

日本統計学会創立75周年記念出版

21世紀の統計科学

国友直人・山本拓 監修

< Vol. III >

数理・計算の統計科学

北川源四郎¹・竹村彰通² 編集

2008年8月(東京大学出版会)

2012年1月(増補HP版)

¹情報・システム研究機構長

²東京大学情報理工学研究所教授

増補HP版・はしがき

シリーズ「21世紀の統計科学」Vol.I, Vol.II, Vol.III は2008年に東京大学出版会より商業出版された。その後、統計学・統計科学に関係するこの種の書籍としては順調に販売が伸び2011年半ばにいたり在庫部数が少なくなってきた。

この書籍は2008年版の前書き・後書きに説明があるように通常の商業出版物とは異なり、日本統計学会の創立75周年を契機に、できるだけ多くの人々に統計学・統計科学の最近の動向を紹介することにある。そこでこれを機会に各原稿を可能な範囲で改訂し、更に2012年増補版として学会HPより無償でダウンロードする形で広く利用して頂くことにした。

もとより本書・2012年HP増補版の各著者は原稿料は要求せず無償で原稿を提供しているわけである。そこで本書の編者・監修者としては各読者にはなるべく本書及び本書の論文を引用等で正確に引用して頂くことを期待したい。

2012年1月
編者・監修者

はしがき

21世紀を迎えて既に8年目を迎えている今日、社会における統計学・統計科学の役割は以前にも増して重要になっている。現代では、社会・経済・経営などの社会科学はもちろん、工学（品質管理、環境、情報、計算機など）、理学（数学・地球物理・天文・化学など）、生命科学（生物・医学・薬学・農学など）、人文科学（教育、心理、言語、文学など）をはじめ、あらゆる学問領域において必要不可欠な基礎的手段として、統計学・統計科学が鍵となる役割を演じている。また、個人・企業・政府、などが直面する様々な問題に適切に対処し、科学的に意志決定を行う多くの状況では、統計データや統計的方法による科学的分析が本質的な役割を演じている。本シリーズは「21世紀の統計科学」と題して、統計学・統計科学における最新の動向を踏まえつつ、現代の日本社会における重要な幾つかの側面にしばって統計学・統計科学の現状を鳥瞰するとともに、統計学・統計科学を専攻する第一線の研究者が挑戦している課題について成果をまとめている。全3巻の中で、この第3巻では

特に数理・計算の統計科学の諸問題について、この間の発展や新たな展開、さらに今後の展望をまとめている。この間、統計学・統計科学の理論分野と計算分野は様々な内的・外的な理由から飛躍的に展開し、進化を続けている。そうした発展そのものは統計学関係者にとってのもめざましいものであるが、これから統計学・統計科学を勉強しようとする学生諸君、あるいは自分が抱えている問題を解決すべく努力を重ね、応用を指向している多くの人々にとっては、縁遠い出来事のように受け取られる可能性もある。またとかく統計学・統計科学の発展は身近な問題には関係のない研究分野の議論と誤解されることも少なくない。こうした危惧を背景として、まず第1巻の第1部ではあえてこの間に展開している統計数理と統計計算への入門的内容を配置した。第1章の竹村論文は統計的分析において特に最近になり有用視されるようになっている確率分布や統計的推測をめぐる議論への導入である。第2章の北川論文はこの間の計算機や計算ソフトウェアの目覚ましい展開とともに登場した新しい分野、計算機統計学 (computational statistics) をめぐる議論への導入である。さらに、第3章の藤越論文は2つの論文による導入の議論を踏まえ、20世紀の統計学のさらなる発展として、21世紀の統計学が取り組むべき課題について解説している。

統計数理もこの間様々な方向への展開を見せている。その主要な内容を展望することは1冊の書物では不可能であるが、第1部では伝統的に既に多様な分野で応用されている統計学の研究分野、統計的多変量統計解析 (statistical multivariate analysis) と統計的時系列統計解析 (statistical time series analysis) における発展の一部分に絞った。第4章の久保川論文では古典的な線形統計モデルが今日なお研究対象であり、政府統計・経済統計における小地域統計という問題において新たな応用可能性があることを示唆している。第5章の塚原論文では伝統的な多次元分布をめぐる議論にコピュラという概念を用いる進展があり、特に保険数理やファイナンス (金融) 分野での新たな展開の可能性を探っている。他方、第6章の田中論文は工学や経済学で必須の統計的分析法である統計的時系列解析における諸問題を鳥瞰している。この研究分野においてもこの間多くの理論的発展や季節調整法などへの応用があることを示している。さらに統計的時系列分析の発展形として、近年では確率過程や確率微分方程式の統計学が盛んになりつつあるが、この種の統計的分析は特にファイナンス (金融) 分野での応用が顕著である。第7章の内田論文はこうした統計的モデル、確率過程における統計的推測の理論的発展をかなり包括的に議論している。

本書の第 部ではこの間の計算機統計学の発展を支えている 4 つの重要な問題に絞って取り上げ、比較的詳しく解説することとした。歴史的にも計算統計の分野が注目されるきっかけの 1 つとしては B.Efron が提唱したブートストラップ (bootstrap) の登場があったが、この話題について第 8 章の下平論文が比較的平易に解説している。次に第 9 章の渡辺論文では統計学における古典的アプローチでは分析が困難であった幾つかの問題を取り上げ、計算統計学の発展により可能となった統計的最適化手法である EM(expectation-maximization) アルゴリズムをめぐる統計的問題を解説している。最後の 2 つの論文はより現代的な統計計算の問題を取り上げている。現代の統計学の潮流としてベイズ統計学 (Bayesian statistics) は無視できない大きなものであるが、こうした研究動向の 1 つのきっかけとして、従来は実際の計算が困難とされていた複雑な事後分布 (posterior distribution) などの確率分布の計算可能性が肯定的に解決されたことが挙げられる。MCMC(マルコフチェーン・モンテカルロ法) と呼ばれる近年のベイズ統計学で多用される統計計算の手法を第 10 章の古澄論文が解説している。最後に、現代の時系列の解析ではフィルタリング手法による分析が不可欠となっている。そこで北川 (編者の一人) などにより開発された統計的フィルタリング (statistical filtering) の方法と実際の適用を第 11 章の生駒論文で解説している。

本書の各章では現代における「数理・計算の統計科学」をめぐる最新の議論がなるべく多くの人々に理解されるよう、学術的な研究論文というよりもかなり分かりやすい形で統計学・統計科学における最新の動向を理解できるように説明している。さらに本巻は本シリーズの他の第 巻「社会・経済の統計科学」および第 巻「自然・生物・健康の統計科学」が扱っている様々な応用分野に関わる諸論考の基礎となるはずである。本書の内容が、統計学・統計科学の理論と応用、特に数理科学や計算科学に関心がある多くの学生、院生、研究者、官庁関係者、実業界の関係者にとって、問題把握や問題の解決へのヒントを与える材料になれば幸いである。

2008 年 5 月

編者

あとがき

本書は2006年5月と同12月に日本統計学会創立75周年を記念して開催された研究集会(東京大学と中央大学にてそれぞれ開催)における招待講演・報告の扱いをめぐる議論を一つのきっかけとして企画された。全体としてバランス良い75周年記念の書籍としてより意味を持たせるため、編集委員より学会において評価の高い研究者に原稿の書き下ろしをも依頼し、全体を編集しこのたび出版する運びとなった。各章の担当者は元原稿に対するコメントを参考にしてさらに加筆、修正を加えて最終原稿を作成した。

2006年度は日本における統計学・統計科学では最古で最大の学術団体である「日本統計学会(Japan Statistical Society, <http://www.jss.gr.jp/ja/>)」の創立75周年に当たっている。草創期における日本での統計学・統計科学の展開や学会周辺の事情については、例えば『日本の統計学50年』(1982年、日本統計学会編、東京大学出版会)により少し知ることができる。創立75周年という節目を迎え、日本統計学会の当時の会長(山本拓・一橋大学経済学部教授)、理事長(竹村彰通・東京大学工学部教授)を中心に75周年記念事業委員会(委員長:杉山高一・中央大学理工学部教授)が組織され、様々な記念事業が企画され、実施された。この記念事業では数多くの招待講演、研究報告が行われ、「一般社会からはかなり専門的」と評されるかもしれないが、日本統計学会自らが発行している英文・和文の専門学術誌上での学術的資料にとどまるにはあまりにも重要な内容が多かった。そこで特別に75周年記念事業委員会の中に編集委員会が組織され、シリーズ「21世紀の統計科学」の出版が企画されることになった。本書はそうした日本統計学会の75周年記念出版事業での成果刊行物である。むろん、専門的な研究報告や統計学会・会員向けの展望報告などは日本統計学会・和文誌上に75周年特集I・II・IIIとして2007年度後半より順次、掲載され始めている。

なお、本企画は意識的に網羅的であることを避け、今回集まった編集委員の判断で内容を厳選し、原稿作成の期限を設けたことをお断りしておく。日本統計学会には今回の全3巻では取り上げることができなかった統計学・統計科学の研究分野や研究・教育などで活躍していると多忙で優れた研究者も少なくない。今回取り上げることの出来なかった多くの研究分野や重要な課題における統計学・統計科学からの貢献・発展・今後の展開についての議論は、別の機会に譲ることとしたい。

ここで第II巻「自然・生物・健康の統計科学」と第III巻「数理・計算の統計科学」の編集にあたっては、寄稿された諸論文の編集に際して評価者・助言者として、(敬称略)大森裕浩、大屋幸輔、川崎能典、間瀬茂、下平英寿、宮田敏、小林景、田中研太郎、二宮嘉行、清水邦夫、清智也、南美穂子、汪金芳、樋口知之の各先生方のご協力を得た。これら諸先生方のご協力に特に感謝したい。また、本シリーズ「21世紀の統計科学」の出版にあたって、東京大学出版会の黒田拓也氏のご協力にも感謝したい。この第II巻が第I巻と第III巻とともに多くの方々にとっての座右の書となることを期待したい。

2008年5月
監修者

目次

はしがき

第 I 部：数理統計と計算統計への誘い

第 1 章「数理統計への誘い」竹村彰通

1. はじめに：数理統計の視点
2. 一様分布と正規分布に関する話題
 - 2.1 一様分布のたたみこみの密度関数
 - 2.2 正規分布の基準化定数
3. 正規分布を拡張した分布
 - 3.1 非対称正規分布
 - 3.2 一般化双曲分布
 - 3.3 安定分布
 - 3.4 星形分布
 - 3.5 凸関数による変数変換法

第 2 章「計算統計学への誘い」北川源四郎

1. はじめに
2. 最尤法数値的最適化
3. 状態空間モデルと逐次フィルタリング
4. いろいろな統計計算手法
 - 4.1 EM アルゴリズム
 - 4.2 ブートストラップ法
 - 4.3 ベイズモデリングと MCMC
 - 4.4 おわりに：実際の統計計算について

第 3 章「21 世紀の統計学への挑戦的課題と展望」藤越康祝

1. まえがき
2. 統計学の未来について
3. 挑戦的課題
4. 高次元多変量解析

- 4.1 $p \sim n$ & $p < n$ の場合
- 4.2 $p \sim n$ & $n < p$ および $n \ll p$ の場合
- 4.3 高次元グラフ表現
- 5. 付録：多変量解析の基礎

第 II 部：統計数理の展開と統計科学

第 4 章「線形混合モデルの理論と応用」久保川達也

- 1. はじめに
- 2. 線形混合モデルとその特徴
 - 2.1 線形混合モデル
 - 2.2 混合モデル方程式と最良線形不偏予測量 (BLUP)
 - 2.3 変量効果と共通母数の役割
 - 2.4 分散成分の推定と経験最良線形不偏予測量 (EBLUP)
- 3. 線形混合モデルを利用した小地域推定と誤差評価
 - 3.1 地域レベルのモデル
 - 3.2 平均 2 乗誤差の推定
 - 3.3 信頼区間の構成
 - 3.4 応用例：地価公示価格の小地域推定
- 4. 線形混合モデルの様々な応用
 - 4.1 経時測定データのモデル
 - 4.2 モデルの修正と小地域推定への応用
 - 4.3 一般化線形混合モデル
 - 4.4 経験ベイズモデルと階層ベイズモデル
- 5. おわりに

第 5 章「接合分布関数 (コピュラ) の理論と応用」塚原英敦

- 1. はじめに
- 2. 定義と基本的性質
 - 2.1 接合関数の基本的性質
 - 2.2 2次元接合関数の 1パラメータ族の例
 - 2.3 アルキメデス型接合関数
 - 2.4 関連性尺度
 - 2.5 従属性の諸概念
 - 2.6 擬似乱数発生の方法
 - 2.7 経験接合関数
- 3. 応用
 - 3.1 多変量生存時間解析
 - 3.2 ファイナンス
 - 3.3 保険
 - 3.4 確率論への応用

4. おわりに

数学付録

第6章「時系列分析の理論と応用」田中勝人

1. はじめに
2. 一変量時系列分析の基礎
 - 2.1 時系列データと定常性
 - 2.2 短期記憶的な時系列モデル
 - 2.3 短期記憶過程のスペクトラム
 - 2.4 ARMA モデルの予測
 - 2.5 ARMA モデルの推定
3. ARIMA モデルと ARFIMA モデル
4. 単位根検定
5. 構造変化を考慮した時系列モデル
6. ウェーブレット解析
7. 多変量時系列
8. おわりに

第7章「確率微分方程式の母数推定」内田雅之

1. 序
2. 離散観測における確率微分方程式の母数推定
 - 2.1 統計モデル
 - 2.2 離散観測データ
 - 2.3 サンプルパスの発生 (simulation)
 - 2.4 統計モデルの仮定
 - 2.5 尤度関数とマルチンゲール推定関数
 - 2.6 疑似対数尤度関数
3. 小さな拡散をもつ拡散過程の母数推定
 - 3.1 疑似最尤推定量
4. 結論と展望
5. 補遺
 - 5.1 Wiener 過程とマルチンゲール
 - 5.2 確率積分と伊藤の公式
 - 5.3 確率微分方程式

第III部：統計計算の展開と統計科学

第8章「ブートストラップ」下平英寿

1. まえがき
2. リサンプリング
3. パラメータの信頼区間

- 4. プラグイン推定量
- 5. 構造のあるデータ
- 6. 精度の高い信頼区間
- 7. ブートストラップ確率
- 8. マルチスケール・ブートストラップ

第9章「EM アルゴリズム」渡辺美智子

- 1. はじめに
- 2. EM アルゴリズムの考え方
- 3. EM アルゴリズムの理論と一般形
- 4. EM アルゴリズムと適用例
 - 4.1 血液型遺伝子に関する発生確率の推定
 - 4.2 多変量データにおける欠測：多変量正規モデル
 - 4.3 重回帰モデル：目的変数に欠測がある場合
 - 4.4 中途打ち切りデータに基づく例
 - 4.5 混合分布モデル
- 5. EM を利用した漸近分散共分散行列の評価
 - 5.1 Louis の方法
 - 5.2 Oakes の方法
 - 5.3 SEM(Supplemented EM)
- 6. EM の特性
- 7. GEM アルゴリズムとその他の拡張型 EM
 - 7.1 GEM(Generalized EM)
 - 7.2 ECM アルゴリズム
 - 7.3 ECME アルゴリズム - Contaminated Normal
 - 7.4 加速化を意識した完全データの探索 - optimal EM アルゴリズム

第10章「マルコフ連鎖モンテカルロ法入門」古澄英男

- 1. はじめに
 - 1.1 モンテカルロ積分
 - 1.2 モンテカルロ法からマルコフ連鎖モンテカルロ法へ
- 2. マルコフ連鎖
 - 2.1 マルコフ連鎖と推移行列
 - 2.2 マルコフ連鎖の性質
 - 2.3 詳細釣り合い条件
- 3. メトロポリス ヘイスティングス法
 - 3.1 メトロポリス ヘイスティングスアルゴリズム
 - 3.2 MH アルゴリズムの収束
 - 3.3 MH アルゴリズムの組み合わせ
- 4. ギブス・サンプリング

- 4.1 ギブス・サンプリングアルゴリズム
- 4.2 多重ブロック MH アルゴリズムとギブス・サンプリング
- 4.3 データ拡大法
- 5. 実際の利用について
 - 5.1 収束の判定
 - 5.2 効率性
 - 5.3 混合の改善
- 6. 応用例
 - 6.1 ロジットモデルのベイズ推定
 - 6.2 隠れマルコフモデルのベイズ推定

第 11 章「逐次モンテカルロ法とパーティクルフィルタ」生駒哲一

- 1. 状態空間モデルと状態推定
- 2. 逐次モンテカルロ法
 - 2.1 パーティクルフィルタ
 - 2.2 モンテカルロフィルタ
- 3. ラオ - ブラックウェル化
- 4. 応用事例
 - 4.1 非線形モデル
 - 4.2 非ガウスモデル
 - 4.3 ターゲット・トラッキング
 - 4.4 動画像におけるビジュアル・トラッキング

第1章 数理統計への誘い

竹村彰通¹

(東京大学大学院情報理工学系研究科 教授)

ここでは数理統計に関するいくつかのやさしいトピックを紹介する中から, 数理統計の考え方を紹介したい. 特に一様分布や正規分布などの基本的な事項について, 通常の数理統計の教科書とは違った側面から説明する. また正規分布を拡張したいいくつかの確率分布についても最近の話題を紹介する.

¹takemura@stat.t.u-tokyo.ac.jp

1 はじめに：数理統計の視点

本シリーズの多くの論文に見られるように，統計科学の方法の応用は多方面にわたっている．数理統計学は，それらの方法の基礎となる理論を数学的な観点から研究する学問である．時代とともに統計科学の手法が応用される問題も変化しつつあり，それにもなって数理統計学の内容も変化してきている．つまり，数理統計学は，純粋数学の理論のように公理系からはじまって演繹的に理論を展開していくというよりは，統計的手法の応用で現れる諸問題から数学的な部分を抽出して理論の形に整理していく帰納的な側面が大きい．そのために，数理統計学はひとつの体系というより，さまざまな理論の集合であり，それらを統一しているものは言わば「統計学的な視点」とでもいうべきものであると考えられる．実際，数理統計学の研究で用いられる数学的手法としては，最近では幾何学や代数学の高度な手法も用いられており，数理統計学とは数学のさまざまな分野の道具を駆使して統計的な問題を解く分野と言ってもよいと思われる．

もちろん数理統計学のなかから演繹的な部分を取り出して，形式の整った数学的な理論を作る努力も精力的におこなわれてきた．20世紀は，ブルバキの著作に代表されるように，数学の研究全体が抽象性・形式性を重視した時代であり，数理統計学もそのような形式化が重視される傾向があった．数理統計学の形式化としては，Wald にはじまる統計的決定理論の枠組が有力であり，ひとつの代表的な教科書としては Lehmann の検定論の教科書 ([13]) があげられる．

しかしながら，最近の急速な計算機技術およびネットワーク技術の発展により，統計的方法を応用して解析すべきデータの質と量が大きく変化してきた．それにもない「理論」よりも「計算」が重要となってきた．このことは，本シリーズの多くの論文に見られる通りである．具体的なデータが与えられた場合には，そのデータの解析に対して 1) まず確率モデルを設定し，2) その確率モデルに基づく統計的推測をおこなう，という伝統的な統計的方法以外にも，さまざまなアプローチが可能である．その意味で，統計的な方法自体が他のさまざまな方法論と競争的な状況におかれているのである．そのようななかで，統計的方法の意義にはどのようなものがあるだろうか．

これについてもさまざまな論点があるが，筆者が強調したいのは次の点である．それは統計的推測の形式にのっとることにより，推測の結果の精度自体が評価できるという点である．すなわち確率モデルを設定することによって，推測の結論の信頼度を確率的に評価することができる．簡単な

例としては、統計的検定において、有意水準を設定することによって、帰無仮説を棄却すると結論する際の妥当性を確率的に保証することができる。また点推定と区間推定を組み合わせることによって、点推定の信頼度を評価することができる。もちろんこのような場合でも、あくまで設定した確率モデルが妥当であることが前提であり、確率モデルの妥当性が疑われる場合には統計的推測にも意味がなくなってしまう。その意味では、形式的な推論より、確率モデルの妥当性をどのように担保するかのほうがより本質的ということもできる。例えば、確率モデルの妥当性の担保のために、無作為化をとる実験を計画する実験計画法の考え方は、現在でも臨床試験などの分野で本質的な役割を果たしている。また、特定の確率分布の仮定無しに統計的推測をおこなうためのノンパラメトリック・セミパラメトリック法や経験尤度法 (Owen[18]) なども、最近の統計推測理論の中で大きな位置をしめている。

以上のような論点を、「見かけの相関」に関するシンプソンのパラドックスとからめて簡単な例で説明してみよう。離散データの場合で、第3の変数によって二つの変数間に見かけの相関が逆方向に現れるような場合を、シンプソンのパラドックスとよぶ。ここでシートベルトの着用と自動車事故の関係を考えてみよう。シートベルトの着用を義務づけることにより自動車事故自体が減るであろうか。いま、自動車事故を起こしたドライバーと、起こしていないドライバーを比較して、後者のほうがシートベルトの着用率が高かったとしてみよう。この事実から、シートベルトの着用を義務づければ自動車事故が減ると結論できるだろうか。これについては次のような反論が想定される。すなわち、ドライバーの中には安全指向のドライバーとそうでないドライバーがいて、そもそも安全指向のドライバーはシートベルトを着用する傾向にあり、安全指向でないドライバーにシートベルトの着用を強制しても事故は減らない、という反論である。この反論の論拠をより統計学的に説明すると次のようになる。各ドライバーには「安全指向度」という潜在的な特徴があり、この潜在的な特徴が事故数とシートベルト着用率の双方と相関しているために、事故数とシートベルト着用率に見られる相関は見かけの相関であるという議論である。ここで問題なのは「安全指向度」という潜在的な変数が直接には観測できないことにある。このように、後知恵で観測できない変数を持ち込まれて反論されると、統計的なデータ解析からは何も結論が出せないことになってしまう。

従って後知恵をできるだけ排除するような統計的解析をおこなう必要がある。例えば、考え得る潜在変数を最初から考慮しそれらの因果関係の方

向が識別できるような確率モデルを構成する手法は，因果推論あるいはグラフィカルモデルなどによられており，その方法論は近年急速に発展している(宮川[17])．また，潜在変数の値を間接的に推測しこれを調整する方法として，傾向スコア(propensity score)法が注目されている．これらはいずれも，さまざまな要因がからむ複雑なデータに対して，統計的なデータ解析の妥当性を確保するための方法論と考えることができる．なお，見かけの相関や因果については，計量経済学の分野ではすでに1950年代より詳しい議論が展開されている(Simon[22], Blalock[3])．

ここまでは数理統計の考え方について一般的なことを述べてきたが，以下では「計算」という観点から正規分布などの基本的な確率分布についていくつかトピックを紹介する．計算機が急速に発展するなかで，従来は扱うことが難しかった複雑なモデルの計算がおこなえるようになってきており，例えば正規分布に基づく小標本理論などにはあまり意味がなくなっているようにも思える．しかしながら大規模な階層ベイズモデルのような複雑なモデルでも，部品としては正規分布などの伝統的な部品が利用されることが多い．それは，例えば部分的な積分計算などを考えても，不定積分が明示的に求められる関数であれば数値積分やモンテカルロ積分の手間が不要となり，その分の計算資源をモデルの精緻化などに用いることができるからである．このような事情から，非常に古典的な特殊関数の知識などが最近になって再認識される傾向も出てきている．計算機が無かった頃の数学者の計算の蓄積の中に，今後も役に立つものは多いはずである．

2 一様分布と正規分布に関する話題

2.1 一様分布のたたみこみの密度関数

0 と 1 の間の一様分布は連続分布のもっとも簡単な例である．その密度関数は

$$f(x) = f_1(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{その他} \end{cases}$$

と与えられる．いま X_1, \dots, X_n を互いに独立な 0 と 1 の間の一様乱数とし，それらをたたみこんだ $Y = X_1 + \dots + X_n$ の密度関数を考えよう．まず $n = 2$ の場合には (x_1, x_2) の単位正方形 $[0, 1]^2$ において $x_1 + x_2 \leq c$ と

なる部分の面積を考えることによって， $Y = X_1 + X_2$ の密度関数 $f_2(y)$ は

$$f_2(y) = \begin{cases} y, & \text{if } 0 \leq y \leq 1 \\ 2 - y, & \text{if } 1 < y \leq 2 \\ 0, & \text{その他} \end{cases} \quad (2.1)$$

と三角形の形をしていることがわかる． $X_1 + X_2 + X_3$ の密度関数は，たたみこみの密度関数の計算の規則により

$$f_3(y) = \int_0^1 f_2(y-x)f_1(x)dx = \int_0^1 f_2(y-x)dx = \int_{y-1}^y f_2(u)du \quad (2.2)$$

となるから，(2.1) 式の三角形の密度関数で長さ 1 の区間の面積を求めればよいことがわかる．一般の n のとき時の結果は，累積分布関数と密度関数の明示的な形が Feller の第 2 巻 ([6]) の I.9 節に以下のように与えられている．

$$P(Y \leq x) = U_n(x) = \frac{1}{n!} \sum_{\nu=0}^n (-1)^\nu \binom{n}{\nu} (x - \nu)_+^n,$$

$$u_{n+1}(x) = U'_{n+1}(x) = \frac{1}{n!} \sum_{\nu=0}^{n+1} (-1)^\nu \binom{n+1}{\nu} (x - \nu)_+^n.$$

ただし $x_+ = \max(x, 0)$ である．答が与えられてしまえば，帰納法によってこれが正しいことを確認するのは容易なので，確認していただきたい．Feller はこの結果について過去の文献を引用していないが，Laplace によつてすでに知られていたとする文献もある．

なお，(2.2) 式と同様の密度のたたみこみのルールから

$$u_{n+1}(x) = U_n(x) - U_n(x-1) = P(x-1 < X_1 + \cdots + X_n \leq x) \quad (2.3)$$

となっていることに注意する．特に x として整数点 $x = 1, \dots, n$ を考えると，

$$u_{n+1}(1) + u_{n+1}(2) + \cdots + u_{n+1}(n) = P(0 < X_1 + \cdots + X_n \leq n) = 1$$

という関係式が導かれる．(2.3) 式より，整数 $k = 1, \dots, n$ について， $u_{n+1}(k)$ は $X_1 + \cdots + X_n$ の整数部分が $k-1$ に一致する確率になっていることがわかる．幾何的に解釈すると，この確率は n 次元の立方体 $[0, 1]^n$ を $x_1 + \cdots + x_n = k-1$ および $x_1 + \cdots + x_n = k$ の 2 枚の超平面で切った時に，それらの超平面にはさまれる部分の体積に一致することもわかる．

ここで密度関数の $n!$ 倍の値を $x = k, k = 1, \dots, n$ で評価してみると

$$n!u_{n+1}(k) = A_n(k-1) = \sum_{j=0}^{k-1} (-1)^j \binom{n+1}{j} (k-j)^n$$

となる．実際に計算してみると，

$$\begin{array}{ccccccc} A_1(0) & & & & & & 1 \\ A_2(0) & A_2(1) & & & & & 1 & 1 \\ A_3(0) & A_3(1) & A_3(2) & & & & 1 & 4 & 1 \\ A_4(0) & A_4(1) & A_4(2) & A_4(3) & & & 1 & 11 & 11 & 1 \end{array} =$$

のように求まる．

$A_n(j)$ は実は組合せ論において「オイラー数」(Eulerian number) として知られているものである (日比 [9] の 1 章参照)．いま $1, 2, \dots, n$ の置換を a_1, \dots, a_n として, $a_i > a_{i+1}$ と下降しているような箇所 $i = 1, \dots, n-1$ の数を「下降数」とよぶ．例えば $n = 7$ で並べ替え 3517426 では, $5 > 1, 7 > 4, 4 > 2$ と 3 箇所の下降が見られるから下降数は 3 である．この定義のもとで

$$A_n(j) = 1, \dots, n \text{ の置換の中で下降数が } j \text{ のものの個数} \quad (2.4)$$

となることが知られている．例えば $n = 3$ の場合は $3! = 6$ 個の置換があるが, 下降数が 0 のものは 123, 下降数が 2 のものは 321 と一意的に定まるから, 残りの 4 個の置換は下降数が 1 であり, 上の数値とあっていることがわかる．

ここで自然な疑問となるのは, (2.3) 式のような確率と (2.4) 式のような組合せ的な量がなぜ対応しているかである．そのひとつの答は Stanley[23] の僅か 1 ページの論文に初等的に与えられている．この対応は次のようなものである． X_1, \dots, X_n を独立な一様乱数とすると, それらは確率 1 で互いに異なる． $R(X_i)$ を X_i のランク (すなわち X_i が $\{X_1, \dots, X_n\}$ の中で小さいほうから何番目であったか) を表すとすれば, $R(X_1), \dots, R(X_n)$ は $1, \dots, n$ の置換となる．しかもどの置換も同様に確からしく確率 $1/n!$ で現れる．さて, $X_i > X_{i+1}$ となる i が置換の下降に対応している．そこで $[0, 1]^n$ から $[0, 1]^n$ への写像 $\psi(x_1, \dots, x_n) = (y_1, \dots, y_n)$ を

$$y_i = \begin{cases} 1 + x_{i+1} - x_i, & \text{if } x_i > x_{i+1} \\ x_{i+1} - x_i, & \text{if } x_i < x_{i+1} \end{cases}$$

とおく．ただし $x_{n+1} \equiv 0$ とする．この時 ψ は確率 0 の領域 (すなわち x_i の中に互いに等しいものがある領域) を除いて 1 対 1 となっている．そして y_i の定義は, $x_i > x_{i+1}$ となり下降がある時に 1 を加えており, 下降数を数えるものとなっている．そこで $y_1 + \dots + y_n$ を求めると

$$y_1 + \dots + y_n = (R(x_1), \dots, R(x_n) \text{ の下降数}) + 1 - x_1$$

という関係式が成り立つ．さらに, ψ は区分的には線形でヤコビアンは 1 であるから, 確率変数としては Y_1, \dots, Y_n も互いに独立な一様乱数である．そこで $Y_1 + \dots + Y_n$ の整数部分が $j = 0, \dots, n-1$ となる確率を考えると

$$P(j < Y_1 + \dots + Y_n \leq j+1) = P(R(X_1), \dots, R(X_n) \text{ の下降数} = j) = \frac{A_n(j)}{n!}$$

であり, 下降数と一様乱数のたたみこみの対応が確認できる．以上のような初等的な事実は意外に知られていないようであり, Schmidt and Simion[20] やその後の最近の論文でもいくつかの方向で発展が見られる．

2.2 正規分布の基準化定数

標準正規分布の密度関数の基準化定数

$$\sqrt{2\pi} = c = \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

は解析のいろいろな場面に現れる．この基準化定数の導出の一番やさしい方法は, いったん 2 次元の正規分布の極座標表示を経由するものである． X, Y を互いに独立に標準正規分布に従う確率変数として

$$X = r \cos \theta, Y = r \sin \theta$$

とおく．この時ヤコビアンが

$$\det \begin{pmatrix} \partial x / \partial r & \partial x / \partial \theta \\ \partial y / \partial r & \partial y / \partial \theta \end{pmatrix} = r$$

となることにより, (r, θ) の同時密度関数が

$$\frac{1}{c^2} r e^{-r^2/2}, \quad r > 0, 0 \leq \theta \leq 2\pi$$

となる．そこで全積分が1となることを用いると $c^2 = 2\pi$ となることがわかる．またこの密度関数の計算により， r と θ は互いに独立で， θ は0と 2π の間の一様分布に従うこともわかる．

以上は2次元の標準正規分布を経由する方法であるが，1次元の積分のみを用いて $c = \sqrt{2\pi}$ を導く方法もある(高木 [24] 第3章 35節，小平 [12] 4.4節)．それは，スターリングの公式

$$n! \sim \sqrt{2\pi n} n^{n+1/2} e^{-n}$$

やワリスの公式

$$\frac{\pi}{2} = \prod_{n=1}^{\infty} \frac{(2n)^2}{(2n-1)(2n+1)} = \frac{2 \cdot 2}{1 \cdot 3} \frac{4 \cdot 4}{3 \cdot 5} \frac{6 \cdot 6}{5 \cdot 7} \cdots$$

に関連している．

以下では積分のラプラス近似，あるいはラプラス法，の考え方を応用することによってスターリング公式において $c = \sqrt{2\pi}$ が現れる理由を説明しよう．積分のラプラス近似は数値的にも簡便であり，理論的解析のためにも有用である．統計学における漸近理論での利用については Barndorff-Nielsen and Cox [2] の 3.3 節に解説されている．またベイズ法においては事後密度関数をそのモード(最大値)の回りで近似するために用いられたり，さらにチューブ法(福水・栗木 [7]) とよばれる方法との関連も深い．いま，部分積分によって得られるガンマ関数に関する漸化式より

$$n! = \int_0^{\infty} x^n e^{-x} dx$$

である．ガンマ分布を基準化して正規分布で近似することとの関連を念頭におくと，積分の変数変換として

$$x = n + \sqrt{n}z, \quad dx = \sqrt{n}dz$$

とすることが考えられる．この変換により

$$\begin{aligned} n! &= \int_{-\sqrt{n}}^{\infty} (n + \sqrt{n}z)^n e^{-n - \sqrt{n}z} \sqrt{n} dz \\ &= n^{n+1/2} e^{-n} \int_{-\sqrt{n}}^{\infty} \left(1 + \frac{z}{\sqrt{n}}\right)^n e^{-\sqrt{n}z} dz \end{aligned}$$

が導かれる．ここで被積分関数の自然対数をとると

$$\log \left[\left(1 + \frac{z}{\sqrt{n}}\right)^n e^{-\sqrt{n}z} \right] = -\sqrt{n}z + n \left(\frac{z}{\sqrt{n}} - \frac{z^2}{2n} + \dots \right) = -\frac{z^2}{2} + o(1)$$

となる．ただし $o(1)$ は $n \rightarrow \infty$ とともに 0 に収束する剰余項である．このことから

$$n!/(n^{n+1/2}e^{-n}) \rightarrow \int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$$

を得る．

以上の変数変換は関数 $\log(x^n e^{-x})$ を，関数値が最大となる $x = n$ の回りで 2 次まで展開することによって得られている．このような積分の近似法をラプラス近似とよぶ．一般には，正の被積分関数 $g(x) > 0$ の対数 $h(x) = \log g(x)$ をその最大値を与える $x = x^*$ の回りでテーラー展開して

$$h(x) \doteq h(x^*) - \frac{1}{2}(x - x^*)^2(-h''(x^*))$$

と近似すると， $g(x)$ はその最大値の回りで

$$g(x) \doteq e^{h(x^*)} e^{-\frac{1}{2}(x-x^*)^2(-h''(x^*))}$$

と近似できる．いま最大値でのピークが高く， x^* を内点に含む区間 $[a, b]$ での積分が最大値での回りで積分で近似できるとすれば

$$\begin{aligned} \int_a^b g(x) dx &\doteq e^{h(x^*)} \int_a^b e^{-\frac{1}{2}(x-x^*)^2(-h''(x^*))} dx \\ &\doteq e^{h(x^*)} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-x^*)^2(-h''(x^*))} dx \\ &\doteq \sqrt{\frac{2\pi}{-h''(x^*)}} e^{h(x^*)} \end{aligned}$$

となる．これがラプラス近似である．

3 正規分布を拡張した分布

正規分布は確率論および統計学においてなくてはならない分布である．それは主に正規分布の数学的な扱いやすさによるものである．一方で，正規分布に基づく確率モデルの妥当性が疑われる場合も多い．まず，正規分布は左右対称な分布であるが，実際のデータには左右の歪みが見られることがある．また，正規分布は裾の確率が急激に現象する裾の軽い分布であるが，実際のデータにはより裾の重い分布が適合することが多い．このような理由から，正規分布を拡張するような形で，しかも扱いやすい分布があれば

好都合である．1節で述べたように，数学的な扱いやすさは計算の効率という点からもやはり重要なのである．このような確率分布の研究は古くからおこなわれているが，最近になっていくつかの新しい確率分布が実際に使われるようになってきている．また筆者および筆者の周りの研究者による新たな提案もここで紹介したい．なお近年盛んに研究されている「コピュラ」を用いた分布の拡張については本書の第5章を参照されたい．

3.1 非対称正規分布

非対称正規分布 (skew-normal distribution) は Azzalini [1] に始まるものであるが，いわばコロンプスの卵のように簡単な仕組みで正規分布を拡張している．その後のさまざまな発展は例えば Genton[8] にサーベイされている．いま標準正規分布の密度関数と累積分布関数を

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \quad \Phi(x) = \int_{-\infty}^x \phi(u)du$$

とおく時，非対称正規分布の密度関数の標準形は

$$f(x) = 2\phi(x)\Phi(\alpha x) \tag{3.5}$$

と表される． α は分布の歪みを表す実パラメータである．

より一般に $g(x)$, $h(x)$ を原点对称な密度関数， $G(x) = \int_{-\infty}^x g(u)du$ を g の累積分布関数， $w(x)$ を奇関数とする時

$$f(x) = 2h(x)G(w(x)) \tag{3.6}$$

とおくと $f(x)$ は密度関数となる．これを確認するには $f(x)$ が非負であることから，全積分が1となることを示せばよい．さて， X を密度 $g(\cdot)$ に従う確率変数， Y を密度 $h(\cdot)$ に従う確率変数とし，これらを独立とする．この時 $X - w(Y)$ は原点对称な連続確率変数であるから $1/2 = P(X - w(Y) \leq 0)$ である．これを，順次積分 (あるいは条件つき期待値の繰り返し) によって評価すれば

$$\begin{aligned} \frac{1}{2} = P(X - w(Y) \leq 0) &= \int \int_{x-w(y) \leq 0} h(y)g(x)dx dy \\ &= \int_{-\infty}^{\infty} h(y) \left(\int_{-\infty}^{w(y)} g(x)dx \right) dy \\ &= \int_{-\infty}^{\infty} h(y)G(w(y))dy \end{aligned}$$

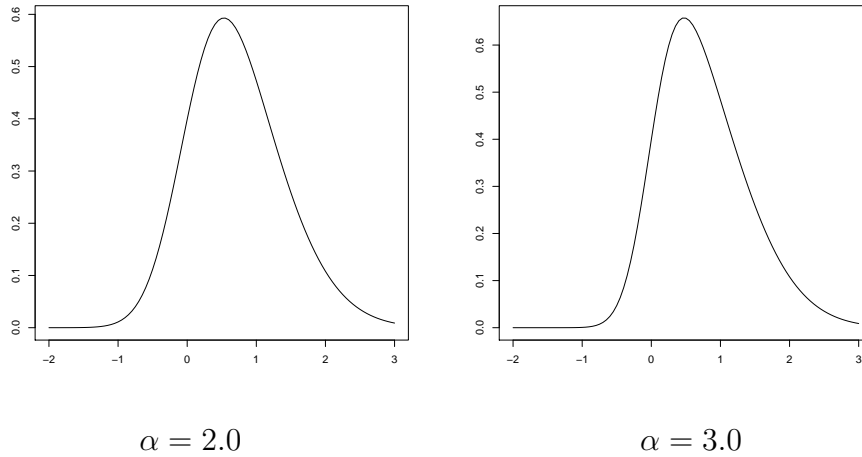


図 1: 非対称正規分布の密度関数の例

となり, (3.6) 式が密度関数を表すことが確認できる. $\alpha > 0$ の場合, $h(x) = \phi(x)$, $G(x) = \Phi(\alpha x)$, $w(x) = x$ としたものが (3.5) 式となる.

X_0, X_1 をそれぞれ独立に標準正規分布に従う確率変数とすると,

$$Y = \frac{\alpha}{\sqrt{1+\alpha^2}}|X_0| + \frac{1}{\sqrt{1+\alpha^2}}X_1 \quad (3.7)$$

の密度関数が (3.5) 式に一致することが示される. これより非対称正規分布は, $\alpha = 0$ の時が標準正規分布, $\alpha \rightarrow \infty$ の時が $|X_0|$ の分布, $\alpha \rightarrow -\infty$ の時が $-|X_0|$ の分布に一致する. したがって α が歪みを表していることがわかる. $\alpha = 2$ および $\alpha = 3$ の場合の非対称正規分布の密度関数を図 1 に示している. (3.7) 式の確率的な表現により, 非対称正規分布のモーメント $E(Y^k)$ も明示的に表現することができる.

非対称正規分布は (3.5) 式の形では裾の軽さは正規分布と同様であるが, (3.6) 式にあるように非常に簡単な仕組みで歪みのある分布を構成することができるので, t -分布などのより裾の重い分布や, 多変量分布への拡張も比較的容易であり, 多数の論文でさまざまな拡張が提案されている.

3.2 一般化双曲分布

正規分布 $N(\mu, \sigma^2)$ を拡張する方法の一つは, 母数の μ, σ^2 を確率変数として分布の混合をとるものである. そのような混合分布のなかでさまざま

まな密度関数の形状を表現できるものとして一般化双曲分布 (generalized hyperbolic distribution) が注目されている．特にファイナンスデータへの応用が重要である (Eberlein[4])．この分布については増田弘毅氏による解説 ([15]) が参考になる．その密度関数は

$$f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = a(\lambda, \alpha, \beta, \delta) (\delta^2 + (x - \mu)^2)^{(\lambda - \frac{1}{2})/2} K_{\lambda - \frac{1}{2}} \left(\alpha \sqrt{\delta^2 + (x - \mu)^2} \right) \exp(\beta(x - \mu))$$

と書かれる．ただし

$$a(\lambda, \alpha, \beta, \delta) = \frac{(\alpha^2 - \beta^2)^{\lambda/2}}{\sqrt{2\pi} \alpha^{\lambda - \frac{1}{2}} \delta^\lambda K_\lambda(\delta \sqrt{\alpha^2 - \beta^2})}, \quad K_\lambda(z) = \frac{\pi I_{-\lambda}(z) - I_\lambda(z)}{2 \sin(\lambda\pi)},$$

$$I_\lambda(z) = (z/2)^\lambda \sum_{i=0}^{\infty} \frac{(z/2)^{2i}}{i! \Gamma(\lambda + i + 1)}$$

である． λ が整数の時は $K_\lambda(z)$ の右辺は極限で定義する． I_λ は変形ベッセル関数， K_λ は変形第3種ベッセル関数とよばれる．ただし，文献によって用語のゆれが見られる．パラメータの範囲は $-\infty < \mu, \lambda < \infty, \alpha > |\beta|$ とその境界 $\{\delta = 0, \lambda > 0\}$ ， $\{\alpha = |\beta|, \lambda < 0\}$ である．パラメータが5個あるためにさまざまな形状の密度関数を表現することができる (Eberlein[4])．

一般化双曲分布は正規分布 $N(\mu, \sigma^2)$ において μ と σ^2 を単一の確率変数で同時に混合することによって得られる．具体的には確率変数 Y を次の密度を持つ一般化逆ガウス分布に従う確率変数とする．

$$f_{GIG}(y; \lambda, \delta, \gamma) = \left(\frac{\gamma}{\delta}\right)^\lambda \frac{1}{2K_\lambda(\delta\gamma)} y^{\lambda-1} \exp\left(-\frac{1}{2}\left(\frac{\delta^2}{y} + \gamma^2 y\right)\right), \quad y > 0.$$

そして， $N(\mu + \beta Y, Y)$ の形で正規分布の混合をおこなうと一般化双曲分布の密度関数が次の形で得られる．

$$f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = \int_{y>0} \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{1}{2y}(x - \mu - \beta y)^2\right) f_{GIG}(y; \lambda, \delta, \sqrt{\alpha^2 - \beta^2}) dy.$$

3.3 安定分布

独立同一分布 (i.i.d., independently and identically distributed) に従う確率変数 X_1, X_2, \dots の部分 and $S_n = X_1 + \dots + X_n$ の分布については，よく

知られているように，もし各 X_i の分散が有限ならば中心極限定理が成り立つ．すなわち任意の x について

$$P\left(\frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} \leq x\right) \rightarrow \Phi(x) \quad (n \rightarrow \infty)$$

である．しかしながら， X_i の分散が無限大の場合には， S_n を基準化した分布は必ずしも正規分布に収束するとは限らず，正規分布を含む安定分布とよばれる分布に収束する．安定分布の一つの例はコーシー分布である．コーシー分布の密度関数と特性関数は

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad \phi(t) = e^{-|t|}$$

であり，特性関数による簡単な計算から，各 X_i がコーシー分布に従う時 S_n/n の分布がやはりコーシー分布となる．

対称な密度関数を持つ場合に限ると，安定分布の特性関数の標準形は

$$\phi(t) = \exp(-|t|^\alpha), \quad 0 < \alpha \leq 2$$

で与えられる． α は安定分布の指数とよばれ，特に $\alpha = 2$ の場合が正規分布である． $\alpha < 2$ の安定分布の分散は無限大であり， α が小さくなるほど分布の裾が重くなる．

以上のように特性関数は簡明であるが，特性関数から密度関数を求めようとしても，例外的ないくつかの α の値以外には密度関数が明示的に求められない．このことが安定分布の実用上の問題点となっている．しかしながら中心極限定理の一般化として安定分布が得られることは理論的には重要な事実であり，裾の重い分布の研究のなかで安定分布はやはり中心的な役割を果たしている (Rachev[19])．なお松井宗也氏と筆者は安定分布の密度関数の数値計算について一連の詳しい研究をおこない ([16] 等)，数値計算上の工夫によって安定分布の推定や検定も十分実用的に実行できることを確認している．

3.4 星形分布

多変量正規分布の一般化として，楕円形分布が以前より多くの研究者によって研究されている (Fang, Kotz, and Ng[5] 等)．原点を中心とする p 次元多変量正規分布 $N(0, \Sigma)$ の密度関数は

$$f(x) = \frac{1}{(2\pi)^{p/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}x'\Sigma^{-1}x\right)$$

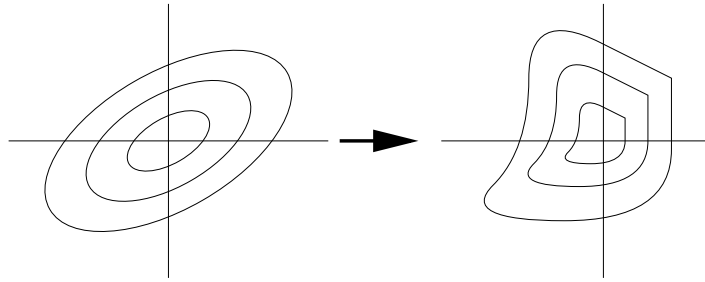


図 2: 楕円形分布から星形分布への一般化

であるが，指数関数の部分を任意の 1 変数関数に置き換えた密度関数

$$f(x) = h(x'\Sigma^{-1}x) \quad (3.8)$$

が楕円形分布の密度関数である．楕円形分布とは，原点を中心とする同心楕円上で密度が一定の分布である．特に $\Sigma = I$ の場合を球面对称分布という．球面对称分布の特徴は p 次元確率ベクトル x をその長さ $\|x\|$ と「角度」 $x/\|x\|$ にわける時に，これらが独立に分布することにある．また角度は超球面 $S^{p-1} = \{x \mid \|x\| = 1\}$ 上一様分布する．

紙屋英彦氏，栗木哲氏と筆者は，不変性の観点から楕円形分布などを含む一般的な分布族を提案し，星形分布とよんでいる (Kamiya, Takemura and Kuriki[10]; Kamiya and Takemura[11])．星形分布は図 2 にあるように，原点を中心とする同心星形集合 (の境界) 上で密度関数が一定となるような分布である．ここでは，原点を含む集合 A が星形集合であることを，原点から出る任意の半直線と A の境界 ∂A がちょうど 1 点で交わると定義することとする．星形分布においても，距離と角度の独立性が成り立つ．ただし距離とは，原点から見てどの星形集合の境界に乗るかということで定義する．

例えば 2 次元で L_1 距離を考えて，密度関数を

$$f(x, y) = h(|x| + |y|)$$

とすると，密度関数は原点を中心とする (45° 回転した) 正方形上で一定であり，長さとしては L_1 距離 $|x| + |y|$ を考えれば，長さ と 角度 が 独立 に 分布 することは直観的にも明らかであろう．

3.5 凸関数による変数変換法

ここでは多次元の連続確率分布として，凸関数の勾配写像から導かれる分布を紹介する．この節の内容は清智也氏による．連続分布として最も重

要なものは言うまでもなく正規分布である。しかし、正規分布には、単峰性や原点周りの対称性など、実際のデータには仮定できないような強い制約が存在する。このような制約を取り除く一つの方法として、凸関数による変数変換法を紹介する。

いま、簡単のため2次元の確率分布のみを考える。2次元の標準正規分布に従う確率変数ベクトルを $Y = (Y_1, Y_2)$ とおく。また、 $\psi(x_1, x_2)$ を2次元の凸関数とする。ここで凸関数とは、その2階偏導関数で作られる行列(ヘッセ行列)の固有値が正となる関数であった。このとき、連立方程式

$$\frac{\partial \psi}{\partial x_1}(x_1, x_2) = Y_1, \quad \frac{\partial \psi}{\partial x_2}(x_1, x_2) = Y_2 \quad (3.9)$$

の一意解 $X = (X_1, X_2)$ を考えることができる。凸関数 ψ をうまく選ぶと、 X の分布は正規分布とは全く異なる性質を持つようにできる(数学的には任意の連続分布が構成可能であることが知られている。McCann [14])。例として、

$$\psi(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2) - 0.9 \sin x_1 \sin x_2$$

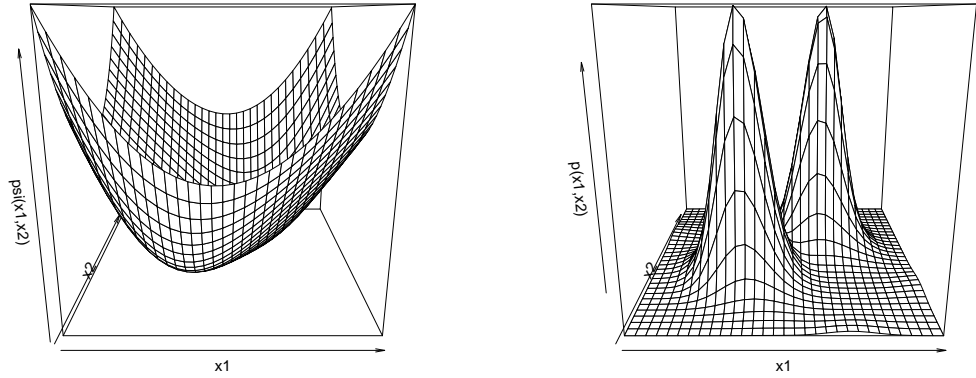
を考えよう(図3(a))。この関数は凸関数であることが確認でき、(3.9)式の解として得られる X の確率分布は図3(b)のようになる。密度関数は、変数変換の公式から

$$p(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}((x_1 - 0.9c_1s_2)^2 + (x_2 - 0.9s_1c_2)^2)} ((1 + 0.9s_1s_2)^2 - (0.9c_1c_2)^2)$$

と陽に書ける(ただし $c_i = \cos x_i$, $s_i = \sin x_i$)。凸関数による変数変換法の詳細は Sei[21] に与えられている。

参考文献

- [1] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171-178.
- [2] Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.
- [3] Blalock, H. M. (1973). *Causal Models in the Social Sciences*. Aldine, Chicago.



