_n に依存する任意の可測関数である。この推定量のクラスは、カーネル法、フーリエ法、スプライン法、ガウス過程法などの多くの推定量を含んでいる。

非滑らか関数を推定する問題について、過去の研究が線形推定量が最適性を達成しないことを示している。それを用いることで、以下の結果を得ることが出来る。

Corollary 1. (DNN の理論的優位性)

$\alpha D / (2\alpha + 2D - 2) \leq \beta$ が成立するとする。この時、ある f^* が存在し、そのもとで DNN による推定量 \hat{f} と任意の線形推定量 \hat{f}^{lin} に関して、十分大きな n のもとで以下が成立する：

$$\mathbb{E}_{f^*} [\|\hat{f} - f^*\|_{L^2(P_X)}^2] < \mathbb{E}_{f^*} [\|\hat{f}^{\text{lin}} - f^*\|_{L^2(P_X)}^2].$$

この結果により、線形推定量のクラスに分類される推定量は最適性を達成しないため、最適性を持つ DNN による推定量を優越できないことが理論的に示されている。

参 考 文 献

- Imaizumi, M., & Fukumizu, K. (2018). Deep Neural Networks Learn Non-Smooth Functions Effectively. arXiv preprint arXiv:1802.04474.