(a) 凸関数 $\psi(x_1, x_2)$.

(b) 密度関数 $p(x_1, x_2)$.

図 3: 凸関数による変数変換による密度関数の例

- [4] Eberlein, E. (2001). Application of generalized hyperbolic Lévy motions to finance. in *Lévy Processes*, Barndorff-Nielsen and Mikosch, T. editors, 319–336, Birkhäuser, Boston.
- [5] Fang, K.T., Kotz, S. and Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London.
- [6] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications – Volume 2*, 2nd edition, Wiley, New York. 和訳: 「確率論とその応用」ト部舜一他訳, 紀伊國屋書店.
- [7] 福水健次・栗木哲 (2004). 特異モデルの統計学. 統計科学のフロンティア第7巻. 1–230. 岩波書店.
- [8] Genton, M.G. (2004). *Skew-Elliptical Distributions and Their Applications: a Journey beyond Normality*. Chapman & Hall, Boca Raton.
- [9] 日比孝之 (1997). 数え上げ数学. 朝倉書店.
- [10] Kamiya, H., Takemura, A. and Kuriki, S. (2008). Star-shaped distributions and their generalizations. *Journal of Statistical Planning and Inference* (Ogawa memorial volume), 138, 3429–3447.

- [11] Kamiya, H. and Takemura, A. (2008). Hierarchical orbital decompositions and extended decomposable distributions. *Journal of Multivariate Analysis*, **99**, 339–357.
- [12] 小平邦彦 (2003). 解析入門 I. 岩波書店 .
- [13] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York.
- [14] McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps, *Duke Math. J.*, **80**, 309–323.
- [15] 増田弘毅 (2002). GIG 分布と GH 分布に関する解析 . 統計数理, 第 50 巻第 2 号, 165–199.
- [16] Matsui, M. and Takemura, A. (2006). Some improvements in numerical evaluation of symmetric stable density and its derivatives. *Communications in Statistics, Theory and Methods*, **35**, 149–172.
- [17] 宮川雅巳 (2004). 統計的因果推論 – 回帰分析の新しい枠組 – . 朝倉書店 .
- [18] Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall, Boca Raton.
- [19] Rachev, S.T. (2003). *Handbook of Heavy Tailed Distributions in Finance*. Elsevier, Amsterdam.
- [20] Schmidt, F. and Simion, R. (1997). Some geometric probability problems involving the Eulerian numbers, *Electronic Journal of Combinatorics*, **4**, Research Paper 18.
- [21] Sei, T. (2011). Gradient modeling for multivariate quantitative data. *Annals of the Institute of Statistical Mathematics*. **63**, 675–688.
- [22] Simon, H. A. (1953). Causal ordering and identifiability. in *Studies in Econometric Method*, Hood W. C. and Koopmans, T. C., editors. pp.49–74. Wiley, New York.
- [23] Stanley, R. (1977). Eulerian partitions of a unit hypercube. in *Higher Combinatorics*, M.Aigner ed., NATO Adv. Study Inst. Series, D.Reidel, Dordrecht, p. 49.
- [24] 高木貞治 (1983). 解析概論 , 改訂第 3 版 , 軽装版 , 岩波書店 .

第2章 計算統計学への誘い

北川 源四郎¹

(情報・システム研究機構 機構長)

情報技術と統計計算法の急速な進歩によって、従来はほとんど不可能であった正規性や線形性を仮定しない自由なモデリングが現実のものとなりつつある。本章では、このような計算統計学の方法について概観する。

¹kitagawa@rois.ac.jp

1 はじめに

計算統計学は特定の対象領域を想定した統計学ではない。急速に発達した計算基盤および計算アルゴリズムを活用して統計科学の可能性を広げようとするものであり、本質は柔軟なモデリングの実現を目指すところにある。統計計算法と控えめに表現するのが穏当だが、分布の記述によってあらゆる現象を科学の土俵に載せようとした記述統計学、データに基づいて真の分布に関する推論を目指した推測統計学と同様に、計算に基づいて柔軟なモデリングを実現しようとする立場を表現しようとしたものである。

情報化の進展に伴って、統計科学の立場も役割も大きく変化している。20世紀に飛躍的に発展した数理統計学では、データは未知の真の分布 $f(x|\theta)$ に従うとの前提のもとで、実験あるいは観測に基づいて獲得した比較的少数のデータから真の分布に関する推論を行うことが志向された。この状況では、少数パラメータで規定されるモデルを想定することが現実的である。いわゆるパラメトリックモデリングである。特に、線形性や正規性を仮定した“硬い”モデルを前提とすることによって、解析解を求めることができる。

しかし、1970年代に情報量規準 AIC の登場によって、モデリングの局面は大きく転換する。AIC は多数のモデル比較を実用的にした一方で、より現実 に即したモデルの開発を強く促していた。情報量規準 AIC の適用には最尤推定が不可欠であるが、現実の構造を取り入れたモデルの多くは、与えられたパラメータに対して尤度関数の計算はできても、最尤推定量を解析的に求めることができない。こうして、統計的モデリングにおいて数値的最適化のアルゴリズムは不可欠のツールとなった。

1980年代以降は、社会や科学研究の現場における大量データの蓄積を背景にして、より“柔軟な”モデルの利用が始まった。少数データの時代に用いられた簡単な“硬い”モデルでは、現実にそぐわないことが多くなったのである。AIC が示唆するように、少数データの場合には小規模なモデルが適切な場合でも、大量データが利用できる場合には、より複雑で現実の構造を反映したモデルの利用が可能かつ必要となる。時系列解析においても、AR モデル、ARMA モデルのパラメトリックモデルから、状態空間モデル、非線形・非ガウス型状態空間モデル、一般状態空間モデルへと一般化が進んだ。これらの一般化において、ARMA モデル以降は非線形最適化のアルゴリズムが、また非線形・非ガウス型状態空間モデルでは数値的なフィルタリングの利用が不可欠となった。

ベイズモデルにおいては更に著しい進展が見られた。ベイズモデルはその原理的優越性は認められながらも、近年に至るまで実際のモデリングに利用されることは少なかった。確率の解釈等の哲学的な問題や、事前分布の設定

の問題に加えて、特別なモデル以外には計算困難が付きまとったからである。しかしながら、20世紀終盤における計算機の飛躍的高速化と普及および統計計算アルゴリズムの開発によって、いまやベイズモデルは統計科学のみならず知的情報処理の主流とってよい状態になっている。

このように計算統計学の進展によって、強い仮定を課した“硬い”モデルに関して解析的方法を適用する従来の数理統計学の方法に替わって、計算アルゴリズムと計算機を最大限活用することによって、現実世界により対応した“柔軟な”モデリングを実現しようとするのが可能となりつつある。計算統計学はこれまでの仮想世界における推論から現実世界における予測と知識発見へ一歩進めようとする挑戦である。

第2節では、最尤法に基づく統計的モデリングにおける数値的最適化法の必要性について、第3節では、時系列モデリングにおける状態空間モデルと逐次フィルタリングのアルゴリズムの役割について説明する。第4節では、本書の他の著者が取り上げる統計計算手法を概観する。最後にまとめて代えて、関連する統計ソフトについて紹介する。

2 最尤法と数値的最適化

モデルに基づく統計的推論の方法においては、よいモデルの利用が成功の鍵である。良いモデルを利用すれば、データの持つ情報を適切に利用して見事な結果を導くことができる反面、不適当なモデルを利用するとはっきりした結果が得られなかったり、場合によっては全く的外れな結論が導かれることもある。情報量規準

$$\text{AIC} = -2(\text{最大対数尤度}) + 2(\text{パラメータ数}) \quad (2.1)$$

はさまざまな次数や種類の統計的モデルの比較を可能にし、良いモデルを客観的に求めることを可能にした (Akaike 1973, 坂元ほか 1983, 小西・北川 2004)。

n 個の独立な観測値 y_1, \dots, y_n が得られたとき、パラメータ θ を持つ確率密度関数 $f(y|\theta)$ で規定される統計的モデルの尤度は

$$L(\theta) = \prod_{i=1}^n f(y_i|\theta) \quad (2.2)$$

と定義され、対数尤度はその対数をとって

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(y_i|\theta) \quad (2.3)$$

で求められる。したがって、どのような確率分布であれ上記の密度関数 $f(y|\theta)$ が与えられれば、対数尤度 $\ell(\theta)$ 自体は比較的簡単に定義できることが多い。問題はこの尤度関数の最大化である。AIC の定義には最大対数尤度が必要である、未知パラメータ θ は対数尤度を最大とする最尤法によって推定することが前提となっている。

例えば、データが平均 μ 、分散 σ^2 の正規分布に従うと仮定した場合には $\theta = (\mu, \sigma^2)^T$ とするとき

$$f(y_i|\theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\} \quad (2.4)$$

と表される。したがって、(2.3) の対数尤度は

$$\ell(\theta) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \quad (2.5)$$

となる。この場合には、この対数尤度を最大とする最尤推定量 $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)^T$ は簡単に求められる。対数尤度のパラメータ μ と σ^2 に関する偏微分を 0 として得られる尤度方程式

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) = 0 \\ \frac{\partial \ell(\theta)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (y_i - \mu)^2 = 0 \end{aligned} \quad (2.6)$$

は簡単に解くことができ、

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2 \quad (2.7)$$

と求められる。線形回帰モデルで残差が正規分布に従うと仮定したモデル

$$y_i = a_0 + \sum_{j=1}^k a_j x_{ij} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \quad (2.8)$$

についても同様に尤度方程式が連立一次方程式となって回帰係数の最尤推定量が簡単に求められる。

しかし、非線形モデルやノイズの分布が正規分布でない場合には事情が異なり、一般にはこのように最尤推定量を解析的には求められない。例えば、回帰曲線の最小値が 0 と仮定した多項式回帰モデル

$$y_i = a(x_i + b)^2 + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \quad (2.9)$$

では、尤度方程式は非線形方程式となる。また、(2.4)の代わりにコーシー分布

$$f(y_i|\theta) = \frac{1}{\pi} \frac{\tau}{(y_i - \mu)^2 + \tau^2} \quad (2.10)$$

を想定した場合にも尤度方程式は非線形方程式となる(北川 2005)。

このように、比較的簡単なパラメトリックモデルの範囲でも線形性・正規性を仮定しないモデルの多くでは、最尤推定値を解析的に求めることができない。このような場合でも、与えられたパラメータ θ に対して尤度関数あるいは対数尤度関数 $\ell(\theta)$ を計算することができさえすれば、数値的最適化のアルゴリズムを利用すれば、最尤推定値 $\hat{\theta}$ は原理的には自動的に求められる。数値的最適化のアルゴリズムとしては多くのものが提案されているが、代表的な擬似ニュートン法では、適当に定めたパラメータの推定値 θ_0 およびヘッセ行列の逆行列の推定値 H_0^{-1} を初期値として

$$\theta_k = \theta_{k-1} + \lambda_k H_{k-1}^{-1} g(\theta_k) \quad (2.11)$$

を反復することによって、最尤推定値に収束させる。ただし、 $g(\theta_k) = \frac{\partial \ell(\theta_{k-1})}{\partial \theta}$ 、 λ_k は直線探索によって求めるスケールパラメータ、 H_{k-1}^{-1} は $s_k = \theta_k - \theta_{k-1}$ 、 $z_k = g(\theta_k) - g(\theta_{k-1})$ とするとき、下記の更新式によって自動的に調整される。

$$H_k^{-1} = H_{k-1}^{-1} + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}} - \frac{H_{k-1}^{-1} y_{k-1} y_{k-1}^T H_{k-1}^{-1}}{y_{k-1}^T H_{k-1}^{-1} y_{k-1}} \quad (2.12)$$

この更新式はDFP(Davidon-Fletcher-Powell)公式と呼ばれる最も基本的なものであり、このほかBFGS公式など多くの更新式が提案されている(今野・山下 1978)。

いったん最尤推定量を解析的に求めるという立場を離れ、数値的最適化による最尤法の実現に移行すると、むしろモデリングは自在に行えるようになる。情報量規準AICは、より良いモデル族を提案することの重要性を明確に示している。その意味では、解析解が得られるモデルに限定することなく、よいモデルを追求する姿勢が必要である。数値的最適化のアルゴリズムは統計的モデリングにおいて、自由なモデリングを可能とする不可欠なツールとなったのである。

3 状態空間モデルと逐次フィルタリング

計算統計学の手法の重要性は、時系列解析においてより顕著である。定常時系列モデルの最も基本的なモデルとして自己回帰 (AR) モデル

$$y_n = \sum_{j=1}^m a_j y_{n-j} + v_n, \quad v_n \sim N(0, \sigma^2) \quad (3.1)$$

がよく知られている。AR モデルは線形回帰モデルの時系列版ともいえる。未知パラメータ a_1, \dots, a_m の最尤推定値が近似的に Yule-Walker 法あるいは最小二乗法によって求められるという特長がある。

ところが、AR モデルの枠組みを超えるとたちまち最尤推定値を求めることが簡単ではなくなる。例えば、Box-Jenkins 法で有名な自己回帰移動平均 (ARMA) モデル

$$y_n = \sum_{j=1}^m a_j y_{n-j} + v_n - \sum_{j=1}^{\ell} b_j v_{n-j} \quad (3.2)$$

の場合には最尤推定値はもちろん、与えられたパラメータに対する対数尤度の値を求めることも容易ではなかった。しかし、ARMA モデルの状態空間モデル表現とカルマンフィルタを用いると、尤度計算は極めて簡単に実現できる (Akaike 1978, 北川 2005)。

このような解析的手法からアルゴリズムに依拠した方法への移行は、工学の分野では 1960 年ごろから逐次フィルタリングとして実現されていた。状態空間モデルでは時系列 y_n に対して k 次元状態ベクトル x_n を導入し、下記のようなモデルを想定する。

$$\begin{aligned} x_n &= F_n x_{n-1} + G_n v_n, & v_n &\sim N(0, Q_n) \\ y_n &= H_n x_n + w_n, & w_n &\sim N(0, R_n) \end{aligned} \quad (3.3)$$

この線形・正規性の仮定の下では、形式的には未知数 x_1, \dots, x_N に関する解析解が得られるが、その計算量は $O(k^3 N^3)$ となり、現在の計算環境ではともかく、当時は人工衛星のガイダンスなど現実の問題への適用は全く実行不可能であった。

時刻 $n-1$ までの観測値 $Y_{n-1} \equiv \{y_1, \dots, y_{n-1}\}$ が得られたときの状態 x_n の一期先予測分布の平均を $x_{n|n-1}$ 、分散共分散行列を $V_{n|n-1}$ 、時刻 n までの観測値 Y_n が得られたときの状態 x_n のフィルタ分布の平均を $x_{n|n}$ 、分散共分散行列を $V_{n|n}$ と表すものとする。このとき、カルマンフィルタは一期先予測とフィルタを下記のように交互に繰り返す (片山 1983, 北川 2005)。

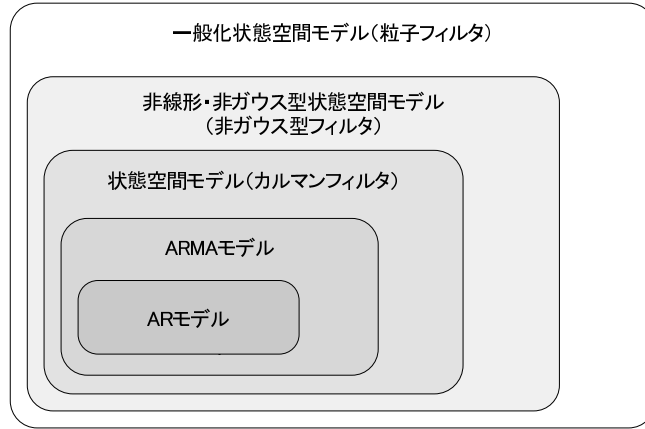


図 1: 時系列モデルの拡張

[一期先予測]

$$\begin{aligned} x_{n|n-1} &= F_n x_{n-1|n-1} \\ V_{n|n-1} &= F_n V_{n-1|n-1} F_n^T + G_n Q_n G_n^T \end{aligned} \quad (3.4)$$

[フィルタ]

$$\begin{aligned} K_n &= V_{n|n-1} H_n^T (H_n V_{n|n-1} H_n^T + R_n)^{-1} \\ x_{n|n} &= F_n x_{n-1|n-1} \\ V_{n|n-1} &= F_n V_{n-1|n-1} F_n^T + G_n Q_n G_n^T \end{aligned} \quad (3.5)$$

このカルマンフィルタを利用すると、それまで殆ど不可能であった巨大な計算を $O(k^3)$ の計算を N 回繰り返すだけで極めて効率よく実行できる。

ARMA モデルは状態空間モデルの形で表現できるので、カルマンフィルタによって状態推定が可能である (図 1, 北川 2005)。したがって、厳密な計算が困難であった ARMA モデルの対数尤度は y_n の予測分布が平均 $H_n x_{n|n-1}$ 、分散 $H_n V_{n|n-1} H_n^T + R_n$ となることを利用して簡単に求められるようになった。パラメータの最尤推定値は、前節に示した数値的最適化の方法を適用することによって求めることができる。このように、ARMA モデルの最尤推定は、カルマンフィルタと数値的最適化の二つの統計計算法によって実現できる。

ARMA モデルの最尤推定を契機に、時系列モデリングに状態空間モデルが利用されるようになったが、この方法を用いれば状態空間モデルで表現できるモデルでありさえすれば、カルマンフィルタと数値的最適化のアルゴリ

ズムを利用して、統一的な取り扱いができるようになる。推定量の解を解析的に求め、そこにデータを代入して推定値を求めるという従来の方法を離れると、自由なモデリングが可能になる。

現時点から振り返ってみると、この状態空間モデルの導入は、計算を高速化するという特長以上に重要な、以下のような特徴と可能性を持っていた。

1. 大規模パラメトリックモデルの導入： ARモデルやARMAモデルなどの少数のパラメータで規定されるパラメトリックモデルと比較すると、状態空間モデルには行列 F_n , G_n , H_n や共分散行列 Q_n , R_n に含まれる構造パラメータの他に、 k 次元の状態 x_n が存在する。この x_n をパラメータと解釈すると、時間に比例する数のパラメータを持つ大規模パラメトリックモデルを自然に表現することになる。
2. 時変パラメトリックモデルの導入： 状態空間モデルの第2式を回帰モデルと解釈すると、状態空間モデルは回帰係数が時間変化する時変モデルを表現していることになる。第1式は時変パラメータの時間変化をモデル化したものである。
3. ベイズモデリングの典型： カルマンフィルタのアルゴリズムは、前の時点までの情報に基づく状態の事前分布の計算と新しいデータに基づく事後分布の計算を繰り返しているものと解釈できる。
4. データ同化モデル： カルマンフィルタはモデルによる時間発展とデータによる修正・更新を繰り返しており、環境シミュレーションで用いられるデータ同化の手法を自然に実現したものと解釈できる。
5. 時系列解析の汎用ツール： ほとんどの時系列モデルを状態空間モデルで表現できる。したがって、状態空間モデルに対する予測、補間、推論、パラメータ推定、モデル選択の方法を確立すれば、殆どの時系列モデルに適用可能な方法が得られることになる。

さて、このように大きな特長をもった状態空間モデルであるが、線形性・正規性の制約を取り除くには大きな障害があった。計算効率的なカルマンフィルタのアルゴリズムから逸脱するとたちまち、その計算量は非現実的となった。一方、線形近似を行う拡張カルマンフィルタでは、本質的な非線形・非正規システムでは良い結果は得られなかった。しかし、計算機の発展は新しい可能性を拓くことになった。Kitagawa(1987)は数値積分に基づいて非正規状態空間モデルの厳密な逐次フィルタリング

$$p(x_n|Y_{n-1}) = \int p(x_n|x_{n-1})p(x_{n-1}|Y_{n-1})dx_{n-1}$$

$$p(x_n|Y_n) = \frac{p(y_n|x_n)p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})} \quad (3.6)$$

を数値積分によって実現可能であることを示した．ここで， $p(x_n|Y_{n-1})$ と $p(x_n|Y_n)$ はそれぞれ，一期先予測分布およびフィルタ分布， $p(x_n|x_{n-1})$ と $p(y_n|x_n)$ はそれぞれ状態空間モデルから求められる状態 x_{n-1} が与えられたときの x_n ，および状態 x_n が与えられたときの時系列 y_n の条件付分布である．

この数値積分に基づく方法も，カルマンフィルタの特長であった計算量が $O(n)$ の性質を保持し，大型計算機の計算速度が現在のパソコンの 100 分の 1 程度の当時でも，4 次元程度の問題まで適用可能であった．非正規モデルと正規モデルでは解析結果に著しい違いがあり，非正規モデルの利用によって自由なモデリングが現実のものとなり，トレンドにジャンプが存在する場合，観測値が異常値を含んだり非対称分布に従う場合，あるいはシステムが非線形性を伴うむ場合にも対応できるようになった．

しかし，大規模データの集積にともなって，大規模モデリングの動きは益々進展し，この数値積分に基づく方法には適用できる状態ベクトルの次元の限界があった．1990 年代に入ると，この問題を緩和する逐次モンテカルロ法の方法が飛躍的に発展した．この方法は，カルマンフィルタ以来，従来のフィルタリング法が状態の密度関数を計算あるいは近似しようとしたのに対して，多数の粒子で近似するものである（Gordon et al. 1993, Kitagawa 1996, Doucet et al. 2001, 本書の 11 章）．この方法は，モンテカルロ近似に伴うサンプリング誤差の問題は避けられないが

1. 非常に複雑なシステムに対しても実装が容易
2. 比較的高次元のシステムに適用可能

という著しい特長があり，理論，アルゴリズム，応用に関して多くの研究が行われている．粒子フィルタのアルゴリズムおよびその応用に関しては，本書の第 11 章を参照されたい．

4 いろいろな統計計算手法

これまでは，時系列解析を中心に非線形最適化と逐次フィルタリングの統計計算手法について紹介したが，本節では本書の統計計算編で取り上げるそのほかの統計計算法，すなわち EM アルゴリズム，ブートストラップ法および MCMC について概観しておく．

4.1 EM アルゴリズム

非線形最適化のアルゴリズムの適用による汎用的パラメータ推定とやや方向を異にする方法としてEMアルゴリズムが提案されている (Dempster et al. 1979, 本書の10章). EMアルゴリズムは, 欠測値を含む不完全データのモデリングにおいて, 非測定データの期待値計算 (E) と完全データが与えられた下での最適化 (M) の繰り返し計算によって, 最尤推定値を得るための統一的手法である.

時系列の場合には前節で紹介したように, 状態空間モデルに基づく統一的な尤度計算法があり, 時系列構造を利用して欠測値が存在する場合にも完全データの場合と同じ手順で尤度計算を実行できる. EMアルゴリズムはこのような特殊なデータ構造が利用できない場合についても, 最尤推定を可能とするものである. EMアルゴリズムは, 収束のオーダーは最適化法に比べて遅いが, 収束の安定性や簡単な二つのステップの繰り返しで構成され実装が簡単であることから, 欠測値だけではなく, 打ち切りデータ, 混合分布モデル, 潜在変数モデルなどの多くの問題に適用されてきている. 状態空間モデルと同様に, ここでも汎用性に大きな特徴がある.

4.2 ブートストラップ法

統計的推論の難しさは真の分布が未知であることに由来する. その未知の分布のパラメータの推定の良さを評価する場合には, 分布が未知であることがいっそう問題となる. 従来の数理統計学の方法では, 分布形に仮定を置き, その仮定の下で, バイアスの評価や, 信頼区間の構成が行われる.

これに対して, ブートストラップ法 (Efron 1979, 本書の8章) では, 真の分布関数 $G(x)$ を, データ $y = \{y_1, \dots, y_n\}$ から定義される経験分布関数 $\hat{G}(x)$ で置き換える. この経験分布関数から生成されたデータ $y^* = \{y_1^*, \dots, y_n^*\}$ はブートストラップサンプルと呼ばれるが, ブートストラップサンプルの生成はデータ y の復元抽出と同等である.

ここで多くの場合, 真の分布関数 G と経験分布関数 \hat{G} の関係が, 経験分布関数とブートストラップサンプルで定義される経験分布関数 \hat{G}^* に反映することを利用すると, 推定量のバイアスや信頼区間などをブートストラップサンプルに基づく推定量の性質を調べることによって評価できる. 詳しくは本書の第8章を参照.

この関係を利用して, ブートストラップ情報量規準 EIC が提案されている (Ishiguro et al. 1997, 小西・北川 2004). 情報量規準 AIC は図2において, 対数尤度のバイアス (対数尤度と平均対数尤度の差 D の期待値) を解析的

に近似したものである．ブートストラップ法を利用すると D^* を繰り返し計算し，その平均を求めることによって面倒な仮定をおくことなくバイアスの評価が可能であり，ブートストラップ情報量規準EICが定義できる．EICでは解析的評価で用いられる解析的近似を使わず，また最尤推定値を用いることを前提とせずに情報量規準を得ることができる．

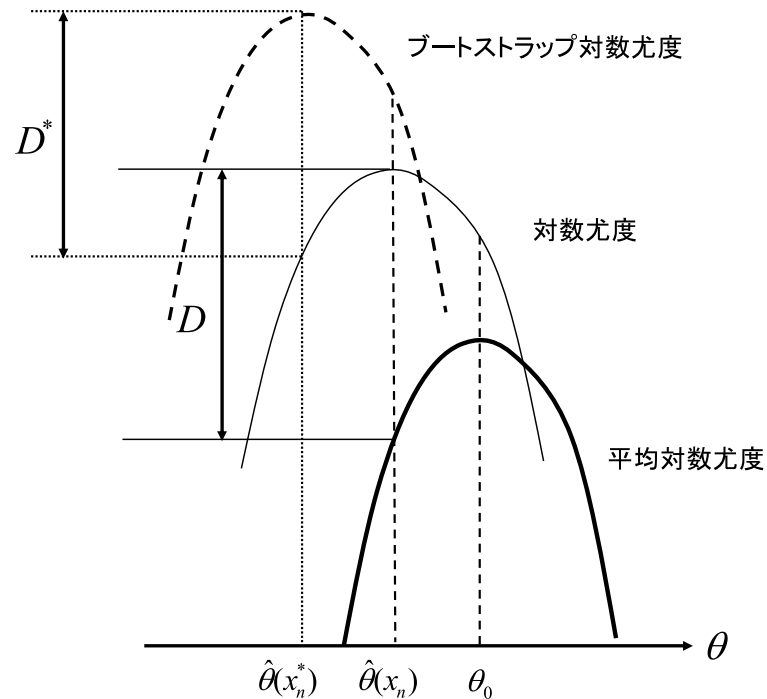


図 2: ブートストラップ法による対数尤度のバイアス推定. 実際には D と D^* をそれぞれ 3 分割する分散減少法を適用する. (小西・北川 2004 より再掲)

4.3 ベイズモデリングと MCMC

現在の知的情報処理においては、モデルはデータを生成する真の構造の忠実な複製というよりは、情報処理のための道具と考えられる。一旦、真の構造という呪縛から開放されると、モデルは対象に関する理論、経験的知識、データ、さらには解析の目的など様々な情報を統合して構築されることになる。これを実現する方法として、ベイズモデルが統計解析の主流になりつつある。

ベイズモデルの実用化における障害のひとつは計算困難の問題であった。正規モデルや共役事前分布を仮定する特殊な場合を除き、事後分布あるいは事後分布からの標本を求めることが困難な場合が多かった。統計科学の分野では 1980 年代以降 MCMC(マルコフ連鎖モンテカルロ)法が急速に普及した。MCMC は正規分布などの特定の分布だけではなく広範な分布に適用で

き、しかも高次元の分布にも適用できる。これによって、事後分布の計算や規格化定数の計算が困難な複雑かつ大規模なベイズモデルが急速に実用化され、計量経済、ゲノム解析、空間統計など様々な問題に適用されている(本書9章参照)。

4.4 おわりに：実際の統計計算について

従来の科学は理論科学と実験科学の方法を車の両輪として発展してきたが、現在では非線形系や複雑系のシミュレーションあるいは大量データの解析の必要性から、いずれの方法においても計算への依存が高まっている。統計科学においても全く同様である。計算によって非線形性や確率分布の評価に伴う困難な問題を解決しようとする計算統計学には計算機の利用が不可欠であることは言うまでもない。

近年とくに重要になってきた超大量・大規模な時系列あるいは時空間データのベイズモデリングにおいては、FORTRANあるいはCで書かれたプログラムを並列計算機で実行することが前提となるが、本書の統計計算編で取り上げた方法自体はのPCでも十分実行可能である。計算統計学においては、それぞれの問題に応じて独自に多様なモデルを構築することが必要になり、従来型の統計ソフトウェアを使いこなすだけでは実現できないことが多い。統計言語RはS言語をオープンソース化したもので、世界中の専門家によって新しいモデルや分析法に対応する関数が開発され続けている。ユーザが書いたFORTRANやCのプログラムを呼び出すこともできるので、Rの関数と組み合わせることで、さまざまなモデリングを効率的に行い、結果を効果的に表示することができる。Rに関する情報源としては以下のようなWebサイトがある。

<http://www.okada.jp.org/RWiki/>

<http://aoki2.si.gunma-u.ac.jp/R/>

一方、逐次モンテカルロ法(粒子フィルタ)に関しては下記のポータルサイトが参考になる。

<http://www-sigproc.eng.cam.ac.uk/smc/index.html>

また、状態空間モデルを用いた非定常時系列解析に関しては、Web上の解析ソフトWeb-Decompがある(佐藤1997)。

<http://ssnt.ism.ac.jp/inets2/title.html>

このソフトは、ユーザとのインターフェイス部分はR言語で構築したシステムで処理し、計算本体はFORTRANプログラムを呼び出して実行する。計算はすべてサーバサイドで行われるため、適当なブラウザさえあればソフ

トウェアのインストール等一切の操作を必要とせず、簡単に解析を実現できることに特長がある。

参考文献

- [1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proc. 2nd International Symposium on Information Theory* (B. N. Petrov and F. Csaki eds.) Akademiai Kiado, Budapest, (1973) 267–281.
- [2] Akaike, H. (1978), Covariance matrix computation of the state variable of a stationary Gaussian process. *Ann. Inst. Statist. Math.* **30-B**, 499–504.
- [3] Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, **B39**, 1–38.
- [4] Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*, Springer, New York.
- [5] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**, 1–26.
- [6] Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings-F*, **140**(2), 107–113.
- [7] 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**(3), 411–434.
- [8] 片山 哲 (1983). 応用カルマンフィルタ, 朝倉書店.
- [9] Kitagawa, G. (1987). Non-Gaussian state space modeling of nonstationary time series (with discussion). *Journal of the American Statistical Association* **82**, 1032–1063.

- [10] 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.
- [11] 北川源四郎 (2005). 時系列解析入門, 岩波書店.
- [12] 小西貞則・北川源四郎 (2004). 情報量規準, シリーズ「予測と発見の科学」, 朝倉書店.
- [13] 今野浩, 山下浩 (1978). 非線形計画法, OR ライブラリー 6, 日科技連.
- [14] 坂元慶行, 石黒真木夫, 北川源四郎 (1983). 情報量統計学, 共立出版.
- [15] 佐藤整尚 (1997). Web Decom の紹介, 統計数理, 第 45 卷第 2 号, 233-243.

第3章 21世紀の統計学への挑戦的 課題と展望

藤越康祝¹

(広島大学大学院理学研究科 名誉教授)

まず, 統計学の本質を概観し, 未来を展望する. 次に, 統計理論と方法に関する研究において, Lindsay et al. (Ed.)(2003) で取り上げられている, 将来直面するであろう6つの挑戦的課題について解説する. さらに, それらの中の1つと関係する高次元多変量解析問題に関しては, いくつかの話題について最近の発展を紹介する. 本原稿は原論文の改訂版である.

¹fujikoshi@yahoo.co.jp

1 はじめに

統計学は、データから有用な情報を抽出することを目的にしている。ところで、データには通常不確実な要因が含まれており、それらから導かれる結論にも不確実性が伴うが、これは避けられない。このため、長い間統計学が市民権を得ることができなかった。しかし、20世紀の初め、導かれる結論に含まれる不確実性の程度を明らかにすることによって、データから厳密に推論が行えるようになり、新たな局面が開けて来た。そして、統計学は不確実な状況の下で最適な決定を研究する学問分野として位置づけられ、今日利用されている統計学の基盤が構築された。それは、未知な量の推定（点推定、区間推定）、仮説検定、あるいは、将来の出来事についての予測であったりする。

統計学においては、これまで主として小標本の場合に、精密理論、あるいは、標本数を大にしたときの漸近理論が展開され、様々な分野へ適応されている。このような推測理論は古典的推測理論ともよばれ、竹内(1998)はその展開を概観している。

1960年代になりコンピュータが普及し始め、その発展が統計学の発展に大きな影響を与えたことは周知の通りである。とくに、計算機統計手法が開発されたり、最近では、電子化された測定技術の進歩により科学の諸分野で、大規模なデータが蓄積され、それらの解析法の開発が重要な課題になっている。

また、最近の注目すべき傾向の1つとして、ゲノム、金融工学、環境問題、ニューラルネット、パターン認識、信号処理、データマイニングなどの分野において、統計学の立場から取り組むべき新しい研究課題が台頭していることが挙げられる。

統計学における、このような大きな動きに対して、最近、研究集会、雑誌などにおいて、統計学の今後の方向が活発に議論されている。とくに、米国国立科学財団(The National Science Foundation, NSF)は2002年5月6-8日世界各国から50名以上の著名な統計学者を集めてワークショップ「統計学：21世紀に対する挑戦と機会」を開催し、Lindsay et al. (Ed.)(2003)による報告書を出版している。伊藤(2007)は、この報告書全般にわたって、その概要を紹介している。

ここでは、まず、統計学の本質を概観し、この観点から最近の動向を見つめ、将来を展望する。次に、統計理論と方法に関する研究において、Lindsay et al. (Ed.)(2003)で取り上げられている、将来直面するであろう6つの挑戦的課題について解説する。ここでは、主として複雑な構造をもつ大規模な

データへの対処法が論じられている。さらに、それらの中の1つと関係する高次元多変量解析問題に関しては、いくつかの話題について最近の発展を紹介する。なお、最後の節において伝統的な多変量解析の方法を簡単に紹介している。

統計学における個々の分野の理論・方法や応用については、高次元多変量解析問題を除き触れていない。多くの分野における展望論文を取り上げたものとして、Raftery et al. (Ed.)(2002)がある。この展望論文集には多変量解析などいくつかの分野は取り上げられていないことを注意しておこう。多変量解析に関しては、やや古くなるが、Rao (Ed.)(1993) および Cuadras and Rao(Ed.)(1995) に多くの展望論文がまとめられている。

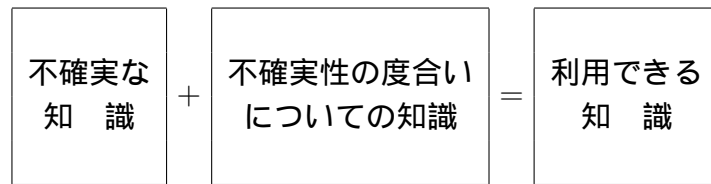
なお、統計学とは何かの記述においては、藤越等(1993, 2010)、田栗等(2007)から引用した部分があることを断っておきたい。

2 統計学とは何か・その未来

はじめの冒頭に述べているように、統計学は、不確実性を含むデータに基づいて、それらが得られた集団、あるいは、それらを発生させるメカニズムについて何らかの決定を行うことを目的にしている。これは、結論からその前提の是非を問題にする帰納的推論である。このように、私たちは帰納的推論によって新しい知識を作り上げているが、データと仮説とは1対1の関係でないため、それは不確実性をもった知識にならざるを得ないのである。しかし、1930年代になって、結論に含まれる不確実性の程度が明らかにされるようになり、帰納的推論が厳密化されると共に新たな知識の獲得法が確立されたのである。

リスクに対処する効果的方法があみ出されたが、その本質は次のような最適決定の問題として定式化される。まず、与えられたデータに基づいて、仮説の1つあるいはそのいくつかを選び出す規則を作り、次に、その規則にしたがって、特定の仮説が選ばれたときの不確実性の度合いを評価する。このとき、誤った決定をする割合を最小にする、もしくは誤りによる損失を最小にするような決定を行う規則を見出すものとして位置づけられた。このプロセスにおいては、データに含まれる不確実性を数学的に記述し、演繹的推論が展開されていることに注意しておこう。

上記のようにして、特定のものから一般化を行うという規則によって作り出された知識は、種類は異なるが、確かな知識となる。このような新しい知識の獲得法を Rao (1977) は次の論理方程式として表している。



上のことを、ある事象 A の起こる確率 θ を推測する問題で具体的に考えてみよう。統計学ではデータに基づいて結論を下すが、今の場合独立に何回か実験を行い事象 A の生起を調べたデータがあるとする。例えば、 n 回の実験で事象 A が x 回生じたとしよう。また、興味ある仮説は $H_1: \theta \neq 1/2$ であるとする。このため、 H_1 とそれを否定した仮説 $H_0: \theta = 1/2$ のいずれが真であるかを判定することを考える。これは仮説検定の問題であり、定数 c を適当に定め、 $|x - 1/2| \geq c$ ならば H_1 を真とする方法が考えられる。このとき、仮説 H_0 が真のとき仮説 H_1 が真であるとする誤り、すなわち第 1 種の誤りや、逆に仮説 H_0 が真でなかったときにそれを採択してしまう誤り、すなわち第 2 種の誤りを犯す確率を与えることにより、仮説が選ばれたときの不確実性の程度を表そうとしているのである。このプロセスで用いられているのは、演繹的な推論方法である。検定法においては、興味ある仮説 $H_1: \theta \neq 1/2$ を確信をもって主張できるようにするため、第 1 種の誤りを犯す確率を一定以下に抑え、第 2 種の誤りを犯す確率を最小にする決定方式を見出すことが行われる。

統計学の本質は上に述べたような考え方にあるが、このほかの視点からも考えてみよう。ここでは、Rao (1997) およびその翻訳書である藤越等 (1993, 2010) で述べられているなかから、いくつかを紹介することにする。

まず、物理学、化学、生物、数学などの多くの学問は、それぞれ“ 独自の問題 ” とそれを解くための“ 独自の方法論 ” をもち、そのための独自の学問として確立している。一方、統計学は対象とする固有の問題をもたない特殊の学問分野であって、他の学問分野の問題を解くことによって存在し、成長していることを注意しておこう。これに関連して、L. J. Savage は次のように述べている (藤越等 (1993, 2010) を参照)。

「統計学は基本的には寄生虫である。すなわち、他分野の研究の上に存在している。このことは統計学への軽蔑を意味するものではない。なぜなら、多くの宿主は、寄生虫がいなくなると死んでしまうからである。また、動物によっては、食べ物を消化することができなくなってしまうからである。統計学は人間と関わ

りをもつ多くの分野と関係しており、それらの分野は、統計学がなければ滅びることはしないものの、かなり弱いものとなることは確実であろう。」

統計学には様々な側面があり、「科学なのであろうか、それとも技術または芸術なのであろうか」という疑問が生じる。これに対して、統計学はこれら三者を組み合わせたものであると考えられる。まず、統計学は、いくつかの基本原則から導出された、幅広いレパートリーからなる独自の手法により、真理の探究を目指しているという意味で科学である。また、工業生産における品質管理のプログラムのように、要求された水準と安全性を維持するための操作システムのなかに、統計的方法論を組み入れることができるという意味では技術でもある。このような応用は、個人や社会の行動の最適化に対しても考えられるであろう。統計学では帰納的推論が行われるが、その方法論が十分には明文化されていなし、また、議論がない訳でもない。このため、データからその背後にある現象のメカニズムを解明する際に、統計家の熟練度や経験に依存する点があり、このことから統計学が芸術であると言われる。

さて、統計学における最近の顕著な傾向として、複雑な構造をもつ大規模データの出現とそれへの取り組みであると述べてきたが、これについて Rao (2006) は次のように述べている。

「統計学は、これまで主として、小標本に基づく枠組みのなかで発展して来ている。今日、実験に関して自動記録装置や情報源の増大により、大量なデータが利用可能になって来ている。これによって、データベースの管理、蓄積や検索に関して新しい問題が提起されている。また、技術の急速な発展により、遺伝子研究におけるマイクロアレーデータ、顔照合に関する画像データ、テロ攻撃に対する郵便番号認識や初期警告システム、のように新たなタイプの実験や測定が求められている。さらに、食料店、銀行などにおいては膨大なデータが蓄積され、これらの中から有用な情報を抽出することが望まれている。統計家にとって、問題解決のための、新しい方法を創造したり、あるいは、統計学の広がり発展させたり、多くのわくわくする可能性が与えられている。」

ここで生じている新しい問題は、新しいタイプのデータの出現によるものであることに注意しよう。統計学の未来は、大規模データのような新しいタ

タイプのデータの出現に係わっていて、とりわけ、他の学問分野における研究者と統計家のコミュニケーションに係っていると言える。

ところで、統計学においては、統計的モデルの導入や推測法において、度数論的接近法とベイズ的接近法があり、必ずしも議論がない訳でもない。しかし、新たなタイプのデータへの取り組みを通して、これらの問題点が明らかにされることが期待される。

3 挑戦的課題

統計学の研究には、データによる現象解明を体系的に行うための基礎的なものから、その方法を種々の学問分野に適用し、当該分野へ本質的に貢献するものまである。これらの研究活動は、いわゆる統計理論・方法と応用に大別される。ワークショップ「統計学：21世紀に対する挑戦と機会」におけるLindsay et al. (Ed.)(2003)による報告書の中では、統計の理論・方法に関する研究活動を「統計学の中核」(the core of statistics)とよび、これに関して6つの挑戦的課題を挙げている。これらの課題の概要は、伊藤(2006)によっても紹介されているが、ここでは若干のコメントを加えながら、より詳しく紹介する。

今後の重要課題をリストするに当たって、いくつかのコメントが与えられている。まず統計学における課題は、他の学問分野におけるものとは異なることが注意されている。例えば、数学の場合には、これまで長く挑戦されている有名な問題をリストすることに焦点が当てられるが、統計学の場合の課題は常に、新たなデータ構造の出現とか、あるいは、新しい計算ツールの発展とに関係して現れている。また、実験科学の場合は問題解決に向けて、巨大な費用が絡んでくるが、この点も異なっている。統計学の分野において今後の最重要課題が何であるかを予測することは、他の分野と比べいっそう困難であろうと指摘している。

このような理由から、統計学の今後の研究において重要なことは、変化に対して柔軟に対応する基本原理を維持し、同時に本質的に異なった技法の集まりにならないようにすることであると報告書は指摘している。

現代の統計理論・方法の研究分野を促進するような、いくつかの課題を取り上げることができるであろう。とくに、多くのパラメータ、種々の規模、複雑な従属性をもつ大規模データを扱うための、基本的考え方と漸近的近似理論の発展にチャレンジすべきであるとしている。具体的には、以下の6つの

重要課題を取り上げ、簡単な説明が与えられている。

(i) 統計データの規模の拡大

従来統計データ解析は、小標本に基づくものが中心であったが、これからはデータの規模が爆発的に増加し、いわゆる大規模データの解析が重要になると指摘している。統計家の増加は1次的で、統計解析家の増加は2次的であるのに比べ、データの規模の増加は指数的であると考えられる。Huber (1994) はデータの規模を

極小 (tiny) 10^2 , 小 (Small) 10^4 , 中 (medium) 10^6 ,
大 (large) 10^8 , 巨大 (huge) 10^{10}

と分類しているが、すでにより大規模な分類も必要になってきていると指摘している。

また、100以下のデータセットさえ、あらゆる問題が解決されている訳ではないが、今後は、各規模の問題に挑戦する必要があるとしている。この場合、各規模での研究に加え、理解の深さのみならず、一般化可能性、スケーラビリティやロバストネスに関する問題およびそれらの混合が、データの規模や与えられた条件によってどのように変わるかも問題になるであろう。さらに、大規模データを扱う際の計算機上などの問題についても挑戦する必要があることも指摘している。

(ii) 統計データの縮小 (reduction) と圧縮 (compression)

統計データを当面の問題の解析に必要な少数個の統計量に縮小する方法には、Fisherによる、十分統計量、補助統計量、条件付推測、変数変換、枢軸法、漸近的最適性、不変性等の概念が考えられている(詳しくは本書収録の竹村論文を参照)。しかし、新たな縮小法が重要になると指摘している。

例えば、モデル選択、予測、分類等の分野では新しい考え方が必要であり、その一つがデータの圧縮であると指摘している。それはデータの構造をより良く理解するためデータを出来るだけコンパクトに圧縮して保存し、必要ならばそれを還元してもとの情報を殆ど完全に復元するという考え方である。

なお、多変量の多くの手法はデータの縮小法と関係して、これに関連した圧縮問題も重要であることを補足しておきたい。

(iii) 機械学習とニューラル・ネットワーク

工学技術者などによって開発された特定目的のための計算機手法が発展している。とくに、機械学習研究グループによって開発されている大量で複雑なデータの解析法を統一して統計学の理論・方法の研究課題として取り上げなければならないと指摘している。ここで、学習と言う用語は、システムやモデルに含まれるパラメータをデータに適合するように逐次更新するという意味に用いられている。この種の研究においては、モデルあるいは構造の構築に基づくべきであって、それらは与えられたデータに対する評価のみならず、リスクを用いて評価されるべきであると指摘している。

例えば、回帰型問題における階層型ニューラル・ネットワークモデルについて考えてみよう。まず、脳の情報処理を模したモデルが用いられているが、これはある種の非線形回帰の問題に他ならないこと注意しておこう。しかし、その回帰式には非常に多くのパラメータが含まれ、また、パラメータ空間には同一の回帰式を定義する領域があるなど、典型的な非線形回帰モデルとは異なっていることが指摘される。したがって、このような点を考慮した統計理論の展開が期待されている。

(iv) 高次元小標本多変量解析

多くの重要な応用において、標本数 n より次元数 p が大きいデータが生じている。例えば、DNA マイクロアレイ、関数データ、スペクトル、画像、などである。DNA マイクロアレイのデータでは、種々の遺伝子を変数となるのでその次元数は数千のオーダーになるが、標本数は高々100程度である。このため、 $p \gg n$ の場合の多変量解析の手法が必要になるであろうと指摘している。一般に、 $n < p$ であると、標本共分散行列が特異になり、方法によっては伝統的な手法が使えなくなるなどの問題が生じる。

この種の問題は、数理物理学の分野で過去40年に渡り発展してきている確率行列論と関係していて、そこでの考えを高次元問題において役立たせる必要があると述べている。高次元問題の1つとして、変数の数と標本数が共に大きなデータの場合、通常の主成分や正準相関などにおける次元縮小法がどんな意味をもつかを明らかにする問題がある。

p と n の両方を大きくした高次元漸近的枠組での近似が、 p を固定し n を大にする伝統的な漸近的枠組での近似より、より役立ち、また、多くの情報を含んでいると言う結果もある。この例として、主成分分析における最大固有値の帰無分布、すなわち、 S がウィシャート分布 $W_p(n, \sigma^2 I_p)$ に従うときの S の最大固有値の分布 (Johnstone (2001)) を挙げている。

高次元小標本問題に関しては、最近かなりの発展が見られるので、次節で

詳しく取り上げている。

(v) ベイズ的接近法と“ 偏りのある (biased) ”推定

最近、ベイズ法が著しく発展して来ている。これは、1990年代に開発された革命的な計算技術と計算能力により、広範囲にわたるモデルに対してベイズ的接近法の適用が可能になったことによる。これに関連して、来るべき研究課題として、ベイズ的手法と現代的ノンパラメトリック・セミパラメトリック手法との関係を探求することが重要であるが、とくにベイズ的接近法と度数論的接近法を組み合わせた手法も視野に入れる必要がある。

また、多くの変数をもつ大規模データのモデルにおいては従来の不偏推定の意味は重要でなくなり、新しい“ 偏りのある ”推定の理論の開発が必要となるであろうと指摘している。これは、非常に大きな変数をもつ大規模データの場合における不偏推定量は複雑になり、また大きく変動するので、それほど重要でなくなるであろうと指摘している。

(vi) 論理的証明と計算機実験の中間点

論理的証明のペースは余りにも遅いが、一方、計算機実験による検証は、一般的に余りにも自由で説得力に欠けた点もあり、沼地のようなものである。両者の間に何らかの中間点を見出すことが重要であると指摘している。統計的方法の開発においては、論理的に証明されていない問題が多く残されている。これは、証明が困難であることと、証明自体は2次的関心であることによる。その一例として、混合分布モデルの識別問題を挙げている。混合分布モデルの場合、パラメーターの値が異なっても混合分布モデルとしては同じであるという識別不能が生じるが、このような場合の統計理論の開発には多くの未解決問題が残されている。

一般に、論理的な厳密証明を強く要求すれば、統計学の種々の分野において発展が遅れる可能性が生じる。一方、若干の数値的な検証で結果の妥当性を認めるとなると、多くの混乱を来すことになるであろう。したがって、両者の中間点を見出す必要があるとしている。

論理的証明が遅れた別な例として、多変量正規線形仮説の尤度比基準の漸近展開がある。その結果は1949年に求められ、雑誌にも掲載されている。しかし、その漸近展開の妥当性の論理的証明が与えられたのは約30年後である。

4 高次元多変量解析

多変量データの変数の数を p とし、標本数を n とする。多変量解析においては、検定統計量などの分布が必要になるが、その基礎分布が多変量正規分布でも正確に求めるのが困難である場合が多い。このため、漸近理論に基づく推測法が考えられている。その際、 p を固定し、 n を大にしたときの大標本の枠組みで調べられている。

一般に、変数の数 p が大きい場合には高次元データあるいは高次元多変量解析とよんでいる。狭い意味の高次元データは、 $n \ll p$ の場合であって、例えば、第3節の(iv)高次元小標本多変量解析において述べたように、DNA マイクロアレイデータにおいては、変数は遺伝子であるが、その数は数千のオーダーで、標本数は高々100である。ファイナンスデータにおいても、種々の銘柄の株価を扱う場合、変数の数は千程度になる。一方、高次元データの場合、冗長変数を除いた分析も行うので、“ $p \sim n$ & $p < n$ ” や、“ $p \sim n$ & $n < p$ ” の場合も重要となる。

最近、このような高次元の状況で、伝統的な推測法の振る舞いを調べたり、あるいは、高次元特有の推測法が提案されている。以下では、いくつかの話題を取り上げ紹介する。

4.1 $p \sim n$ & $p < n$ の場合

多変量解析の伝統的推測法においては、 p を固定し n を大きくした、いわゆる大標本の枠組での漸近理論が展開されて来た。このような結果は $p \ll n$ の場合には有効であるが、 p が増えるにつれて利用できなくなると言う問題点がある。例えば、統計量の漸近分布について大雑把であるが、 $n = 100$ の場合、 $p \leq 10$ であれば有効であろう。しかし、以下で数値例で見ると、例えば、正準相関係数の信頼区間を構成する場合、 p が10より大きくなると近似は段々と悪くなり、 $p = 20$ では使えなくなる場合もある。このような場合

$$c = p/n \rightarrow c_0 \in (0, 1)$$

とした高次元の枠組での近似を求めると、非常によい近似が得られることが分かって来ている。なお、 p/n がある一定値に近づくという高次元漸近理論の枠組みの始まりは古く40年前に遡る (Bai (1999) を参照)。ここでは、標本共分散行列の固有値の経験分布関数についてその漸近的挙動が研究されている。 p 次元対称確率行列 S_n の固有値を $\ell_1 \geq \dots \geq \ell_p$ とし、これらの固有値

に関する経験分布関数を

$$F_n(x) = \frac{1}{p} \#\{\ell_i : \ell_i \leq x\}$$

とする. ここに, $\#\{\cdot\}$ は指示された集合内の要素の数を表す. 経験分布関数 $F_n(x)$ が F に収束するとき, F は S_n の極限スペクトル分布関数とよばれる.

行列 S_n は

$$S_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i'$$

で与えられるものとする. ここに, $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})'$, X_{ij} は独立同一分布に従う確率変数で, 平均 0, 分散 σ^2 をもつものとする. 行列 S_n の定義においては, 各データから標本平均が引かれていなく, 標本分散行列とは異なるものである. しかし, X_{ij} が正規分布に従う場合は, S_n の分布は自由度 1 の違いがあるが, 本質的には標本共分散行列の分布と同じであることを注意しておく. このとき, $c = p/n \rightarrow c_0 \in (0, 1)$ のもとで

$$F_n \rightarrow F, \quad \text{a.s.}$$

が成り立つ. ここに

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} \frac{1}{2\pi c x \sigma^2} \sqrt{(x-a)(b-x)}, & (a < x < b), \\ 0, & (\text{その他}) \end{cases}$$

で, 定数 a, b は $a = (1 - \sqrt{c})^2 \sigma^2$, $b = (1 + \sqrt{c})^2 \sigma^2$ で与えられる. また, a.s. はその収束性を表し, 概収束あるいは確率 1 での収束である. このような研究は, 数理物理学の分野における確率行列に関する理論に端を発し, その後, 多変量研究者によってさらに発展している. 詳細は, Bai (1999) による総合報告を参照されたい.

スペクトル分布は, 経験分布関数を用いて表せる統計量

$$\begin{aligned} T_n &\equiv \frac{1}{p} \{\phi(\ell_1) + \dots + \phi(\ell_p)\} \\ &= \int_0^\infty \phi(x) dF_n(x) \end{aligned}$$

の高次元枠組みのもとでの挙動を求めるのに利用できる. すなわち, 統計量 T_n は適当な正則条件のもとで漸近的に

$$\int_0^\infty \phi(x) dF(x)$$

に概収束することがわかる.

基礎分布が正規分布 $N(0, \sigma^2)$ に従うとき, nS_n はウィシャート分布 $W_p(n, \sigma^2 I_p)$ に従う. このとき, Johnstone (2001) は最大固有値 ℓ_1 の極限分布を与えている. 極限分布関数は Painlevé II 型微分方程式の解を用いた積分表示として表されている. このような分布関数は Tracy and Widom (1996) によって発見された. また, MANOVA の最大固有値の極限分布に関しても同様な極限分布が導出されている (Johnstone (2008)).

次に, 主成分分析における小さい固有値の同等性に関する尤度比統計量の分布を考える. S を $N(\mu, \Sigma)$ からの大きさ N の標本に基づく標本共分散行列とし, S および Σ の固有値をそれぞれ $\ell_1 > \dots > \ell_p > 0$, $\lambda_1 \geq \dots \geq \lambda_p > 0$ とする. このとき, 仮説検定問題

$$H_0 : \lambda_{q+1} = \dots = \lambda_p = \lambda$$

の検定に対する尤度比基準は, $m = p - q$ とおくと

$$V = \prod_{j=q+1}^p \ell_j / \left\{ \frac{1}{m} \sum_{j=q+1}^p \ell_j \right\}^m$$

で与えられる.

p を固定して N を大きくすると, 統計量 $-c_{m,n} \log V$ は漸近的には自由度 $m(m+1)/2 - 1$ の χ^2 分布に従う (Anderson (2003)). ここで, $c_{m,n} = \{n - (2m+1 + 2/m)/6 + \lambda^2 \sum_{j=1}^q (\lambda_j - \lambda)^{-2}\}$, $n = N - q - 1$ である.

n は m より大きいという状況において, n と m を同時に大きくしたときの尤度比統計量 $\log V$ の漸近分布については, Fujikoshi et al. (2007) による次の結果がある. 高次元漸近的枠組として, 次を仮定する.

$$C(0): c = m/n \rightarrow c_0 \in (0, 1).$$

$$C(1): \lambda_j = O(n), \rho_j = \lambda_j / \text{tr} \Sigma \in (0, 1), j = 1, \dots, q.$$

このとき, $Z_{m,n} = (\log V - \mu_{m,n}) / \sigma_{m,n}$ の帰無分布は標準正規分布に収束する. ここで

$$\begin{aligned} \mu_{m,n} &= m \log m - m \psi \left(\frac{mn}{2} \right) + \psi_m \left(\frac{n}{2} \right), \\ \sigma_{m,n}^2 &= \psi'_m \left(\frac{n}{2} \right) - m^2 \psi' \left(\frac{mn}{2} \right) \end{aligned}$$

であり, $\psi(\cdot)$ はディガンマ関数, $\psi_m(a) = \sum_{j=1}^m \psi(a - \frac{1}{2}(j-1))$ である.

両者の近似の精度を把握するため、シミュレーション実験が試みられている。帰無仮説 H_0 の下では、一般性を失うことなく共分散行列の共通な最小固有値を 1 とし、最初の q 個の固有値を λ_j/λ , $j = 1, \dots, q$ としよく、以下では λ_j/λ を単に λ_j としている。シミュレーションでは、 $\lambda_j = \rho_j m / (1 - \sum_{k=1}^q \rho_k)$, $j = 1, \dots, q$ としている。また、 $N = 100, q = 2, \rho_1 = 0.56, \rho_2 = 0.24$ としたときに、1,000,000 回のシミュレーションからパーセント点を求め、それを正規近似およびカイ 2 乗近似に当てはめたものである。

表 4.1. $Z_{m,n}$ の正規近似を用いたときの実際の確率

真の確率	0.01	0.05	0.5	0.95	0.99
$p = 10$	0.004221	0.040958	0.539795	0.934835	0.976473
$p = 20$	0.008586	0.055513	0.567381	0.965133	0.994144
$p = 30$	0.009937	0.057897	0.569441	0.969330	0.995520
$p = 40$	0.010770	0.059437	0.568486	0.970095	0.995647
$p = 50$	0.010816	0.059819	0.564065	0.969164	0.995477
$p = 60$	0.011114	0.058948	0.558578	0.967603	0.994904
$p = 70$	0.010960	0.057867	0.551711	0.965076	0.994227

表 4.2. $-c_{m,n} \log V$ の χ^2 近似を用いたときの実際の確率

真の確率	0.01	0.05	0.5	0.95	0.99
$p = 10$	0.009442	0.048169	0.492684	0.947871	0.989367
$p = 20$	0.003045	0.029127	0.455016	0.941857	0.988261
$p = 30$	0.003502	0.031534	0.474136	0.949255	0.990242
$p = 40$	0.005802	0.043735	0.534003	0.963756	0.993736
$p = 50$	0.013744	0.081291	0.654368	0.982864	0.997706
$p = 60$	0.051688	0.200582	0.831919	0.996581	0.999688
$p = 70$	0.254373	0.551445	0.974300	0.999999	1

これらの結果より高次元近似は $p = 10$ の場合を除けば χ^2 近似より優れていることがわかる。また、大標本近似は、固定された N に対し p の値を増やしていくと悪くなっていくことがわかる。

正準相関係数の関数の分布についても高次元漸近分布が求められている (Fujikoshi and Sakurai (2009)). 大きさ $N = n + 1$ の正規標本に基づく、 p

次元変数と q 次元変数との間の第 i 標本正準相関係数を r_i とし, 対応する母集団正準相関係数を ρ_i とする. ρ_i は単根であるとする. よく知られているように, 次元数 p, q を固定し, 標本数 $N \rightarrow \infty$ とすると

$$\sqrt{n} \{f(r_i^2) - f(\rho_i^2)\} \rightarrow N(0, \sigma^2(\rho_i^2)), \quad \text{in dist.}$$

となる. ここに, "in dist." は分布収束を意味し, $\sigma^2(\rho_i^2) = 4\rho_i^2(1 - \rho_i^2)^2 f'(\rho_i^2)^2$ である.

高次元漸近枠組として: q ; 固定, $p \rightarrow \infty$, $m = n - p \rightarrow \infty$, $c = p/n \rightarrow c_0 \in (0, 1)$ を仮定する. このとき

$$\sqrt{n} \{f(r_i^2) - f(\tilde{\rho}_i^2)\} \rightarrow N(0, \tau^2(\tilde{\rho}_i^2)) \quad \text{in dist.}$$

となる. ここに, $\tilde{\rho}_i^2 = \rho_i^2 + c(1 - \rho_i^2)$, $\tau^2(\tilde{\rho}_i^2) = 2(1 - c)(1 - \rho_i^2)^2 [2\rho_i^2 + c(1 - 2\rho_i^2)] f'(\tilde{\rho}_i^2)^2$ である. 高次元の結果において, $c = 0$ とすれば $\tilde{\rho}_i^2 = \rho_i^2$, $\tau(\tilde{\rho}_i^2) = \sigma(\rho_i^2)$ となり, 大標本の結果と一致することがわかる.

漸近分散がパラメータに依存しなく 1 となる変換は

$$z = \frac{1}{2} \log \frac{1 + \sqrt{r_i^2 - c(1 - r_i^2)/(2 - c)}}{1 - \sqrt{r_i^2 - c(1 - r_i^2)/(2 - c)}}$$

である. 実際, z の定義式において, r_i^2 を $\tilde{\rho}_i^2$ で置き換えたものを ζ とすると

$$\sqrt{n(1 - c)/(1 - c/2)}(z - \zeta) \rightarrow N(0, 1) \quad \text{in dist.}$$

である. この結果において, $c = 0$ とすれば, よく知られた Fisher の z -変換とその大標本漸近結果

$$\sqrt{n} \left\{ \frac{1}{2} \log \frac{1 + r_i}{1 - r_i} - \frac{1}{2} \log \frac{1 + \rho_i}{1 - \rho_i} \right\} \rightarrow N(0, 1) \quad \text{in dist.}$$

を得る. これらの漸近的結果を用いて信頼区間を構成することができる. 近似の精度をシミュレーションによって数値的に調べた結果が表 4.3 に与えられている. 数値実験は, $q = 3$, $\rho_1 = 0.9$, $\rho_2 = 0.5$, $\rho_3 = 0.3$ で, N, p は表 4.3 で与えられる. 高次元の場合の z -変換では, $r_i^2 < c/2$ のとき根号内が負になるが, そのときには絶対値で置き換えることにし, 表では * 印を付けている. この結果, N, p および母集団正準相関係数に関するほとんどの領域において, 高次元近似が大標本近似よりよいことが見て取れる.

表 4.3. z -変換を用いたときの 95 信頼区間に対する実際の信頼係数

N	p	$\rho_1 = 0.9$		$\rho_2 = 0.5$		$\rho_3 = 0.3$	
		大標本	高次元	大標本	高次元	大標本	高次元
50	3	0.917	0.939	0.933	0.956*	0.970	0.915*
	7	0.852	0.935	0.813	0.958*	0.937	0.839*
	17	0.495	0.925	0.173	0.947	0.252	0.890*
	27	0.095	0.902	0.001	0.917	0.003	0.929*
	37	0.001	0.809	0.000	0.849	0.000	0.958*
100	3	0.935	0.945	0.946	0.956	0.956	0.867*
	7	0.909	0.945	0.895	0.958	0.949	0.886*
	17	0.764	0.943	0.565	0.959	0.544	0.910*
	37	0.233	0.939	0.007	0.953	0.002	0.931*
	47	0.061	0.934	0.000	0.945	0.000	0.941*
	67	0.000	0.912	0.000	0.916	0.000	0.955

高次元の枠組での漸近分布あるいは漸近展開に関する結果として、この他、線形判別関数、MANOVA 検定統計量・固有値、共分散行列の検定に関する結果もある (Fujikoshi (2000), Fujikoshi et al. (2008), Ledoit and Wolf (2002), Wakaki, Fujikoshi and Ulyanov (2003), など).

4.2 $p \sim n$ & $n < p$ および $n \ll p$ の場合

次元が標本数より大きくなると、多くの場合伝統的な推測法は適用できなくなる。その主な要因は標本共分散行列や群内平方和積和行列が特異行列になることによる。したがって、高次元推測法の開発には、特異行列になる行列を S とするとき

- (1) $|S|$ に関連した部分を $\text{tr } S$ に置き換える (Dempster (1958)),
- (2) リッジ法・罰則付法, 正則化法を利用 (Friedman (1989), Hastie, Buja and Tibshirani (1994), Ghosh (2003)),
- (3) S の逆行列を Moore-Penrose 逆行列 S^+ に置き換える (Srivastava (2006)),

などの工夫が行われている。このような推測法の性質を調べるには、高次元の枠組 $p/n \rightarrow c \in (1, \infty)$ での漸近理論が重要になる。なお

$$\lim_{p/n \rightarrow c} \quad \text{と} \quad \lim_{n \rightarrow \infty} \lim_{p \rightarrow \infty}$$

は必ずしも等しくはなく、注意を要する。一方、高次元の場合、多くの変数の中から有効な変数を如何に見出すかの変数選択の問題は益々重要となる。

Fisher による正準判別分析法を高次元の場合へ拡張する方法を紹介してみよう。今 p 次元ベクトル変数 $X = (X_1, \dots, X_p)'$ に基づく q 群の判別分析を考える。 n 個の独立な観測値 x_1, \dots, x_n について、最初の n_1 個は第 1 群に属し、以下同様に最後の n_q 個は第 q 群に属するものとする。Fisher による最初の $m (\leq \min(k-1, p))$ 個の判別関数 $h_i'X$ の係数ベクトル $H = [h_1, \dots, h_m]$ は回帰の枠組みにおいて最適数量化法を考えることによって次のように求められる。 Z を n 個の個体がどの群に属しているかを表している $0, 1$ 要素からなる $n \times q$ の計画行列とする。このとき、各群の m 次元数量化は Θ を $q \times m$ のパラメータとして $Y = Z\Theta$ と表せる。そこで、基準量

$$A = A(\Theta, H) = \text{tr}(Z\Theta - XH)'(Z\Theta - XH)$$

を考える。ここで、データの平均ベクトルはゼロになるように基準化されて、 $Y = Z\Theta$ も中心化され各列は互いに直交しているとする。このとき、基準量 $A(\Theta, H)$ を Θ, H に関して最小化することによって Fisher の判別関数が求められる。このような回帰のアプローチを利用すると、 $X = (X_1, \dots, X_p)$ の代わりに変換変量

$$T: X \rightarrow T = (T_1(X), \dots, T_s(X))'$$

に基づく判別関数が導入される。また、罰則付き基準

$$A_* = A_*(\Theta, H, \lambda) = \text{tr}(Z\Theta - XH)'(Z\Theta - XH) + \text{tr} H' \Omega H$$

を用いれば、正則化判別法になる。ここに、 $\Omega = \lambda I_p$ 。応用等については Ghosh (2003) などを参照されたい。

高次元の場合の 2 群の判別法に関しては、線形判別関数の修正を含め多くの方法が提案されている。Dudoit et al. (2002) は数値実験により比較検討し、対角線形判別法がよいふるまいをすることを指摘している。対角線形判別法は、線形判別関数において合併共分散行列 $S = (S_{ij})$ を

$$D_S = \text{diag}(S_{11}, \dots, S_{pp})$$

で置き換えた判別関数を用いる判別法である。Srivastava and Kubokawa (2007) は \mathbf{S} を経験ベイズ推定量

$$\hat{\Sigma}_B = c \left(\mathbf{S} + \frac{\text{tr } \mathbf{S}}{\min(n, p)} \mathbf{I}_p \right)$$

で置き換えた判別法がよりよいふるまいをすることを指摘している。ここに、 $\hat{\Sigma}_B^{-1}$ は Σ^{-1} の経験ベイズ推定量で、 c は定数である。

検定問題については、Fujikoshi, Himeno and Wakaki (2004), Fujikoshi et al. (2010), Ledoit and Wolf (2002), Schoot (2005, 2005), Srivastava (2007) などを参照されたい。

4.3 高次元グラフ表現

高次元小標本データを p 次元空間における n 個のベクトルあるいは点とみなすことにする。このとき、通常の漸近理論とは異なって、標本数 n を固定し、次元数 p を大きくした場合の高次元小標本データベクトルの漸近的挙動を問題にする。Hall et al. (2005) はこれらの n 個の点は正則な単体の頂点に近づき、ランダムネスは頂点間の回転に集約されることを示している。より正確には、 p 次元変量 $\mathbf{X} = (X_1, \dots, X_p)'$ を考える。ここに、 X_1, \dots, X_p は互いに独立で $N(0, 1)$ に従うものとする。 \mathbf{X} についての大きさ n 個の無作為標本を

$$\mathbf{X}_i = (X_{i1}, \dots, X_{ip})', \quad i = 1, \dots, n$$

としたときに、 \mathbf{X}_i の長さ、 \mathbf{X}_i と $\mathbf{X}_j (i \neq j)$ との距離、また、 \mathbf{X}_i と \mathbf{X}_j との成す角に対して、 $p \rightarrow \infty$ における挙動について、次のことが示されている。

- (1) $\|\mathbf{X}_i\| = \sqrt{p} + O(1), \quad i = 1, \dots, n.$
- (2) $\|\mathbf{X}_i - \mathbf{X}_j\| = \sqrt{2p} + O(1), \quad i, j = 1, \dots, n, i \neq j.$
- (3) $\text{ang}(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{2}\pi + O(p^{-1/2}), \quad i, j = 1, \dots, n, i \neq j.$

ただし、 $\|\cdot\|$ は、ユークリッド距離とする。(2) より、無作為標本における各対の長さは近似的に等しく、また、(3) より各対の角度は近似的に 90 度であること示している。

この結果は、次の 3 つの条件のもとで非正規の場合に拡張される。

- (C1) すべての成分に関して、4 次モーメントは一様有界。

(C2) 定数 σ^2 に対して, $\frac{1}{p} \sum_{k=1}^p \text{Var}(x_k) \rightarrow \sigma^2$.

(C3) ρ mixing 条件;

$$r \rightarrow \infty \text{ とき, } \sup_{|i-j| \geq r} |\text{E}(X_i X_j)| \leq \rho(r) \rightarrow 0$$

をみたす.

条件 (C3) をさらに緩めた結果については, Ahn et al. (2007) を参照されたい.

これらの性質は, 2 群の p 次元判別問題に応用できる. 母集団 Π_1, Π_2 における p 次元変数を X, Y とする. X は上記の条件をみたし, 同様に Y もその条件をみたすとする. ただし, (2) におけるパラメータ σ^2 は τ^2 とする. さらに, 2 群は

$$\frac{1}{p} \sum_{k=1}^p \{\text{E}(X_k) - \text{E}(Y_k)\}^2 \rightarrow \mu^2$$

をみたしているとする. 今 X についての大きさ n の標本 X_1, \dots, X_n と, Y についての大きさ m の標本 Y_1, \dots, Y_m が与えられているとする. このとき, $\sigma^2/n \geq \tau^2/m$ を仮定すると,

$$\mu^2 > \sigma^2/n - \tau^2/m$$

の場合, 新しいデータが線形な超平面によって正しく母集団 Π_1 に属していると分類される確率は $p \rightarrow \infty$ のときに, 1 に収束する.

各群の確率分布が複雑な非正規分布をしている場合の判別法の 1 つとして, サポートベクターマシン法がある (詳しくは Hastie, Buja. and Tibshirani (1994) などを参照). これは p 次元多変量変数 $\mathbf{X} = (X_1, \dots, X_p)'$ から新たな変数 $h_1(\mathbf{X}), \dots, h_m(\mathbf{X})$ を構成し, これらの 1 次式

$$b_0 + b_1 h_1(\mathbf{X}) + \dots + b_m h_m(\mathbf{X})$$

をもとにした判別法である. ここで, 係数 b_0, b_1, \dots, b_m は初期データをできるだけ分離する考え方の 1 つであるマージン基準にもとづいて数値的に決められる. このような判別法における変数の数は m であって, 高次元になると考えられる. したがって, 上記のような判別に関する高次元漸近的結果をサポートベクターマシン法へ応用することが期待される.

5 付録: 多変量解析の基礎

多くの個体 (被験者) について, 2つ以上の変数の測定値が与えられたときの分析法は多変量解析とよばれる. 多変量解析においては, 各変数間の相関関係を利用して, 個々の変数だけの分析では得られない新たな情報を得ること目指している. ここでは, 第4節の話題と関連するいくつかの伝統的な多変量解析法について解説する. これらについてのより詳細な記述に関しては, 多変量解析の専門書 Anderson (2003), Siotani et al.(1985), 塩谷 (1990), Fujikoshi et al.(2010)などを参照されたい.

5.1 主成分分析

p 個の変数 X_1, \dots, X_p をまとめて p 次元確率ベクトル $\mathbf{X} = (X_1, \dots, X_p)'$ として表す. \mathbf{X} の平均ベクトル, 共分散行列はそれぞれ

$$E(\mathbf{X}) = \boldsymbol{\mu} = (\mu_1, \dots, \mu_p)', \quad \text{Var}(\mathbf{X}) = \Sigma = (\sigma_{ij})$$

として定義される. ここに, $E(X_i) = \mu_i, \text{Cov}(X_i, X_j) = \sigma_{ij}$ である.

主成分分析は, \mathbf{X} の変動をできるだけ少数個の主成分とよばれる変換変数で説明することを目的にしている. 主成分は元の変数の平均のまわりの1次結合 $\gamma_1(X_1 - \mu_1) + \dots + \gamma_p(X_p - \mu_p)$ を考え, その分散が最大になるように決められる. このとき, 係数を大きくすると分散はいくらでも大きくなるので, 係数の2乗和が1であるという条件のもとで分散が最大になるようにする. このようにして第1主成分が定まる. 次に, この主成分とは無相関なものの中から分散が最大になるものを定める. 以下, この手続きを繰り返して p 個の主成分が定義されるが, これらは次のように与えられる. Σ の固有値を $\lambda_1 \geq \dots \geq \lambda_p > 0$ とし, 対応する正規直交固有ベクトルを $\gamma_1, \dots, \gamma_p$ とする. すなわち

$$\Sigma \gamma_i = \lambda_i \gamma_i, \quad \gamma_i' \gamma_j = \delta_{ij}, \quad i, j = 1, \dots, p$$

をみtas. ここで, δ_{ij} はクロネッカーのデルタとよばれる記号であって, $i = j$ のとき1, $i \neq j$ のとき0を意味している. このとき, 第 i 母集団主成分は

$$Y_i = \gamma_{1i}(X_1 - \mu_1) + \dots + \gamma_{pi}(X_p - \mu_p) = \boldsymbol{\gamma}_i'(\mathbf{X} - \boldsymbol{\mu}), \quad i = 1, \dots, p$$

で定義される. ここに, $\boldsymbol{\gamma}_i = (\gamma_{1i}, \dots, \gamma_{pi})'$ である. 第 i 固有値 λ_i が単根であると, 第 i 主成分は符号を除いて一意的に定義される. しかし, 重根の場合

合にはそれに対応する主成分は一意的ではない．主成分について

$$E(Y_i) = 0, \text{Var}(Y_i) = \lambda_i, \text{Cov}(Y_i, Y_j) = 0 \quad (i \neq j)$$

が成り立つ．第 i 番目の固有値 λ_i は第 i 主成分 Y_i の重要度を表し， $\lambda_i / (\lambda_1 + \dots + \lambda_p)$ は第 i 主成分の寄与率とよばれる．最初の q 個の主成分によって説明される割合は

$$(\lambda_1 + \dots + \lambda_q) / (\lambda_1 + \dots + \lambda_p)$$

であって，これは第 q 主成分までの累積寄与率とよばれる．第 i 主成分と元の変数との関連の度合いは，基準化変数の係数の絶対値 $|\gamma_{1i}| \sqrt{\sigma_{11}}, \dots, |\gamma_{pi}| \sqrt{\sigma_{pp}}$ が目安になる．他の基準としては，主成分 Y_i と変数 X_j の相関係数

$$\rho(Y_i, X_j) = \frac{\text{Cov}(Y_i, X_j)}{\sqrt{\text{Var}(Y_i)\text{Var}(X_j)}} = \frac{\sqrt{\lambda_j} \gamma_{ji}}{\sqrt{\sigma_{jj}}}$$

が用いられる． $\rho(Y_i, X_j)$ は主成分 Y_i の変数 X_j への因子負荷量とよばれる．

$N = n + 1 (\geq p)$ 個の標本に基づく標本ベクトル，標本共分散行列をそれぞれ \bar{X}, S とする． S の固有値を ℓ_1, \dots, ℓ_p ($\ell_1 \geq \dots \geq \ell_p > 0$) とし，対応する正規直交固有ベクトルを h_1, \dots, h_p とする．このとき，第 i 標本主成分は

$$Y_i = h_{1i}(X_1 - \bar{X}_1) + \dots + h_{pi}(X_p - \bar{X}_p) = \mathbf{h}'_i(\mathbf{X} - \bar{\mathbf{X}}), \quad i = 1, \dots, p$$

で定義される．ここに， $\mathbf{h}_i = (h_{1i}, \dots, h_{pi})'$ である．これまで母集団主成分の性質を述べてきたが，標本主成分の場合も，平均や分散を標本平均や標本分散に置き換えることによって，同様な性質が導出される．また，標本の場合の累積寄与率は λ_i を ℓ_i で置き換えればよい．

各個体について主成分の値 (得点) が定義される．個体 i の第 j 主成分得点は

$$\begin{aligned} Y_{ij} &= h_{1j}(X_{i1} - \bar{X}_1) + \dots + h_{pj}(X_{ip} - \bar{X}_p) \\ &= \mathbf{h}'_j(\mathbf{X}_i - \bar{\mathbf{X}}), \quad i = 1, \dots, n; \quad j = 1, \dots, p \end{aligned}$$

で定義される．主成分得点を利用して，各個体を低次元空間へ配置し，そこからデータの背後に潜む構造を探ることが試みられる．例えば，第 1 主成分と第 2 主成分を用いた 2 次元空間への配置の場合，個体 j の座標は (Y_{j1}, Y_{j2}) である．

主成分分析においては、何個の主成分を用いればよいかを決めることは重要な課題である。1つの目安は標本累積寄与率を求め、それが例えば0.8を越えたときの主成分を用いることである。また、小さい固有値の同等性仮説 $\lambda_{q+1} = \dots = \lambda_p$ を、与えられた q ($0 \leq q < p - 1$) に対して検定することが考えられている。さらに、このような仮説に関して、適切な q を決めることも試みられている。

5.2 正準相関分析

2つの多変量変数 $\mathbf{X} = (X_1, \dots, X_p)'$ と $\mathbf{Y} = (Y_1, \dots, Y_q)'$ との相関関係を最も簡潔に表す方法として、正準相関分析法がある。以下では一般性を失うことなく、 $p \leq q$ とし

$$\text{Var}\left(\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}\right) = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

とする。 \mathbf{X}, \mathbf{Y} の1次結合

$$\xi = \alpha_1 X_1 + \dots + \alpha_p X_p = \boldsymbol{\alpha}' \mathbf{X}, \quad \eta = \beta_1 Y_1 + \dots + \beta_q Y_q = \boldsymbol{\beta}' \mathbf{Y}$$

を考える。ここに、 $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)'$ 、 $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)'$ である。このような1次結合のうちで、以下の(1), (2), (3)をみたす1次結合

$$\xi_i = \boldsymbol{\alpha}'_i \mathbf{X}, \quad (i = 1, \dots, p), \quad \eta_j = \boldsymbol{\beta}'_j \mathbf{Y}, \quad (j = 1, \dots, q)$$

を定めることができる。ここに、 $\boldsymbol{\alpha}_i = (\alpha_{1i}, \dots, \alpha_{pi})'$ 、 $\boldsymbol{\beta}_j = (\beta_{1j}, \dots, \beta_{qj})'$ である。このとき、一般性を失うことなく、 $\text{Var}(\xi) = 1, \text{Var}(\eta) = 1$ としてよいので、これを仮定する。

$$(1) \quad \rho(\xi_1, \eta_1) = \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \rho(\xi, \eta).$$

$$(2) \quad k \leq p \text{ のとき, 条件 } \rho(\xi, \xi_i) = \rho(\eta, \eta_i) = 0, \quad i = 1, \dots, k-1 \text{ のもとで } \rho(\xi, \eta) \text{ が } \boldsymbol{\alpha}, \boldsymbol{\beta} \text{ に関して最大になるのは, } \boldsymbol{\alpha} = \boldsymbol{\alpha}_k, \boldsymbol{\beta} = \boldsymbol{\beta}_k \text{ のときである.}$$

$$(3) \quad k > p \text{ のとき, } \rho(\eta_k, \eta_i) = 0, \quad i = 1, \dots, k-1.$$

このような $\xi_i = \boldsymbol{\alpha}'_i \mathbf{x}$ 、 $\eta_i = \boldsymbol{\beta}'_i \mathbf{y}$ に対して、 $\rho(\xi_i, \eta_i) = \rho_i$ とおく。このとき、 $\rho_1 \geq \dots \geq \rho_p \geq 0$ で、 ρ_i を第 i 正準相関係数、 (ξ_i, η_i) を第 i 正準相関変

数という。 $p = q = 1$ の場合の正準相関係数は相関係数の絶対値に等しく、
 $p = 1 < q$ の場合には重相関係数に等しい。

正準相関係数 2 乗 $\rho_1^2 \geq \dots \geq \rho_p^2$ は固有方程式

$$|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \rho^2\Sigma_{11}| = 0$$

の解である。また、係数ベクトル α_i, β_j はそれぞれ固有値問題

$$\begin{aligned}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\alpha_i &= \rho_i^2\Sigma_{11}\alpha_i, & \alpha_i'\Sigma_{11}\alpha_j &= \delta_{ij}, \\ \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\beta_j &= \rho_j^2\Sigma_{22}\beta_j, & \beta_j'\Sigma_{22}\beta_i &= \delta_{ij}\end{aligned}$$

の解である。ここに、 $\rho_{p+1} = \dots = \rho_q = 0$ とする。

ゼロでない正準相関係数の個数は、 $\text{rank}(\Sigma_{12})$ に等しく、正準相関の次元とよばれる。一般に、最初の k 個の正準相関変数によって説明される割合は

$$(\rho_1^2 + \dots + \rho_k^2) / (\rho_1^2 + \dots + \rho_p^2)$$

で定義され、最初の k 個の正準相関変数の累積寄与率とよばれる。

2つの多変量変数 $X = (X_1, \dots, X_p)'$ と $Y = (Y_1, \dots, Y_q)'$ について、大きさ $N = n + 1$ の標本に基づく標本共分散行列を S 、標本相関行列を R とし、次のように分割する。

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \quad R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}.$$

ただし、 $S_{12} : p \times q$, $R_{12} : p \times q$. 標本正準相関変数は母集団正準相関変数において、平均ベクトル μ 、共分散行列 Σ をそれぞれ標本平均ベクトル \bar{X} 、標本共分散行列 S で置き換えて定義される。なお、基準化変数の標本正準相関変数は S を R で置き換えて定義されるが、標本正準相関係数 $r_1 > \dots > r_p > 0$ は不変である。したがって、標本正準相関係数の 2 乗は固有方程式

$$|S_{12}S_{22}^{-1}S_{21} - r^2S_{11}| = 0 \quad \text{または} \quad |R_{12}R_{22}^{-1}R_{21} - r^2R_{11}| = 0$$

の解である。

5.3 判別分析

ある個体の複数個の変数の観測値にもとづいて、その個体が 2 つ (あるいはそれ以上) の群のいずれに属するかを決定する問題は判別問題とよばれる。

判別分析においては、このような判別問題と関連して、複数個の群の間の差異を決定付ける少数個の変数あるいは変換変数を見つけることも重要な課題である。判別の問題には、例えば、次のような例がある。

- (1) 主婦の調査データを利用して、ある属性をもつ主婦が2つのブランド品のうちどちらを志向するかを判断したい。
- (2) ある患者の臨床所見を基にして、その患者が病気を患っているか否かを判別したい。
- (3) A氏, B氏, C氏, D氏の4人の中の誰かが書いたことは確かである小切手がある。そこに書かれている筆跡から、誰が書いたかを知りたい。
- (4) 作者不明のある作品の著者として、可能性のある人物が2人存在する場合、真の作者がそのどちらであるかを判定したい。
- (5) 国語, 数学, 社会が得意で、理科, 英語が不得意な高校生が、大学の理系, 文系のどちらへ志願したらよいかを判定したい。
- (6) ある患者のマイクロアレイデータにもとづいて、その患者がガンであるかどうかを判定したい。

一般に、判別分析においては複数個 (p 個) の変数が用いられる。ここで、標本の大きさ n は、 p に比べてかなり大きいのが普通である。しかし、上の (6) のような場合には、例えば、 $p = 3,000$ で $n = 100$ のように、逆転現象が生じていることを注意したい。このような場合の統計的問題は高次元問題、あるいは“ $n \ll p$ 問題”とよばれ、最近大きな関心が寄せられている。これについての最近の発展については、第4節を参照のこと。

2群の判別法について考えてみよう。群 $G_i (i = 1, 2)$ からの大きさ n_i の初期標本

$$G_1; \mathbf{X}_1^{(1)}, \mathbf{X}_2^{(1)}, \dots, \mathbf{X}_{n_1}^{(1)}, \quad G_2; \mathbf{X}_1^{(2)}, \mathbf{X}_2^{(2)}, \dots, \mathbf{X}_{n_2}^{(2)}$$

が与えられて、新たな観測値 \mathbf{X} が群 G_1 と G_2 のいずれに属するかを判定したいとしよう。ここで、新たな観測値は群 G_1 あるいは G_2 のいずれかに属するものとする。初期標本より計算される、各群の標本平均ベクトル、標本共分散行列を $\bar{\mathbf{X}}^{(i)}, \mathbf{S}^{(i)} (i = 1, 2)$ とし、合併標本共分散行列を $\mathbf{S} = (1/n)\{(n_1 - 1)\mathbf{S}^{(1)} + (n_2 - 1)\mathbf{S}^{(2)}\}$ とする。ここに、 $n = n_1 + n_2 - 2$ である。代表的な判別法の一つは、線形判別関数

$$W = (\bar{\mathbf{X}}^{(1)} - \bar{\mathbf{X}}^{(2)})' \mathbf{S}^{-1} \left\{ \mathbf{X} - \frac{1}{2}(\bar{\mathbf{X}}^{(1)} + \bar{\mathbf{X}}^{(2)}) \right\}$$

を用いて, $W \geq 0$ ならば観測値 \mathbf{X} は G_1 に属すると判別し, $W < 0$ ならば観測値 \mathbf{X} は G_2 に属すると判別する方法である. このような判別の有効性は誤判別確率によって評価される. 誤判別確率には, 本来 G_1 に属するものを誤って G_2 に属すると判別する確率 $p(2|1)$ と, 本来 G_2 に属するものを誤って G_1 に属すると判別する確率 $p(1|2)$ があるが, これらは

$$p(2|1) = P(W < 0 \mid \mathbf{X} \in G_1), \quad p(1|2) = P(W \geq 0 \mid \mathbf{X} \in G_2)$$

と表せる. $p(2|1)$ を推定するナীবな方法は, 群 G_1 の n_1 個の標本に対して W の符号を調べ, それらの中で負になった割合として推定することである. これを改良した推定法としては, 初期標本のある個体を判別するとき, 判別関数はその個体を除いた $(n_1 + n_2 - 1)$ 個の標本から求め, これによってその個体を判別する方法である. このような考え方はクロスバリデーション (cross-validation, 交差検証法) とよばれ, 最近よく利用されている. 観測値の各群での分布が共通な共分散行列をもつ多変量正規分布である場合, 誤判別確率の n_1 と n_2 を大にしたときの漸近展開が求められている. また, このような結果を利用して, $p(2|1)$ に対する漸近的不偏推定量

$$\Phi\left(-\frac{1}{2}D\right) + \phi\left(-\frac{1}{2}D\right) \left[\frac{p-1}{n_1 D} + \frac{D}{32(n-2)} \{4(4p-1) - D^2\} \right]$$

が提案されている. ここに, Φ は標準正規分布の分布関数, ϕ は標準正規分布の確率密度関数, D は 2 群間の標本マハラノビスの距離であって $D^2 = (\bar{\mathbf{X}}^{(1)} - \bar{\mathbf{X}}^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{X}}^{(1)} - \bar{\mathbf{X}}^{(2)})$ である. $p(1|2)$ については n_1 と n_2 を入れ替えればよい.

その他の判別法として, 観測値 \mathbf{X} から群 G_i の中心までのマハラノビス平方距離

$$d_i^2 = (\mathbf{X} - \bar{\mathbf{X}}^{(i)})' (\mathbf{S}^{(i)})^{-1} (\mathbf{X} - \bar{\mathbf{X}}^{(i)})$$

を考え, それらの平方距離が最小となる群に判別する最小距離判別法がある. この方法は 2 群の共分散行列が異なる場合にも適用することができる. とくに, $\mathbf{S}^{(i)}$ として合併共分散行列 \mathbf{S} を用いると, 最小距離判別法は線形判別法と同値になる.

上で述べた方法は 3 群以上の判別に対しても拡張される. しかし, 多群の場合には, 次元縮小を伴う判別関数を用いた正準判別分析が利用される場合が多い. 2 群の場合と同様に, q 個の群 $G_i (i = 1, \dots, q)$ について大きさ n_i の初期標本が与えられているとする. このデータに関する群間平方和積和

行列を S_b , 群内平方和積和行列を S_w とする. これらは 1 元配置計画における群間平方和および群内平方和の多次元版であって

$$S_b = n_1(\bar{X}^{(1)} - \bar{X})(\bar{X}^{(1)} - \bar{X})' + \cdots + n_q(\bar{X}^{(q)} - \bar{X})(\bar{X}^{(q)} - \bar{X})',$$

$$S_w = (n_1 - 1)S^{(1)} + \cdots + (n_q - 1)S^{(q)}$$

として与えられる. ここに, \bar{X} は全標本の平均ベクトルである.

p 次元変数 $X = (X_1, \dots, X_p)'$ の 1 次結合

$$Z = a_1X_1 + \cdots + a_pX_p = \mathbf{a}'X$$

を考える. ここに $\mathbf{a} = (a_1, \dots, a_p)'$ である. このとき, $Z = \mathbf{a}'X$ の群間平方和, 群内平方和はそれぞれ $\mathbf{a}'S_b\mathbf{a}$, $\mathbf{a}'S_w\mathbf{a}$ と表せる. したがって, これらの比

$$(\mathbf{a}'S_b\mathbf{a})/(\mathbf{a}'S_w\mathbf{a})$$

が最大になる係数ベクトル \mathbf{a} が, 群間を最もよく分離する方向と考えられる. この最適化問題の解は, $S_w^{-1}S_b$ の最大固有値に対応する固有ベクトルである. 一般に, $S_w^{-1}S_b$ は s ($\leq \min(p, q-1)$) 個のゼロでない固有値 $\ell_1 > \dots > \ell_s > 0$ をもち, 対応する固有ベクトルを $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_s$ とする. したがって, これらは固有値問題

$$S_b\mathbf{a}_i = \ell_i S_w\mathbf{a}_i, \quad \mathbf{a}_i'S_w\mathbf{a}_j = n\delta_{ij}$$

の解である. ここに, $n = n_1 + \cdots + n_q$ である. このとき, 線形関数 $Z_i = \mathbf{a}_i'X$, $i = 1, \dots, s$ は第 i 正準判別変数とよばれる. とくに, $\mathbf{a} = \mathbf{a}_1$ は $(\mathbf{a}'S_b\mathbf{a})/(\mathbf{a}'S_w\mathbf{a})$ を最大にしている. 一般に, $Z_k = \mathbf{a}_k'X$ は, それ以前の z_1, \dots, z_{k-1} とは

$$\mathbf{a}'S_w\mathbf{a}_i = 0, \quad i = 1, \dots, k-1$$

を満たすという意味で無相関であって, かつ, このような性質をみたすものなかで $(\mathbf{a}'S_b\mathbf{a})/(\mathbf{a}'S_w\mathbf{a})$ を最大にするものである.

最初の k 個の判別関数を用いたときの各観測値 $X_j^{(i)}$ の判別得点は

$$Z_j^{(i)} = \begin{pmatrix} Z_{j1}^{(i)} \\ \vdots \\ Z_{jk}^{(i)} \end{pmatrix} = \begin{pmatrix} \mathbf{a}'_1 \\ \vdots \\ \mathbf{a}'_k \end{pmatrix} X_j^{(i)}, \quad j = 1, \dots, n_i; \quad i = 1, \dots, q$$

で定義される. この判別得点は, 例えば $k = 2$ として, 各個体を 2 次元空間へ配置するときの座標として利用できる. また, 所属不明の観測値 X を判

別するには, 判別得点

$$\begin{aligned} \mathbf{Z} &= (Z_1, \dots, Z_k) \\ &= (\mathbf{a}_1, \dots, \mathbf{a}_k)' \mathbf{X} = \mathbf{A}\mathbf{X} \end{aligned}$$

を求め, この点から各群の中心の判別得点 $\bar{\mathbf{Z}}^{(i)} = \mathbf{A}\bar{\mathbf{X}}^{(i)}$ までの距離を比較し, 最も近い群に分類すればよい. このとき, Z_1, \dots, Z_k は互いに無相関であるので, 距離としては通常距離を用いてよい. すなわち, $d_i = \|\mathbf{Z} - \bar{\mathbf{Z}}^{(i)}\|$, $i = 1, \dots, q$ とおくと

$$\min\{d_1, \dots, d_q\} = d_i \Rightarrow \mathbf{X} \in G_i$$

と判別する. $q = 2$ の場合の判別法は, 線形判別関数による方法と同じである.

5.4 ウィシャート分布と多変量特性量の分布

多変量解析の推測理論の展開においては, 基礎になる分布として多変量正規分布を想定する場合が多い. このような場合, カイ 2 乗分布の多変量への拡張であるウィシャート分布が基本的となる. 一般に, p 次の確率行列 \mathbf{W} が n 個の独立な p 次元正規確率変数 $U_j \sim N_p(\boldsymbol{\mu}_j, \Sigma)$, $j = 1, \dots, n$ を用いて

$$\mathbf{W} = \sum_{j=1}^n \mathbf{U}_j \mathbf{U}_j'$$

と表せるとき, \mathbf{W} は自由度 n , 非心行列 $\Delta = \boldsymbol{\mu}_1 \boldsymbol{\mu}_1' + \dots + \boldsymbol{\mu}_n \boldsymbol{\mu}_n'$ の p 次元非心ウィシャート分布に従うといい, $\mathbf{W} \sim W_p(n, \Sigma; \Delta)$ とかく. とくに, $\Delta = \mathbf{0}$, すなわち, $\boldsymbol{\mu}_j = \mathbf{0}$, $j = 1, \dots, n$ のとき, \mathbf{W} は単にウィシャート分布に従うといい, $\mathbf{W} \sim W_p(n, \Sigma)$ とかく.

主成分分析において, 大きさ $N = n + 1$ の標本が正規母集団 $N_p(\boldsymbol{\mu}, \Sigma)$ からのものであるとしよう. このとき, 標本共分散行列 \mathbf{S} の分布について, $n\mathbf{S} \sim W_p(n, \Sigma)$ となる. 第 i 主成分の係数は \mathbf{S} の第 i 固有値に対応するベクトルである. また, 第 i 主成分の重要度あるいは説明力は \mathbf{S} の第 i 固有値で測られる. したがって, ウィシャート分布に従う確率行列の固有値・固有ベクトルの分布は統計的推測において基本的となる. 正準相関分析の場合, $(p + q)$ 次元の標本共分散行列 \mathbf{S} から出発し, これを $\mathbf{S}_{11} : p \times p$, $\mathbf{S}_{12} : p \times q$, $\mathbf{S}_{22} : q \times q$ と分割したとき, $\mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1} \mathbf{S}_{21}$ あるいは $\mathbf{S}_{22}^{-1} \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12}$ の固有値・固有ベクトルの分布が基本的になる.

正準判別分析の場合, 群間平方和積和行列 S_b と群内平方和積和行列 S_w が用いられる. 正準判別変数の係数ベクトルやその重要度あるいは判別力は, $S_w^{-1}S_b$ の固有値・固有ベクトルで与えられる. 群 G_i からの標本が平均ベクトル μ_i で共通な共分散行列 Σ をもつ正規母集団から得られたとしよう. このとき, S_b と S_w は互いに独立で, $S_b \sim W_p(q-1, \Sigma; \Omega)$, $S_w \sim W_p(n-q, \Sigma)$ である. ここに, $n = n_1 + \dots + n_q$, 非心行列は $\bar{\mu} = (1/n)(n_1\mu_1 + \dots + n_1\mu_q)$ とおくと $\Omega = \sum_{i=1}^q n_i(\mu_i - \bar{\mu})(\mu_i - \bar{\mu})'$ である. 群間の有意差検定, すなわち, 仮説 $\mu_1 = \dots = \mu_q$ の検定に対して, 次の検定統計量が提案されている.

(i) 尤度比統計量; $T_{LR} = -(n + d_1) \log(|S_w|/|S_w + S_b|)$.

(ii) ローレイ・ホテルグ基準; $T_{LH} = (n + d_2) \text{tr } S_b S_w^{-1}$.

(iii) バートレット・ナンダ・ピライ基準;

$$T_{BNP} = (n + d_3) \text{tr } S_b (S_w + S_b)^{-1}.$$

ここに, d_j は各標本が正規分布に従うときのカイ 2 乗近似を改良するためのバートレット補正項であり, それぞれ, $d_1 = -(p+q+2)/2$, $d_2 = -(p+q+1)$, $d_3 = -1$ で与えられる. この他最大固有値基準 ℓ_1 も用いられる.

多変量特性量の分布は, 基礎分布が多変量正規分布であっても特別な場合を除き, 正確に求めるのは困難である. このため, 標本数が大のときの漸近展開が導出されて来ている. また, これらの漸近理論は多変量非正規モデルの場合にも求められるようになっている (Anderson (2003), 藤越 (2003), Fujikoshi et al. (2010) などを参照).

謝辞

編集者, 査読者から最初の原稿に対して多くの有益なコメント頂きました. ここに記して謝意を表します.

参考文献

- [1] Ahn, J., Marron, J. S., Muller, K. M. and Chi, Y.-Y. (2007). The high-dimensional, low-sample-size geometric representation holds under mild conditions. *Biometrika*, **94**, 760-766.
- [2] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed.). John Wiley & Sons, New York.

- [3] Bai, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, **9**, 611-677.
- [4] Cuadrass, C. M. and Rao, C. R. (Ed.) (1995). *Multivariate Analysis 2; Future Direction*. North-Holland, Amsterdam.
- [5] Dempster, A. P. (1958). A high dimensional two sample significance test. *Ann. Math. Statist.*, **29**, 995-1010.
- [6] Dudoit, S., Fridltano, J. and Speed, T. P. (2002). Comparison of discrimination methods for classification of Tumors using gene expression data. *J. Amer. Stat. Assoc.*, **97**, 78-87.
- [7] Friedman, J. H. (1989). Reguralized discriminant analysis. *J. Amer. Statist. Assoc.*, **84**, 165-175.
- [8] 藤越 康祝, 柳井 晴夫, 田栗 正章 (訳)(1993). 統計学とは何か. 丸善株式会社. 原著: 「Rao, C. R. (1997). *Statistics and Truth*. World Scientific」のタイプ原稿.
- [9] 藤越 康祝, 柳井 晴夫, 田栗 正章 (訳)(2010). 統計学とは何か. 筑摩書房株式会社. 原著: Rao, C. R. (1997). *Statistics and Truth*. World Scientific.
- [10] Fujikoshi, Y. (2000). Error bounds for asymptotic approximations of the linear discriminant function when the sample size and dimensionality are large. *J. Multivariate Anal.*, **73**, 1-17.
- [11] 藤越 康祝 (2003). 多変量解析へのチャレンジ: 現状と展望. 日本統計学会誌, **33**, 273-306.
- [12] Fujikoshi, Y., Himeno, T. and Wakaki, H. (2004). Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large. *J. Japan Statist. Soc.*, **34**, 19-26.
- [13] Fujikoshi, Y., Yamada, T., Watanabe, D. and Sugiyama, T. (2007). Asymptotic distribution of the LR statistic for equality of the smallest eigenvalues in high-dimensional principal component analysis. *J. Multivariate Anal.*, **98**, 2002-2008.
- [14] Fujikoshi, Y., Himeno, T. and Wakaki, H. (2008). Asymptotic results in MANOVA model when the dimension is large compared to the sample size. *J. Statist. Plann. Inf.*, **138**, 3457-3466

- [15] Fujikoshi, Y. and Sakurai, T. (2009). High-dimensional asymptotic expansions of the distributions of canonical correlations. *J. Multivariate Anal.*, **100**, 231-242.
- [16] Fujikoshi, Y., Ulyanov, V. V. and Shimizu, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley, Hoboken, New Jersey.
- [17] Ghosh, D. (2003). Penalized discriminant methods for the classification of tumors from gene expression data. *Biometrics*, **59**, 992-1000.
- [18] Hall, P., Marron, J. S. and Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *J. R. Statist. Soc. B*, **67**, 427-444.
- [19] Hastie, T., Buja, A. and Tibshirani, R. (1994). Penalized discriminant analysis. *Ann. Statist.*, **23**, 173-102.
- [20] 伊藤 孝一 (2007). 統計学の現状と課題—統計教育の視点から—. 日本統計学会和文誌, **36**, 231-249.
- [21] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *Ann. Statist.*, **29**, 295-327.
- [22] Johnstone, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist.*, **36**, 2638-2716.
- [23] Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.*, **30**, 1081-1102.
- [24] Raftery, A. E., Tanner, M. A. and Wells, M. T. (Ed.)(2002). *Statistics in the 21st Century*. Chapman & Hall/CRC.
- [25] Rao, C. R. (1997). *Statistics and Truth (2nd Ed.)*. World Scientific.
- [26] Rao, C. R. (Ed.) (1993). *Multivariate Analysis; Future Direction*. North-Holland, Amsterdam.
- [27] Rao, C. R. (2006). The past, present and future of statistics. *IMS Bulletin*, **35-2**, 4-5.
- [28] Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, **92**, 951-956.

- [29] Schott, J. R. (2006). A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *J. Multivariate Anal.*, **97**, 827-843.
- [30] 塩谷實 (1990). 多変量解析概論. 朝倉書店.
- [31] Siotani, M., Hayakawa, T., and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*. American Sciences Press, Columbus, Ohio.
- [32] Srivastava, M. (2007). Multivariate analysis for analyzing high dimensional data. *J. Japan Statist. Soc.*, **37**, 53-86.
- [33] Srivastava, M. S. and Kubokawa, T. (2007). Comparison of discrimination methods for high dimensional data. *J. Japan Statist. Soc.*, **37**, 123-134.
- [34] 竹内 啓 (1998). 統計的推測理論の展開. 「20世紀の数学」(数理学部編集) , 123-128, サイエンス社.
- [35] 田栗 正章, 藤越 康祝, 柳井 晴夫, ラオ, C. R. (2007). やさしい統計入門. 講談社ブルーバックス.
- [36] Tracy, C. A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, **177**, 727-754.
- [37] Wakaki, H., Fujikoshi, Y. and Ulyanov, V. (2003). Asymptotic expansions of the distributions of MANOVA test statistics when the dimension is large. TR 02-9, Statistical Research Group, Hiroshima Univ., Japan.

第4章

線形混合モデルの理論と応用 - 特に小地域推定を巡って -

久保川 達也¹

(東京大学・大学院経済学研究科・教授)

線形混合モデルの特徴は、観測値を共変量を用いて回帰するとき、個体や地域の違いを共変量として組み入れ、それらの背後に共通な確率分布を想定して個体や地域の差異を推定している点である。全体の特性値だけでなく個体や地域ごとの特性値への関心が高まるにつれ、個々の差異を共変量として捉えた線形混合モデルについての研究が盛んになり、このモデルの研究が始まった家畜育種学の分野はもとより医学・生物学分野から経済・教育など社会科学の分野、特に官庁統計分野での小地域推定において利用されている。本稿では、線形混合モデルとそこから導かれる経験最良線形不偏予測量について解説し、そのモデルがもっている予測精度を高めるための仕組みや経験最良線形不偏予測量の予測誤差の評価について小地域推定に焦点を当てて説

¹tatsuya@e.u-tokyo.ac.jp

明する。また経時測定データを解析するための線形混合モデルについても紹介し、地価公示価格データへの応用例を与える。

1 はじめに

線形混合モデル (Linear Mixed Model, LMM) と最良線形不偏予測量 (Best Linear Unbiased Predictor, BLUP) についての研究は C.R. Henderson の論文以来 50 年以上にわたって発展してきた。当初は、家畜育種学の分野で個体のもつ遺伝的能力などの推定を行うために研究されたが、次第に線形混合モデルの有用性が広く認識され、またベイズモデルとの関連においてベイズ推測の理論と計算方法についての顕著な発展に伴って、現在では実に広い分野で利用されている。LMM の離散分布への拡張である一般化線形混合モデル (Generalized Linear Mixed Model, GLMM) を含めれば、線形混合モデルに関する文献はかなりの量になっていることから、理論と応用の両面から関心が高いことがわかる。

LMM の応用例の一つに小地域推定の問題がある。これは標本調査に関連した問題で、通常は調査区全体の特性を調べるために標本調査が行われるが、そのデータを利用して地域ごとの特性値を推定したい状況がしばしば生ずる。例えば、得られたデータから各地域への予算配分の仕方を決めたり、政策を決定したりする場合がある。そのとき、狭い地域や人口が粗な地域に対しては十分なデータがとられていないため、その地域のデータだけでは特性値の十分な推測ができない。このような状況での推定問題を小地域推定という。この問題を解決する方法は、周辺地域のデータを組み込んで推定精度を高めることであり、どのような形でデータを取り込むかがポイントになる。そのために利用されるのが LMM であり、そのモデルから導かれる経験最良線形不偏予測量 (Empirical Best Linear Unbiased Predictor, EBLUP) が小地域の安定した推定値を与えるのに役立つ。では、LMM がそのような性質をもつのはなぜであろうか。LMM は、基本的に共通母数に基づいて回帰する項と地域の差異を表す変量効果の項及び誤差項とから構成されている。すべての地域を通して回帰係数を共通に設定することによってすべてのデータをプールして安定した推定値を与えることができる。しかし、これだけでは地域の特徴や地域による差異を引き出すことができない。そこで地域の差異を変量効果としてモデルに取り込む。この効果を予測してやることにより、標本平均を縮小する作用が生ずることになる。LMM は、母数の共通化によるデータのプーリングと変量効果による標本平均の

縮小作用を生み出すことのできるモデルであり，その結果生ずる予測量が EBLUP となる。したがって，EBLUP は，各々の地域の標本平均とプールされた回帰推定量との加重平均になっており，データ数が少ないときには標本平均をプールされた推定値の方向へ縮小することにより，推定精度の改善が図られている。

本稿では LMM の理論と小地域推定への応用について解説する。2 節では，LMM の紹介，混合モデル方程式と BLUP の説明，変量効果と共通母数の役割，分散成分を推定するための最尤法と制限最尤法についての解説を行う。3 節では，小地域推定の問題に焦点をしばり，予測精度を高めるために導出された EBLUP が実際どの程度推定誤差を改善しているのかについて，平均 2 乗誤差とその推定方法について説明する。また信頼区間の構成を行い，これらを用いた地価公示価格データへの応用を与える。4 節では，LMM の様々な応用や拡張について紹介する。経時測定データを解析するための LMM の紹介を行い，小地域推定のためのモデルの修正と上記の価格データへの適用結果について説明する。最後に GLMM への拡張，階層ベイズモデルへの拡張について若干の説明を与える。

なお，LMM や GLMM の解説書については，広津 (1992)，McCulloch and Searle (2001)，McCulloch (2003)，佐々木 (2007)，特に本格的なものとして Searle, Casella and McCulloch (1992)，Demidenko (2004)，小地域推定に関するものとして Rao (2003) が挙げられるので参照してほしい。

2 線形混合モデルとその特徴

2.1 線形混合モデル

[1] 枝分かれ誤差回帰モデル. 線形混合モデルが使われている分野の一つに小地域推定がある。線形混合モデルをわかりやすく説明するために，その分野の啓蒙的な論文として知られる Battese, Harter and Fuller (1988) が取り上げた推定問題について紹介しよう。アイオワ州の k 個の郡について穀物（とうもろこし，大豆等）の作付面積の調査がなされた。 k 個の郡それぞれをさらに約 250h の農作区画 (segment) に細分し，その中から n_i 個の区画をランダムに抽出し，直接農家にインタビュー調査を行うことによってそれぞれの区画におけるトウモロコシの作付面積についてのデータをとった。 i 番目の郡における j 番目の区画に対するこのデータを y_{ij} で表すことにする。他方，人工衛星 LANDSAT からの観測により，約 0.45h のピクセ

ル (picture element, 画像から識別する単位) に対していずれの穀物が作付けされているのかが識別され, k 個の郡すべてにわたってこうした衛星データが補助情報として利用可能である。トウモロコシ, 大豆に関してピクセル単位での識別が可能であるとして, i 番目の郡における j 番目の区画に対して, トウモロコシ, 大豆に識別されたピクセルの個数を x_{1ij}, x_{2ij} とすると, y_{ij} と (x_{1ij}, x_{2ij}) との間に線形関係が認められるため,

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i$$

なるモデルが想定できる。ここで $\mathbf{x}_{ij} = (1, x_{1ij}, x_{2ij})'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, $\mathbf{x}'_{ij}\boldsymbol{\beta} = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2$ である。また u_{ij} は y_{ij} を \mathbf{x}_{ij} で説明したときの誤差項であるが, これが郡に依存する項 v_i と郡の差異に依らない項 e_{ij} に加法的に分解されて

$$u_{ij} = v_i + e_{ij} \tag{2.1}$$

と表されるとする。 v_i は地域の差異を表しているので地域効果と呼ばれる。これは未知母数もしくは確率変数として扱われ, それぞれ母数効果, 変量効果と呼ばれる。どちらで扱うかは実際の問題に依存しており応用する際に検討する必要がある。地域の差異 v_i についてその背後に共通な分布が想定できるか否かに従って判断するのがよい。 v_i の背後に共通な分布が想定できる場合, すなわち v_i が変量効果として扱えるときには, v_i, e_{ij} はすべて互いに独立な確率変数とし, それぞれ正規分布

$$v_i \sim \mathcal{N}(0, \sigma_v^2), \quad e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$$

に従うとする。すると, 線形モデルは

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + v_i + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \tag{2.2}$$

と表される。 σ_v^2, σ_e^2 を一般に分散成分といい, 特にこのモデルについては σ_v^2 を群間成分, σ_e^2 を群内成分という。 $\boldsymbol{\beta}$ は回帰係数の未知母数ベクトルなので, このモデルの主要項は (母数効果) + (変量効果) の形に書かれており, (2.2) のようなモデルを一般に線形混合モデル (LMM) という。特に, y_{ij} の分散が分散成分の線形結合で書かれるときには分散成分モデル (Variance Component Model) という。(2.2) は誤差項が枝分かれ配置しているので, 枝分かれ誤差回帰モデル (Nested Error Regression Model) ともいわれる。

[2] モデルの行列表現. モデルを行列を用いて表現しておくとう便利である。上の例では $\mathbf{x}_{ij}, \boldsymbol{\beta}$ は 3×1 のベクトルであるが, これ以降は $p \times 1$ の

ベクトルとして一般的な場合を扱うことにする。ベクトルと行列の大きさを適当に揃えて

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{pmatrix},$$

$$\mathbf{x}'_i = \begin{pmatrix} \mathbf{x}'_{i1} \\ \vdots \\ \mathbf{x}'_{in_i} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_k \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$

とし、 \mathbf{e} も \mathbf{y} と同様に $\mathbf{e}_i = (e_{i1}, \dots, e_{in_i})'$, $\mathbf{e} = (\mathbf{e}'_1, \dots, \mathbf{e}'_k)'$ と定義する。また、すべての成分が 1 の $n_i \times 1$ ベクトルを \mathbf{j}_{n_i} で表し、ブロック対角行列 $\text{block diag}(\cdot)$ を用いて $\mathbf{Z} = \text{block diag}(\mathbf{j}_{n_1}, \dots, \mathbf{j}_{n_k})$ とおき、 $\mathbf{v} = (v_1, \dots, v_k)'$ とおく。このときモデル (2.2) は、 $N = \sum_{i=1}^k n_i$ に対して

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \quad (2.3)$$

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I}_k), \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I}_N)$$

と表すことができる。 \mathbf{y}_i の共分散行列は $\text{Cov}(\mathbf{y}_i) = \boldsymbol{\Sigma}_i(\sigma_e^2, \sigma_v^2) = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 \mathbf{J}_{n_i}$ と表される。ここで、 \mathbf{I}_{n_i} は $n_i \times n_i$ の単位行列、 $\mathbf{J}_{n_i} = \mathbf{j}_{n_i} \mathbf{j}'_{n_i}$ はすべての要素が 1 の $n_i \times n_i$ 行列である。従って、 \mathbf{y} の分散共分散行列は

$$\text{Cov}(\mathbf{y}) = \boldsymbol{\Sigma}(\sigma_e^2, \sigma_v^2) = \text{block diag}(\boldsymbol{\Sigma}_1(\sigma_e^2, \sigma_v^2), \dots, \boldsymbol{\Sigma}_k(\sigma_e^2, \sigma_v^2))$$

と書ける。

v_i を母数効果とした場合には $\text{Cov}(\mathbf{y}_i) = \sigma_e^2 \mathbf{I}_{n_i}$ となって \mathbf{y}_i の成分は互いに独立に分布するのに対して、変量効果とした場合 $\sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 \mathbf{J}_{n_i}$ という相関構造をもつことがわかる。一般に相関関係を利用してより推定精度の高い推定手法を導出することが可能になることは次の節で説明する。すなわち相関構造をどのように組み入れるかがより優れた推定を行う上で重要である。その際、誤差項の共分散行列に直接相関構造を埋め込んでしまえばよいわけであるが、変量効果を導入することによってモデルの意味が理解しやすくなるだけでなく、より豊かで複雑なモデルの構築も可能になる。さらに、変量効果の入っているモデルはベイズモデルの枠組みで捉えることができるため、たとえ複雑なモデルを作ったとしても、マルコフチェーン・モンテカルロ (MCMC) 法を利用して数値的な計算が可能になる。

[3] 一般的な線形混合モデル. (2.3) で記述されたモデルは, より一般的な線形混合モデル

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \\ \mathbf{v} &\sim \mathcal{N}_q(\mathbf{0}, \mathbf{G}), \quad \mathbf{e} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{R}) \end{aligned} \quad (2.4)$$

に拡張することができる。ここで, \mathbf{y} は $N \times 1$ の観測データのベクトル, \mathbf{X} は $N \times p$ の共変量からなる既知の行列, \mathbf{Z} は $N \times q$ の既知の計画行列である。 \mathbf{y} の共分散行列は

$$\text{Cov}(\mathbf{y}) = \boldsymbol{\Sigma} = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}' \quad (2.5)$$

と表される。共分散行列 \mathbf{G} , \mathbf{R} は一般に分散成分を含む母数を用いて表されるので, それらを $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)$ として $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$ と書く。

2.2 混合モデル方程式と最良線形不偏予測量 (BLUP)

[1] BLUP. 一般的な線形混合モデル (2.4) において観測できない母数効果 $\boldsymbol{\beta}$ と変量効果 \mathbf{v} の推定を考えよう。 \mathbf{v} は変量効果なので推定というよりは予測という方がふさわしい。共分散行列 \mathbf{G} , \mathbf{R} が既知の場合には \mathbf{v} の標準的な予測量は \mathbf{y} の線形関数として与えられる。線形で不偏な予測量の中で最良なものを最良線形不偏予測量 (BLUP) といい $\hat{\mathbf{v}}$ で表すことにする。また $\boldsymbol{\beta}$ の最良線形不偏推定量を $\hat{\boldsymbol{\beta}}$ で表すと, それらは次の連立方程式の解として与えられることが Henderson (1950) によって示された。

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \quad (2.6)$$

これは, 混合モデル方程式と呼ばれ, その解は

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}, \quad \hat{\mathbf{v}} = \mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (2.7)$$

で与えられる。ここで $(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$ は \mathbf{X} のランクが落ちている場合を考慮した $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$ の一般化逆行列を表している。 $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の一般化最小 2 乗推定量 (GLS) であることがわかる。既知のベクトル $\mathbf{a} \in \mathbf{R}^p$, $\mathbf{b} \in \mathbf{R}^q$ に対して $\mu = \mathbf{a}'\boldsymbol{\beta} + \mathbf{b}'\mathbf{v}$ を推定したいときには, その BLUP は

$$\hat{\mu} = \mathbf{a}'\hat{\boldsymbol{\beta}} + \mathbf{b}'\mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (2.8)$$

で与えられることになる。

ここで，連立方程式 (2.6) の解が (2.7) で与えられることを確かめよう。まず，(2.6) の 2 番目の方程式 $Z'R^{-1}X\hat{\beta} + (Z'R^{-1}Z + G^{-1})\hat{v} = Z'R^{-1}y$ より

$$\hat{v} = (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - X\beta) \quad (2.9)$$

と書ける。ここで， $(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}$ を変形すると，

$$\begin{aligned} & (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \\ &= GZ'R^{-1} - G \{ (Z'R^{-1}Z + G^{-1}) - G^{-1} \} (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \\ &= GZ'R^{-1} - GZ'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \\ &= GZ' \{ R^{-1} - R^{-1}Z(G^{-1} + Z'R^{-1}Z)^{-1}Z'R^{-1} \} \\ &= GZ'\Sigma^{-1} \end{aligned}$$

と書けることがわかる。最後の等式は，逆行列の計算でしばしば用いられる等式

$$\Sigma^{-1} = (ZGZ' + R)^{-1} = R^{-1} - R^{-1}Z(G^{-1} + Z'R^{-1}Z)^{-1}Z'R^{-1} \quad (2.10)$$

から従う。これを (2.9) に代入すると (2.7) の \hat{v} が得られることがわかる。次に，いま求めた \hat{v} を (2.6) の 1 番目の方程式 $X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{v} = X'R^{-1}y$ に代入して整理すると，

$$X'R^{-1}X\hat{\beta} + X'R^{-1}ZGZ'\Sigma^{-1}(y - X\hat{\beta}) = X'R^{-1}y$$

より，

$$X'R^{-1}(\Sigma - ZGZ')\Sigma^{-1}X\hat{\beta} = X'R^{-1}(\Sigma - ZGZ')\Sigma^{-1}y$$

となる。 $\Sigma = ZGZ' + R$ より， $R^{-1}(\Sigma - ZGZ') = I$ となるので，結局， $X'\Sigma^{-1}X\hat{\beta} = X'\Sigma^{-1}y$ となり，(2.7) の $\hat{\beta}$ が得られることが確かめられる。

[2] 混合モデル方程式の導出. 混合モデル方程式 (2.6) の導出に関してはいくつかのアプローチが知られている。代表的なものに最尤法に基づいた方法と経験ベイズ法によるものがある。 y と v の同時密度関数は，基準化定数を除くと

$$\begin{aligned} & |G|^{-1/2}|R|^{-1/2} \\ & \times \exp \left\{ -\frac{1}{2} \begin{pmatrix} v \\ y - X\beta - Zv \end{pmatrix}' \begin{pmatrix} G^{-1} & 0 \\ 0 & R^{-1} \end{pmatrix} \begin{pmatrix} v \\ y - X\beta - Zv \end{pmatrix} \right\} \end{aligned}$$

と書ける。exp{·} の中身を (-2) 倍したものを

$$h(\beta, v) = v'G^{-1}v + (\mathbf{y} - X\beta - Zv)'R^{-1}(\mathbf{y} - X\beta - Zv)$$

とおく。これを β と v に関して最小化するために β, v に関して偏微分すると

$$\begin{aligned}\frac{\partial h(\beta, v)}{\partial \beta} &= -2X'R^{-1}(\mathbf{y} - X\beta - Zv), \\ \frac{\partial h(\beta, v)}{\partial v} &= 2G^{-1}v - 2Z'R^{-1}(\mathbf{y} - X\beta - Zv),\end{aligned}$$

となり, $\partial h(\beta, v)/\partial \beta = 0, \partial h(\beta, v)/\partial v = 0$ の連立方程式を行列で表すと, (2.6) が得られる。これが最尤法に基づいた導出方法である。

もう1つの方法は, \mathbf{y} を与えたときの v の条件付き分布に基づいている。 (\mathbf{y}, v) の共分散行列は

$$\text{Cov}(\mathbf{y}, v) = \begin{pmatrix} \Sigma & ZG \\ GZ' & G \end{pmatrix} \quad (2.11)$$

で与えられるので, \mathbf{y} を与えたときの v の条件付き期待値は多変量正規分布の基本的な性質から

$$E[v|\mathbf{y}] = GZ'\Sigma^{-1}(\mathbf{y} - X\beta)$$

となる。この条件付き分布は, ベイズの枠組みでは, \mathbf{y} を与えたときの v の事後分布に相当しており, $v|\mathbf{y} \sim \mathcal{N}_q(GZ'\Sigma^{-1}(\mathbf{y} - X\beta), G - GZ'\Sigma^{-1}ZG)$ で与えられる。また (2.10) を用いて \mathbf{y} の周辺分布を計算すると, $\mathbf{y} \sim \mathcal{N}_N(X\beta, \Sigma)$ となることがわかる。周辺分布の密度は定数項を除いて

$$|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - X\beta)'\Sigma^{-1}(\mathbf{y} - X\beta)\right\} \quad (2.12)$$

と表されるので, 周辺分布に基づいた β の最尤推定量は一般化最小2乗推定量 $\hat{\beta}$ に一致する。また v のベイズ推定量は事後分布の平均で与えられるので $GZ'\Sigma^{-1}(\mathbf{y} - X\beta)$ がベイズ推定量になる。これに $\hat{\beta}$ を代入したものの $GZ'\Sigma^{-1}(\mathbf{y} - X\hat{\beta})$ は経験ベイズ推定量と呼ばれるが, 混合モデル方程式の解 \hat{v} に一致している。従って混合モデル方程式の解は経験ベイズ解として導出されることがわかる。

上の2つの方法の違いは、後者が事後分布の平均で推定するのに対して前者は事後分布のモードで推定している点である。正規分布の場合には両者が一致するので同じ解が得られたことになるが、一般には異なったものになり、前者はベイズ的最尤法と呼ばれる手法である。

観測できない変量を予測するためには (2.11) で与えられる相関関係が本質的であることを上で説明した。観測できなくても相関関係を利用して条件付き期待値で予測可能なわけである。このことは、広く用いられている考え方で、例えば、欠測値がある場合には条件付き期待値を用いることによって補完することができる。EM アルゴリズムや有限母集団の予測問題でも同様な方法が用いられている。

[3] 枝分かれ誤差回帰モデルにおける BLUP. 2.1 節で紹介したモデル (2.2) において、各郡におけるとうもろこしの平均的作付面積 (農作区画単位)

$$\mu_i = \bar{\mathbf{x}}_i' \boldsymbol{\beta} + v_i$$

に対して BLUP を求めてみよう。ここで、 $\bar{\mathbf{x}}_i = \sum_{j=1}^{n_i} \mathbf{x}_{ij} / n_i$ である。この場合、 $\mathbf{G}(\sigma_v^2) = \sigma_v^2 \mathbf{I}_k$, $\boldsymbol{\Sigma}_i(\sigma_e^2, \sigma_v^2) = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 \mathbf{J}_{n_i}$,

$$\boldsymbol{\Sigma}(\sigma_e^2, \sigma_v^2) = \text{block diag}(\boldsymbol{\Sigma}_1(\sigma_e^2, \sigma_v^2), \dots, \boldsymbol{\Sigma}_k(\sigma_e^2, \sigma_v^2))$$

となる。

$$\boldsymbol{\Sigma}_i^{-1} = \frac{1}{\sigma_e^2} \left(\mathbf{I}_{n_i} - \frac{\sigma_v^2}{\sigma_e^2 + n_i \sigma_v^2} \mathbf{J}_{n_i} \right)$$

に注意し、 $\theta = \sigma_v^2 / \sigma_e^2$ とおくと、 μ_i の BLUP $\hat{\mu}_i(\theta)$ は、(2.8) から

$$\hat{\mu}_i(\theta) = \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\theta) + \frac{\theta n_i}{1 + \theta n_i} \left\{ \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\theta) \right\} \quad (2.13)$$

となる。ただし、 $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}$ であり、 $\boldsymbol{\beta}$ の GLS は次で与えられる。

$$\hat{\boldsymbol{\beta}}(\theta) = \left\{ \sum_{i=1}^k \left(\mathbf{x}_i \mathbf{x}_i' - \frac{n_i^2 \theta}{1 + n_i \theta} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i' \right) \right\}^{-1} \sum_{i=1}^k \left(\mathbf{x}_i \mathbf{y}_i' - \frac{n_i \theta}{1 + n_i \theta} \bar{\mathbf{x}}_i \bar{y}_i \right)$$

2.3 変量効果と共通母数の役割

線形混合モデル (2.4) の特徴は、

$$(\text{データ}) = (\text{共通母数}) + (\text{変量効果}) + (\text{誤差項})$$

なる構造をしている点であり（共通母数）と（変量効果）がデータ解析においてどのように働くのか調べてみたい。モデル (2.2) において平均的作付面積 $\mu_i = \bar{x}'_i\beta + v_i$ の予測問題を取り上げてみると、各郡の標本平均 \bar{y}_i を用いて予測するのが基本であるが、 n_i が 1~5 程度であるため予測誤差が大きいという問題がある。これに対して、(2.13) で与えられる BLUP $\hat{\mu}_i(\theta)$ は \bar{y}_i と $\bar{x}'_i\hat{\beta}(\theta)$ との加重平均になっていることがわかる。 \bar{y}_i は直接取られたデータに基づいた平均値であるので、個々の郡の特徴を反映している。しかし n_i が小さいときには、 \bar{y}_i の予測誤差が問題となる。他方、 $\bar{x}'_i\hat{\beta}(\theta)$ は全データに基づいて構成されているので安定しているが、郡の特徴は \bar{y}_i ほど強くは現れないと考えられる。BLUP は、これらの点を考慮した方法であり、 n_i もしくは θ が小さければ \bar{y}_i を $\bar{x}'_i\hat{\beta}(\theta)$ の方向へ縮小することによって安定化を図っている。すなわち、 n_i が小さければ、データの不足を周辺もしくは全体のデータで補うことによって予測精度を高めていると解釈される。このようにして BLUP が小地域の予測問題に役立つことがわかる。言い換えれば、BLUP を生み出すところのモデルの形に、小地域の予測を効果的に行う仕組みが備わっていることになる。

[1] 変量効果と縮小推定. v_i を母数効果とし $\beta = 0$ としたときには、 μ_i の最適な推定量は \bar{y}_i となる。これに対して、 v_i を変量とすると、 (\bar{y}_i, v_i) の共分散行列が

$$\text{Cov}(\bar{y}_i, v_i) = \begin{pmatrix} \sigma_v^2 + \sigma_e^2/n_i & \sigma_v^2 \\ \sigma_v^2 & \sigma_v^2 \end{pmatrix}$$

となることからわかるように、 \bar{y}_i と v_i の間に相関が生ずる。このことから条件付き期待値が $E[v_i|\bar{y}_i] = \theta n_i(1 + \theta n_i)^{-1}(\bar{y}_i - \bar{x}'_i\beta)$ となり、 \bar{y}_i が $\bar{x}'_i\beta$ の方向へ縮小される。こうして、線形混合モデルにおいて変量効果が \bar{y}_i を縮小する作用を生むことがわかる。

[2] 共通母数によるデータのプーリング. モデル (2.2) からわかるように \bar{y}_i の期待値は $E[\bar{y}_i] = \bar{x}'_i\beta$ であり、これは i に依存しない共通な母数 β に基づいている。 β は全データ $\bar{y}_1, \dots, \bar{y}_k$ の加重平均 $\hat{\beta}(\theta)$ により推定される。すなわち、母数を共通にとることによってデータをプーリングする作用が働き、結果として安定した推定が可能になる。

以上述べてきたように、主要項の母数に等号制約や順序制約などの母数制約をいれることによってデータのプーリングがなされて安定した推定値が得られ、また変量効果を組み入れることによって \bar{y}_i をその安定化された推定値の方向へ縮小することができ、その結果、推定精度を高めることが

できる。この考え方は経験ベイズ法の枠組みで Efron and Morris (1975) の論文の中で示されたものであり、ベイズ的アプローチの現実的な有用性はこのような考え方に基づいている。

2.4 分散成分の推定と経験最良線形不偏予測量 (EBLUP)

[1] 最尤法 (ML) と制限最尤法 (REML) 線形混合モデル (2.4) において、一般に共分散行列 G , R は分散成分等の未知母数に依存している。それらの母数を $\alpha = (\alpha_1, \dots, \alpha_m)$ とし、 $\Sigma(\alpha) = R(\alpha) + ZG(\alpha)Z'$ と書くことにすると、BLUP (2.8) は $\hat{\mu}(\alpha) = a'\hat{\beta}(\alpha) + b'G(\alpha)Z'\{\Sigma(\alpha)\}^{-1}\{y - X\hat{\beta}(\alpha)\}$ と表される。 α をその推定量 $\hat{\alpha}$ で置き換えたもの $\hat{\mu}(\hat{\alpha})$ を経験最良線形不偏予測量 (EBLUP) という。

α を推定する代表的な方法に最尤法 (Maximum Likelihood, ML) と制限付き最尤法 (Restricted Maximum Likelihood, REML) がある。 y の周辺分布は、(2.12) より $y \sim \mathcal{N}_N(X\beta, \Sigma(\alpha))$ で与えられるが、この尤度に基づいた推定法が ML である。 β を GLS $\hat{\beta}(\alpha)$ で推定すると、 α の ML は、 $\log |\Sigma(\alpha)| + (y - X\hat{\beta}(\alpha))'\Sigma(\alpha)^{-1}(y - X\hat{\beta}(\alpha))$ を最小にする解として与えられる。一方、 X のランクを r とし K を $K'X = 0$ なる $N \times (N - r)$ 行列とすると、 $K'y \sim \mathcal{N}_{N-r}(0, K'\Sigma(\alpha)K)$ となり、この分布に基づいた最尤法が REML である。従って、 $\log |K'\Sigma(\alpha)K| + y'K(K'\Sigma(\alpha)K)^{-1}K'y$ の最小値を達する解が REML になる。

$$P(\alpha) = \Sigma(\alpha)^{-1} - \Sigma(\alpha)^{-1}X \{X'\Sigma(\alpha)^{-1}X\}^{-1}X'\Sigma(\alpha)^{-1}$$

とおくとき、

$$\begin{aligned} y'P(\alpha)y &= (y - X\hat{\beta}(\alpha))'\Sigma(\alpha)^{-1}(y - X\hat{\beta}(\alpha)), \\ P(\alpha) &= K(K'\Sigma(\alpha)K)^{-1}K' \end{aligned}$$

が成り立つことに注意する。また

$$\begin{aligned} (\partial/\partial\alpha_i) \log |\Sigma| &= \text{tr}(\Sigma^{-1}\partial\Sigma/\partial\alpha_i), \\ \partial P/\partial\alpha_i &= -P(\partial\Sigma/\partial\alpha_i)P, \\ (\partial/\partial\alpha_i) \log |K'\Sigma K| &= \text{tr}(P\partial\Sigma/\partial\alpha_i) \end{aligned}$$

が成り立つので，ML と REML は次の方程式の解として得られることがわかる。

$$[\text{ML}] \quad \text{tr} \left(\Sigma(\alpha)^{-1} \frac{\partial \Sigma(\alpha)}{\partial \alpha_i} \right) = \mathbf{y}' \mathbf{P}(\alpha) \frac{\partial \Sigma(\alpha)}{\partial \alpha_i} \mathbf{P}(\alpha) \mathbf{y} \quad (2.14)$$

$$[\text{REML}] \quad \text{tr} \left(\mathbf{P}(\alpha) \frac{\partial \Sigma(\alpha)}{\partial \alpha_i} \right) = \mathbf{y}' \mathbf{P}(\alpha) \frac{\partial \Sigma(\alpha)}{\partial \alpha_i} \mathbf{P}(\alpha) \mathbf{y} \quad (2.15)$$

右辺は

$$\begin{aligned} & \mathbf{y}' \mathbf{P}(\alpha) \{ \partial \Sigma(\alpha) / \partial \alpha_i \} \mathbf{P}(\alpha) \mathbf{y} \\ &= (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\alpha))' \Sigma(\alpha)^{-1} \{ \partial \Sigma(\alpha) / \partial \alpha_i \} \Sigma(\alpha)^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\alpha)) \\ &= - (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\alpha))' \{ \partial \Sigma(\alpha)^{-1} / \partial \alpha_i \} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\alpha)) \end{aligned}$$

と変形できるので，いずれか計算しやすいものを利用すればよい。ML と REML とではどちらがよいかについて McCulloch and Searle (2001) の 6.10 節で議論されているが，REML は分散成分の推定が不偏に近くなるように自由度が調整されている点で優れていると思われる。ML と REML の具体的な例が 3.1 節で取り上げられているので両者の違いを見ることができる。

[2] 枝分かれ誤差回帰モデルにおける推定。一般には ML, REML を明示的に導くことができないため (2.14), (2.15) の方程式を数値的に解く必要がある。枝分かれ誤差回帰モデル (2.2) についても， n_1, \dots, n_k が異なる場合には明示的な解をもたない。そこで，ここでは分散成分の明示的な推定量を与えておこう。まず， σ_e^2 の不偏推定量は，

$$\hat{\sigma}_e^{2UB} = \frac{S_1}{N - k - p + \lambda}, \quad S_1 = \sum_{i=1}^k \sum_{j=1}^{n_i} \left\{ (y_{ij} - \bar{y}_i) - (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \hat{\boldsymbol{\beta}}_1 \right\}^2 \quad (2.16)$$

で与えられる。ただし， $p - \lambda$ は行列 $\sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'$ のランクを表しており，通常は線形モデル (2.2) が定数項をもつときには $\lambda = 1$ ，もたないときには $\lambda = 0$ となる。また $\hat{\boldsymbol{\beta}}_1$ は $\sum_{i=1}^k \sum_{j=1}^{n_i} \{ (y_{ij} - \bar{y}_i) - (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \boldsymbol{\beta} \}^2$ における $\boldsymbol{\beta}$ の最小 2 乗推定量を表している。一方， $\bar{y}_1, \dots, \bar{y}_k$ は $\hat{\sigma}_e^{2UB}$ と独立に分布し， $\bar{y}_i \sim \mathcal{N}(\bar{\mathbf{x}}_i' \boldsymbol{\beta}, \sigma_e^2/n_i + \sigma_v^2)$ ， $i = 1, \dots, k$ に従う。 σ_v^2 を推定するために，ヘンダーソンの方法 III が使われる。 $\hat{\boldsymbol{\beta}}_0 = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$ に対して， $S = (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_0)' (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_0)$ とおき， $N = \sum_{i=1}^k n_i$ ， $N_* = N - \text{tr}(\mathbf{X}' \mathbf{X})^{-1} \sum_{i=1}^k n_i^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'$ とおくと， S の期待値は $E[S] = (N - p)\sigma_e^2 + N_*\sigma_v^2$ と書ける。従って， σ_v^2 の 1 つの不偏推定量は

$$\hat{\sigma}_v^{2UB} = N_*^{-1} \{ S - (N - p)\hat{\sigma}_e^{2UB} \}$$

となる。分散成分 σ_v^2 の不偏推定量 $\hat{\sigma}_v^{2UB}$ の欠点は正の確率で負の値をとってしまうことである。Kubokawa (2000) は σ_e^2, σ_v^2 の不偏推定量に修正を加えた打ち切り推定量を提案し不偏推定量を優越することを示した。ここでは分散成分の比 $\theta = \sigma_v^2/\sigma_e^2$ の打ち切り推定量として,

$$\hat{\theta} = \max\left\{\frac{1}{N_*}\left\{\frac{S}{\hat{\sigma}_e^{2UB}} - (N - p)\right\}, \frac{1}{k^{2/3}}\right\} \quad (2.17)$$

を用いることにする。結局, この $\hat{\theta}$ を (2.13) に代入することにより EBLUP $\hat{\mu}_i(\hat{\theta})$ が得られる。

3 線形混合モデルを利用した小地域推定と誤差評価

2.3 節で説明されたように EBLUP は予測精度を高める手法として用いられ, 特に官庁統計の分野では小地域推定のための実用的な方法として利用可能である。その際, EBLUP がどの程度 予測誤差を改善しているのかを見積もる必要がある。この節では, 小地域推定に焦点をしばって EBLUP の平均 2 乗誤差の推定と EBLUP に基づいた信頼区間の構成を行い, 実際の応用例に当てはめる。

3.1 地域レベルのモデル

官庁から発行される数値には集計データが多いことから, Fay and Herriot (1979) は,

$$\bar{y}_i = \bar{x}_i' \beta + v_i + e_i, \quad i = 1, \dots, k, \quad (3.1)$$

なる形のモデルを扱った。ただし, 誤差項 e_i は $e_i \sim \mathcal{N}(0, \sigma_e^2/n_i)$ に従う。個々のデータに基づいたモデル (2.2) が個体レベルのモデルであるのに対して, これは地域レベルのモデルとなっている。また, 特に Fay-Herriot モデルとも呼ばれる。このモデルのもとでは誤差分散 σ_e^2 は既知として扱われるが, 実際上は何らかの形で σ_e^2 を推定してやる必要がある。しかし解析上扱いやすいので, 小地域推定の論文ではこのモデルを設定することが少なくない。モデル (2.2) との関係については記号上 $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i, \bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i$ に対応しているが, 観測されるデータは個々のデータではなく集計データ \bar{y}_i, \bar{x}_i

である点に注意する。 $\bar{\mathbf{y}} = (\bar{y}_1, \dots, \bar{y}_k)'$, $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_k)'$, $\mathbf{e} = (e_1, \dots, e_k)'$ として行列表現すると,

$$\begin{aligned}\bar{\mathbf{y}} &= \bar{\mathbf{X}}\boldsymbol{\beta} + \mathbf{v} + \mathbf{e}, \\ \mathbf{v} &\sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I}_k), \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})\end{aligned}\quad (3.2)$$

と表されることがわかる。ただし \mathbf{D} は対角行列 $\mathbf{D} = \text{diag}(\sigma_e^2/n_1, \dots, \sigma_e^2/n_k)$ である。これ以降は, このモデルを用いて説明する。モデル (3.1) において σ_v^2 及び分散成分比 $\theta = \sigma_v^2/\sigma_e^2$ は $(\bar{y}_1, \dots, \bar{y}_k)$ に基づいて推定される。 θ の推定量を $\hat{\theta} = \hat{\theta}(\bar{y}_1, \dots, \bar{y}_k)$ で表すことにすると, $\mu_i = \bar{\mathbf{x}}_i'\boldsymbol{\beta} + v_i$ の EBLUP は (2.13) よりこの $\hat{\theta}$ を用いて

$$\hat{\mu}_i(\hat{\theta}) = \bar{y}_i - \hat{\gamma}_i(\bar{y}_i - \bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}(\hat{\theta})), \quad \hat{\gamma}_i = \gamma_i(\hat{\theta}) = (1 + n_i\hat{\theta})^{-1} \quad (3.3)$$

なる形で表現できる。このモデルでは $\boldsymbol{\beta}$ の GLS は次のようになる。

$$\hat{\boldsymbol{\beta}}(\theta) = \left(\sum_{i=1}^k \frac{n_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'}{1 + n_i\theta} \right)^{-1} \sum_{i=1}^k \frac{n_i \bar{\mathbf{x}}_i \bar{y}_i}{1 + n_i\theta} \quad (3.4)$$

θ の推定についてはいくつかの方法が知られている。ML, REML は (2.14), (2.15) より,

$$\begin{aligned}[\text{ML}] \quad \sigma_e^2 \sum_{i=1}^k \frac{n_i}{1 + n_i\theta} &= \sum_{i=1}^k \frac{n_i^2 \{\bar{y}_i - \bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}(\theta)\}^2}{(1 + n_i\theta)^2} \\ [\text{REML}] \quad \sigma_e^2 \sum_{i=1}^k \frac{n_i}{1 + n_i\theta} - \sigma_e^2 \text{tr} \left(\sum_{i=1}^k \frac{n_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'}{1 + n_i\theta} \right)^{-1} \sum_{i=1}^k \frac{n_i^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'}{(1 + n_i\theta)^2} \\ &= \sum_{i=1}^k \frac{n_i^2 \{\bar{y}_i - \bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}(\theta)\}^2}{(1 + n_i\theta)^2}\end{aligned}$$

の非負の解もしくは 0 として与えられる。それらを $\hat{\theta}^{ML}$, $\hat{\theta}^{REML}$ で表す。また Fay and Herriot (1979) はモーメント推定量として次の方程式の解で推定することを提案した。

$$[\text{FH}] \quad \sigma_e^2(k - p) = \sum_{i=1}^k \frac{n_i \{\bar{y}_i - \bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}(\theta)\}^2}{1 + n_i\theta}$$

これは、 $\hat{\theta}^{FH}$ で表される。 $\hat{\beta}_2 = (\sum_{j=1}^k n_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j')^{-1} \sum_{j=1}^k n_j \bar{\mathbf{x}}_j \bar{y}_j$ に対して、 $S_2 = \sum_{i=1}^k n_i (\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\beta}_2)^2$ とおくと、(2.16) の下で与えられた議論と同様にして

$$[\text{TR}] \quad \hat{\theta}^{TR} = \max \left\{ \frac{1}{n_*} \left\{ \frac{S_2}{\sigma_e^2} - (k-p) \right\}, \frac{1}{k^{2/3}} \right\}$$

なる形の明示的な推定量が得られる。ただし、

$$n_* = N - \text{tr} \left(\sum_{i=1}^k n_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i' \right)^{-1} \sum_{i=1}^k n_i^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'$$

である。 $n_1 = \dots = n_k = n$ のときには、 θ の ML は $n^{-1} \max \{ n \sum_{i=1}^k \{ \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\beta}(\theta) \}^2 / (k\sigma_e^2) - 1, 0 \}$ 、REML は $n^{-1} \max \{ n \sum_{i=1}^k \{ \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\beta}(\theta) \}^2 / ((k-p)\sigma_e^2) - 1, 0 \}$ となり、REML は不偏に近いことがわかる。この場合、FH は REML と同じであり、また $\hat{\theta}^{TR}$ も近い推定値を与えている。

3.2 平均 2 乗誤差の推定

さて EBLUP の予測誤差を見積もるために EBLUP の平均 2 乗誤差の推定量を求めよう。EBLUP の $\mu_i = \bar{\mathbf{x}}_i' \beta + v_i$ に対する推定誤差として、

$$M_i(\theta, \hat{\mu}_i(\hat{\theta})) = E \left[\{ \hat{\mu}_i(\hat{\theta}) - \mu_i \}^2 \right] / \sigma_e^2 \quad (3.5)$$

を用いる。これは標準化平均 2 乗誤差と呼ぶべきものであるが、ここでは混乱がない限りこの誤差を平均 2 乗誤差 (Mean Squared Error, MSE) と呼ぶことにする。この MSE を推定することによって EBLUP の誤差を評価することができる。

まず、 \bar{y}_i が $\mathcal{N}(\bar{\mathbf{x}}_i' \beta, \sigma_e^2/n_i + \sigma_v^2)$ に従うことに注意し、縮小推定の理論において知られている Stein の等式を用いると、EBLUP の MSE に対して正確な不偏推定量が得られる。詳しい形は久保川 (2007) で与えられているが、これは多くの項から構成されているのでそれだけ分散が大きくなってしまおうという弊害をもつことになる。実際、Datta, Kubokawa, Rao and Molina (2011) は正確な不偏推定量の推定誤差が大きいことを数値実験を通して検証している。データを当てはめてみると MSE 推定値が負の値をとってしまうこともおこる。そこで、MSE を漸近的に近似した推定量を求めることにしよう。

小地域推定の問題を扱っているので各郡の標本サイズ n_i が小さいため、郡の個数 k が大きい場合を考えて k についての漸近近似を求める。 $\hat{\theta}$ の

バイアスと分散を $Bias_{\theta}(\hat{\theta}) = E[\hat{\theta} - \theta]$, $Var_{\theta}(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2]$ で表す。 $Bias_{\theta}(\hat{\theta}) = O_p(k^{-1})$ となることを仮定する。 $\hat{\theta}$ が ML, REML, FH, TR で与えられるときにはこの条件がみたされる。

$$\begin{aligned} g_{1i}(\theta) &= n_i^{-1} - n_i^{-1}\gamma_i(\theta), \\ g_{2i}(\theta) &= \{\gamma_i(\theta)\}^2 \bar{\mathbf{x}}_i' \left\{ \sum_{j=1}^k \gamma_j n_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j' \right\}^{-1} \bar{\mathbf{x}}_i, \\ g_{3i}(\theta) &= n_i \{\gamma_i(\theta)\}^3 Var_{\theta}(\hat{\theta}), \end{aligned}$$

とおくと, EBLUP の MSE の k に関する 2 次近似は

$$M_i(\theta, \hat{\mu}_i(\hat{\theta})) = g_{1i}(\theta) + g_{2i}(\theta) + g_{3i}(\theta) + O(k^{-3/2})$$

となる。この近似式を用いると MSE の 2 次近似に基づいて MSE の 2 次漸近不偏推定量を構成することができ,

$$\widehat{M}_i^U(\hat{\theta}) = g_{1i}(\hat{\theta}) + g_{2i}(\hat{\theta}) + 2g_{3i}(\hat{\theta}) - Bias_{\hat{\theta}}(\hat{\theta}) \left\{ \gamma_i(\hat{\theta}) \right\}^2 \quad (3.6)$$

で与えられる。実際, $E[\widehat{M}_i^U(\hat{\theta})] = M_i(\theta, \hat{\mu}_i(\hat{\theta})) + O(k^{-3/2})$ が成り立つ。

前節で与えられた θ の推定量 $\hat{\theta}^{ML}$, $\hat{\theta}^{REML}$, $\hat{\theta}^{FH}$, $\hat{\theta}^{TR}$ の分散については, $Var_{\theta}(\hat{\theta}^{ML}) = Var_{\theta}(\hat{\theta}^{REML}) = 2 / \sum_{i=1}^k (n_i \gamma_i)^2 + O(k^{-3/2})$, $Var_{\theta}(\hat{\theta}^{FH}) = 2k / (\sum_{i=1}^k n_i \gamma_i)^2 + O(k^{-3/2})$, $Var_{\theta}(\hat{\theta}^{TR}) = 2 \sum_{i=1}^k \gamma_i^{-2} / N^2 + O(k^{-3/2})$ となる。またバイアスについては, $Bias_{\theta}(\hat{\theta}^{REML}) = Bias_{\theta}(\hat{\theta}^{TR}) = O(k^{-3/2})$,

$$\begin{aligned} Bias_{\theta}(\hat{\theta}^{ML}) &= - \frac{\text{tr}(\sum_i n_i \gamma_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i')^{-1} \sum_i (n_i \gamma_i)^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'}{\sum_i (n_i \gamma_i)^2} + O(k^{-3/2}), \\ Bias_{\theta}(\hat{\theta}^{FH}) &= 2 \frac{k \sum_i (n_i \gamma_i)^2 - (\sum_i n_i \gamma_i)^2}{(\sum_i n_i \gamma_i)^3} + O(k^{-3/2}) \end{aligned}$$

となるので, これらを用いて MSE の 2 次漸近不偏推定値が計算できる。EBLUP の MSE $M_i(\theta, \hat{\mu}_i(\hat{\theta}))$ において $\hat{\theta}$ の影響は, $g_{3i}(\theta)$ の中の $Var_{\theta}(\hat{\theta})$ に現れるので, 分散の小さい推定量 $\hat{\theta}$ を用いれば MSE の 2 次近似が小さくなる。 $Var_{\theta}(\hat{\theta}^{ML}) = Var_{\theta}(\hat{\theta}^{REML}) \leq Var_{\theta}(\hat{\theta}^{FH})$ となることが示されるので, FH よりも ML, REML を用いた方が漸近的には MSE が小さくなるのがわかる。また (3.6) の中の項の数が増えれば推定量 $\widehat{M}_i^U(\hat{\theta})$ のバラツキが大きくなるので, $Bias(\hat{\theta}) = 0$ となる推定量を用いた方がよい。その点から θ の推定量として REML を用いるのがよいと思われる。以上の内容は,

Datta *et al.* (2011), Rao (2003), Datta, Rao and Smith (2005), Kubokawa (2011b) など詳しく説明されている。また、パラメトリック・ブートストラップ法を用いた2次不偏推定も導出されており、詳しくは、Butar and Lahiri (2003), Hall and Maiti (2006a,b), Kubokawa and Nagashima (2011) を参照してほしい。

3.3 信頼区間の構成

EBLUP の誤差を評価するもう一つの方法は EBLUP に基づいた信頼区間を構成することで、その誤差の程度が信頼区間の幅として表されるので実用上有用である。前節と同様、 i 番目の小地域の平均 $\mu_i = \bar{\mathbf{x}}_i' \boldsymbol{\beta} + v_i$ の推定に興味があるとし、 σ_e^2 が既知として話を進める。まず、 μ_i の標本平均に基づいた信頼区間は、 μ_i を与えたときの \bar{y}_i の条件付分布が $\bar{y}_i | \mu_i \sim \mathcal{N}(\mu_i, \sigma_e^2/n_i)$ であることに注意すると、

$$I_i^* : \bar{y}_i \pm z_{\alpha/2} \sqrt{\sigma_e^2/n_i} \quad (3.7)$$

で与えられる。ただし、 $z_{\alpha/2}$ は標準正規分布の上側 $\alpha/2$ 点を表す。これは信頼係数 $1 - \alpha$ の信頼区間となるが、 n_i が小さいときには \bar{y}_i のバラツキが大きくなってしまおうとともに信頼区間の長さが長くなってしまおう。

そこで、線形混合モデル (3.1) の下での信頼区間の構成について考えてみよう。そのモデルをベイズモデルの枠組みで捉えると、 μ_i に $\mu_i \sim \mathcal{N}(0, \sigma_v^2 \mathbf{I}_k)$ なる事前分布を仮定していることになる。 μ_i のベイズ推定量は、 $\gamma_i = (1 + n_i \theta)^{-1}$ に対して $\hat{\mu}_i^B(\boldsymbol{\beta}, \theta) = \bar{\mathbf{x}}_i' \boldsymbol{\beta} + (1 - \gamma_i)(\bar{y}_i - \bar{\mathbf{x}}_i' \boldsymbol{\beta})$ で与えられるので、 \bar{y}_i を所与としたときの μ_i の事後分布は、

$$\mu_i | \bar{y}_i \sim \mathcal{N}(\hat{\mu}_i^B(\boldsymbol{\beta}, \theta), (\sigma_e^2/n_i)(1 - \gamma_i))$$

となり、 μ_i に対する信頼係数 $1 - \alpha$ のベイズ的信頼区間は、

$$I_i^B(\boldsymbol{\beta}, \theta) : \hat{\mu}_i^B(\boldsymbol{\beta}, \theta) \pm z_{\alpha/2} \sqrt{(\sigma_e^2/n_i)(1 - \gamma_i)}$$

と書ける。これは未知母数 $\boldsymbol{\beta}, \theta$ を含んでいるので、それらの推定量を代入した信頼区間が考えられる。一般に $\hat{\theta}$ を $\bar{y}_1, \dots, \bar{y}_k$ に基づいた θ の推定量とし、 $\boldsymbol{\beta}$ を (3.4) で与えられる一般化最小2乗推定量 $\hat{\boldsymbol{\beta}}(\hat{\theta})$ で推定すると、 μ_i の経験ベイズ推定量は

$$\hat{\mu}_i^{EB}(\hat{\theta}) = \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\hat{\theta}) + (1 - \hat{\gamma}_i) \left(\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\hat{\theta}) \right), \quad \hat{\gamma}_i = (1 + n_i \hat{\theta})^{-1}$$

と書けるので，経験ベイズ的信頼区間は $I_i^B(\beta, \theta)$ から

$$I_i^{EB}(\hat{\theta}) : \hat{\mu}_i^{EB}(\hat{\theta}) \pm z_{\alpha/2} \sqrt{(\sigma_e^2/n_i)(1 - \hat{\gamma}_i)}$$

となる。経験ベイズ推定量 $\hat{\mu}_i^{EB}(\hat{\theta})$ は \bar{y}_i に比べて推定精度が高いだけでなく，信頼区間の長さも短くなっている。しかし，その被覆確率が $1 - \alpha$ 以上になることは保証されていない。 $k = 50, p = 3, 1 - \alpha = 0.95$ とし β, n_i, \bar{x}_i 等を適当に与えたときの被覆確率をシミュレーション実験によって求め，横軸に θ の値をとって描いたものが図 1 で示されている。この図からわかるように， $I_i^{EB}(\hat{\theta})$ の被覆確率は 95% をかなり下回ってしまう。

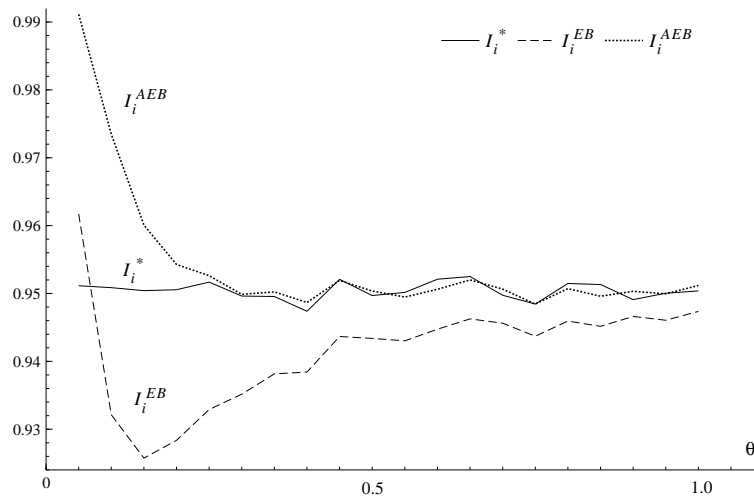


図 1: 標本平均に基づく信頼区間 I_i^* ，経験ベイズ信頼区間 I_i^{EB} ，補正後の信頼区間 I_i^{AEB} の被覆確率の比較

そこで，3.2 節の議論を用いて，被覆確率が k に関して 2 次漸近的に $1 - \alpha$ で近似できるような信頼区間を構成する。このような信頼区間は， $n_1 = \dots = n_k$ で σ_e^2 が既知のときには Basu, Ghosh and Mukerjee (2003) などによって求められ， n_1, \dots, n_k が等しいことを仮定せず， σ_e^2 が未知という一般的な設定の下では笹瀬-久保川 (2005), Kubokawa (2010) によって得られた。Basu *et al.* (2003) の論法に従って， $z_{\alpha/2}$ の代わりに $z_{\alpha/2} \{1 + (2k)^{-1}h(\hat{\theta})\}$ を用いた信頼区間

$$I_i^{AEB} : \hat{\mu}_i^{EB}(\hat{\theta}) \pm z_{\alpha/2} \left[1 + (2k)^{-1}h(\hat{\theta}) \right] \sqrt{(\sigma_e^2/n_i)(1 - \hat{\gamma}_i)} \quad (3.8)$$

を考える。ここで，補正関数 $h(\hat{\theta})$ は，

$$h(\hat{\theta}) = -\frac{kn_i\hat{\gamma}_i^2}{1-\hat{\gamma}_i}Bias_{\hat{\theta}}(\hat{\theta}) + (1+z_{\alpha/2}^2)\frac{kn_i^2\hat{\gamma}_i^4}{4(1-\hat{\gamma}_i)^2}Var_{\hat{\theta}}(\hat{\theta}) \\ + \frac{kn_i\hat{\gamma}_i^2}{1-\hat{\gamma}_i}\left\{\bar{\mathbf{x}}_i'\left(\sum_{j=1}^k\frac{n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}_j'}{1+n_j\hat{\theta}}\right)^{-1}\bar{\mathbf{x}}_i + 2n_i\hat{\gamma}_iVar_{\hat{\theta}}(\hat{\theta})\right\} \quad (3.9)$$

で与えられる。このとき， $Bias_{\hat{\theta}}(\hat{\theta}) = O_p(k^{-1})$ ， $\partial\hat{\theta}/\partial\bar{y}_i = O_p(k^{-1})$ を仮定すると，

$$P[\mu_i \in I_i^{AEB}] = 1 - \alpha + o(k^{-1}), \quad (k \rightarrow \infty)$$

が成り立つことが示される。(3.1) で与えられた推定量 $\hat{\theta}^{TR}$ に対しては， $Bias_{\hat{\theta}}(\hat{\theta}^{TR}) = o(k^{-1})$ ， $Var_{\hat{\theta}}(\hat{\theta}^{TR}) = 2\sum_{i=1}^k(1+n_i\theta)^2/N^2 + o(k^{-1})$ となるので，これを (3.9) に代入すれば補正項が得られる。特に $n_1 = \dots = n_k = n$ のときには次のようになる。

$$h(\hat{\theta}) = \frac{1+z_{\alpha/2}^2}{2n\hat{\theta}^2} + \frac{1}{n\hat{\theta}}\left\{k\bar{\mathbf{x}}_i'\left(\sum_{j=1}^k\bar{\mathbf{x}}_j\bar{\mathbf{x}}_j'\right)^{-1}\bar{\mathbf{x}}_i + 4\right\}$$

図 1 からわかるように，2 次補正後の経験ベイズ信頼区間 I_i^{AEB} の被覆確率は 95% より大きく， $\theta > 0.25$ に対してはほぼ 95% になっている。図 2 は同じ設定のもとでの信頼区間の長さの平均値を描いたものであり， I_i^{AEB} の長さが I_i^* より短くなっていることがわかる。このことから，補正した経験ベイズ信頼区間 I_i^{AEB} が被覆確率と区間の長さの意味で優れているといえよう。

一般的な LMM における信頼区間の構成については Kubokawa (2011b) を，パラメトリック・ブートストラップ法を用いた信頼区間については Chatterjee, Lahiri and Li (2008), Hall and Maiti (2006b), Kubokawa and Nagashima (2011) を参照してほしい。

3.4 応用例：地価公示価格の小地域推定

さて，具体的な小地域に関するデータを用いて EBLUP の特徴，平均 2 乗誤差推定及び EBLUP に基づいた信頼区間の挙動を調べてみよう。ここで用いるデータは，京浜急行電鉄本線及び久里浜線沿いの宅地物件について 2001 年に公表された $1m^2$ 当たりの地価公示価格である。各駅を 1 つの小地

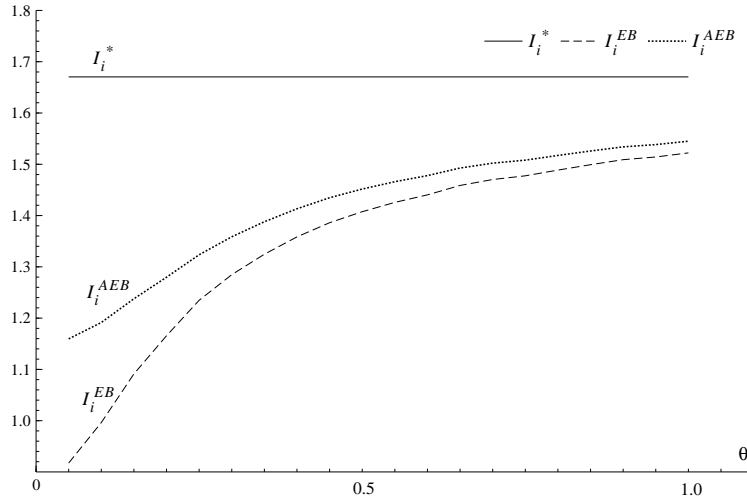


図 2: 標本平均に基づく信頼区間 I_i^* , 経験ベイズ信頼区間 I_i^{EB} , 補正後の信頼区間 I_i^{AEB} の長さの比較

域と考え, また i 番目の駅を最寄り駅とする物件のデータをその小地域からとられたデータと考えて, その個数を n_i で表す。小地域の総数は $k = 48$ であり, n_i は 1 から 12 まで不均一な値をとっているが平均 3.73 程度である。各地価公示価格を対数変換したものを y_{ij} とし, (2.2) に対応するモデル

$$y_{ij} = \beta_0 + x_{1i}\beta_1 + x_{2ij}\beta_2 + x_{3ij}\beta_3 + v_i + e_{ij}$$

を想定してみる。ここで, 共変量 x_{1i} は i 番目の駅から品川駅に到着するのに要する時間, x_{2ij} は物件 (i, j) から最寄り駅 (i) までの時間距離, x_{3ij} は物件 (i, j) の容積率を表している。 $\mathbf{x}_{ij} = (1, x_{1i}, x_{2ij}, x_{3ij})'$, $\bar{\mathbf{x}}_i = (1, x_{1i}, \bar{x}_{2i}, \bar{x}_{3i})'$ とおくと, 行列 $\sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'$ のランクは 2 となるので, (2.16) の $\hat{\sigma}_e^2$ における λ は $\lambda = 2$ となることに注意する。このモデルにおいて, 駅 (小地域) ごとに平均的な地価公示価格

$$\mu_i = \beta_0 + x_{1i}\beta_1 + \bar{x}_{2i}\beta_2 + \bar{x}_{3i}\beta_3 + v_i$$

の予測に関して, これまで説明してきた手法の挙動を調べてみよう。

$\hat{\theta}$ として 3.1 節の打ち切り推定量 $\hat{\theta}^{TR}$ を用いて, データから推定値 $\hat{\theta}^{TR}$, $\hat{\beta}(\hat{\theta}^{TR})$, $\hat{\sigma}_e^2$ を計算すると, $\hat{\theta}^{TR} = 0.551775$, $\hat{\beta}(\hat{\theta}^{TR}) = (12.927, -0.014251, 9.38 \times 10^{-6}, 0.001444)$, $\hat{\sigma}_e^2 = 0.020936$ となる。 β_1 の推定値が負の値であることから, 品川駅から遠くなるにつれて地価公示価格は低くなる傾向にあり, 合

表 1: 京浜急行線沿線の物件の $1m^2$ 当たりの駅ごとの平均価格の予測値と予測誤差
($EBLUP_i^*$ は時系列データを組み込んだ予測量 (4.6) による値)

No.	最寄り駅	n_i	\hat{v}_i^*	\bar{y}_i	予測値		予測誤差		時系列予測値
					$EBLUP_i$	$\hat{\beta}(\hat{\theta}^{TR})$	$1/n_i$	\widehat{M}_i^U	$EBLUP_i^*$
1	北品川	1	1.73	607000	470301	408538	1.000	0.428	465078
2	新馬場	3	0.29	536217	528604	516239	0.333	0.229	565306
3	青物横丁	1	-0.02	484000	485410	486190	1.000	0.402	535321
4	鮫洲	1	1.38	569000	463645	414112	1.000	0.405	470527
5	立会川	2	1.08	508199	469206	429639	0.500	0.291	482257
6	大森海岸	2	0.77	525882	496574	466124	0.500	0.293	528852
7	平和島	2	0.16	470500	464950	458901	0.500	0.289	502266
13	京急川崎	4	0.13	604995	601955	595299	0.250	0.197	650621
14	鶴見市場	7	-1.28	319928	328736	365095	0.143	0.119	353112
18	京急新子安	2	-1.41	321134	356534	400146	0.500	0.286	407568
19	子安	1	1.70	525000	408045	355071	1.000	0.391	417252
27	弘明寺	6	0.02	243983	243815	243260	0.167	0.136	251979
32	能見台	12	1.42	244788	240536	214183	0.083	0.075	239566
33	金沢文庫	10	-0.29	236221	237245	242974	0.100	0.088	244637
34	金沢八景	2	-1.05	210000	226995	247350	0.500	0.282	253148
35	追浜	7	-1.92	189859	197708	231197	0.143	0.119	214788
39	汐入	2	-1.65	174379	196973	225318	0.500	0.284	226449
41	県立大学	6	-0.03	208107	208260	208766	0.167	0.137	219065
45	京急久里浜	6	2.31	240698	227358	188249	0.167	0.137	224582
46	Y R P野比	6	2.80	200674	187314	149112	0.167	0.138	181381

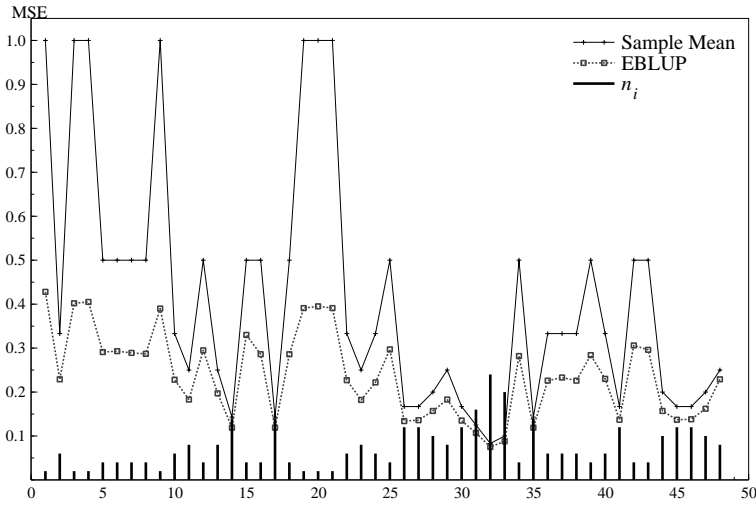


図 3: 標本平均 \bar{y}_i の MSE と EBLUP $_i$ の MSE 推定値 \widehat{M}_i^U の比較 (No.1 から No.48 までの各駅を横軸に並べている。)

理的な符号を示している。表 1 は、京浜急行沿線の物件 $1m^2$ 当たりの駅ごとの平均価格の予測値と予測誤差を与えている。No.1 北品川から No.48 津久井浜までの 48 の駅のうち 20 の駅だけを表 1 に載せている。 n_i は利用可能なデータ数、 \bar{y}_i は標本平均値、EBLUP $_i$ は (3.3) から計算される予測値、 $\hat{\beta}(\hat{\theta}^{TR})$ はプールされた推定値に基づいた回帰推定値を示している。これらの推定値を指数変換した値が、表 1 の「予測値」の欄で与えられている。この表から、EBLUP $_i$ は \bar{y}_i を $\hat{\beta}(\hat{\theta}^{TR})$ の方向へ縮小しており、特に n_i が小さいときに縮小の度合いが大きくなり n_i が大きくなるにつれて縮小の程度は小さくなるのがわかる。表 1 の「予測誤差」の欄は、(3.5) で与えられた平均 2 乗誤差 (MSE) の推定値が与えられている。 $1/n_i$ は、 v_i を母数としたときの \bar{y}_i の MSE であり、 \widehat{M}_i^U は EBLUP $_i$ の MSE の 2 次漸近不偏推定値を表しており、(3.6) から計算されたものである。図 3 は \bar{y}_i の MSE と \widehat{M}_i^U と n_i の値を、駅を No.1 から No.48 まで横軸にとって描いた図である。これらの数値的な結果をながめてみると、EBLUP $_i$ の予測誤差は \bar{y}_i よりもかなり小さく、特に n_i が小さいときには著しい改善がなされていることがわかる。 n_i が大きいところでは両者の差は小さいようである。表 1 では与えられていないが、MSE の正確な不偏推定量の値を計算したところ負の値になってしまった。全体的にかなり縮小がなされているため MSE の正確な不偏推定値が 0 を越えて負の値を取ってしまったのかもしれないが、

正確な不偏推定値が好ましくないことを意味している。

表 1 中の \hat{v}_i^* は変量効果の予測値を標準化した値を示している。変量である地域効果の分布が駅毎に与えられており、共変量の影響を取り除いた後の地域的な差をみるのに役に立つ。例えば、能見台、京急久里浜、Y R P 野比はそれぞれ 1.42, 2.31, 2.80 と高く、追浜、汐入は -1.92, -1.65 と低いが、これらの要因を調べてみるのも興味深いであろう。 \hat{v}_i^* の分布のヒストグラムを描いてみると正規分布を仮定したことは悪くなかったことが確かめられる。

図 4 は、2 次補正した経験ベイズ的信頼区間 I_i^{AEB} と \bar{y}_i に基づいた信頼区間 I_i^* の両端の値を、駅を No.1 から No.48 まで横軸にとって描いた図である。 I_i^{AEB} の値は (3.8) から計算できる。 I_i^* の信頼区間の動きが n_i が小さいときに不安定になるのに比べ、 I_i^{AEB} は安定した信頼区間を与えている。 I_i^{AEB} の動きをながめてみると、全体として東京から遠くなるにつれて価格が減少するという合理的な傾向がみられる。図 5 は、信頼区間の長さ とデータ数 n_i との関係を示したものである。 I_i^* の長さは、 n_i が大きいときには I_i^{AEB} と同程度であるものの、 n_i が小さいときには I_i^{AEB} に比べてかなり大きくなってしまふことがわかる。

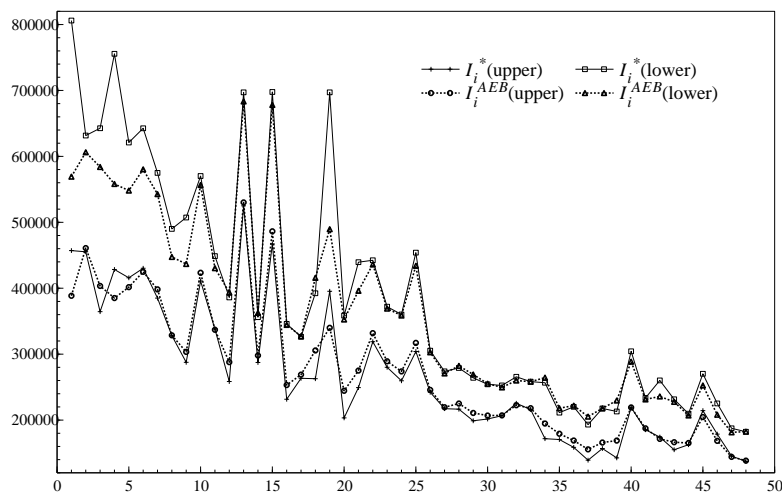


図 4: 補正後の信頼区間 I_i^{AEB} と 標本平均に基づく信頼区間 I_i^* の両端の値の比較 (No.1 から No.48 までの各駅を横軸に並べている。)

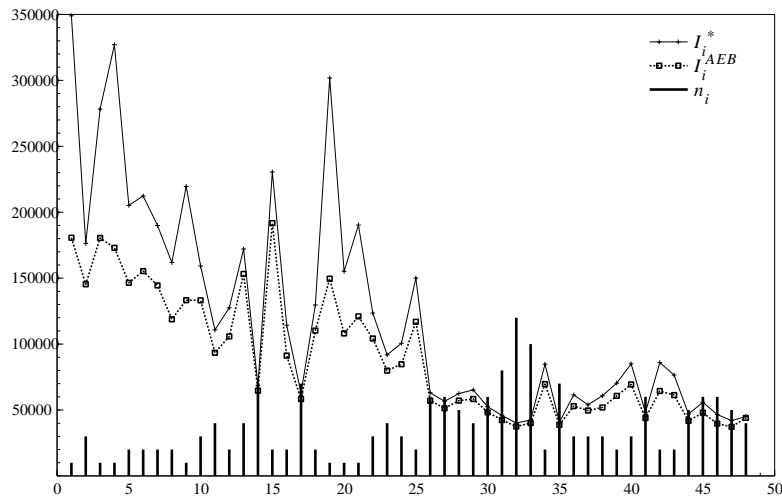


図 5: 補正後の信頼区間 I_i^{AEB} と 標本平均に基づく信頼区間 I_i^* の長さの比較とデータ数 n_i との関係 (n_i のスケールは縦軸 1000 が 1 個のデータを表している。No.1 から No.48 までの各駅を横軸に並べている。)

4 線形混合モデルの様々な応用

線形混合モデル (LMM) の理論と小地域推定への応用について紹介してきたが、線形混合モデルは長い歴史と幅広い応用分野があり、様々な拡張や変形、推測手法及び計算方法などが活発に研究されてきた。2.3 節で説明したように、LMM が小地域推定を行う上で優れた予測量を導くことのできる要因は、共通母数と変量効果を組み込んでいる点である。したがって、共通母数と変量効果を巧みに組み入れることによって、変量係数モデル、分散変動モデルなど応用例に即した様々なモデルを構築することができる。ここでは、経時測定データのモデル、一般化線形混合モデル、経験ベイズと階層ベイズモデルについて説明する。なお、久保川 (2007) でも詳しく解説されているので参照してほしい。

4.1 経時測定データのモデル

線形混合モデルの最近の中心的话题は、空間的かつ時系列的に取られたデータを扱うモデルであり、Laird and Ware (1982), Tsimikas and Ledolter (1997), Das, Jiang and Rao (2004) などの論文の他、Diggle, Liang and Zeger

(1994), Verbeke and Molenberghs (2000), McCulloch and Searle (2001), Demidenko (2004), Fitzmaurice, Laird and Ware (2004), Molenberghs and Verbeke (2006) などの解説書でも章を割いて扱っている。またこのモデルは、計量経済学の分野ではパネル計量モデルと呼ばれ、経済データを解析するために用いられている。パネルデータの解析については、Hsiao (2003) で詳しく解説されている。

ここでは、地域レベルのモデル (3.1) に関連して地域レベルの集計データが T 時点で時系列的にとられている場合を考える。これは、経時測定データ (repeated measures data, longitudinal data) という。本来ならば標本サイズはデータがとられた時点と小地域により異なるのであるが、簡単のために同じ小地域ならば各時点で同じサイズの標本が得られているという設定のもとで考える。いま時点 $t = 1, \dots, T$ において $\bar{y}_{i1}, \dots, \bar{y}_{iT}$ なるデータが観測され、そのときの共変量を $\bar{x}_{i1}, \dots, \bar{x}_{iT}$ とする。 $\bar{\mathbf{y}}_i = (\bar{y}_{i1}, \dots, \bar{y}_{iT})'$, $\bar{\mathbf{X}}_i = (\bar{x}_{i1}, \dots, \bar{x}_{iT})$ とおき、 \mathbf{y}_i は

$$\bar{\mathbf{y}}_i = \bar{\mathbf{X}}_i' \boldsymbol{\beta} + \mathbf{j}_T v_i + \mathbf{e}_i \quad (4.1)$$

に従うとする。ただし、 \mathbf{e}_i と v_i は独立に分布し、 $\mathbf{e}_i \sim \mathcal{N}_T(\mathbf{0}, (\sigma_e^2/n_i)\mathbf{Q})$, $v_i \sim \mathcal{N}(0, \sigma_v^2)$ に従うとする。 $\mathbf{e}_i = (e_{i1}, \dots, e_{iT})'$ として成分で表すと

$$\bar{y}_{it} = \bar{\mathbf{x}}_{it}' \boldsymbol{\beta} + v_i + e_{it}, \quad i = 1, \dots, k, \quad t = 1, \dots, T,$$

となり、多くの文献で扱われてきた。また e_{is} と e_{it} , $s \neq t$, の間には相関関係が入っており、その関係は共分散行列 \mathbf{Q} の中に時間的な相関構造として埋め込まれることになる。通常は一様共分散構造もしくは自己共分散構造 AR(1) を仮定する。例えば、 $T = 4$ の場合、それらは $|\rho| < 1$ に対して

$$\mathbf{Q}_1 = \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix} = (1 - \rho)\mathbf{I}_4 + \rho\mathbf{J}_4,$$

$$\mathbf{Q}_2 = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} = (1 - \rho^2)^{-1} (\rho^{|i-j|})$$

と表される。いまこのようなモデルを k 個の地域について考えることにし、

$$\bar{y} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_k \end{pmatrix}, \bar{X} = \begin{pmatrix} \bar{X}'_1 \\ \vdots \\ \bar{X}'_k \end{pmatrix}, v = \begin{pmatrix} v_1 \\ \vdots \\ v_k \end{pmatrix}, e = \begin{pmatrix} e_1 \\ \vdots \\ e_k \end{pmatrix},$$

$Z = \text{block diag}(j_T, \dots, j_T)$ とおくと、

$$\bar{y} = \bar{X}\beta + Zv + e \quad (4.2)$$

と表される。 $p \times q$ 行列 $A = (a_{ij})$, $r \times s$ 行列 B のクロネッカー積は $A \otimes B = (a_{ij}B)$ で定義されるが、これを用いると Z は $Z = I_k \otimes j_T$ と書ける。 v の共分散行列を $\text{Cov}(v) = G$ とし、 $D = \text{diag}(\sigma_e^2/n_1, \dots, \sigma_e^2/n_k)$ とおくと、 \bar{y} の共分散行列は、

$$\Sigma = ZGZ' + D \otimes Q$$

となる。さらに $ZGZ' = (I_k \otimes j_T)(G \otimes 1)(I_k \otimes j_T)' = G \otimes J_T$ に注意すると、結局

$$\Sigma = G \otimes J_T + D \otimes Q$$

と書ける。 $G = \text{Cov}(v)$ には空間的相関構造を入れることができ、例えば一様共分散構造

$$G = \sigma_v^2 \{(1 - \rho_v)I_k + \rho_v J_k\}$$

を仮定すると、他の地域の情報を有効に利用する予測が作れる。空間的な広がり方は 2 次元的なので、局所的にこの相関構造を用いて隣接する地域の情報を利用する手法を構成することができるが、広い範囲にわたってこの相関構造をおくことは好ましくない。 G として空間的な自己回帰モデルを仮定する方が自然である。

モデル (2.2), (3.1) では空間的に無相関 $\text{Cov}(v) = \sigma_v^2 I_k$ を仮定している。その代わり β を共通にとることによってデータの空間的プーリングがなされている。以降は v が無相関として議論する。このとき、 $\Sigma = \text{Cov}(\bar{y})$ は

$$\Sigma = \sigma_v^2 I_k \otimes J_T + D \otimes Q = \text{diag}(\Sigma_1, \dots, \Sigma_k),$$

$$\Sigma_i = \sigma_v^2 J_T + (\sigma_e^2/n_i)Q, \quad i = 1, \dots, k$$

と書ける。 $\Sigma^{-1} = \text{diag}(\Sigma_1^{-1}, \dots, \Sigma_k^{-1})$ であり、 $\theta = \sigma_v^2/\sigma_e^2$ に対して

$$\Sigma_i^{-1} = \frac{n_i}{\sigma_e^2} \left\{ Q^{-1} - \frac{n_i \theta Q^{-1} j_T j_T' Q^{-1}}{1 + n_i \theta j_T' Q^{-1} j_T} \right\}$$

となることから, $\hat{v} = (\hat{v}_1, \dots, \hat{v}_k)' = \mathbf{GZ}'\Sigma^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})$ は

$$\begin{aligned}\hat{v}_i &= \sigma_v^2 \mathbf{j}'_T \Sigma_i^{-1} (\bar{\mathbf{y}}_i - \bar{\mathbf{X}}'_i \hat{\beta}) \\ &= \frac{n_i \theta}{1 + n_i \theta \mathbf{j}'_T \mathbf{Q}^{-1} \mathbf{j}_T} \mathbf{j}'_T \mathbf{Q}^{-1} (\bar{\mathbf{y}}_i - \bar{\mathbf{X}}'_i \hat{\beta})\end{aligned}$$

となる。また, β の GLS は

$$\hat{\beta} = \left(\sum_{i=1}^k \bar{\mathbf{X}}_i \Sigma_i^{-1} \bar{\mathbf{X}}'_i \right)^{-1} \sum_{i=1}^k \bar{\mathbf{X}}_i \Sigma_i^{-1} \bar{\mathbf{y}}_i$$

となる。従って, 小地域の T 期間における平均値 $\mu_i = \mathbf{j}'_T \bar{\mathbf{X}}'_i \beta / T + v_i$ を推定したいときには, その推定量は $\hat{\mu}_i = \mathbf{j}'_T \bar{\mathbf{X}}'_i \hat{\beta} / T + \hat{v}_i$ で与えられる。

\mathbf{Q} に一様共分散構造 $\mathbf{Q}_1 = (1 - \rho)\mathbf{I}_T + \rho\mathbf{J}_T$ を仮定すると,

$$\mathbf{Q}_1^{-1} = \frac{1}{1 - \rho} \left\{ \mathbf{I}_T - \frac{\rho}{1 + (T - 1)\rho} \mathbf{J}_T \right\}$$

より, $\mathbf{j}'_T \mathbf{Q}_1^{-1} = \{1 + (T - 1)\rho\}^{-1} \mathbf{j}'_T$ となることに注意して,

$$\hat{v}_i = \frac{n_i \theta}{1 + (T - 1)\rho + n_i \theta T} \sum_{t=1}^T (\bar{y}_{it} - \bar{\mathbf{x}}'_{it} \hat{\beta})$$

と書ける。 \mathbf{Q} に自己共分散構造 $\mathbf{Q}_2 = (1 - \rho^2)^{-1} (\rho^{|i-j|})$ を仮定すると, 例えば $T = 4$ の場合

$$\mathbf{Q}_2^{-1} = \begin{pmatrix} 1 & -\rho & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho \\ 0 & 0 & -\rho & 1 \end{pmatrix}$$

と書けることから, $\mathbf{j}'_T \mathbf{Q}_2^{-1} = (1 - \rho)(1, 1 - \rho, \dots, 1 - \rho, 1) = (1 - \rho)^2 \{ \mathbf{j}'_T + \rho(1 - \rho)^{-1}(1, 0, \dots, 0, 1) \}$, $\mathbf{j}'_T \mathbf{Q}_2^{-1} \mathbf{j}_T = (1 - \rho)^2 \{ T + 2\rho / (1 - \rho) \}$ となることに注意する。従って,

$$\begin{aligned}\hat{v}_i &= \frac{n_i \theta}{(1 - \rho)^{-2} + n_i \theta \{ T + 2\rho / (1 - \rho) \}} \\ &\quad \times \left\{ \sum_{t=1}^T (\bar{y}_{it} - \bar{\mathbf{x}}'_{it} \hat{\beta}) + \frac{\rho}{1 - \rho} (\bar{y}_{i1} - \bar{\mathbf{x}}'_{i1} \hat{\beta} + \bar{y}_{iT} - \bar{\mathbf{x}}'_{iT} \hat{\beta}) \right\}\end{aligned}$$

となる。いずれの場合も, n_i を大きくし $\rho \rightarrow 0$ とすると全体の平均 $\sum_{t=1}^T (\bar{y}_{it} - \bar{\mathbf{x}}'_{it} \hat{\beta}) / T$ に近づくことになる。

4.2 モデルの修正と小地域推定への応用

以上の設定では, 変量効果 v_i を全期間 $t = 1, \dots, T$ にわたって同一としているため v_i の予測量は T 個の時点にわたって合算した量 $\sum_{t=1}^T (\bar{y}_{it} - \bar{x}'_{it}\beta)$ に基づいている。しかし, 1年ごとにデータが取られていて最新の予測値を知りたいときには, 上述の予測量は明らかに好ましくない。第 T 期のデータにウェイトをかけた予測量を導くためにはモデル (4.1) の中の変量効果の部分 $j_T v_i$ を地域と時点に依存させて

$$\bar{y}_i = \bar{X}'_i \beta + v_i + e_i \quad (4.3)$$

と変えてみるとよい。ここで $v_i = (v_{i1}, \dots, v_{iT})'$ は $\mathcal{N}_T(0, \sigma_v^2 \mathbf{I}_T)$ に従う変量とする。すなわち, 変量効果が時点に関して独立に分布すると仮定する。このとき, \bar{y}_i の共分散行列は

$$\Sigma_i = \sigma_v^2 \mathbf{I}_T + (\sigma_e^2/n_i) \mathbf{Q}, \quad i = 1, \dots, k$$

となる。第 T 期の平均値 $\mu_{iT} = \bar{x}'_{iT}\beta + v_{iT}$ を予測したいときには, 条件付き期待値が $E[v_{iT}|\bar{y}_i] = \sigma_v^2(0, \dots, 0, 1)\Sigma_i^{-1}(\bar{y}_i - \bar{x}'_i\beta)$ となることより, 予測量は

$$\hat{\mu}_{iT} = \bar{x}'_{iT}\hat{\beta} + \sigma_v^2(0, \dots, 0, 1)\Sigma_i^{-1}(\bar{y}_i - \bar{X}'_i\hat{\beta})$$

で与えられる。 $\Sigma_i = (\sigma_e^2/n_i)(n_i\theta\mathbf{I}_T + \mathbf{Q})$ と書けるので,

$$n_i\theta\mathbf{I}_T + \mathbf{Q} = \mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & a_{22} \end{pmatrix}, \quad \mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}^{11} & \mathbf{a}^{12} \\ \mathbf{a}^{21} & a^{22} \end{pmatrix}$$

と表すことにする。ただし, a_{22} はスカラーである。 $a_{22.1} = a_{22} - \mathbf{a}_{21}\mathbf{A}_{11}^{-1}\mathbf{a}_{12}$ とおくと $\mathbf{a}^{21} = -\mathbf{a}_{21}\mathbf{A}_{11}^{-1}/a_{22.1}$, $a^{22} = 1/a_{22.1}$ と書けるので, $\hat{\mu}_{iT}$ は

$$\hat{\mu}_{iT} = \bar{x}'_{iT}\hat{\beta} + \frac{n_i\theta}{a_{22.1}} \left\{ (\bar{y}_{iT} - \bar{x}'_{iT}\hat{\beta}) - (\mathbf{a}_{21}\mathbf{A}_{11}^{-1}, 0)(\bar{y}_i - \bar{X}'_i\hat{\beta}) \right\} \quad (4.4)$$

となる。これは, $\bar{y}_{iT} - \bar{x}'_{iT}\hat{\beta}$ を過去の $T-1$ 個のデータに基づいた予測量の方向へ縮小した形をしていることがわかる。

\mathbf{Q} に一様共分散構造 $\mathbf{Q}_1 = (1-\rho)\mathbf{I}_T + \rho\mathbf{J}_T$ を仮定すると, $a_{22.1} = (n_i\theta + 1 - \rho)(n_i\theta + 1 + (T-1)\rho)/(n_i\theta + 1 + (T-2)\rho)$, $\mathbf{a}_{21}\mathbf{A}_{11}^{-1} = \{\rho/(n_i\theta + 1 +$

$(T-2)\rho\}j'_{T-1}$ となることから,

$$\begin{aligned}\hat{\mu}_{iT} = & \bar{\mathbf{x}}'_{iT}\hat{\boldsymbol{\beta}} + \frac{n_i\theta(n_i\theta + 1 + (T-2)\rho)}{(n_i\theta + 1 - \rho)(n_i\theta + 1 + (T-1)\rho)} \left\{ (\bar{y}_{iT} - \bar{\mathbf{x}}'_{iT}\hat{\boldsymbol{\beta}}) \right. \\ & \left. - \frac{\rho}{n_i\theta + 1 + (T-2)\rho} \sum_{t=1}^{T-1} (\bar{y}_{it} - \bar{\mathbf{x}}'_{it}\hat{\boldsymbol{\beta}}) \right\}\end{aligned}\quad (4.5)$$

と表される。 $n_i \rightarrow \infty$ とすると $\hat{\mu}_{iT} \rightarrow \bar{y}_{iT}$, $\rho \rightarrow 0$ とすると $\hat{\mu}_{iT} \rightarrow \bar{\mathbf{x}}'_{iT}\hat{\boldsymbol{\beta}} + \{n_i\theta/(1+n_i\theta)\}(\bar{y}_{iT} - \bar{\mathbf{x}}'_{iT}\hat{\boldsymbol{\beta}})$ に近づくことがわかる。一方, Q に自己共分散構造 $Q_2 = (1-\rho^2)^{-1}(\rho^{|i-j|})$ を仮定したときには一般の次元に対しては予測量 (4.4) を明示的に書き下すことはできないが, 例えば $T=3$ の場合には,

$$\begin{aligned}\hat{\mu}_{iT} = & \bar{\mathbf{x}}'_{iT}\hat{\boldsymbol{\beta}} + \frac{n_i\theta}{n_i\theta + 1} \frac{(n_i\theta + 1)^2 - \rho^2(n_i\theta)^2}{(n_i\theta + 1)^2 - \rho^2 n_i\theta(n_i\theta - 1)} \left\{ (\bar{y}_{iT} - \bar{\mathbf{x}}'_{iT}\hat{\boldsymbol{\beta}}) \right. \\ & - \frac{\rho}{(n_i\theta + 1)^2 - \rho^2(n_i\theta)^2} [n_i\theta\rho(\bar{y}_{i,T-2} - \bar{\mathbf{x}}'_{i,T-2}\hat{\boldsymbol{\beta}}) \\ & \left. + (n_i\theta + 1)(\bar{y}_{i,T-1} - \bar{\mathbf{x}}'_{i,T-1}\hat{\boldsymbol{\beta}})] \right\}\end{aligned}\quad (4.6)$$

と表される。(4.5) が過去のデータを同じ重みでプーリングしているのに対して, $n_i\theta\rho < n_i\theta + 1$ となることからわかるように, (4.6) はより最近のデータをより重くして加重和をとっている。この場合も, $n_i \rightarrow \infty$ とすると $\hat{\mu}_{iT} \rightarrow \bar{y}_{iT}$, $\rho \rightarrow 0$ とすると $\hat{\mu}_{iT} \rightarrow \bar{\mathbf{x}}'_{iT}\hat{\boldsymbol{\beta}} + \{n_i\theta/(1+n_i\theta)\}(\bar{y}_{iT} - \bar{\mathbf{x}}'_{iT}\hat{\boldsymbol{\beta}})$ に近づくことがわかる。

θ, ρ を推定するには ML もしくは REML を数値的に求める必要がある。しかし次のような簡便な推定量も利用可能である。時点 t での β の最小 2 乗推定量を $\hat{\boldsymbol{\beta}}_t = (\sum_{j=1}^k n_j \bar{\mathbf{x}}_{jt} \bar{\mathbf{x}}'_{jt})^{-1} \sum_{j=1}^k n_j \bar{\mathbf{x}}_{jt} \bar{y}_{jt}$ とおいて $\hat{e}_{it} = \bar{y}_{it} - \bar{\mathbf{x}}'_{it} \hat{\boldsymbol{\beta}}_t$ とする。 θ については (2.16) の下で与えられた議論と同様にして, $S_2 = \sum_{t=1}^T \sum_{i=1}^k n_i \hat{e}_{it}^2$,

$$n_* = \sum_{t=1}^T \left\{ \sum_{i=1}^k n_i - \text{tr} \left(\sum_{i=1}^k n_i^2 \bar{\mathbf{x}}_{it} \bar{\mathbf{x}}'_{it} \right) \left(\sum_{i=1}^k n_i \bar{\mathbf{x}}_{it} \bar{\mathbf{x}}'_{it} \right)^{-1} \right\}$$

に対して

$$\hat{\theta}^{TR} = \max \left\{ \frac{1}{n_*} \left(\frac{S_2}{\sigma_e^2} - T(k-p) \right), \frac{1}{k^{2/3}} \right\}$$

で推定できる。 ρ については、 $\hat{\rho} = \sum_{i=1}^k n_i \hat{\rho}_i / N$ で推定できる。ここで、 $\hat{\rho}_i$ は、 Q_1 の場合、 $\bar{e}_i = \sum_{t=1}^T \hat{e}_{it} / T$ に対し、

$$\hat{\rho}_i = 2 \sum_{t=1}^{T-1} \sum_{s=t+1}^T (\hat{e}_{is} - \bar{e}_i)(\hat{e}_{it} - \bar{e}_i)' / \left\{ (T-1) \sum_{t=1}^T (\hat{e}_{it} - \bar{e}_i)^2 \right\}$$

であり、 Q_2 の場合、 $\hat{\rho}_i = \sum_{t=2}^T (\hat{e}_{it} - \bar{e}_i)(\hat{e}_{i,t-1} - \bar{e}_i) / \sum_{t=1}^T (\hat{e}_{it} - \bar{e}_i)^2$ である。いずれも $|\hat{\rho}| < 1$ であり一応これらを用いて推定できるが、それらを初期値として ML, REML を数値的に解くのがよいであろう。

さて、自己共分散構造 Q_2 を仮定するとき、(4.4) もしくは (4.6) によって予測される値を実際のデータに基づいて計算してみよう。実は、3.4 節で取り上げた宅地物件の地価公示価格に関するデータは時系列的に得られており、ここでは 1997 年から 2001 年までの 5 年間 ($T = 5$) のデータを利用することができる。 $\hat{\theta}^{TR} = 0.686732$, $\hat{\rho} = 0.244273$ となり、駅毎の予測値が表 1 の EBLUP $_i^*$ の欄で与えられている。ほとんどの予測値が 2001 年だけのデータに基づいた予測値 EBLUP $_i$ を上方修正していることがわかる。これは、この 5 年間の宅地の価格が年々低下する傾向にあることと、(4.4) が過去のデータに基づいた予測値の方向へ縮小する作用があることに起因している。図 6 は、変量効果の予測値 \hat{v}_{it} を標準化した値 \hat{v}_{it}^* の経年的変化を、Nos.1, 3, 4, 13, 14, 33 の 6 つの駅について描いたものである。地価公示価格から共変量の影響を取り除いたものを縮小した値が \hat{v}_{it} であり、 \hat{v}_{it}^* はその相対的な値を示していることになる。この 5 年間で全体的には価格が低下している状況の中であって、品川駅に近い駅 (No.1, 3, 4) の地域効果が相対的に上昇し、No.13, 14, 33 の地域効果が相対的に下降していることがわかる。地価の下落の影響を強く受けた地域とそれほど影響を受けなかった地域など、小地域の価格変動に関する経年的変化の様子を捉えることができる。

4.3 一般化線形混合モデル

データが死亡数など離散的に変動するときには、2 項分布やポアソン分布などに基づいたモデルを考える必要がある。こうした離散分布に回帰項と変量効果を組み入れてモデルを構築することができ、それを統一的に扱ったモデルが一般化線形混合モデル (GLMM) である。いま、全体で k 個の地域があり、 i 番目の地域から n_i 個のデータ (もしくは集計データ) y_{i1}, \dots, y_{in_i}

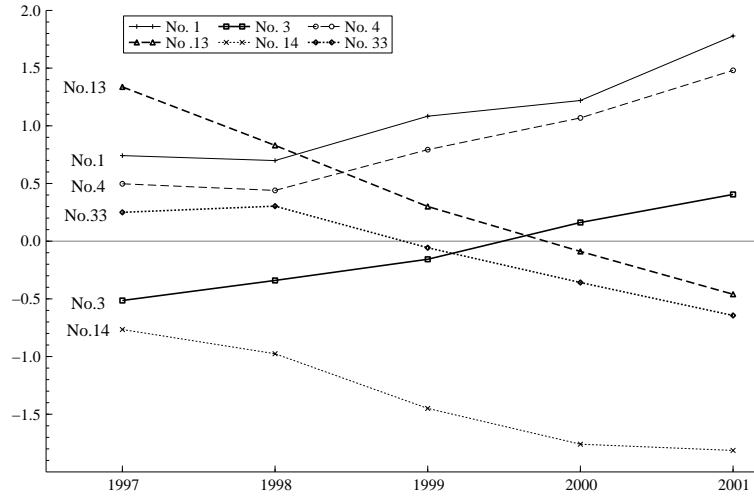


図 6: 変量効果 (地域効果) の予測値 \hat{v}_{it} を標準化した値 \hat{v}_{it}^* の経年的変化

が取られており, v_i を所与としたときの y_{ij} の条件付分布が

$$f(y_{ij}|v_i) = \exp \{ [y_{ij}\theta_{ij} - b(\theta_{ij})] / \tau_{ij} + c(y_{ij}, \tau_{ij}) \},$$

$j = 1, \dots, n_i; i = 1, \dots, k,$ で与えられるとする。これは一般化線形モデルと呼ばれる。この密度関数は自然母数 θ_{ij} と尺度母数 $\tau_{ij} (> 0)$ を用いて表現されており, τ_{ij} は既知と仮定される。 y_{ij} の条件付期待値を $E[y_{ij}|v_i] = \mu_{ij}$ と書くとき, μ_{ij} がリンク関数 $g(\cdot)$ を通して共変量 x_{ij} と関係づけることができることを仮定して

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + v_i$$

なる形で表現できるとする。ここで, $v_i \sim \mathcal{N}(0, \sigma_v^2)$ であり, すべて互いに独立であると仮定する。これを GLMM という。GLMM の詳しい解説については, McCullagh and Nelder (1989, 14.5 節), Fahrmeir and Tutz (2001), McCulloch (2003) で与えられている。また, McCulloch and Searle (2000) はわかりやすく書かれた本であり, 疾病地図など空間モデルを扱った本としては, Lawson, Browne and Vidal Rodeiro (2003), Lawson (2006) などがあるので参照してもらいたい。

共通母数と変量効果を組み入れた GLMM は, 地域の特性値に対して精度の高い安定した推定値を与えることができる。その代表的な例が疾病地

図の作成に用いられる死亡率及び死亡指標の推定である。死亡率の代表的な指標として用いられるのが標準化死亡率 (Standardized Mortality Rate, SMR) で、(観測死亡数)/(期待死亡数) で定義される。地域と年齢階級を細分していくと SMR のバラツキが大きくなってしまう。この問題に対して GLMM が安定した推定値を与えるために有用であることが示されており肺ガンの死亡率地図が作成されている。丹後 (1988) はポアソン・ガンマモデルによる経験ベイズ推定値の導出方法について議論し安定した疾病地図の作成が可能であることを示した。

4.4 経験ベイズモデルと階層ベイズモデル

枝分かれ誤差回帰モデル (2.2) を $y_{ij} = \mu_{ij} + e_{ij}$, $\mu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + v_i$ と分解すると、条件付分布は $y_{ij} | (\mu_{ij}, \sigma_e^2) \sim \mathcal{N}(\mu_{ij}, \sigma_e^2)$ となり、 μ_{ij} の事前分布が $\mu_{ij} \sim \mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}, \sigma_v^2)$ で与えられるベイズモデルの形で捉えることができる。母数 $\boldsymbol{\beta}$, σ_v^2 , σ_e^2 が既知のときが主観的ベイズモデルであり、それらの事前の値に強く依存してしまう。そこで客観性を持たせるために2つの方法が知られている。1つは $\boldsymbol{\beta}$, σ_v^2 , σ_e^2 を未知母数として扱う方法で経験ベイズモデルと呼ばれる。未知母数とすることで事前分布の恣意性を排除できる。経験ベイズ推定量は事前分布の情報を取り入れながらもその情報の誤りに対する実害が生じないという利点をもっている。いいかえると、事前情報に関して頑健であることを意味する。(2.2) などこの論文で扱ってきた線形混合モデルは経験ベイズモデルに対応している。

事前分布に客観性を持たせるもう1つの方法は、 $\boldsymbol{\beta}$, σ_v^2 , σ_e^2 を変量として扱う方法で階層ベイズモデルと呼ばれる。この場合、 (i, j) -成分に μ_{ij} をもつ行列を $\boldsymbol{\mu}$ とおくと、 $(\boldsymbol{\mu}, \sigma_e^2) | (\boldsymbol{\beta}, \sigma_v^2) \sim \pi_1(\boldsymbol{\mu}, \sigma_e^2 | \boldsymbol{\beta}, \sigma_v^2)$ を第1段階事前分布、 $(\boldsymbol{\beta}, \sigma_v^2) \sim \pi_2(\boldsymbol{\beta}, \sigma_v^2)$ を第2段階事前分布という。客観的なベイズ推定を構成するためには、一般に、第1段階事前分布は正確に、第2段階事前分布は無情報的に設定するのがよいとされている。例えば、第1段階事前分布 $\pi_1(\boldsymbol{\mu}, \sigma_e^2 | \boldsymbol{\beta}, \sigma_v^2)$ として

$$\begin{aligned} \mu_{ij} | (\boldsymbol{\beta}, \sigma_v^2) &\sim \mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}, \sigma_v^2), \\ \sigma_e^2 &\sim \sigma_e^{-2} d\sigma_e^2 \end{aligned}$$

がとられる。 $\sigma_e^{-2} d\sigma_e^2$ は尺度変換に関して不変な測度で無情報事前分布を表している。また第2段階事前分布 $\pi_2(\boldsymbol{\beta}, \sigma_v^2)$ としては σ_v^2 には無情報事前分布 $\sigma_v^{-2} d\sigma_v^2$ を想定し、 $\boldsymbol{\beta}$ に対しては、(1) 一様分布 $d\boldsymbol{\beta}$, (2) $\boldsymbol{\beta} | \sigma_v^2 \sim \mathcal{N}(\boldsymbol{\beta}_0, \sigma_v^2 \mathbf{A})$,

(3) $\beta | (\sigma_v^2, \lambda) \sim \mathcal{N}(\beta_0, \lambda \sigma_v^2 A)$, $\lambda \sim \pi_3(\lambda)$, などの設定が考えられる。ここで β_0 , A は既知の値とする。このような階層ベイズ推定量の理論的な性質については Kubokawa and Strawderman (2007) とその中の参考文献を参照してほしい。また階層ベイズを用いた空間データの解析については, Banerjee, Carlin and Gelfand (2004) に述べられている。

5 おわりに

線形混合モデルと経験最良線形不偏予測量について基本的な性質を解説し, 小地域推定や経時測定データへの応用について主に説明してきた。最後に, 線形混合モデルと関係のある推定問題について若干説明を加えて本稿を終えることにする。

まず, 家畜育種学という分野で生まれ育ってきた線形混合モデルの手法が, 数理統計学において有名なスタイン問題と関係している点を指摘しておこう。C. Stein が 1956 年に発見した理論は, 「3 次元以上の正規分布の平均を同時に推定する問題において標本平均よりも平均 2 乗誤差 (MSE) を一様に小さくする縮小推定量が存在する」ということであった。そこで登場した縮小推定量を $n_1 = \dots = n_k = n$ の地域モデル (3.1) に当てはめると, 平均 $\mu_i = \bar{x}_i' \beta + v_i$ からなるベクトル $\mu = (\mu_1, \dots, \mu_k)'$ の縮小推定量は

$$\hat{\mu}^S = \bar{X} \hat{\beta} + \max \left\{ 1 - \frac{(k-p-2)\sigma_e^2}{n \|\bar{y} - \bar{X} \hat{\beta}\|^2}, 0 \right\} (\bar{y} - \bar{X} \hat{\beta})$$

という形で表される。ここで $\hat{\beta} = (\bar{X}' \bar{X})^{-1} \bar{X}' \bar{y}$ は β の OLS である。Stein の理論によると, $k-p \geq 3$ ならば $\hat{\mu}^S$ の MSE は \bar{y} の MSE より一様に小さいことになる。Stein 問題については篠崎 (1991), 下平, 伊藤, 久保川, 竹内 (2004) に詳しく説明されているが, この問題は理論家の興味の対象となり 20 世紀の数理統計学の一分野に発展した。一方, $n_1 = \dots = n_k = n$ の地域モデル (3.1) に対して μ の EBLUP は, θ を REML で推定すると (3.3) より, 縮小推定量 $\hat{\mu}^S$ において $(k-p-2)$ を $(k-p)$ に置き換えたものに一致する。Henderson の BLUP についての論文が出されたのが 1950 年であり, その数年後に全く違う分野で Stein が EBLUP と同等な手法を考案していたことになる。しかも, Henderson は家畜育種という応用分野, Stein は統計的決定理論という理論分野であり, 応用・理論の双方向から研究されてきたトピックであるといえよう。

次に、2.4 節で扱ったように、LMM の分散成分に関する推定は LMM の主要な研究テーマの一つであるが、この問題が分散の不等式制約に関係している点を指摘しておこう。例えば、 $n_1 = \dots = n_k = n$ の枝分かれ誤差回帰モデル (2.2) を取り上げてみると、(2.16) 周辺の議論から S_1 と S_2 が独立になり、 $S_1/\sigma_1^2 \sim \chi_{m_1}^2$ 、 $S_2/\sigma_2^2 \sim \chi_{m_2}^2$ に従うことがわかる。ここで m_1, m_2 は対応する自由度を表し、 $\sigma_1^2 = \sigma_e^2$ 、 $\sigma_2^2 = \sigma_e^2 + n\sigma_v^2$ である。従って分散 σ_1^2, σ_2^2 の間に $\sigma_1^2 < \sigma_2^2$ なる不等式制約が入っていることがわかる。一般に多元配置の変量モデルにおいては分散の間に複雑な順序関係が入ることになり、母数制約のもとでの推定問題として定式化される。また多変量へ拡張したモデルを考えると、共分散行列の間に不等式制約が入った推定問題になり、Srivastava and Kubokawa (1999), Kubokawa and Tsai (2006) などで議論されている。

LMM の変数選択などのモデル選択規準の導出は非常に興味ある問題であり、Jiang, Rao, Gu and Nguyen (2008) の Fence 法や Kubokawa and Srivastava (2010) の経験ベイズ情報量規準は一致性をもつ変数選択法になっている。LMM の赤池情報量規準 (AIC) には縮小推定量が規準の中に現れないため、小地域の推定に関わる変数選択規準としてはあまり好ましくない。そこで、Vaida and Blanchard (2005) は条件付き AIC 規準を導入し、縮小推定量に基づいた規準を導出した。Kubokawa (2011a), Kubokawa and Nagashima (2011) はこの結果を一般的な LMM に拡張している。

本稿で取り上げてきた内容からわかるように、筆者が線形混合モデルの魅力に惹かれたのは小地域推定という問題を通してであった。カナダのオタワにある Carleton 大学へ滞在していた 1989 年、91 年当時、その数学・統計学科の教授 John N.K. Rao の研究テーマが LMM を利用した小地域推定であった。Rao はオタワ市内にあるカナダ統計局 (Statistica Canada) のスタッフとつねに連携を取り統計調査の現場において何が必要とされているのかに関心をもちながら研究を進めていた。当時の PhD 学生の論文テーマもその現場にモチベーションをもつ内容であり、PhD を取得した後にカナダ統計局へ就職して活躍している研究者もいる。Rao のこのような研究姿勢に、応用と理論の両輪で発展する統計学の生きた姿をみた思いがして大変感銘したことが思い出される。

謝辞. 編集者及び査読者の方々から貴重なコメントを頂きました。ここに深く感謝申し上げます。本研究は、科学研究費補助金 16500172, 21540114 及び東京大学大学院経済学研究科 21 世紀 COE プログラムから研究助成

を受けております。

参考文献

- [1] Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall, New York.
- [2] Basu, R., Ghosh, J.K., and Mukerjee, R. (2003). Empirical Bayes prediction intervals in a normal regression model: higher order asymptotics. *Statist. Prob. Letters*, **63**, 197-203.
- [3] Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **83**, 28-36.
- [4] Butar, F.B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *J. Statist. Plan. Inf.*, **112**, 63-76.
- [5] Chatterjee, S., Lahiri, P., and Li, H. (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *Ann. Statist.*, **36**, 1221-1245.
- [6] Das, K., Jiang, J. and Rao, J.N.K. (2004). Mean squared error of empirical predictor. *Ann. Statist.*, **32**, 818-840.
- [7] Datta, G.S., Kubokawa, T., Rao, J.N.K., and Molina, I. (2011). Estimation of mean squared error of model-based small area estimators. *Test, an Official Journal of the Spanish Society and Operations Research*, **20**, 367-388.
- [8] Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, **92**, 183-196.
- [9] Demidenko, E. (2004). *Mixed Models: Theory and Applications*. Wiley.
- [10] Diggle, P., Liang, K.-Y., and Zeger, S.L. (1994). *Longitudinal Data Analysis*. Oxford Univ. Press.

- [11] Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.*, **70**, 311-319.
- [12] Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. 2nd ed. Springer, New York.
- [13] Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269-277.
- [14] Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley.
- [15] Hall, P. and Maiti, T. (2006a). Nonparametric estimation of mean-squared prediction error in nested-error regression models. *Ann. Statist.*, **34**, 1733-1750.
- [16] Hall, P. and Maiti, T. (2006b). On parametric bootstrap methods for small area prediction. *J. Royal Statist. Soc.*, **68**, 221-238.
- [17] Henderson, C.R. (1950). Estimation of genetic parameters. *Ann. Math. Statist.*, **21**, 309-310.
- [18] Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge University Press. (「ミクロ計量経済学の方法：パネル・データ分析」(2007) 国友直人訳, 東洋経済新報社)
- [19] Jiang, J., Rao, J.S., Gu, Z., and Nguyen, T. (2008). Fence methods for mixed model selection. *Ann. Statist.*, **36**, 1669-1692.
- [20] Kubokawa, T. (2000). Estimation of variance and covariance components in elliptically contoured distributions. *J. Japan Statist. Soc.*, **30**, 143-176.
- [21] Kubokawa, T. (2010). Corrected empirical Bayes confidence intervals in nested error regression models. *J. Korean Statist. Soc.*, **39**, 221-236.
- [22] Kubokawa, T. (2011a). Conditional and unconditional methods for selecting variables in linear mixed models. *J. Multivariate Analysis*, **102**, 641-660.

- [23] Kubokawa, T. (2011b). On measuring uncertainty of small area estimators with higher order accuracy. *J. Japan Statist. Soc.*, to appear.
- [24] Kubokawa, T., and Nagashima, B. (2011). Parametric bootstrap methods for bias correction in linear mixed models. Discussion Paper Series, CIRJE-F-801.
- [25] Kubokawa, T., and Srivastava, M.S. (2010). An empirical Bayes information criterion for selecting variables in linear mixed models. *J. Japan Statist. Soc.*, **40**, 111-130.
- [26] Kubokawa, T. and Strawderman, W.E. (2007). On minimaxity and admissibility of hierarchical Bayes estimators. *J. Multivariate Analysis*, **98**, 829-851.
- [27] Kubokawa, T. and Tsai, M.-T. (2006). Estimation of covariance matrices in fixed and mixed effects linear models. *J. Multivariate Analysis*, **97**, 2242-2261.
- [28] Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.
- [29] Lawson, A.B. (2006). *Statistical Methods in Spacial Epidemiology*. 2nd ed. Wiley, England.
- [30] Lawson, A.B., Browne, W.J. and Vidal Rodeiro, C.L. (2003). *Disease Mapping with WinBUGS and MLwiN*. Wiley, England.
- [31] McCulloch, C.E. (2003). *Generalized Linear Mixed Models*. NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 7. IMS, USA.
- [32] McCulloch, C.E. and Searle, S.R. (2001). *Generalized, Linear and Mixed Models*. Wiley, New York.
- [33] Molenberghs, G. and Verbeke, G. (2006). *Models for Discrete Longitudinal Data*. Springer.
- [34] Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, New Jersey.

- [35] Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components*, Wiley, New York.
- [36] Srivastava, M.S. and Kubokawa, T. (1999). Improved nonnegative estimation of multivariate components of variance. *Ann. Statist.*, **27**, 2008-2032.
- [37] Tsimikas, J.V. and Ledolter, J. (1997). Mixed model representation of state space models: New smoothing results and their application to REML estimation. *Statistica Sinica*, **7**, 973-991.
- [38] Vaida, F., and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, **92**, 351-370.
- [39] Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.
- [40] 広津千尋 (1992). 実験データの解析—分散分析を超えて—. 共立出版.
- [41] 久保川達也 (2007). 線形混合モデルと小地域の推定. 応用統計学, **35**, 139-161.
- [42] 佐々木義之 (2007). 変量効果の推定と BLUP 法. 京都大学学術出版会.
- [43] 下平英寿, 伊藤秀一, 久保川達也, 竹内啓 (2004). モデル選択: 予測・検定・推定の交差点. 岩波書店
- [44] 笹瀬吉隆, 久保川達也 (2005). 経験ベイズ信頼区間の漸近補正と小地域推定への応用. 日本統計学会誌 (和文誌), **35**, 27-54.
- [45] 篠崎信雄 (1991). Stein タイプの縮小推定量とその応用. 応用統計学, **20**, 59-76.
- [46] 丹後俊郎 (1988). 死亡指標の経験的ベイズ推定量について. 応用統計学, **17**, 81-96.

第5章 接合分布関数（コピュラ） の理論と応用

塚原 英敦¹

（成城大学経済学部 教授）

この章では、接合関数（コピュラ）の理論と応用について概観する。まずその定義から始め、基本的な結果を述べた後、2次元接合関数の例を挙げる。また重要なクラスであるアルキメデス型接合関数について詳述する。次に、関連性尺度やその他の従属性概念、裾従属性との関連について説明し、周辺分布と接合関数が与えられた場合の乱数発生方法を解説する。接合関数モデルにおける統計的推定についても簡単に言及する。応用としては、生存解析やファイナンス、そして保険の分野における応用例を紹介し、さらに多変量極値理論やマルコフ過程論といった確率論の分野で接合関数が果たす役割を簡潔に説明する。

¹tsukahar@seiyo.ac.jp . 本稿は、最新の情報に基づいて塚原 (2003) に大幅な加筆・修正を施したものである。

1 はじめに

接合分布関数とは copula (コピュラ) の訳語である。接合分布とは、1次元周辺分布を接合して多次元分布を構成する基本的標準分布という意味合いであり、特に連続の場合、多次元確率変数の各成分を確率積分変換したものの分布に等しい。その接合分布の分布関数を接合分布関数と呼ぶのであるが、以下では短く接合関数と書くことにする。

接合関数、あるいはその変形は多次元分布や従属性の研究に古くから現れていたが、Sklar (1959) により ‘copula’ と命名された。しかし、その後も名前の認知度が低く、いくつかの論文中で ‘再発見’ されている (日本人研究者による例としては、Sibuya (1960) や Yanagimoto (1970) がある)。正確な定義は次節で述べるが、接合関数は周辺分布を所与とした確率分布の研究 (例えば、Rüschendorf, Schweizer and Taylor (1996) 参照)、および分布関数から成る空間上の2項演算族との関連で確率距離空間 (probabilistic metric spaces) の理論 (Schweizer and Sklar (1983) 参照) において古くから用いられてきた道具であった。

しかし近年、統計学の立場から、複数の確率変数間の従属性を理解し、それをモデル化するための有用かつ柔軟な道具として、また1次元周辺分布が与えられたときの (あるいは特定化しないままの) 多変量モデルの構成法として、様々な応用分野で接合関数が脚光を浴びている (Fisher (1997) 参照)。数学的に扱いやすい分布族が多変量正規分布以外には見当たらないということから、伝統的な多変量推測統計の分野では、母集団分布として多変量正規分布を前提とすることが圧倒的であったが (例えば、竹村 (1991) を参照)、多変量正規分布の仮定が理論的、または経験的に問題のある状況もまた少なからず存在する。そうした場合には、接合関数によるモデリングが有効である。さらに、接合関数は非正規多変量データを擬似発生させる場合にも便利である。この接合関数を主題として詳しく扱っている書物としては、Drouot Mari and Kotz (2001), Joe (1997), Nelsen (2006) の3冊があるが、Nelsen によるものが最も読み易い好著であり、本稿を書く上でもかなり参考にした書物である。また、McMeil, Frey and Embrechts (2005) の第5章にもよくまとまった接合関数の解説がある。

本稿の目的は、この接合関数の理論と応用について概観することである。まず次節で接合関数の定義から始めて、その基本的な性質を述べた後、2次元接合関数の具体例をいくつか挙げる。そして、関連性尺度やその他の従属性概念とのつながりについて説明し、周辺分布と接合関数が与えられた場合の乱数発生方法を解説する。さらに、データに基づく接合関数の推定

量を定義し、その性質を考察する。また、接合関数モデルを多変量データに適用した場合のパラメータ推定法についても簡単に解説する。3節では、生存解析やファイナンス、保険などいくつかの分野における接合関数の応用例を紹介し、さらに多変量極値理論やマルコフ過程論といった確率論の分野で接合関数が果たす役割を簡潔に説明する。

2 定義と基本的性質

定義 2.1 d 次元接合関数 (copula function) とは、すべての1次元周辺分布が $[0, 1]$ 上の一様分布である d 次元分布関数のことである。通常、次元は文脈から明らかであることがほとんどであるから、以下では単に接合関数と呼ぶことにする。

接合関数に関する最も基本的な結果は次の定理である：

定理 2.2 (Sklar (1959)) 任意の d 次元分布関数 F に対して、

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \quad (2.1)$$

となる d 次元接合関数 C が存在する。ここで、 F_i は F の i 番目の1次元周辺分布関数である ($i = 1, \dots, d$)。

この定理、そして2.1節で述べる他の定理について、その証明は数学付録にまとめられているので、適宜参照されたい。

定理 2.2 は接合関数が多次元同時分布とその1次元周辺分布をつなぐ役割をもっていることを示す。よって、多次元分布をモデル化するときに周辺分布と確率変数間の従属構造を表す接合関数とを別々に特定化することが可能となり非常に便利である。

F が連続である場合には C は一意的に定まり、 F の接合関数と呼ばれる。この場合、 C は

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \quad (2.2)$$

と与えられることが容易にわかる。

2.1 接合関数の基本的性質

次の不等式は、 $d = 2$ の場合に Hoeffding (1940) と Fréchet (1951) により独立に発見された。

定理 2.3 (Frechét-Hoeffding の不等式) 任意の d 次元接合関数 C に対して,

$$\max\left(\sum_{i=1}^d u_i - d + 1, 0\right) \leq C(u_1, \dots, u_d) \leq \min(u_1, \dots, u_d). \quad (2.3)$$

注意 2.4 上限 $M(u_1, \dots, u_d) := \min(u_1, \dots, u_d)$ は常に接合関数であるが, 下限 $W(u_1, \dots, u_d) := \max(\sum_{i=1}^d u_i - d + 1, 0)$ は $d \geq 3$ のとき接合関数にならないことに注意しよう.

定義 2.5 (X_1, \dots, X_d) の同時分布関数を $F(x_1, \dots, x_d)$ とする. 記号 $\stackrel{\mathcal{L}}{=}$ は右辺と左辺の確率ベクトルが同じ分布をもつことを意味する.

- (i) F の Sklar 分解 (2.1) において, $C(u_1, \dots, u_d) = M(u_1, \dots, u_d)$ ととれるとき, (X_1, \dots, X_d) は共単調 (comonotone) であるという. このとき, 確率変数 Z と非減少関数 f_i ($i = 1, \dots, d$) が存在して, $(X_1, \dots, X_d) \stackrel{\mathcal{L}}{=} (f_1(Z), \dots, f_d(Z))$ となることを示せる. これはある意味で, X_1, \dots, X_d の間に存在しうる最も強い正の従属関係である.
- (ii) $d = 2$ とする. F の Sklar 分解 (2.1) において, $C(u_1, u_2) = W(u_1, u_2)$ ととれるとき, (X_1, X_2) は反単調 (countermonotone) であるという. このとき, 確率変数 Z , 非減少関数 f_1 と非増加関数 f_2 が存在して, $(X_1, X_2) \stackrel{\mathcal{L}}{=} (f_1(Z), f_2(Z))$ となることを示せる. これも X_1 と X_2 の間に存在しうる最も強い負の従属関係であると言えよう.

共単調性の理論と応用について良くまとまった展望論文としては, Dhaene et al. (2002a, 2002b) がある.

以下では, (X_1, \dots, X_d) の同時分布関数を $F(x_1, \dots, x_d)$ とし, その Sklar 分解が (2.1) で与えられているとする. まず次の Hoeffding の公式を思い起こそう: $E|X_1|, E|X_2|, E|X_1 X_2|$ がすべて有限であるとき, X_1 と X_2 の共分散は

$$\text{cov}(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F(x_1, x_2) - F_1(x_1)F_2(x_2)) dx_1 dx_2 \quad (2.4)$$

と書ける (証明は数学付録参照).

この公式と Frechét-Hoeffding の不等式 (2.3) を合わせて用いると, 次のことが容易にわかる.

定理 2.6 分散が有限かつ非ゼロである 2 つの確率変数 X_1 と X_2 の周辺分布がそれぞれ F_1, F_2 であるとするとき, X_1 と X_2 の相関係数がとり得る値の集合は閉区間 $[\underline{\rho}, \bar{\rho}]$ となる ($\underline{\rho} < 0 < \bar{\rho}$). $\underline{\rho}$ が達成されるための必要十分条件は X_1 と X_2 が反単調となることであり, $\bar{\rho}$ が達成されるための必要十分条件は X_1 と X_2 が共単調となることである.

このことから, 2 つの確率変数の周辺分布だけが指定されている場合においても, 従属性を自由に設定することによって相関係数の値がとり得る範囲を $[-1, 1]$ とできるとは限らないことがわかる. 実際, McNeil, Frey and Embrechts (2005) の例 5.26 では, 周辺分布を対数正規分布としたとき, 区間 $[\underline{\rho}, \bar{\rho}]$ が非常に狭くなる可能性があることが示されている.

定理 2.7 (独立性の必要十分条件) F が連続であると仮定する. このとき, X_1, \dots, X_d が独立であることと $C(u_1, \dots, u_d) = \prod_{i=1}^d u_i =: \Pi(u_1, \dots, u_d)$ が成り立つことは同等である.

この事実は, データから接合関数を推定して独立性を検定できることを示唆している. これについては Tsukahara (2000) に要約と検定統計量の漸近理論が与えられている. また, Genest and Rémillard (2004) も参照のこと.

定理 2.8 (単調変換の下での不変性) F が連続であるとする. h_i が X_i の値域上で a.s. 狭義単調増加ならば, (X_1, \dots, X_d) と $(h_1(X_1), \dots, h_d(X_d))$ は同一の接合関数をもつ.

この結果は, 接合関数が X_i の単調増加変換により不変であり, 接合関数が確率変数間の順序的な (ordinal) 従属性を支配していると解釈できる.

次の性質は接合関数に関する技術的事実を示す場合等に役立つ.

定理 2.9 (Lipschitz 連続性) $(u_1, \dots, u_d), (v_1, \dots, v_d) \in [0, 1]^d$ に対して,

$$|C(u_1, \dots, u_d) - C(v_1, \dots, v_d)| \leq \sum_{i=1}^d |u_i - v_i|.$$

この定理と Sklar の定理から, d 次元分布関数 F とその 1 次元周辺分布関数 F_1, \dots, F_d に対して,

$$|F(x_1, \dots, x_d) - F(y_1, \dots, y_d)| \leq \sum_{i=1}^d |F_i(x_i) - F_i(y_i)|$$

が成り立つことがわかる. よって, d 次元分布関数 F の連続性とそのすべての 1 次元周辺分布関数 F_1, \dots, F_d の連続性は同等である.

2.2 2次元接合関数の1パラメータ族の例

応用上よく用いられている2次元接合関数の1パラメータ族をいくつか挙げよう.

- (i) Clayton-Cook-Johnson 族 (Clayton (1978), Cook and Johnson (1981))

$$C_\theta(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \quad \theta \in [-1, \infty) \setminus \{0\}$$

- (ii) Gumbel-Hougaard 族 (Gumbel (1960), Hougaard (1986))

$$C_\theta(u_1, u_2) = \exp \left\{ - [(-\log u_1)^\theta + (-\log u_2)^\theta]^{1/\theta} \right\}, \quad \theta \geq 1$$

- (iii) Frank 族 (Frank (1979))

$$C_\theta(u_1, u_2) = \frac{1}{\theta} \log \left(1 + \frac{(e^{\theta u_1} - 1)(e^{\theta u_2} - 1)}{e^\theta - 1} \right), \quad \theta \in \mathbb{R} \setminus \{0\}$$

- (iv) Plackett 族 (Plackett (1965))

$$C_\theta(u_1, u_2) = \begin{cases} \frac{1 + (\theta - 1)(u_1 + u_2) - \sqrt{\{1 + (\theta - 1)(u_1 + u_2)\}^2 - 4u_1u_2\theta(\theta - 1)}}{2(\theta - 1)}, & \theta > 0, \theta \neq 1 \\ u_1u_2, & \theta = 1 \end{cases}$$

- (v) Gauss 族

$$C_\theta(u_1, u_2) = \Phi_\theta(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$$

ここで, Φ_θ は平均 $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, 共分散行列 $\begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$ をもつ2変量正規分布の分布関数, Φ は1次元標準正規分布の分布関数である.

これらのうち, Gumbel-Hougaard 族を除く4つの族は Frechét-Hoeffding の不等式における上限と下限, さらに独立接合関数 Π を含む(このような族は包括的 (comprehensive) と呼ばれる). よって, 幅広い従属関係を表すことのできるモデルとして多用されている.

下の図1-3では, 上記の(i)-(iii), (v)について, Kendallの τ (2.4節参照)がそれぞれ0.2, 0.5, 0.8になるようにパラメータ θ の値を定めた上で, 周辺分布はすべて標準正規分布として1,000個の擬似乱数を発生させプロットしたものである. これらの図から, 異なる接合関数によって表現される従属性の性質はかなり違ったものになるということが直観的に理解できよう.

なお，これらの乱数発生には自前の MATLAB コードを用いたが，現在では S+FinMetrics™ 2.0 に接合関数のためのオブジェクト・クラスが装備されており，大変有用である．その用例については，Zivot and Wang (2006) の第 19 章を参照せよ．

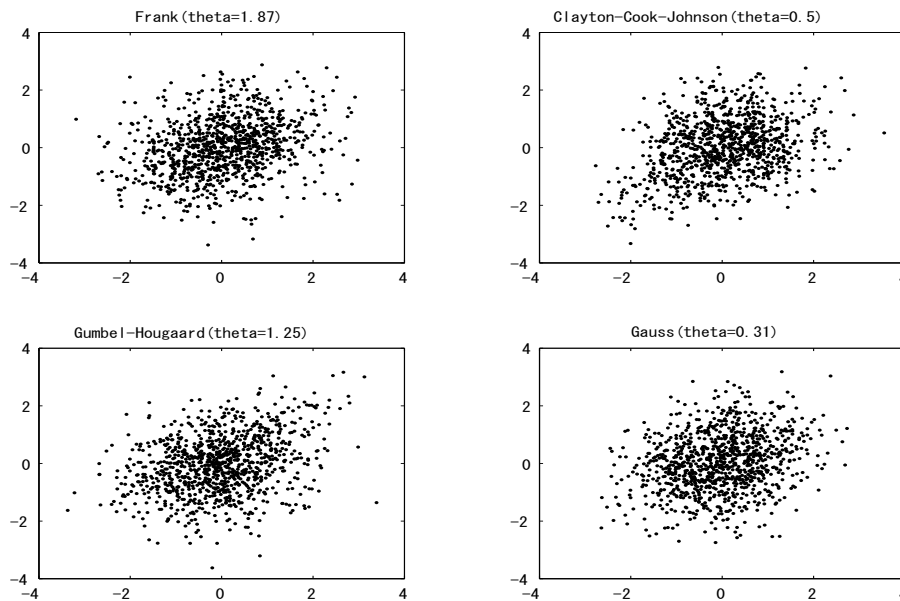


図 1: $\tau = 0.2$ での従属性 (周辺分布は $N(0, 1)$)

2.3 アルキメデス型接合関数

定義 2.10 区間 $(0, 1]$ 上で定義され， \mathbb{R}_+ の値をとる単調減少凸関数 ϕ が $\phi(1) = 0$ を満たすとする．このとき，

$$C_\phi(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2)), \quad (u_1, u_2) \in (0, 1]^2$$

を (2 次元) アルキメデス型接合関数 (Archimedean copula) と呼び， ϕ は C_ϕ の生成素 (generator) という．

このアルキメデス型という名前の由来については，Schweizer and Sklar (1983)，あるいは Nelsen (2006) を参照されたい．これは数学的に扱いやす

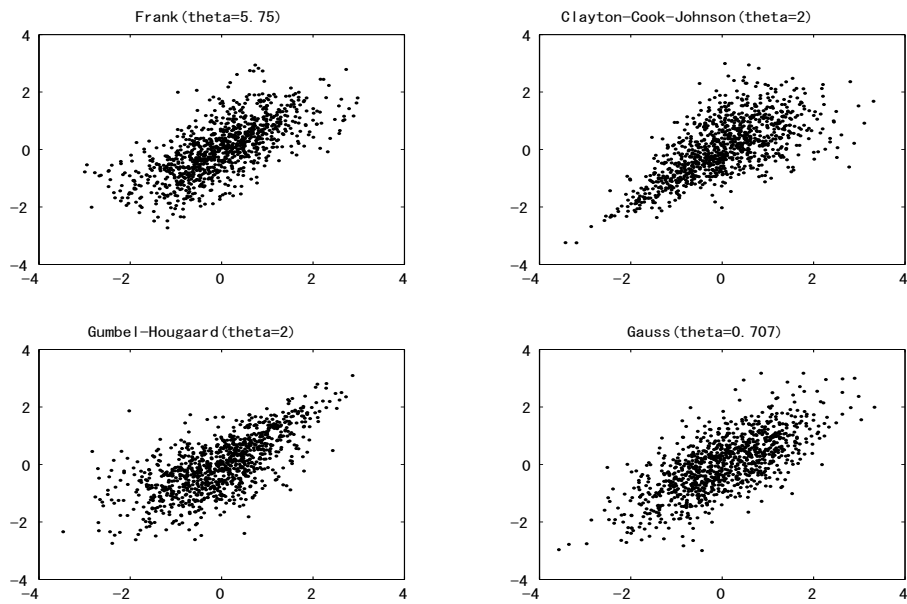


図 2: $\tau = 0.5$ での従属性 (周辺分布は $N(0, 1)$)

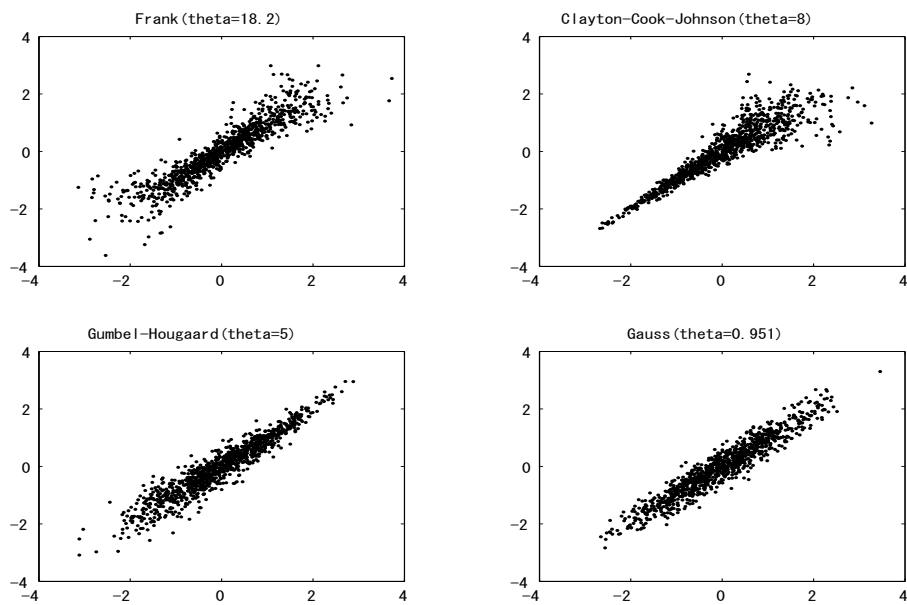


図 3: $\tau = 0.8$ での従属性 (周辺分布は $N(0, 1)$)

い性質をもつ接合関数のクラスであり, 2.2 節に挙げた例の中で (i), (ii), (iii) はこのクラスに属する. 対応する生成素は

$$\text{Clayton 族: } \phi(t) = \frac{t^{-\theta} - 1}{\theta}, \quad \text{Gumbel-Hougaard 族: } \phi(t) = (-\log t)^\theta,$$

$$\text{Frank 族: } \phi(t) = -\log \frac{e^{\theta t} - 1}{e^\theta - 1}$$

である. 独立接合関数 $\Pi(u_1, u_2) = u_1 u_2$ と Frechét-Hoeffding 不等式の下限に対応する接合関数 $W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$ もそれぞれ $\phi(t) = -\log t$, $\phi(t) = 1 - t$ ととればアルキメデス型接合関数であることがわかる.

注意 2.11 d 次元 ($d \geq 3$) への拡張としては当然

$$C_\phi(u_1, \dots, u_d) = \phi^{-1}(\phi(u_1) + \dots + \phi(u_d)), \quad (u_1, \dots, u_d) \in (0, 1]^d \quad (2.5)$$

とすることが考えられるが, これは必ずしも接合関数とはならない (反例: $\phi(t) = 1 - t$). 条件 $\lim_{t \rightarrow 0} \phi(t) = \infty$ が成り立つときには, ϕ^{-1} が \mathbb{R}_+ 上で完全単調 (completely monotone) であることと (2.5) の C_ϕ が任意の $d \geq 2$ に対して接合関数となることが同値となる (証明は Kimberling (1974) 参照). ここで, 関数 $g(t)$ が区間 I 上で完全単調とは, g が I 上で連続であり, かつ I の内点 t と $k = 0, 1, 2, \dots$ に対して, $(-1)^k (d^k / dt^k) g(t) \geq 0$ が成り立つことと定義される.

複数の確率変数間の関連を共通の潜在変数への依存という形で説明しようとする 'random effect' モデル (生存解析では frailty モデルと呼ばれる) において, アルキメデス型接合関数は必然的に現れる. 正の確率変数 W を潜在変数 (frailty) とする. X_1, \dots, X_d は, W が与えられたとき, 条件付きで独立な確率変数であり, X_i の条件付き分布関数が

$$H_i(x_i|W) = (H_i(x_i))^W$$

という形になっていると仮定する. ここで H_i はある分布関数である. さらに, $\psi(t) := E[e^{-tW}]$ を W の Laplace 変換とすると, X_i の周辺分布関数は $F_i(x_i) = \psi(-\log H_i(x_i))$ となり, X_1, \dots, X_d の同時分布関数 $F(x_1, \dots, x_d)$ は

$$E \left[(H_1(x_1) \cdots H_d(x_d))^W \right] = \psi \left[\psi^{-1}(F_1(x_1)) + \dots + \psi^{-1}(F_d(x_d)) \right]$$

で与えられる．この分布関数の接合関数はアルキメデス型であり，その生成素は W の Laplace 変換の逆関数で与えられることになる． W の分布として，ガンマ分布をとると Clayton-Cook-Johnson 族，正の狭義安定 (positive strictly stable) 分布をとると Gumbel-Hougaard 族が得られる (Hougaard (1986) 参照)．

2.4 関連性尺度

確率変数 X_1 と X_2 の同時分布関数を $F(x_1, x_2)$ ，それぞれの周辺分布関数を F_1, F_2 とする．さらに， F は連続と仮定し， F の Sklar 分解が (2.1) で与えられているとする．以下では，Kruskal (1958) に述べられているような関連性の順序尺度 (ordinal measures of association) が接合関数 C のみの関数になることを示す．

X_1 と X_2 の間の最も単純な関連性尺度としては，まず (x_1^*, x_2^*) を固定した点として

$$P[(X_1 - x_1^*)(X_2 - x_2^*) > 0] = 1 - F_1(x_1^*) - F_2(x_2^*) + 2F(x_1^*, x_2^*)$$

が考えられる．これは一方向しか考えていないので対称化すると，

$$\begin{aligned} & P[(X_1 - x_1^*)(X_2 - x_2^*) > 0] - P[(X_1 - x_1^*)(X_2 - x_2^*) < 0] \\ &= 2P[(X_1 - x_1^*)(X_2 - x_2^*) > 0] - 1 = E[\text{sgn}(X_1 - x_1^*)\text{sgn}(X_2 - x_2^*)] \end{aligned}$$

となる．特に (x_1^*, x_2^*) を X_1 と X_2 のミディアンとした場合，Blomqvist (1950) により提案された関連性尺度となる．しかしながら (x_1^*, x_2^*) をどう選ぶかは恣意性を伴うため，特定の (x_1^*, x_2^*) の影響をなくするのが望ましい．このための方法として次の 2 つがある．

(i) (x_1^*, x_2^*) に重み $F(x_1, x_2)$ をつけて積分すると，これはよく知られた Kendall の τ となる．すなわち， (X_1, X_2) と $(\tilde{X}_1, \tilde{X}_2)$ が独立であり，同一分布 F をもつとすると，

$$\begin{aligned} & P[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0] - P[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0] \\ &= 2P[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0] - 1 = E[\text{sgn}(X_1 - \tilde{X}_1)\text{sgn}(X_2 - \tilde{X}_2)] \end{aligned}$$

であり，右辺は Kendall の τ の定義式に等しい．この Kendall の τ が C を通してのみ F に依存することは容易にわかる： (X_1, X_2) と $(\tilde{X}_1, \tilde{X}_2)$ が交換可

能であることから ,

$$\begin{aligned}
 \tau &= 2\text{P}[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0] - 1 = 4\text{P}(X_1 < \tilde{X}_1, X_2 < \tilde{X}_2) - 1 \\
 &= 4\text{E}\left(\text{P}(X_1 < \tilde{X}_1, X_2 < \tilde{X}_2 \mid \tilde{X}_1, \tilde{X}_2)\right) - 1 \\
 &= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{P}(X_1 < x_1, X_2 < x_2) dF(x_1, x_2) - 1 \\
 &= 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1.
 \end{aligned}$$

また , $-1 \leq \tau \leq 1$ も定義よりすぐわかる .

(ii) (x_1^*, x_2^*) に重み $F_1(x_1)F_2(x_2)$ をつけて積分すると , これもまたよく用いられる Spearman の ρ (の定数倍) となる . まず , $(X_1, X_2), (\tilde{X}_1, \tilde{X}_2), (\hat{X}_1, \hat{X}_2)$ が独立かつ同一分布 F をもつとすると , 重み $F_1(x_1)F_2(x_2)$ をつけた積分の結果は , (i) での計算と同様にして ,

$$\begin{aligned}
 &\text{P}[(X_1 - \tilde{X}_1)(X_2 - \hat{X}_2) > 0] - \text{P}[(X_1 - \tilde{X}_1)(X_2 - \hat{X}_2) < 0] \quad (2.6) \\
 &= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{P}(X_1 < x_1, X_2 < x_2) dF_1(x_1)dF_2(x_2) - 1 \\
 &= 4 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 1 = 4 \int_0^1 \int_0^1 [C(u_1, u_2) - u_1 u_2] du_1 du_2
 \end{aligned}$$

と表される . 一方 , Spearman の ρ は定義から $F_1(X_1)$ と $F_2(X_2)$ の相関係数であるが , それらは共に一様分布 $U(0, 1)$ に従い , その分散が $1/12$ であることから , $\rho = 12 \text{cov}(F_1(X_1), F_2(X_2))$ となる . さらに , $F_1(X_1)$ と $F_2(X_2)$ の同時分布関数は C であるから , (2.4) を用いると ,

$$\rho = 12 \int_0^1 \int_0^1 [C(u_1, u_2) - u_1 u_2] du_1 du_2$$

となる . よって , この ρ も C を通してのみ F に依存し , (2.6) の 3 倍となっていることがわかる .

2.2 節で挙げた 1 パラメータ接合関数族に対する Kendall の τ と Spearman の ρ の値はすべて θ の関数となるが , それらを表 1 としてまとめておく . ここで , $D_k(x)$ は次のように定義される k 次の Debye 関数である .

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$$

他の関連性尺度については , Schweizer and Wolff (1981) や Yanagimoto (1970) を参照せよ .

接合関数族	Kendall の τ	Spearman の ρ
Clayton-Cook-Johnson	$\theta/(\theta + 2)$	—
Gumbel-Hougaard	$(\theta - 1)/\theta$	—
Frank	$1 - (4/\theta)\{1 - D_1(\theta)\}$	$1 - (12/\theta)\{D_1(\theta) - D_2(\theta)\}$
Plackett	—	$(\theta + 1)/(\theta - 1) - 2\theta \log \theta / (\theta - 1)^2$
Gauss	$(2/\pi) \arcsin \theta$	$(6/\pi) \arcsin(\theta/2)$

表 1: Kendall の τ と Spearman の ρ (— : θ の明示的な関数として書けない場合)

2.5 従属性の諸概念

上記の関連性尺度は、2つの確率変数間の関連の強さを1つの数値として表そうというものであるが、一方、定性的な従属性の概念も数多く存在する。そこでも、特に連続分布を考える場合には、接合関数が本質的な役割を果たす。最も基本的な例は定理 2.7 であり、独立性は対応する接合関数が Π であるということによって特徴付けられるのである。

まず、大域的な従属性を表す代表的な例を2つ紹介する。記号は前節と全く同様であるが、ここでも F は連続であると仮定する。

象限従属性： X_1 と X_2 が正に象限従属である (positively quadrant dependent) とは、すべての $(x_1, x_2) \in \mathbb{R}^2$ に対して、

$$F(x_1, x_2) \geq F_1(x_1)F_2(x_2)$$

が成り立つことである。負の象限従属性は不等号を逆向きにして定義される。この Lehmann (1966) による正の象限従属性は、明らかに、すべての $(u_1, u_2) \in [0, 1]^2$ に対して、 $C(u_1, u_2) \geq u_1u_2$ ということと同等である。

確率的単調性： X_2 が X_1 について確率的に増加 (stochastically increasing) であるとは、固定された各 x_2 に対して、 $P(X_2 > x_2 | X_1 = x_1)$ が x_1 の関数として増加関数であることをいう。この概念も Lehmann (1966) に端を発するが、その論文では正の回帰従属性 (positive regression dependence) と呼ばれている。 X_2 が X_1 について確率的に増加であるための必要十分条件を接合関数を用いて述べると、ほとんどすべての u_1 に対して、 $\partial C(u_1, u_2)/\partial u_1$ が u_1 の関数として減少関数であるこ

ととなる（接合関数は各引数に関して増加だから，その偏微分係数はほとんどいたるところ存在することに注意せよ）．もちろん，これは $C(u_1, u_2)$ が u_1 の関数として凹であることと同等である．

次に， X_1 と X_2 の極端に大きな（小さな）値同士が関連する度合を表すことを考える．

定義 2.12 X_1 と X_2 の上裾従属性係数 (coefficient of upper tail dependence) は，

$$\lambda_U := \lim_{u \uparrow 1} P(X_2 > F_2^{-1}(u) \mid X_1 > F_1^{-1}(u))$$

と定義される．ただし，右辺の極限は存在するものとする．同様に， X_1 と X_2 の下裾従属性係数 (coefficient of lower tail dependence) は，

$$\lambda_L := \lim_{u \downarrow 0} P(X_2 \leq F_2^{-1}(u) \mid X_1 \leq F_1^{-1}(u))$$

と定義される．ただし，右辺の極限は存在するものとする．

$\lambda_U \in (0, 1]$ のとき， X_1 と X_2 は上裾従属であるといい， $\lambda_U = 0$ のときには， X_1 と X_2 は（漸近的に）上裾独立であるという． λ_L についても同様である．

F_1 と F_2 が連続である場合，この裾従属性係数は F の接合関数 C のみに依存する：

$$\lambda_U = \lim_{u \uparrow 1} \frac{\bar{C}(u, u)}{1 - u}, \quad \lambda_L = \lim_{u \downarrow 0} \frac{C(u, u)}{u}$$

が成り立つ．ここで， $\bar{C}(u_1, u_2) := 1 - u_1 - u_2 + C(u_1, u_2)$ である．

若干の計算により，Frank 族，Plackett 族，Gauss 族については $\lambda_U = \lambda_L = 0$ ，Gumbel-Hougaard 族については， $\lambda_U = 2 - 2^{1/\theta}$ ， $\lambda_L = 0$ ，Clayton-Cook-Johnson 族については， $\lambda_U = 0$ ， $\lambda_L = 2^{-1/\theta}$ ($\theta > 0$)， $\lambda_L = 0$ ($\theta \leq 0$) となることがわかる．これらの結果は，図 1-3 から直観的に納得できよう．

これらの他にも多くの従属性概念が存在する．それらの詳しい解説については，Nelsen (2006) の第 5 章，Joe (1997) の第 2 章，Drouot Mari and Kotz (2001) を参照せよ．また，関連する多変量分布の確率順序に関しては，Müller and Stoyan (2002) が詳しく扱っている．

2.6 擬似乱数発生の方法

まず 2 次元の場合を考える．分布関数 C からの乱数 (u_1, u_2) を発生させることができれば，任意の周辺分布関数 F_1 と F_2 に対して， $(F_1^{-1}(u_1), F_2^{-1}(u_2))$

と変換することにより，同時分布関数 $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ からの乱数が得られる．

いま (U_1, U_2) が分布 C に従うとする． $(\partial/\partial u_1)C(u_1, u_2)$ は， $U_1 = u_1$ が与えられたときの U_2 の条件付き分布関数 $C_{2|1}(u_2 | u_1) := P(U_2 \leq u_2 | U_1 = u_1)$ に等しいから，一般の接合関数 C に従う (u_1, u_2) を生成するには次のステップに従えばよい：

- (i) $U(0, 1)$ に従う一様乱数 u_1 と z を独立に発生させる．
- (ii) $u_2 = C_{2|1}^{-1}(z | u_1)$ を計算する．

$C_{2|1}^{-1}$ が解析的に求まる場合，この方法は非常に容易である．

C が 2.3 節で述べたアルキメデス型接合関数の場合には，次の 2 つの方法がある．

- (I) $C(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2))$ を微分して，

$$C_{2|1}(u_2 | u_1) = \frac{\phi'(u_1)}{\phi'(C(u_1, u_2))} \Rightarrow C(u_1, u_2) = (\phi')^{-1} \left(\frac{\phi'(u_1)}{C_{2|1}(u_2 | u_1)} \right).$$

よって，次の手順に従えば C からの乱数 (u_1, u_2) を生成できる：

- (i) $U(0, 1)$ に従う一様乱数 u_1 と z を独立に発生させる．
- (ii) $w = (\phi')^{-1} \left(\frac{\phi'(u_1)}{z} \right)$ を計算する (w は $C(U_1, U_2)$ の分布からの乱数と考えられる)．
- (iii) $u_2 = \phi^{-1}(\phi(w) - \phi(u_1))$ を計算する．

この方法が有効であるためには $(\phi')^{-1}$ が明示的に求まる必要がある．Clayton-Cook-Johnson 族や Frank 族については $(\phi')^{-1}$ が簡単に求まるが，Gumbel-Hougaard 族については $(\phi')^{-1}$ が解析的には求まらない．

- (II) Genest and Rivest (1993) によれば， (U_1, U_2) が C に従うとき，

$$S := \frac{\phi(U_1)}{\phi(U_1) + \phi(U_2)}, \quad T := C(U_1, U_2)$$

で定義される S と T は独立であり， S の分布は $U(0, 1)$ ， T の分布関数は $K_C(t) := t - \frac{\phi(t)}{\phi'_+(t)}$ で与えられる．ただし， ϕ'_+ は ϕ の右導関数である．この結果を用いると，

- (i) $U(0, 1)$ に従う一様乱数 s と z を独立に発生させる .
- (ii) $t = K_C^{-1}(z)$ を計算する (t は $C(U_1, U_2)$ の分布からの乱数と考えられる) .
- (iii) $u_1 = \phi^{-1}(s\phi(t))$, $u_2 = \phi^{-1}((1-s)\phi(t))$ とおく .

という手順で C からの乱数 (u_1, u_2) を発生させることも可能である .
 しかし , 一般に K_C^{-1} は明示的に求まらず , このためにステップ (ii) で数値的に方程式を解くことが必要となる .

しかし , ステップ (iii) を見ればわかるように , 最終的に必要なのは $\phi(T)$ の分布からの乱数発生である . Gumbel-Hougaard 族については , この点を考慮して , K_C^{-1} が求まらないという問題をうまく回避できる . この接合関数族に対しては $\phi(t) = (-\log t)^\theta$ であり , $-\log T$ の分布が密度関数 $(1 - 1/\theta)e^{-x} + (1/\theta)xe^{-x}$ をもつガンマ分布の混合分布となることが容易に確認できる (Lee (1979) を参照せよ) . よって ,

- (i) $U(0, 1)$ に従う一様乱数 z と , それとは独立に密度関数 $(1 - 1/\theta)e^{-s} + (1/\theta)se^{-s}$ をもつ分布からの乱数 s を発生させる .
- (ii) $u_1 = \exp[-z^{1/\theta}s]$, $u_2 = \exp[-(1-z)^{1/\theta}s]$ とおく .

という手順に従えばよい .

次元が 3 以上の場合には , Gauss 族などの特別な接合関数の場合を除いて , この節の最初に述べた基本的な考え方を適用する . (U_1, \dots, U_d) が分布 C に従うとする . まず , (U_1, \dots, U_k) ($k = 1, \dots, d$) の周辺分布関数を

$$C_k(u_1, \dots, u_k) = \begin{cases} u_1 & k = 1 \text{ のとき} \\ C(u_1, \dots, u_k, 1, \dots, 1) & k = 2, \dots, d-1 \text{ のとき} \\ C(u_1, \dots, u_d) & k = d \text{ のとき} \end{cases}$$

としたとき , $U_1 = u_1, \dots, U_{k-1} = u_{k-1}$ が与えられたときの U_k の条件付分布関数が帰納的に

$$C_{k|1, \dots, k-1}(u_k | u_1, \dots, u_{k-1}) = \frac{\partial^{k-1} C_k(u_1, \dots, u_k)}{\partial u_1 \cdots \partial u_{k-1}} \bigg/ \frac{\partial^{k-1} C_{k-1}(u_1, \dots, u_{k-1})}{\partial u_1 \cdots \partial u_{k-1}}$$

と C の偏導関数で計算できることに注意する . C からの乱数 (u_1, \dots, u_d) を生成するには ,

- (i) $U(0, 1)$ からの乱数 u_1 を発生させる .
- (ii) $U(0, 1)$ からの乱数 z_2 を発生させ , $u_2 = C_{2|1}^{-1}(z_2 | u_1)$ を計算する .
- ⋮
- (d) $U(0, 1)$ からの乱数 z_d を発生させ , $u_d = C_{d|1, \dots, d-1}^{-1}(z_d | u_1, \dots, u_{d-1})$ を計算する .

という手順を踏めばよい .

2.7 経験接合関数

$\mathbf{X}^k = (X_1^k, \dots, X_d^k)$, $k = 1, \dots, n$ を d 次元連続分布関数 F からの iid 標本とし , F_i を i 番目の周辺分布 , C を F の接合関数とする . 同時経験分布関数と周辺経験分布関数はそれぞれ

$$\mathbb{F}_n(x_1, \dots, x_d) := \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_1^k \leq x_1, \dots, X_d^k \leq x_d\}} = \frac{1}{n} \sum_{k=1}^n \prod_{i=1}^d \mathbf{1}_{\{X_i^k \leq x_i\}},$$

$$\mathbb{F}_{ni}(x_i) := \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_i^k \leq x_i\}}$$

で定義される . (2.2) 式より , $\mathbf{X}^1, \dots, \mathbf{X}^k$ に基づく経験接合関数 (empirical copula) を

$$\mathbb{C}_n(u_1, \dots, u_d) := \mathbb{F}_n(\mathbb{F}_{n1}^{-1}(u_1), \dots, \mathbb{F}_{nd}^{-1}(u_d))$$

で定義するのは自然な考え方である . この \mathbb{C}_n の確率分布は C を通してのみ F に依存することが容易にわかるので , \mathbb{C}_n の確率的挙動を調べるためには , C からの iid 標本 $\boldsymbol{\xi}^k = (\xi_1^k, \dots, \xi_d^k)$, $k = 1, \dots, n$ に基づく経験接合関数を考えれば十分である . よって以下では , $\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^n$ に基づく経験分布関数と経験接合関数をそれぞれ \mathbb{G}_n , \mathbb{C}_n で表すことにする . そして , 通常経験過程 (empirical process) \mathbb{U}_n^C と経験接合過程 (empirical copula process) \mathbb{D}_n^C を

$$\mathbb{U}_n^C := \sqrt{n}(\mathbb{G}_n - C), \quad \mathbb{D}_n^C(u) := \sqrt{n}(\mathbb{C}_n - C)$$

で定義する . 次の Bahadur-Kiefer 型漸近表現は Stute (1984) によって述べられた (その証明の細部は Tsukahara (2000) に与えられている) .

定理 2.13 F の接合関数 C が $(0, 1)^d$ 上 2 回連続微分可能で, 2 次導関数が $[0, 1]^d$ 上で連続であると仮定する. このとき, 確率 1 で (u_1, \dots, u_d) について一様に

$$\begin{aligned} \mathbb{D}_n^C(u_1, \dots, u_d) &= \mathbb{U}_n^C(u_1, \dots, u_d) - \sum_{i=1}^d C^i(u_1, \dots, u_d) \mathbb{U}_n^C(\mathbf{1}, u_i, \mathbf{1}) \\ &\quad + O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}) \end{aligned}$$

が成り立つ. ここで, $\mathbf{1}$ は 1 のみからなる適当な長さのベクトルであり, $C^i = \partial C / \partial u_i$, $i = 1, \dots, d$ である.

\mathbb{D}_n^C の弱収束はこの定理から容易に従うが, より弱い条件の下で示すことが可能である (Tsukahara (2005) 参照).

経験接合関数は Deheuvels (1979) によって初めて定義された. 彼はその漸近理論を考察し, 独立性の検定への応用についていくつかの結果を残している.

セミパラメトリック・モデル 上記の結果を応用して接合関数を用いたセミパラメトリック・モデルの推定に関する漸近理論を導くことができる. $\mathbf{X}^k = (X_1^k, \dots, X_d^k)$, $k = 1, \dots, n$ が d 次元連続分布関数 F からの iid 標本としたとき, F の接合関数としてパラメータ族 $\{C_\theta: \theta \in \Theta \subset \mathbb{R}^m\}$ を考える:

$$F(x_1, \dots, x_d) = C_\theta(F_1(x_1), \dots, F_d(x_d)).$$

このモデルで θ が興味のあるパラメータであり, この θ を周辺分布 F_i が未知のまま推定したいとする. この問題に対して Genest, Ghoudi and Rivest (1995) は, 尤度方程式において現れる F_i を経験分布関数で置き換えた推定方程式の解を推定量として提案し, その性質を述べた. この推定方程式は経験接合関数の関数となっており, 定理 2.13 を用いて推定量の漸近的性質を導くことが可能である. Tsukahara (2005) では, 彼らの推定量を一般化したものの漸近分布を厳密に求め, また経験接合関数を用いた最小距離法による推定量を提案し, さらにその一致性と漸近正規性を示し, 頑健性について調べた.

この接合関数を用いたセミパラメトリック・モデルに対する一般論が Bickel et al. (1993) に試みられているが, $\{C_\theta\}$ が Gauss 族である場合以外には θ のセミパラメトリックな意味での有効スコアや情報量の計算は難しい問題であり, 従って効率性についてはまだあまりわかっていない. Sieve 最尤法を

用いたノンパラメトリックな推定については, Chen, Fan and Tsyrennikov (2006) を見よ. Hoff (2007) では, F が必ずしも連続ではない場合において, 順位尤度に基づく推定方法が提案されている.

3 応用

3.1 多変量生存時間解析

簡単のため, 2次元の場合を考える. T_1 と T_2 を生存時刻 (survival time) とし, 我々の関心が T_1 と T_2 の間の従属構造にあるとする. 例えば, 疫学における先駆的研究である Clayton (1978) では, 父と息子のある病気による死亡時刻をそれぞれ T_1 と T_2 としてそれらの従属関係が分析されている.

いま, T_i の周辺分布関数を F_i , T_1 と T_2 の同時分布関数を F とする. F の接合関数を C で書き表すと同時生存関数 \bar{F} は

$$\bar{F}(t_1, t_2) = P(T_1 > t_1, T_2 > t_2) = \hat{C}(\bar{F}_1(t_1), \bar{F}_2(t_2))$$

と書ける. ここで, $\bar{F}_i = 1 - F_i$ は周辺生存関数であり, 生存接合関数 (survival copula) \hat{C} は

$$\hat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$$

で定義されている.

Oakes (1989) によるクロス比は

$$\theta^*(t_1, t_2) := \frac{\lambda_1(t_1 | T_2 = t_2)}{\lambda_1(t_1 | T_2 > t_2)}$$

で定義される. ここで, $\lambda_1(t_1 | T_2 = t_2)$ は $T_2 = t_2$ が与えられたときの T_1 の条件付分布に関するハザード関数であり, $\lambda_1(t_1 | T_2 > t_2)$ は $T_2 > t_2$ が与えられたときの T_1 の条件付分布に関するハザード関数である. $\bar{F}^{(1,2)}(t_1, t_2) = (\partial^2 / \partial t_1 \partial t_2) \bar{F}(t_1, t_2)$, $\bar{F}^{(i)}(t_1, t_2) = (\partial / \partial t_i) \bar{F}(t_1, t_2)$ ($i = 1, 2$) とおくと, 簡単な計算により,

$$\theta^*(t_1, t_2) = \frac{\bar{F}(t_1, t_2) \bar{F}^{(1,2)}(t_1, t_2)}{\bar{F}^{(1)}(t_1, t_2) \bar{F}^{(2)}(t_1, t_2)}$$

となることがわかる. Oakes (1989) は微分可能性に関する適当な条件の下で, $\theta^*(t_1, t_2) = \theta(\bar{F}(t_1, t_2))$ がある関数 θ に対して成り立つ (すなわち, ク

ロス比が同時生存関数を通してのみ時刻 (t_1, t_2) に依存することと、 T_1 と T_2 の生存接合関数がアルキメデス型であることの同等性を示した。この場合、関数 θ は $\theta(v) = -v\phi''(v)/\phi'(v)$ で与えられる。もし $\theta^*(t) = c$ 、すなわちクロス比が一定ならば、満たされるべき関数関係から $c = \theta + 1$ とした Clayton 族が導かれる。

いま、 n 個の独立で同一分布をもつ生存時刻の組 (T_{1k}, T_{2k}) に対して、打切り時刻の組 (C_{1k}, C_{2k}) があり、実際に観測されるのは $(X_{1k}, X_{2k}, \delta_{1k}, \delta_{2k})$ であるとする。ただし、 $i = 1, 2$ と $k = 1, \dots, n$ に対して、 $X_{ik} = T_{ik} \wedge C_{ik}$ 、 $\delta_{ik} = 1_{\{T_{ik} \leq C_{ik}\}}$ である。ここで 2.7 節と同様に、 (T_{1k}, T_{2k}) の同時分布に対する生存接合関数がパラメータ族 $\{C_\theta: \theta \in \Theta \subset \mathbb{R}^m\}$ に属しているというモデルを考えることができる。このようなモデルの望ましさは Cox and Oakes (1984) に述べられている。このモデルで、周辺分布が未知のまま $(X_{1k}, X_{2k}, \delta_{1k}, \delta_{2k})$ 、 $k = 1, \dots, n$ に基づいて θ を推定する問題は、 $\{C_\theta\}$ が Clayton 族の場合に Clayton (1978) によって初めて考えられた。Oakes (1982) はやはり Clayton 族の場合に、いくつかのセミパラメトリックな推定量を提案し、それらの性質を調べた。Shih and Louis (1995) は $\{C_\theta\}$ が一般のパラメータ族の場合を考察したが、彼らの推定量は Genest, Ghoudi and Rivest (1995) による推定量の打切りがある場合への拡張となっている。Tsukahara (2005) の最小距離推定量も自然にこの場合へ拡張でき、理論的にはどちらの推定量も 3 次元以上の場合への拡張は容易である。その他、多変量生存時間解析における接合関数の応用については、Hougaard (2000) も参照されたい。

3.2 ファイナンス

3.2.1 相互依存性のあるデフォルト事象のモデリング

T_i を証券 i のデフォルト時刻を表す正の確率変数とし ($i = 1, \dots, d$)、 T_i の周辺分布関数を F_i 、 T_1, \dots, T_d の同時分布関数を F とする。CDO (collateralized debt obligation) に代表される信用デリバティブの価格付け、そしてこれらのリスク管理を行うためには、デフォルト時刻 T_1, \dots, T_d のモデリングを実務的に扱いやすい形で行うことが不可欠であり、この目的のために接合関数を用いることができる。

生存時間解析アプローチ このアプローチでは、分布関数 F の接合関数 C を既知の接合関数族を用いて直接モデル化する。 F の同時生存関数 \bar{F} は

$$\bar{F}(t_1, \dots, t_d) = P(T_1 > t_1, \dots, T_d > t_d) = \hat{C}(\bar{F}_1(t_1), \dots, \bar{F}_d(t_d))$$

と書ける．ここで， $\bar{F}_i = 1 - F_i$ は証券 i の周辺生存関数であり，生存接合関数 \hat{C} は

$$\begin{aligned} \hat{C}(u_1, \dots, u_d) &= \sum_{i=1}^d u_i - d + 1 + \sum_{i < j} C(\mathbf{1}, 1 - u_i, \mathbf{1}, 1 - u_j, \mathbf{1}) \\ &- \sum_{i < j < k} C(\mathbf{1}, 1 - u_i, \mathbf{1}, 1 - u_j, \mathbf{1}, 1 - u_k, \mathbf{1}) + \dots + (-1)^d C(1 - u_1, \dots, 1 - u_d) \end{aligned}$$

で定義されている（ $\mathbf{1}$ は成分がすべて 1 であるベクトルである）．

Li (2000) は C として Gauss 接合関数を用いることを提案し，主に 2 次元の場合に接合関数のパラメータをどう推定すればよいかを示し，クレジット・デフォルト・スワップや first-to-default 契約などの信用デリバティブを評価する際の接合関数の用い方を示した．このアプローチの拡張については，例えば Hull and White (2006) を参照せよ．

閾値アプローチ Schönbucher and Schubert (2001) は，デフォルト時刻間の従属性を直接モデル化するのではなく，デフォルト時刻を定義するときの閾値に接合関数を用いる方法を導入した．

各 $i = 1, \dots, d$ に対して，債務者 i のデフォルト時刻 T_i を

$$T_i = \inf\{t: \gamma_i(t) \leq U_i\}$$

によって定義する．ただし， U_i は一様分布に従う確率変数であり，デフォルト・カウントダウン過程 $\gamma_i(t)$ はある正の過程（擬生起度過程） $\lambda_i(t)$ を用いて

$$\gamma_i(t) = \exp\left[-\int_0^t \lambda_i(u) du\right]$$

と定義される．デフォルト時刻間の従属性は (U_1, \dots, U_d) の同時分布関数がある接合関数 C であると仮定することにより導入される．生存確率は C とその導関数に依存した形で書くことができるが，Schönbucher and Schubert (2001) はどの接合関数を選べばよいかについては述べていない．この設定は明らかにモンテカルロ・シミュレーションに適したものとなっている．

上記 2 つのアプローチ以外にも，やはり接合関数をうまく用いた潜在変数アプローチと呼ばれるものがあるが，それについては Frey, McNeil and Nyfeler (2001) を参照されたい．信用リスクモデリングにおける接合関数の応用については，McNeil, Frey and Embrechts (2005) や Schönbucher (2003)，室町 (2007) を推奨する．

3.2.2 リスク管理

リスク管理の分野では、様々な資産の価格、あるいは収益率間の相互依存性が非常に重要な分析対象である。古典的な平均・分散アプローチでは、暗に多変量正規分布が仮定されてきたため、資産価格（収益率）間の従属性は相関係数のみで記述されることになる。

ところが、Embrechts, McNeil and Straumann (2002) は、いくつかの反例やよくある誤信を接合関数を使って例示することにより、Pearson の相関係数は非楕円型分布の世界では誤った従属性の尺度となり得ることを指摘した。これらは、ファイナンスの分野で接合関数が注目されるきっかけを作った論文であるといつてよいだろう。彼らはモンテカルロ・シミュレーションを通じて、互いに関連し合うリスク間の従属関係を理解しモデル化するための新たな道具として接合関数を用いることを提案した。

Embrechts, Höing and Juri (2003), Embrechts and Puccetti (2006) では、接合関数を用いて VaR に対する上限・下限を求めている。また、条件付分布関数に付随する接合関数を用いてファイナンスや経済の時系列を分析しようというアプローチが、Granger, Teräsvirta and Patton (2006) や Patton (2006, 2009) に述べられている。定量的リスク管理全般の優れた教科書としては、McNeil, Frey and Embrechts (2006) があり、2.5 節で説明した裾従属性との関連など、接合関数の応用も包括的に扱っている。接合関数を利用した多変量オプションの価格付けについては、Cherubini, Luciano and Vecchiato (2004) に詳しい。Malevergne and Sornette (2006) も接合関数にかなり重点をおき、極端なリスクのモデリングについて解説している。

3.3 保険

ここでは、保険の分野で接合関数が用いられた初期の応用例を 2 つ簡潔に挙げるに止める。

例 3.1 (死亡率が互いに関連する年金評価) 共同年金 (joint annuity) や last-survivor 年金では、幾人かの個人から成る団体に対して最後の 1 人が死亡するまで、定期的に一定の支払が行われる。終身年金の評価は年金受給者の生存確率に依存するため、上のような年金を正確に評価するためには同時生存確率を推定しなければならない。

この問題に対して、Free, Carriere and Valdez (1996) は Gompertz 分布

$$\bar{F}(t) = \exp \left[e^{-m/\sigma} (1 - e^{t/\sigma}) \right],$$

を周辺分布とし，Frank 族の接合関数を用いてパラメトリック・モデルとし，そのパラメータを最尤法によって推定している．

例 3.2 (損害保険における損失と費用の従属関係) 支払請求に関連する次の 2 つの確率変数を導入する： X_1 は補償金額（損失）， X_2 は allocated loss adjustment expense と呼ばれるものである．保険証券の支払上限と控除条項 (deductibles) の存在から，データは右からの打切り (right-censoring) と左側の切断 (left-truncation) を被る可能性がある．

数学的扱い易さとこの分野の慣習から Klugman and Parsa (1999) は Frank 族を選択し，周辺分布は分布関数が $F(x) = (x^\gamma / (x^\gamma + \theta^\gamma))^\tau$ で与えられる逆 Burr 分布と仮定した．彼らは最尤法によってパラメータを推定し，当てはめたモデルを使って再保険料を計算し，損失金額からそれに関わる費用を予測する回帰関数を推定している．

Frees and Valdez (1998) は，接合関数の保険理論への応用を念頭において書かれた，優れた展望論文である．また最近出版された Denuit et al. (2005) は，接合関数の基礎や従属性，リスク尺度，確率順序について，保険数理の観点から非常に読み易くまとめられた教科書である．

3.4 確率論への応用

3.4.1 多変量極値理論

F を d 次元分布関数とし， $\mathbf{X}_i = (X_{i1}, \dots, X_{id})'$, $i = 1, \dots, n$ を F からの iid 標本とする．このとき，成分毎の最大値 $M_{jn} = \max\{X_{1j}, \dots, X_{nj}\}$ を並べたベクトル $M_n = (M_{1n}, \dots, M_{dn})'$ を基準化したものが非退化分布に分布収束するならば，よく知られた Gnedenko-Fisher-Tippett の定理により，極限分布は

$$C \left(H_\xi \left(\frac{x_1 - \mu_1}{\sigma_1} \right), \dots, H_\xi \left(\frac{x_d - \mu_d}{\sigma_d} \right) \right)$$

の形にならなければならない．ここで， $H_\xi(x)$ は一般化極値分布の分布関数である（渋谷・高橋 (2007) 参照）． C は接合関数であるが，この極限に現れる接合関数のクラスを極値接合関数という．このクラスは実は，すべての $t > 0$ に対して $C(u_1^t, \dots, u_d^t) = C^t(u_1, \dots, u_d)$ が成り立つという性質で特徴付けられる．この代表的な例は 2.2 節で紹介した Gumbel-Hougaard 族である．

この極値接合関数は極値同士の従属性に関する情報をすべてもっていると考えることができるが、この極限に現れる C は、 F の周辺分布とは無関係に F の接合関数のみから定まることも知られており、そのことから接合関数の最大吸引域という概念も導入されている。その他、Pickands による極値接合関数の表現定理など多変量極値理論と極値接合関数については、Galambos (1987) や Joe (1997) の第 6 章が詳しい。また、本論文集所収の渋谷・高橋 (2007) も参照されたい。日本人による 2 変量の場合の先駆的研究としては Sibuya (1960) がよく知られている。極値理論のリスク管理論への応用については、McNeil, Frey and Embrechts (2005) を参照せよ。

3.4.2 マルコフ過程への応用

2 次元接合関数 $C(u_1, u_2)$ に対して、 $D_1C := \partial C / \partial u_1$, $D_2C := \partial C / \partial u_2$ とおく。そして、2 つの 2 次元接合関数 C_1 と C_2 の積 $C_1 * C_2$ を

$$(C_1 * C_2)(u_1, u_2) := \int_0^1 D_2C_1(u_1, t) D_1C_2(t, u_2) dt$$

で定義する。この積について次のことが成り立つ。

- (i) $C_1 * C_2$ は接合関数となる。
- (ii) $\Pi(u_1, u_2) = u_1 u_2$ は任意の接合関数 C に対して、 $C * \Pi = \Pi * C = \Pi$ を満たす。
- (iii) $M(u_1, u_2) = \min(u_1, u_2)$ は任意の接合関数 C に対して、 $C * M = M * C = C$ を満たす。
- (iv) $*$ は可換ではないが ($W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$ に対して、 $W * C \neq C * W$)、結合律は満たす。

いま、マルコフ過程 (X_t) に対して、 $P(x, s, y, t) = P(X_t \leq y | X_s = x)$ とおくと、よく知られた Chapman-Kolmogorov の方程式は

$$P(x, s, y, t) = \int_{-\infty}^{\infty} P(z, u, y, t) \frac{\partial P(x, s, z, u)}{\partial z} dz, \quad s < u < t \quad (3.1)$$

と書かれる。これは (X_t) がマルコフ過程であるための必要条件である。

定理 3.3 (Darsow, Nguyen and Olsen (1992)) C_{st} を X_s と X_t の同時分布に対する接合関数とすると、次の 2 つの条件は同等である：

- (i) すべての $s < u < t$ とほとんどすべての $x, y \in \mathbb{R}$ に対して, $\{P(x, s, y, t)\}$ は Chapman-Kolmogorov の方程式 (3.1) を満たす .
- (ii) すべての $s < u < t$ に対して, $C_{st} = C_{su} * C_{ut}$ が成り立つ .

例 3.4 (X_t) を標準ブラウン運動とすると,

$$P(x, s, y, t) = \Phi\left(\frac{y-x}{\sqrt{t-s}}\right), \quad s < t$$

であり, 対応する接合関数は

$$C_{st}(u_1, u_2) = \int_0^{u_1} \Phi\left(\frac{\sqrt{t}\Phi^{-1}(u_2) - \sqrt{s}\Phi^{-1}(w)}{\sqrt{t-s}}\right) dw$$

で与えられることが若干の計算でわかる . ここから, 適当な周辺分布を選んでこの接合関数と組み合わせることによって, ブラウン運動のもつ順序的従属構造を保持しつつ, 非正規の周辺分布をもつようなマルコフ過程を作ることができる . さらに, 他のよく知られたマルコフ過程から接合関数を取り出しても同様のことが可能である .

4 おわりに

‘copula’ で Google 検索を行えばすぐわかるように, 接合関数の応用は多岐に渡っており, すべてを網羅することはできない . ファイナンスや保険, 生存解析については, これまですでにいくつか紹介してきたが, 近年, 接合関数の応用を扱った多くの書物が出版されているので適宜それらを参照されたい . 本稿で論じなかった応用分野の一例として, Salvadori et al. (2007) は自然災害の研究における接合関数の応用を解説している .

接合関数に関する統計理論を包括的に扱った書物はないが, Genest and Favre (2007) は統計家にとって有益な展望論文である . 今後は, 適合度・独立性の検定やパラメータ推定などの統計的推測理論ばかりではなく, 探索的データ解析の道具としてもより発展させることが必要である .

接合関数に関して過度に批判的だと思われる Mikosch (2006) では, 多変量分布を理解するために接合関数を用いる利点はなく, モデルにおける接合関数の選択は便宜的であり, 時系列的従属性のモデリングには無益であるなどという接合関数不要論が展開されている . 確かに接合関数を用いることでいわゆる次元の呪いが解消される訳ではないが, 多変量モデリング

の一手法として、また従属性の様々な側面を分析する道具として応用上有効であることは間違いない事実であると筆者は考える。また、上記論文で挙げられた接合関数アプローチの欠点は、これからの研究課題であると前向きに考えることもできよう。

接合関数の応用範囲が今後広がるためには、高次元における確率変数間の従属性を幅広く表現できるような接合関数のパラメータ族が見出されることが望ましい。さらに、時空間での従属構造の分析に接合関数がどう用いられうるかというのも興味深い課題である。

謝辞 本研究は科研費(18500217)の助成を受けたものである。本稿は、匿名のレフェリー及び編集者の方々の意見を取り入れることにより格段に読み易さが上昇した。これらの方々が査読の労を執ってくださったことに対して深くお礼を申し上げたい。さらに、原稿の誤りをいくつか指摘してくれた東京大学大学院経済学研究科院生の三崎広海君にも感謝する。

A 増補・HP版への補遺

増補・HP版の出版に際して、接合関数に関する最新の研究結果のなかでもとりわけ重要と思われる2つの話題について解説を試みる。ここで扱えなかった話題については、最近の学会報告論文集である Jaworski et al. (2010)などを参照されたい。

A.1 ヴァイン接合関数

まず、Aas et al. (2009)に従って、ヴァイン接合関数(vine copula)という考え方を導入しよう。これは、2次元接合関数のみを用いて任意の多次元分布を構成しようというアプローチであり、对接合関数構成とも呼ばれる。まず、確率変数 X_1, \dots, X_d の同時密度関数を

$$\begin{aligned} f(x_1, \dots, x_d) \\ = f_1(x_1) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|12}(x_3|x_1, x_2) \cdots f_{d|12\dots d-1}(x_d|x_1, \dots, x_{d-1}) \end{aligned} \quad (\text{A.1})$$

と分解する。2次元分布関数 F_{12} の Sklar 分解(定理 2.2)を

$$F_{12}(x_1, x_2) = C_{12}(F_1(x_1), F_2(x_2))$$

とし，両辺を x_1, x_2 で微分すると，密度関数は

$$f_{12}(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2)$$

である．ただし， $c_{12}(u, v) = \partial^2 C_{12}(u, v) / \partial u \partial v$ は接合密度関数である．よつて， $X_1 = x_1$ を所与としたときの X_2 の条件付き密度関数は

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2)$$

となる．

同様の議論を， $f_{23|1}$ に対して行くと，

$$f_{23|1}(x_2, x_3|x_1) = c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1}(x_3|x_1)$$

ここで， $f_{23|1}(x_2, x_3|x_1) / f_{2|1}(x_2|x_1) = f_{3|12}(x_3|x_1, x_2)$ より，

$$\begin{aligned} f_{3|12}(x_3|x_1, x_2) &= c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \cdot f_{3|1}(x_3|x_1) \\ &= c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot f_3(x_3) \end{aligned}$$

を得る．一般に， $\mathbf{v} = (v_1, \dots, v_k) = (x_{i_1}, \dots, x_{i_k})$ と $x_m \notin \{x_{i_1}, \dots, x_{i_k}\}$ に対して， $\mathbf{v}_{-j} = (v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_k)$ と書くと

$$f(x|\mathbf{v}) = c_{mi_j|i_1, \dots, i_{j-1}, i_{j+1}, \dots, i_k}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})) \cdot f(x|\mathbf{v}_{-j})$$

が成り立つから，これを繰り返し用いると，同時密度関数 $f(x_1, \dots, x_d)$ は，適当な条件付き分布関数を引数とする条件付き 2 次元接合関数と 1 次元密度関数の積として書けることがわかる．もちろん，変数のラベリングも含めてこのような分解の仕方はたくさんあり，うまく整理して考えなくてはならない．そこで，グラフィカルモデリングの分野で Bedford and Cooke (2002) により導入されたヴァイン (vine) という概念が援用される．以下で用いるグラフ理論の用語・定義については，宮川 (1997) を参照せよ．

定義 A.1 d 個の要素 $\{1, 2, \dots, d\}$ 上のヴァイン (vine) とは，木 (tree) T_i ， $i = 1, \dots, d-1$ の集合であり，次の入れ子条件を満たすものである．

- (i) T_1 は $\{1, 2, \dots, d\}$ を頂点集合， E_1 を辺集合とする連結な木である．
- (ii) $i = 2, \dots, d-1$ に対して， T_i は E_{i-1} を頂点集合， E_i を辺集合とする木である．

ヴァイン $\mathcal{V} = \{T_1, \dots, T_{d-1}\}$ が正則 (regular) であるとは, $i = 2, \dots, d-1$ に対して, $\{a, b\} \in E_i$ ならば, $a \triangle b$ の要素の個数が 2 となることである (ここで, \triangle は対称差を表す).

よく取り上げられるヴァインの例を 2 つ挙げよう.

例 A.2 (D ヴァイン) まず D ヴァインに基づく対接合関数構成による同時密度関数を $d = 3, 4, 5$ について書き下してみると,

$$f(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}(F_2(x_2), F_3(x_3)) \\ \cdot c_{13|2}(F(x_1|x_2), F(x_3|x_2))$$

$$f(x_1, x_2, x_3, x_4) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \\ \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot c_{34}(F_3(x_3), F_4(x_4)) \\ \cdot c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot c_{24|3}(F(x_2|x_3), F(x_4|x_3)) \\ \cdot c_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3))$$

$$f(x_1, x_2, x_3, x_4, x_5) \\ = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \cdot f_5(x_5) \\ \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot c_{34}(F_3(x_3), F_4(x_4)) \cdot c_{45}(F_4(x_4), F_5(x_5)) \\ \cdot c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot c_{24|3}(F(x_2|x_3), F(x_4|x_3)) \cdot c_{35|4}(F(x_3|x_4), F(x_5|x_4)) \\ \cdot c_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3)) \cdot c_{25|34}(F(x_2|x_3, x_4), F(x_5|x_3, x_4)) \\ \cdot c_{15|234}(F(x_1|x_2, x_3, x_4), F(x_5|x_2, x_3, x_4))$$

となる. 一般の d に対する D ヴァインに基づく同時密度関数は,

$$\prod_{k=1}^d f_k(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i, i+j|i+1, \dots, i+j-1}(F(x_i|x_{i+1}, \dots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \dots, x_{i+j-1}))$$

と書ける.

D ヴァインは, どの木 T_i においても, 各頂点に接続する辺の本数が 2 を超えないという定義である. このヴァインから構成される分布は各変数を同等に扱っているものとも解釈できよう. $d = 5$ についてグラフ表示は図 4 のようになる.

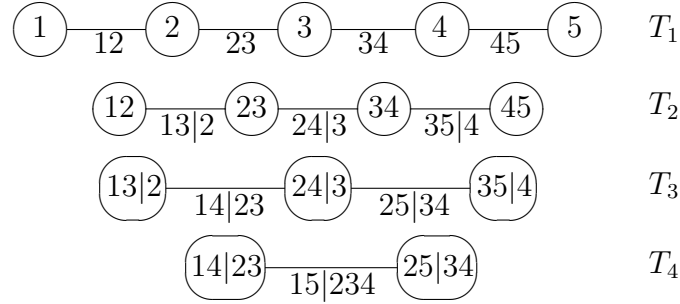


図 4: D ヴァインのグラフ表示 ($d = 5$)

例 A.3 (正準ヴァイン) 上の例と同様に, 正準ヴァインに基づく対接合関数構成による同時密度関数を $d = 3, 4, 5$ について書き下してみると,

$$f(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \\ \cdot c_{23|1}(F(x_2|x_3), F(x_3|x_1))$$

$$f(x_1, x_2, x_3, x_4) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \\ \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot c_{14}(F_1(x_1), F_4(x_4)) \\ \cdot c_{23|1}(F(x_2|x_1), F(x_3|x_1)) \cdot c_{24|1}(F(x_2|x_1), F(x_4|x_1)) \\ \cdot c_{34|12}(F(x_3|x_1, x_2), F(x_4|x_1, x_2))$$

$$f(x_1, x_2, x_3, x_4, x_5) \\ = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \cdot f_5(x_5) \\ \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot c_{14}(F_1(x_1), F_4(x_4)) \cdot c_{15}(F_1(x_1), F_5(x_5)) \\ \cdot c_{23|1}(F(x_2|x_1), F(x_3|x_1)) \cdot c_{24|1}(F(x_2|x_1), F(x_4|x_1)) \cdot c_{25|1}(F(x_2|x_1), F(x_5|x_1)) \\ \cdot c_{34|12}(F(x_3|x_1, x_2), F(x_4|x_1, x_2)) \cdot c_{35|12}(F(x_3|x_1, x_2), F(x_5|x_1, x_2)) \\ \cdot c_{45|123}(F(x_4|x_1, x_2, x_3), F(x_5|x_1, x_2, x_3))$$

となる. 一般の d に対する正準ヴァインに基づく同時密度関数は,

$$\prod_{k=1}^d f_k(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j,j+i|1, \dots, j-1}(F(x_j|x_1, \dots, x_{j-1}), F(x_{j+i}|x_1, \dots, x_{j-1}))$$

である.

定義により，正準ヴァインでは各木 T_i において， $d-i$ 本の辺で結ばれている頂点が唯一つ存在する．上の例では変数 1 からその重要度の順に変数が並んでいるものと考えられる． $d=5$ について対応するヴァインのグラフ表示は図 5 のようになる．

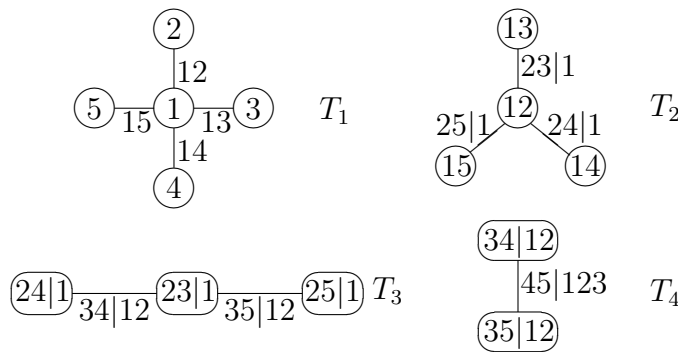


図 5: 正準ヴァインのグラフ表示 ($d=5$)

高次元 ($d \geq 3$) では接合関数の族として使いやすいものは少ない一方で，2.2 節で述べたように，2 次元接合関数については多くの具体的な族が知られている．そのため，上記のような 2 次元接合関数と条件付き分布関数のみを用いて一般の多次元分布を構成するアプローチは非常に有効であると考えられる．しかし，このモデリングでは，条件付き接合関数が条件付け変数に依存しないという形になっており，これがどのくらい一般性を失うことになるのか，その影響についてはわかっていない．また，このクラスの同時分布で表現される従属性についての理解もまだ十分に進んでいるとは言えないのが現状である．

ヴァインの平易な解説と多次元従属性モデリングについては Kurowicka and Cooke (2006) を参照せよ．Aas et al. (2009) は，多次元分布の对接合関数構成に関して，その分布からの乱数発生法やモデル選択，統計的推測の方法について様々な提案を行っている．ヴァイン接合関数全般について，最新の成果をまとめた有益な書物としては Kurowicka and Joe (2011) がある．

A.2 多次元アルキメデス型接合関数

注意 2.11 では，次元 d が 3 以上の場合におけるアルキメデス型接合関数について触れた．ここでは記号を少し変えて，(2.5) を

$$C(u_1, \dots, u_d) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)), \quad (u_1, \dots, u_d) \in [0, 1]^d \quad (\text{A.2})$$

と書いておく．さらに次の定義を置く．

定義 A.4 ψ がアルキメデス生成素であるとは， ψ が単調減少かつ連続であり， $\psi(0) = 1$ ， $\lim_{x \rightarrow \infty} \psi(x) = 0$ を満たし，さらに $[0, \inf\{x: \psi(x) = 0\})$ 上で狭義単調減少となることである．

まず問題になるのは， d を 1 つ固定したとき， C が接合関数となるために ψ が満たすべき条件であるが（すべての d について上の C が接合関数となるための条件は ψ が完全単調の場合である），最近 McNeil and Nešlehová (2009) により必要十分条件が得られた．まず， d 単調性を定義する．

定義 A.5 実数値関数 f が (a, b) 上で d 単調であるとは $(a, b \in \bar{\mathbb{R}}, d \geq 2)$ ， f が $d-2$ 回微分可能であり，導関数がすべての $x \in (a, b)$ に対して

$$(-1)^k f^{(k)}(x) \geq 0, \quad k = 0, 1, \dots, d-2$$

を満たし，かつ $(-1)^{d-2} f^{(d-2)}$ が単調減少な凸関数となることである．

この定義を用いれば，(A.2) の C が接合関数となるための必要十分条件は次のように述べられる．

定理 A.6 ψ をアルキメデス生成素とする．このとき，(A.2) で定義される関数 C が d 次元接合関数であるためには， ψ が d 単調であることが必要十分である．

アルキメデス生成素 ψ が完全単調であるためには， ψ が非負確率変数のラプラス変換となることが必要十分である（Widder (1946) 参照）．これに類似の結果を述べるためには，次の定義が必要である．

定義 A.7 (Williamson(1956)) X を非負確率変数， F をその分布関数とする． F の Williamson d 変換とは，

$$\mathcal{W}_d F(x) = \int_{(x, \infty)} \left(1 - \frac{x}{t}\right)^{d-1} dF(t) = \begin{cases} \mathbb{E} \left(1 - \frac{x}{X}\right)_+^{d-1} & x > 0 \text{ のとき} \\ 1 - F(0) & x = 0 \text{ のとき} \end{cases}$$

で定義される $[0, \infty)$ 上の関数 $\mathcal{W}_d F$ である．

定理 A.8 アルキメデス生成素 ψ が d 単調であるためには， ψ が，0 に正の確率をもたない非負確率変数の分布関数 F の Williamson d 変換であることが必要十分である．

この証明については，Williamson (1956) と McNeil and Nešlehová (2009) を参照せよ．後者の論文では，このクラスの接合関数と ℓ_1 ノルム対称分布の関係についても詳述している (ℓ_1 ノルム対称分布については Fang et al. (1990) も参照せよ)．Genest et al. (2011) は，さらに進んで多変量アルキメデス型接合関数モデルに関する統計的推測と漸近理論を論じており，いくつかの未解決の課題も提示されている．

B 数学付録

定理 2.2 の証明 $X = (X_1, \dots, X_d)$ が分布関数 F をもつとする．また， V_1, \dots, V_d を独立，かつ同一分布 $U(0, 1)$ をもつ確率変数とし， X とは独立であるとする． $F_i(x-) := \lim_{t \uparrow x} F_i(t)$ とするとき，

$$U_i := (1 - V_i)F_i(X_i-) + V_iF_i(X_i), \quad i = 1, \dots, d$$

で定義される確率変数 U_1, \dots, U_d の分布関数を C とおく． U_i が一様分布 $U(0, 1)$ に従うことは容易に確認できる．さらに， $U_i \leq F_i(x_i) \Leftrightarrow F_i^{-1}(U_i) \leq x_i$ であり，確率 1 で $F_i^{-1}(U_i) = X_i$ となることを用いれば，

$$P(U_1 \leq F_1(x_1), \dots, U_d \leq F_d(x_d)) = P(X_1 \leq x_1, \dots, X_d \leq x_d)$$

が得られる．よって，(2.1) が成り立つ． ■

この証明は Moore and Spruill (1975) によるものであるが，与えられた分布をもつ確率変数がある確率空間上に構成できるという Kolmogorov の定理を暗に用いているため，他の直接的な証明（例えば Nelsen (2006) 参照）よりはるかに簡単になっている．

定理 2.3 の証明 同時分布関数 C をもつ確率ベクトル (U_1, \dots, U_d) を考えると，明らかにすべての i に対して

$$C(u_1, \dots, u_d) = P(\{U_1 \leq u_1\} \cap \dots \cap \{U_d \leq u_d\}) \leq P(U_i \leq u_i) = u_i$$

が成り立つから，(2.3) の 2 番目の不等号が示される．次に，確率の劣加法性より

$$\begin{aligned} C(u_1, \dots, u_d) &= P(\{U_1 \leq u_1\} \cap \dots \cap \{U_d \leq u_d\}) \\ &= 1 - P(\{U_1 > u_1\} \cup \dots \cup \{U_d > u_d\}) \\ &\geq 1 - \sum_{i=1}^d P(U_i > u_i) = 1 - \sum_{i=1}^d (1 - u_i) = \sum_{i=1}^d u_i - d + 1 \end{aligned}$$

$C(u_1, \dots, u_d) \geq 0$ は明らかだから，これで (2.3) の下限が示された． ■

(2.4) の証明 分布関数 F をもつ確率ベクトル (X_1, X_2) と，それとは独立に分布関数 F をもつ $(\tilde{X}_1, \tilde{X}_2)$ を考えると，

$$2 \operatorname{cov}(X_1, X_2) = E[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2)]$$

と書ける．ここで，任意の $a \in \mathbb{R}$ と $b \in \mathbb{R}$ に対して，

$$a - b = \int_{-\infty}^{\infty} (\mathbf{1}_{\{b \leq x\}} - \mathbf{1}_{\{a \leq x\}}) dx$$

という恒等式を $(X_1 - \tilde{X}_1)$ と $(X_2 - \tilde{X}_2)$ に適用すると，Fubini の定理から

$$\begin{aligned} & 2 \operatorname{cov}(X_1, X_2) \\ &= E \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\mathbf{1}_{\{\tilde{X}_1 \leq x_1\}} - \mathbf{1}_{\{X_1 \leq x_1\}}) (\mathbf{1}_{\{\tilde{X}_2 \leq x_2\}} - \mathbf{1}_{\{X_2 \leq x_2\}}) dx_1 dx_2 \right] \\ &= 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [P(X_1 \leq x_1, X_2 \leq x_2) - P(X_1 \leq x_1)P(X_2 \leq x_2)] dx_1 dx_2 \end{aligned}$$

を得る（この証明は Lehmann (1966) による）． ■

定理 2.6 の証明 まず定理の後半から示す．定理 2.3 から，

$$\max\{F_1(x_1) + F_2(x_2) - 1, 0\} \leq F(x_1, x_2) \leq \min\{F_1(x_1), F_2(x_2)\}$$

が成り立つが， F_1 と F_2 が固定されていれば， $F(x_1, x_2)$ の接合関数が $C(u_1, u_2) = \min\{u_1, u_2\}$ であるとき，すなわち X_1 と X_2 が共単調の場合に，(2.4) の被積分関数は各点で最大化される．同様に，被積分関数が最小化されるのは， X_1 と X_2 が反単調である場合である．

次に，明らかに $\bar{\rho} \geq 0$ であるが， $\bar{\rho} = 0$ ではあり得ない．なぜならば，すべての x_1, x_2 に対して， $\min\{F_1(x_1), F_2(x_2)\} = F_1(x_1)F_2(x_2)$ が成り立つと， F_1 か F_2 が 1 点に退化した分布となってしまう，それは分散がゼロではないという仮定に反するからである．同様の議論により， $\underline{\rho} < 0$ も得られる．さらに，混合分布 $\lambda W(F_1, F_2) + (1 - \lambda)M(F_1, F_2)$ ， $0 \leq \lambda \leq 1$ を考えることにより， $\underline{\rho}$ と $\bar{\rho}$ の間の任意の値 ρ を相関係数とする同時分布を作ることができる． ■

定理 2.7 の証明 F_i は連続だから $F_i(F_i^{-1}(u_i)) = u_i$ が成り立つ . このことから , X_1, \dots, X_d が独立ならば , $C = \Pi$ となることは明らかである . 逆に $C = \Pi$ とすると , F の連続性より C は (2.2) で一意的に与えられるから , $F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) = u_1 \cdots u_d$ がすべての $(u_1, \dots, u_d) \in [0, 1]^d$ について成り立たなければならない . $x_i = F_i^{-1}(u_i)$ とおくと , $\{(x_1, \dots, x_d) : (u_1, \dots, u_d) \in [0, 1]^d\}$ は F の台となる . $F_i(x_i) = u_i$ だから $F(x_1, \dots, x_d) = F_1(x_1) \cdots F_d(x_d)$ が容易に従う . ■

定理 2.8 の証明 定義を素直に確認していけばよい . ■

定理 2.9 の証明 $u_i \geq v_i, i = 1, \dots, d$ と w_1, \dots, w_d に対して ,

$$\begin{aligned} & C(w_1, \dots, u_i, \dots, w_d) - C(w_1, \dots, v_i, \dots, w_d) \\ &= P(U_1 \leq w_1, \dots, v_i < U_i \leq u_i, \dots, U_d \leq w_d) \leq P(v_i < U_i \leq u_i) = u_i - v_i \end{aligned}$$

ここで , (U_1, \dots, U_d) は分布関数 C をもつ確率ベクトルである . よって , 任意の $u_i, v_i, w_1, \dots, w_d$ に対して , $|C(w_1, \dots, u_i, \dots, w_d) - C(w_1, \dots, v_i, \dots, w_d)| \leq |u_i - v_i|$ が成り立つ . これを繰り返し用いると ,

$$\begin{aligned} & |C(u_1, \dots, u_d) - C(v_1, \dots, v_d)| \\ &\leq |C(u_1, \dots, u_d) - C(u_1, v_2, \dots, v_d)| + |C(u_1, v_2, \dots, v_d) - C(v_1, \dots, v_d)| \\ &\leq |C(u_1, \dots, u_d) - C(u_1, v_2, \dots, v_d)| + |u_1 - v_1| \\ &\leq |C(u_1, \dots, u_d) - C(u_1, u_2, \dots, v_d)| + |C(u_1, u_2, \dots, v_d) - C(u_1, v_2, \dots, v_d)| \\ &\quad + |u_1 - v_1| \\ &\leq |C(u_1, \dots, u_d) - C(u_1, u_2, \dots, v_d)| + |u_2 - v_2| + |u_1 - v_1| \\ &\leq \dots \leq \sum_{i=1}^d |u_i - v_i|. \quad \blacksquare \end{aligned}$$

参考文献

- [1] Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009). Pair-copula constructions of multiple dependence, *Insurance: Mathematics and Economics*, **44**, 182–198.
- [2] Bedford, T. and Cooke, R. M. (2002). Vines — A new graphical model for dependent random variables, *Ann. Statist.*, **30**, 1031–1068.

- [3] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore and London.
- [4] Blomqvist, N. (1950). On a measure of dependence between two random variables, *Ann. Math. Statist.*, **21**, 593–600.
- [5] Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models, *J. Amer. Statist. Assoc.*, **101**, 1228–1240.
- [6] Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*, John Wiley & Sons, Chichester.
- [7] Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, **65**, 141–151.
- [8] Cook, R. D. and Johnson, M. E. (1981). A family of distributions for modelling non-elliptically symmetric multivariate data, *J. Roy. Statist. Soc. B*, **43**, 210–218.
- [9] Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*, Chapman and Hall, London.
- [10] Darsow, W. F., Nguyen, B. and Olsen, E. T. (1992). Copulas and Markov processes, *Ill. J. Math.*, **36**, 600–642.
- [11] Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés, Un test non paramétrique d’indépendance, *Bulletin de la classe des sciences, Académie Royale de Belgique, 5e série*, **65**, 274–292.
- [12] Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*, John Wiley & Sons, Chichester.
- [13] Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R. and Vyncke, D. (2002a). The concept of comonotonicity in actuarial science and finance: Theory, *Insur. Math. Econ.*, **31**, 3–33.

- [14] Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R. and Vyncke, D. (2002b). The concept of comonotonicity in actuarial science and finance: Applications, *Insur. Math. Econ.*, **31**, 133–161.
- [15] Drouot Mari, D. and Kotz, S. (2001). *Correlation and Dependence*, Imperial College Press, London.
- [16] Embrechts, P. and Puccetti, G. (2006). Bounds for functions of dependent risks, *Finance Stochast.*, **10**, 341–352.
- [17] Embrechts, P., Höing, A. and Juri, A. (2003). Using Copulae to bound the Value-at-Risk for functions of dependent risk, *Finance Stochast.*, **7**, 145–167.
- [18] Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls, in: *Risk Management: Value at Risk and Beyond*, pp. 176–223, M. A. H. Dempster (ed.), Cambridge University Press, Cambridge.
- [19] Fang, K.-T., Kots, S. and Ng, K.-W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.
- [20] Fisher, N. I. (1997). Copulas, in: *Encyclopedia of Statistical Sciences, Update Volume 1*, pp. 159–163, S. Kotz, C. B. Read and D. L. Banks (eds.), Wiley-Interscience, New York.
- [21] Frank, M. J. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$, *Aequationes Math.*, **19**, 194–226.
- [22] Fréchet, M. (1951). Sur les tableaux de corr’elation dont les marges sont données, *Ann. Univ. Lyon*, 3^e série, Sciences, Sect. A, **14**, 53–77.
- [23] Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas, *North Amer. Actuar. J.*, **2**, 1–25.
- [24] Frees, E. W., Carriere, J. and Valdez, E. A. (1996). Annuity valuation with dependent mortality, *J. Risk Insurance*, **63**, 229–261.
- [25] Frey, R., McNeil, A. and Nyfeler (2001). Copulas and credit models, *Risk Magazine*, October, 111–114.

- [26] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed., Krieger, Melbourne.
- [27] Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrologic Eng.*, **12**, 347–368.
- [28] Genest, C., Ghoudi, K. and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, **82**, 543–552.
- [29] Genest, C., Nešlehová, J. and Ziegel, J. (2011). Inference in multivariate Archimedean copula models (with discussions), *Test*, **20**, 223–292.
- [30] Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas, *J. Amer. Statist. Assoc.*, **88**, 1034–1043.
- [31] Genest, C. and Rémillard, B. (2004). Test of independence and randomness based on the empirical copula process, *Test*, **13**, 335–369.
- [32] Granger, C. W. J, Teräsvirta, T. and Patton, A. J. (2006). Common Factors in Conditional Distributions, *J. Econometrics*, **132**, 43–57.
- [33] Gumbel, E. J. (1958). *Statistics of Extremes*, Columbia Univ. Press, New York [Reprint (2004), Dover Publications].
- [34] Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions, *Publ. Inst. Statist. Univ. Paris*, **9**, 171–173.
- [35] Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie, *Schriften des Mathematischen Instituts und Instituts für Angewandte Mathematik der Universität Berlin*, **5**, 181–233 [English translation “Scale-invariant correlation theory”, in: *The Collected Works of Wassily Hoeffding*, pp. 57–107, N. I. Fisher and P. K. Sen (eds.), 1994, Springer-Verlag, New York].
- [36] Hoff, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation, *Ann. Appl. Statist.*, **1**, 265–283.

- [37] Hougaard, P. (1986). A class of multivariate failure time distributions, *Biometrika*, **73**, 671–678.
- [38] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*, Springer-Verlag, New York.
- [39] Hull, J. C. and White, A. D. (2006). Valuing credit derivatives using an implied copula approach, *J. Derivatives*, **14** (Winter), 8–28.
- [40] Jaworski, P., Durante, F., Härdle, W. and Rychlik, T. (eds.) (2010). *Copula Theory and Its Applications*, Springer-Verlag, Berlin Heidelberg.
- [41] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.
- [42] Kimberling, C. H. (1974). A probabilistic interpretation of complete monotonicity, *Aequationes Math.*, **10**, 152–164.
- [43] Klugman, S. A. and Parsa, R. (1999). Fitting bivariate loss distributions with copulas, *Insur. Econ. Math.*, **24**, 139–148.
- [44] Kruskal, W. H. (1958). Ordinal measures of association, *J. Amer. Statist. Assoc.*, **53**, 814–861.
- [45] Kurowicka, D. and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*, John Wiley & Sons, Chichester.
- [46] Kurowicka, D. and Joe, H. (eds.) (2011). *Dependence Modeling: Vine Copula Handbook*, World Scientific, Singapore.
- [47] Lee, L. (1979). Multivariate distributions having Weibull properties, *J. Multivariate Anal.*, **9**, 267–277.
- [48] Lehmann, E. L. (1966). Some concepts of dependence, *Ann. Math. Statist.*, **37**, 1137–1153.
- [49] Li, D. X. (2000). On default correlation: A copula function approach, *J. Fixed Income*, **9**, March, 43–54.

- [50] Malevergne, Y. and Sornette, D. (2006). *Extreme Financial Risks: From Dependence to Risk Management*, Springer-Verlag, Berlin Heidelberg.
- [51] McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton University Press, Princeton, New Jersey [邦訳：『定量的リスク管理：基礎概念と数理技法』，塚原英敦訳者代表，共立出版 (2008)].
- [52] McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d -monotone functions and ℓ_1 -norm symmetric distributions, *Ann. Statist.*, **37**, 3059–3097.
- [53] Mikosch, T. (2006). Copulas: Tales and facts (with discussions), *Extremes*, **9**, 3–62.
- [54] 宮川雅巳 (1997). 『グラフィカルモデリング』，朝倉書店．
- [55] Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit, *Ann. Statist.*, **3**, 599–616.
- [56] Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*, John Wiley & Sons, Chichester.
- [57] 室町幸雄 (2007). 『信用リスク計測とCDOの価格付け』，朝倉書店．
- [58] Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed., Springer-Verlag, New York.
- [59] Oakes, D. (1982). A model for association in bivariate survival data, *J. Roy. Statist. Soc., B*, **44**, 414–422.
- [60] Oakes, D. (1989). Bivariate survival models induced by frailties, *J. Amer. Statist. Assoc.*, **84**, 487–493.
- [61] Patton, A. J. (2006). Estimation of multivariate models for time series of possibly different lengths, *J. Appl. Econometrics*, **21**, 147–173.
- [62] Patton, A. J. (2009). Copula-based models for financial time series, in: *Handbook of Financial Time Series*, pp. 767–785, T. G. Andersen et al. (eds.), Springer-Verlag, Berlin Heidelberg.

- [63] Plackett, R. L. (1965). A class of bivariate distributions, *J. Amer. Statist. Assoc.*, **60**, 516–522.
- [64] Rüschendorf, L. Schweizer, B. and Taylor, M. D. (eds.) (1996). *Distributions with Fixed Marginals and Related Topics*, IMS Lecture Notes – Monograph Series, Vol. 28, Institute of Mathematical Statistics, Hayward, California.
- [65] Salvadori, G., De Michele, C., Kottegoda, N. T. and Rosso, R. (2007). *Extremes in Nature: An Approach Using Copulas*, Springer, Dordrecht.
- [66] Schönbucher, P. J. (2003). *Credit Derivatives Pricing Models*, John Wiley & Sons, Chichester [『クレジット・デリバティブ：モデルと価格評価』, 望月衛訳 (2005), 東洋経済新報社].
- [67] Schönbucher, P. J. and Schubert, D. (2001). *Copula-dependent Default Risk in Intensity Models*, Preprint [http://www.defaultrisk.com/_pdf6j4/Copula-Dependent.Default.Rsk.Intnst.Mdls.pdf].
- [68] Schweizer, B. and Sklar, A. (1983). *Probabilistic Metric Spaces*, North-Holland, New York [Reprint (2005), Dover Publications].
- [69] Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables, *Ann. Statist.*, **9**, 879–885.
- [70] Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data, *Biometrics*, **51**, 1384–1399.
- [71] Sibuya, M. (1960). Bivariate extreme statistics I, *Ann. Inst. Statist. Math.*, **11**, 195–210.
- [72] 渋谷政昭・高橋倫也 (2007). 「極値理論, 信頼性, リスク管理」, 小西貞則・国友直人編『21世紀の統計科学 II: 自然・生物・健康の統計科学』(東京大学出版会, 2008), p.p. 89–124.
- [73] Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris*, **8**, 229–231.
- [74] Stute, W. (1984). The oscillation behavior of empirical processes: the multivariate case, *Ann. Probab.*, **12**, 361–379.

- [75] 竹村彰通 (1991). 『多変量推測統計の基礎』, 共立出版 .
- [76] Tsukahara, H. (2000). *Empirical Copulas and Some Applications*, Research Report No. 27, The Institute for Economic Studies, Seijo University [<http://www.seijo.ac.jp/research/keiken/green/green27.pdf>].
- [77] 塚原英敦 (2003). 「接合分布関数とその応用」, 『応用統計学』第 32 巻 2 号, 77–88.
- [78] Tsukahara, H. (2005). Semiparametric Estimation in Copula Models, *Canad. J. Statist.*, **33**, 357–375 [Correction: to appear in *Canad. J. Statist.* (2011)] .
- [79] Widder, D. V. (1946). *The Laplace Transform*, Princeton University Press, Princeton [Reprint (2010), Dover Publications].
- [80] Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms, *Duke Math. J.*, **23**, 189–207.
- [81] Yanagimoto, T. (1970). On measures of association and a related problem, *Ann. Inst. Statist. Math.*, **22**, 57–63.
- [82] Zivot, E. and Wang, J. (2006). *Modeling Financial Time Series with S-PLUS*[®], 2nd ed., Springer Science+Business Media, New York.

第6章 時系列分析の理論と応用

田中勝人¹

(一橋大学大学院経済学研究科 教授)

本章は、時系列分析に関する理論と応用の両面について、最近の研究成果を踏まえて論じたものである。時系列分析の進展はめざましく、その研究成果は多くの分野に影響を与えている。例えば、計量経済学の標準的なテキストでは、時系列分析に関する最新のトピックスが、1変量のみならず多変量の場合についても、基礎知識を前提として述べられることが多くなった。しかし、初学者にとっては、それらの内容をすぐに理解することは至難の技であろう。以下では、初学者の便宜のために、基礎理論を最初に解説した上で、主として、1980年代以降に発展した時系列分析のトピックスについて考察したい。

¹tanaka@stat.hit-u.ac.jp

1 はじめに

時系列分析の本格的な研究は 20 世紀初頭に始まったとされる。当初は、時系列モデルに関連した推定や検定のためのさまざまな時系列統計量の標本分布を導出することがメイン・テーマであった。この点については、例えば、イギリスの学術雑誌 *Journal of the Royal Statistical Society, Series B* の 20 世紀中期のもので知ることができる。その後、1960 年代までは、M.S. Bartlett, M.H. Quenouille, H. Wold, A.M. Walker, P. Whittle, E.J. Hannan, T.W. Anderson らによって、このような理論的な研究が積み重ねられた。

1970 年代に入り、コンピュータの進展という追い風の中で、時系列データを比較的容易に分析できる環境が整い始め、時系列分析は専門家の占有物としての状態を脱し、ユーザーが新規参入することになった。G.E.P Box と G.M. Jenkins により 1970 年に出版された *Time Series Analysis: Forecasting and Control* は、この時代の金字塔である。彼らは、著書の中で、モデルの特定化 (identification)、推定 (estimation)、診断 (diagnostic checking) という手続きを繰り返すことにより、よりよいモデルを見つけるというモデル・ビルディングの方法を提案した。この最後の診断は、統計的な検定の手続きを踏むものであるが、それに対して、Akaike (1973) は、情報量規準の観点から、より実際的なモデル選択の方法として、AIC (Akaike's Information Criterion: 赤池情報量規準) を提案した。これらの方法は、時系列データを扱う多くの分野に及ぶこととなり、1980 年代にかけて時系列分析の隆盛が訪れた。

1980 年代は、それまでの時系列分析のフレームワークでは捉えられない諸現象に対して、新たな理論や方法が提案された時期でもある。「長期記憶的な時系列」、「フラクショナル・モデル」、「閾値モデル」、「GARCH モデル」、「単位根時系列」、「共和分」、「ウェーブレット解析」などは、そのような例である。

1990 年代は、上述した新しい理論や方法を使って、実際の分析が盛んに行われた時代である。さらに、ダイナミックな「パネル・データ・モデル」、空間的な要因も取り入れた「時空間モデル」の分析手法が進展した時代でもある。21 世紀に入った現在は、1980 年代のエポック・メイキングな時代に比べれば、ディテールにこだわった研究が主流を占めている感があるが、いずれ、その中から次のステップへの架け橋となる研究成果が生まれるであろう。

以下では、まず、第 2 節で、一変量時系列に関する基礎理論的な事項を説明する。その応用に関するトピックスに関しては、第 3 節から第 6 節で解説する。第 7 節では、多変量時系列に関する説明を行う。上記のトピックスについては、紙数の制約のために取り上げることができなかったものもあるが、それらについては、最後の第 8 節で簡単に触れることにしたい。

2 一変量時系列分析の基礎

本節では一変量時系列の基礎理論を 5 つの小節に分けて説明する．応用に関しては，第 3 節から第 6 節で述べることにする．

2.1 時系列データと定常性

時系列分析は，端的にいえば，時間の経過とともに得られる時系列データを分析するための方法論的科学的学である．時系列データは，背後にある母集団からの標本であると仮定される．この母集団は確率過程と呼ばれ，時間の添え字をもつ確率変数の全体からなっており，例えば， $\{X_t\}$ のように表現される．時間 t が連続な場合を連続時間確率過程，離散的な場合を離散時間確率過程という．本節では，一変量の確率過程について，主として，離散時間の場合を対象として話を進める．

一変量離散時間確率過程 $\{y_t\}$ は定常過程であるとする．すなわち，次の 2 つの性質をもつとする．

$$(i) \quad E(y_t) = \mu \quad (\text{平均が時間に依存せずに一定})$$

$$(ii) \quad \text{Cov}(y_t, y_{t+h}) = \gamma(|h|) \quad (\text{異時点間の共分散が時間差のみに依存})$$

定常過程に従う時系列データは，一定のレベルの回りで同程度の変動をしているということができる．簡単な例としては， $\gamma(h) = 0$ ($h \neq 0$) となるような平均 0 のホワイト・ノイズがある．ホワイト・ノイズは，分散が一定の無相関過程である．

一般的な定常過程は，次のように表すことができる．

$$y_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \{\varepsilon_t\} \sim \text{i.i.d.}(0, \sigma^2), \quad \alpha_0 = 1, \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty \quad (2.1)$$

ここで， $\text{i.i.d.}(0, \sigma^2)$ は，互いに独立で，平均 0，分散 σ^2 の同一分布を意味する．このような i.i.d. 系列は，平均 0 のホワイト・ノイズでもある．この仮定と，(2.1) の最後に記述した係数列 $\{\alpha_j\}$ に関する収束条件のもとで， $\{y_t\}$ は定常過程となる．実際，

$$E(y_t) = 0, \quad \text{Cov}(y_t, y_{t+h}) = \sigma^2 \sum_{j=0}^{\infty} \alpha_j \alpha_{j+|h|} \quad (2.2)$$

となる．

式 (2.1) の定常過程を線形過程と呼ぶ。分解定理 (例えば, Brockwell-Davis (1991), Fuller (1996) を参照) によれば, 任意の定常過程は, 線形過程 (ただし, i.i.d. 系列ではなく, ホワイト・ノイズから生成される表現) と, 決定論的 (完全な予測が可能) な定常過程の和として表現できる。したがって, 後者の成分がなければ, 線形過程は定常過程の一般的な表現とみなすことができる。

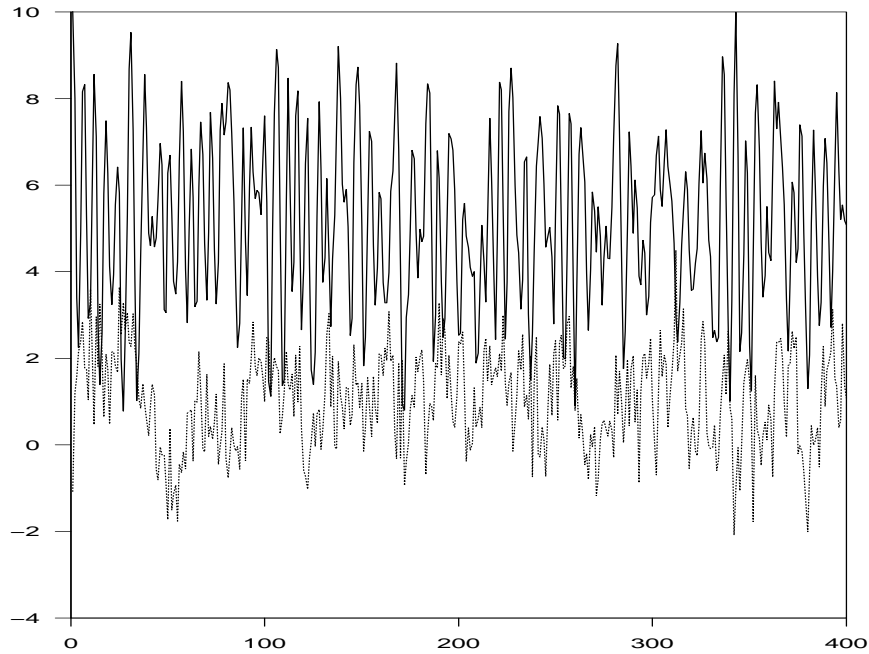
定常過程では, 時点 s と時点 t の共分散 $\text{Cov}(y_s, y_t)$ が, 時差 $|s - t|$ のみに依存するので, $\gamma(|s - t|)$ と表現することができる。これを, 時差 $|s - t|$ の自己共分散という。自己共分散は, 時間的従属性の強さ, あるいは弱さを表す指標であるが, 測定単位に依存するので扱いにくい。それに対して,

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\gamma(h)}{V(y_t)} = \rho(-h) \quad (2.3)$$

は無名数であり, これを時差 h の自己相関という。また, $\rho(h)$ の全体をコレログラムという。 $\rho(h)$ は, いかなる h に対しても, 絶対値が常に 1 以下であり, 1 に近いとき相関の程度が強くなる。

図 1 には, コンピュータにより生成された定常過程に従う時系列データの 2 つの例がプロットされている。これら 2 種類のデータの違いは何か。また, それぞれのデータは, どのような特徴をもっているのか。あるいは, いかなるメカニズムで生み出されたのか, というようなことを統計的に解明するのが時系列分析の仕事である。これらのデータについては, あとで再び取り上げることにするが, 先取りしていえば, 実線のデータは, 次の小節で説明する短期記憶時系列のデータ, 点線のデータは, 第 3 節で説明する長期記憶時系列のデータである。

図 1 時系列データの例



2.2 短期記憶的な時系列モデル

定常過程 $\{y_t\}$ の従属性の程度を測るために、すべての時差の自己共分散の絶対総和

$$S = \sum_{h=-\infty}^{\infty} |\text{Cov}(y_t, y_{t+h})| = \sum_{h=-\infty}^{\infty} |\gamma(h)| \quad (2.4)$$

を考えよう。このとき、 S が有限確定ならば短期記憶的、発散するならば長期記憶的であるという。

式 (2.1) で定義した線形過程が短期記憶的となるか長期記憶的となるかは、係数列 $\{\alpha_j\}$ の $j \rightarrow \infty$ の場合の挙動に依存する。例えば、 $\alpha_j = O(|\lambda|^j)$, ($|\lambda| < 1$) のように幾何級数的に減少する場合は、 $S < \infty$ となるので、短期記憶的である。しかし、例えば、 $\alpha_j = O(j^{d-1})$, ($0 < d < 1/2$) のように、減衰の程度が遅い場合には、 S が発散して、長期記憶的となる。以下、本節では短期記憶的な時系列モデルについて考える。

確率過程 $\{y_t\}$ が次の ARMA(p, q) モデルに従う場合を考えよう .

$$y_t = m + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (2.5)$$

ここで , $\{\varepsilon_t\}$ は i.i.d.($0, \sigma^2$) の誤差項である . 上記の ARMA(p, q) モデルは , ラグ多項式

$$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p, \quad \theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$$

を定義することにより ,

$$\phi(L) y_t = m + \theta(L) \varepsilon_t \quad (2.6)$$

と表すことができる .

ARMA(p, q) モデルが定常となる条件は , AR 部分の特性方程式 $\phi(x) = 0$ の根の絶対値がすべて 1 よりも大きいことである . 定常性のためには , MA 部分の特性方程式 $\theta(x) = 0$ の根に制約は不要である . しかし , モデルの識別性のために , $\theta(x) = 0$ の根の絶対値は , すべて 1 以上であることが仮定される . さらに , 2 つの特性方程式は共通根をもたないものとする . これらの詳細については , 例えば , Anderson (1971) を参照されたい .

式 (2.6) の ARMA(p, q) モデルが定常な場合には , $\{y_t\}$ の平均 μ は , 両辺の期待値を取ることにより , 次のように求めることができる .

$$\mu = \frac{m}{\phi(1)} = \frac{m}{1 - \phi_1 - \cdots - \phi_p}$$

また , このとき , (2.6) の ARMA(p, q) モデルは , 次のように表現できる .

$$y_t = \mu + \phi^{-1}(L) \theta(L) \varepsilon_t = \mu + \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j} \quad (2.7)$$

ここで , α_j は , $\phi^{-1}(L) \theta(L)$ を無限次のラグ多項式で表現した場合の L^j の係数である . 式 (2.7) の最後の表現は , 定常な ARMA モデルが線形過程として表すことができることを示している . なお , 線形過程は MA(∞) 過程であるということもできる .

他方 , ARMA(p, q) モデルは AR(∞) 表現が可能かどうかという問題がある . そのための条件は , MA 部分の特性方程式 $\theta(x) = 0$ の根の絶対値がすべて 1 より大きくなることである . このとき , (2.7) は ,

$$\varepsilon_t = \theta^{-1}(L) \phi(L) (y_t - \mu) = \sum_{j=0}^{\infty} \beta_j (y_{t-j} - \mu)$$

と表現できる . ここで , β_j は , $\theta^{-1}(L) \phi(L)$ を無限次のラグ多項式で表現した場合の L^j の係数である . このように , 誤差項 ε_t が y_t, y_{t-1}, \cdots の線形結合で

表現できるとき，ARMA(p, q) モデルは，反転可能であるという．反転不可能な ARMA モデルとしては，例えば， $y_t = \varepsilon_t - \varepsilon_{t-1}$ のような MA(1) モデルがある．定常な ARMA モデルにおいては，反転可能性は識別性よりも強い条件である．両者の違いは，MA 単位根（MA 部分の特性方程式の根で絶対値が 1 となる根）をもつかどうかということである．反転不可能なモデルでは，パラメータの推定量などの性質が反転可能な場合と異なることから，通常は，反転可能性が仮定される．

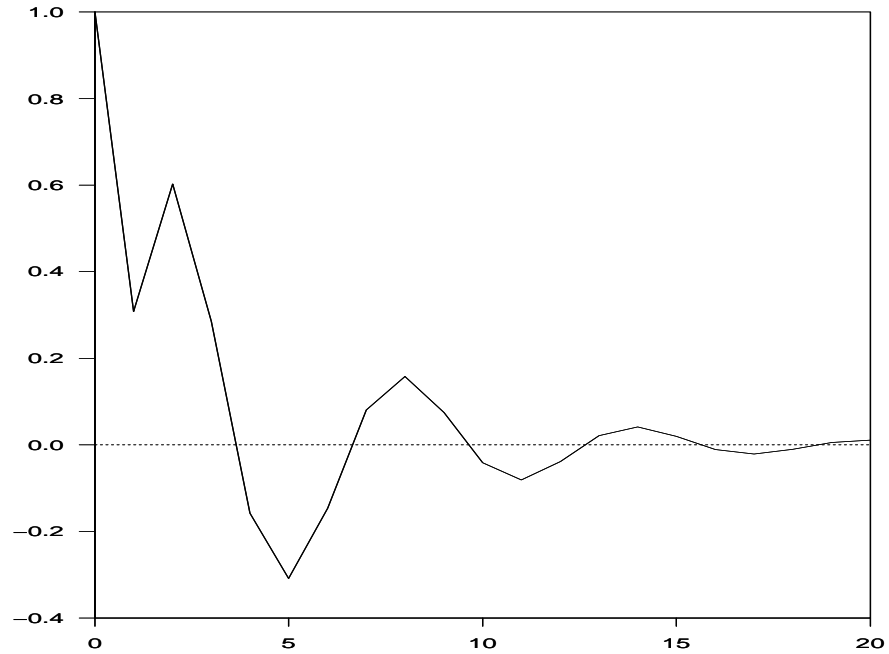
例として，次の ARMA(2, 1) モデルを取り上げよう．

$$\begin{aligned} y_t &= 4.2 + 0.8y_{t-1} - 0.64y_{t-2} + \varepsilon_t - 0.5\varepsilon_{t-1} \\ &= 5 + \frac{1 - 0.5L}{1 - 0.8L + 0.64L^2} \varepsilon_t \end{aligned} \quad (2.8)$$

これは，平均 5 の定常な ARMA(2, 1) モデルである．第 2.1 節の図 1 にある実線の時系列プロットは，このモデルにおいて， $\sigma^2 = 1$ とした場合のデータである．定常性は，AR 部分の特性方程式の 2 根の逆数が， $0.4 \pm i\sqrt{0.48} = 0.8 \exp\{\pm i\pi/3\}$ となることから保証される．MA(∞) 表現の係数は， $(1 - 0.8L + 0.64L^2)(1 + \alpha_1L + \alpha_2L^2 + \dots) = 1 - 0.5L$ の関係から， $\alpha_1 = 0.3$ ， $\alpha_2 = -0.4$ ，一般に $\alpha_j = 0.8\alpha_{j-1} - 0.64\alpha_{j-2}$ となる．

図 2 は，(2.8) の ARMA(2, 1) モデルのコレログラムを図示したものである．図からは，コレログラムが，ほぼ周期 6 で減衰して行く様子が見てとれる．周期をもつことは，AR 部分の特性方程式が複素根をもつことと関連しており，周期が 6 となるのは，その根を極形式 $re^{i\omega}$ で表したときの周波数 ω が $\pi/3$ であることによる．なお，時系列の周期については，別の観点から説明することも可能である．この点については次小節で述べる．

図 2 ARMA モデルのコレログラム



2.3 短期記憶過程のスペクトラム

定常過程 $\{y_t\}$ が短期記憶的な場合，すなわち，式 (2.4) で定義された自己共分散の絶対総和 S が収束する場合には，自己共分散 $\gamma(h)$ を係数とする Fourier 級数

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega} \quad (-\pi \leq \omega \leq \pi) \quad (2.9)$$

が定義されて，右辺の和は一様に絶対収束，そして極限の $f(\omega)$ は連続となる． $f(\omega)$ を $\{y_t\}$ のスペクトル密度関数，あるいはスペクトラムという．

スペクトラムは，原点对称，周期 2π の非負値関数である．このことから， $f(\omega)$ の挙動は $[0, \pi]$ で考えれば十分である．このとき， ω は周波数の意味合いをもち， $2\pi/\omega$ は周期となる．周期は，周波数が π のときに最小値 2 を取り，周波数 0 で無限大となる．

自己共分散の列 $\{\gamma(h)\}$ が与えられれば, (2.9) によりスペクトラム $f(\omega)$ が求まる. これは, 時間領域の情報を周波数領域に変換したものと考えられる. この逆の演算も可能である. すなわち, $f(\omega)$ が与えられれば, (2.9) の両辺に $e^{ih\omega}$ をかけて, $[-\pi, \pi]$ の範囲で ω に関して積分することにより, $\gamma(h)$ を次の定理に述べるように求めることができる.

定理 1 短期記憶過程の自己共分散 $\gamma(h)$ は, スペクトル密度関数 $f(\omega)$ を使って次の形で与えられる.

$$\gamma(h) = \int_{-\pi}^{\pi} f(\omega) e^{ih\omega} d\omega \quad (2.10)$$

以上のことから, 自己共分散とスペクトラムは, 確率密度関数と特性関数の関係と同様に, 1 対 1 に対応する. スペクトラム $f(\omega)$ の意味合いは, (2.10) において, $\gamma(0)$ の場合を考えることにより, 時系列に含まれる変動の周波数 ω における寄与を表すものと解釈できる. そして, $f(\omega)$ の値が大きければ大きいほど, 時系列に含まれる周波数 ω に対応する成分の変動の程度が強いことを意味する.

スペクトラムの最も簡単な例は, $\{u_t\}$ がホワイト・ノイズの場合, すなわち, 分散が σ^2 (一定) の無相関過程の場合である. このときのスペクトラムは, $f(\omega) = \sigma^2/(2\pi)$ と定数になる. このことは, 各周波数が同一の変動をもたらすことを意味し, それは, 白色光の波長としての性質と同じになる. 定常な無相関過程をホワイト・ノイズと呼ぶのは, このことに由来する.

一般の定常な線形過程

$$y_t = \mu + \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \{\varepsilon_t\} \sim \text{i.i.d.}(0, \sigma^2), \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty \quad (2.11)$$

のスペクトラム $f_y(\omega)$ は,

$$f_y(\omega) = \left| \sum_{j=0}^{\infty} \alpha_j e^{ij\omega} \right|^2 f_\varepsilon(\omega) = \frac{\sigma^2}{2\pi} |\alpha(e^{i\omega})|^2 \quad (2.12)$$

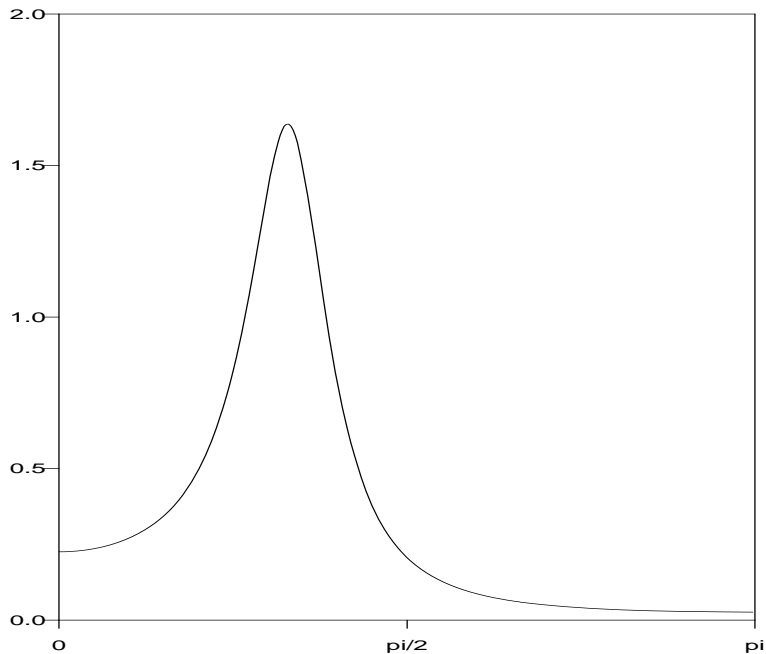
で与えられることがわかる. ここで, $\alpha(x) = \sum_{j=0}^{\infty} \alpha_j x^j$ である.

このことを使えば, 定常な ARMA(p, q) 過程のスペクトラムを求めることは簡単である. 例として, 式 (2.8) で定義した ARMA(2, 1) モデルを考えよう. このとき, スペクトラムは,

$$f_y(\omega) = \frac{\sigma^2}{2\pi} \frac{|1 - 0.5e^{i\omega}|^2}{|1 - 0.8e^{i\omega} + 0.64e^{2i\omega}|^2} \quad (2.13)$$

となる．図 3 には， $\sigma^2 = 1$ の場合のスペクトラムが図示されている．ピークは， $\omega = \pi/3$ (周期は 6) に対応している．この点については，すでに前小節で自己相関の観点から見たが，周波数領域では，スペクトラムにより，このように明瞭に見ることができる．

図 3 ARMA モデルのスペクトラム



スペクトラムは，ARMA モデルのパラメータの推定において，周波数領域における推定を考える際に有用となる．この点については，例えば，Taniguchi-Kakizawa (2000) を参照されたい．スペクトラムの別の有用性は，式 (2.10) の関係が示唆するように，スペクトラムが与えられれば，自己共分散が積分により求められる点である．ただし，積分は複素積分となり，一般には，それほど簡単でない点は，時間領域の場合と同様である．この点については，田中 (2006) を参照されたい．

2.4 ARMA モデルの予測

定常な確率過程に対して ARMA モデルを使うことの利点は、予測を容易にすることである。このことを見るために、まず、 $\{y_t\}$ を平均 μ の任意の定常過程として、無限の過去から時点 T までの観測値 $\mathbf{y}(T) = (y_T, y_{T-1}, \dots)$ が与えられた場合に、 h 時点先の y_{T+h} を予測する問題を考えよう。

そこで、望ましい予測量として、不偏で、かつ、予測誤差の分散を最小にするものを考える。すなわち、予測量を \hat{y}_{T+h} とするとき、次の条件をみたすものを求める。

- (a) 予測誤差 $e_{T+h} = y_{T+h} - \hat{y}_{T+h}$ の平均が 0 となる。
- (b) 予測誤差分散 $V(e_{T+h})$ を最小にする。

予測誤差分散を最小にするような不偏予測量を最良不偏予測量という。そのような最良不偏予測量は、次の形で与えられる。

定理 2 $\mathbf{y}(T) = (y_T, y_{T-1}, \dots)$ に基づいて、 y_{T+h} を予測する場合の最良不偏予測量 y_{T+h}^* と、その予測誤差分散 $V(e_{T+h}^*)$ は、次のように与えられる。

$$y_{T+h}^* = E(y_{T+h} | \mathbf{y}(T)), \quad V(e_{T+h}^*) = E \left[\{y_{T+h} - E(y_{T+h} | \mathbf{y}(T))\}^2 \right]$$

一般に、条件付き期待値の計算は困難であるが、ARMA モデルの場合には、その計算は非常に容易となる。実際、定常かつ反転可能な平均 0 の ARMA(p, q) モデルにおいては、 y_{T+h} の最良不偏予測量 y_{T+h}^* は、次のように表すことができる。

$$y_{T+h}^* = \phi_1 y_{T+h-1}^* + \dots + \phi_p y_{T+h-p}^* + \varepsilon_{T+h}^* - \theta_1 \varepsilon_{T+h-1}^* - \dots - \theta_q \varepsilon_{T+h-q}^*$$

ここで、

$$y_j^* = y_j \quad (j \leq T), \quad \varepsilon_j^* = \begin{cases} \varepsilon_j & (j \leq T) \\ 0 & (j > T) \end{cases}$$

であり、 $\{\varepsilon_t\}$ の条件付き期待値の計算においては、反転可能性が使われている。一般の h 期先の予測量の予測誤差分散を求めるには、MA(∞) 表現を使うと便利である。すなわち、ARMA(p, q) モデルを

$$y_{T+h} = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{T+h-j}, \quad \alpha_0 = 1 \tag{2.14}$$

と MA(∞) 表現する。このとき、次の諸量を得る。

$$y_{T+h}^* = \sum_{j=h}^{\infty} \alpha_j \varepsilon_{T+h-j}, \quad e_{T+h}^* = \sum_{j=0}^{h-1} \alpha_j \varepsilon_{T+h-j}, \quad V(e_{T+h}^*) = \sigma^2 \sum_{j=0}^{h-1} \alpha_j^2$$

このことから，予測誤差分散は，予測時点が遠ざかるにつれて単調に増加して，系列の分散 $V(y_t)$ に近づいて行くことがわかる．

以上の議論では，モデルのパラメータが既知であるとして話を進めてきたが，実際に予測を行う場合は，パラメータを推定する必要がある．この点については，次小節で説明する．

2.5 ARMA モデルの推定

ARMA モデルを推定するためには，まず，AR 部分の次数 p と MA 部分の次数 q を決める必要がある．そのための手がかりとなる統計量としては，標本の自己相関や偏自己相関などがある．しかし，データからモデルの次数を一意的に絞り込むのは困難な場合が普通であるので，ひとまず，候補となるモデルを $ARMA(p, q)$ に特定化したものとしよう．

その上で， $ARMA(p, q)$ モデル

$$\phi(L)y_t = \theta(L)\varepsilon_t, \quad \{\varepsilon_t\} \sim \text{i.i.d.}(0, \sigma^2) \quad (2.15)$$

を推定するものとしよう．このモデルでは，平均を 0 としているが，平均が 0 でない場合は，標本平均を差し引いたあとに，モデルをあてはめるものとする．このモデルに含まれるパラメータは，次の通りである．

$$\phi = (\phi_1, \dots, \phi_p)', \quad \theta = (\theta_1, \dots, \theta_q)', \quad \sigma^2 = V(\varepsilon_t)$$

これらのパラメータを推定するための主要な方法は，最小 2 乗法と最尤法に大きく分けられる．そして，それぞれについて，時間領域と周波数領域における方法が提案されている．ここでは，時間領域における方法について説明する．

2.5.1 最小 2 乗法

標本サイズ T の観測値が与えられたとき，時間領域における ϕ と θ の最小 2 乗法による推定量は，次の関数を最小化することにより得られる．

$$f(\phi, \theta) = \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T \left\{ \frac{\phi(L)y_t}{\theta(L)} \right\}^2 \quad (2.16)$$

この関数は，その形から明らかなように，MA 部分のラグ多項式 $\theta(L)$ が入り込むために複雑となっている．もし，MA 部分を含まないならば， ε_t はパラメータ ϕ の線形関数となり，通常の回帰モデルの場合と同様に，推定も容易である．しかし，MA 部分が存在すると， ε_t は常にパラメータの非線形関数となる．したがって，そこから得られる推定量は NLSE (非線形最小 2 乗推定量) と呼ばれる．式 (2.16) の最小化については，さまざまな方法が提案されており，Box-Jenkins-Reinsel (1994) に詳しい解説がある．

なお, σ^2 は, パラメータ ϕ と θ の NLSE に基づいて計算される残差を使って, 事後的に推定することができる.

2.5.2 時間領域における最尤推定法

定常, 反転可能な ARMA(p, q) モデルからの観測値ベクトル $\mathbf{y} = (y_1, \dots, y_T)'$ が, 多変量正規分布 $N(0, \sigma^2 \Sigma)$ に従うものとしよう. ここで, Σ は, T 次の正値定符号行列であり, 各要素は ϕ と θ の関数である. このとき, 対数尤度は,

$$L(\phi, \theta, \sigma^2) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log |\Sigma| - \frac{1}{2\sigma^2} \mathbf{y}' \Sigma^{-1} \mathbf{y} \quad (2.17)$$

となる. パラメータ ϕ と θ が与えられれば, (2.17) を最大にする σ^2 は, $\mathbf{y}' \Sigma^{-1} \mathbf{y} / T$ であるので, これを σ^2 に代入することにより, 集約対数尤度関数

$$l(\phi, \theta) = -\frac{T}{2} \log(\mathbf{y}' \Sigma^{-1} \mathbf{y}) - \frac{1}{2} \log |\Sigma| \quad (2.18)$$

が定義される. これを最大化することにより, ϕ と θ の MLE (最尤推定量) が得られる. そして, σ^2 は, $\hat{\sigma}^2 = \mathbf{y}' \hat{\Sigma}^{-1} \mathbf{y} / T$ により推定できる. ここで, $\hat{\Sigma}$ は, Σ に含まれる ϕ と θ を MLE で置き換えたものである.

以上の方法で, 時間領域における MLE が得られるが, T が大きくなると, 行列式 $|\Sigma|$ や逆行列 Σ^{-1} の計算量は莫大なものとなるので, 計算上の工夫が必要となる. 詳しくは, Box-Jenkins-Reinsel (1994), Brockwell-Davis (1991) などを参照されたい.

一般に, MLE は, 前節で述べた NLSE よりも望ましい性質をもっている. しかし, 漸近的には, 両者は同等であり, 次のことが成り立つ.

定理 3 定常, 反転可能な ARMA モデルにおいては, パラメータ $\beta = (\phi', \theta)'$ の NLSE および MLE は, 漸近的に同一の正規分布に従い, 次のことが成り立つ.

$$\sqrt{T}(\hat{\beta} - \beta) \Rightarrow N(0, \Omega^{-1})$$

ここで, $\hat{\beta}$ は β の MLE あるいは NLSE である. また, Ω は, パラメータ β に関する Fisher の情報行列であり, 次のように求めることができる.

$$\Omega = \frac{1}{\sigma^2} \begin{pmatrix} E(\mathbf{u}_t \mathbf{u}_t') & E(\mathbf{u}_t \mathbf{v}_t') \\ E(\mathbf{v}_t \mathbf{u}_t') & E(\mathbf{v}_t \mathbf{v}_t') \end{pmatrix}$$

ここで, $\mathbf{u}_t = (u_t, \dots, u_{t-p+1})'$, $\mathbf{v}_t = (v_t, \dots, v_{t-q+1})'$ であり, $\{u_t\}$, $\{v_t\}$ は, それぞれ, AR(p) 過程 $\phi(L)u_t = -\varepsilon_t$, AR(q) 過程 $\theta(L)v_t = \varepsilon_t$ に従う.

2.5.3 モデルの診断

推定した ARMA モデルでは、次数の特定化が本当に適切かどうかを検討することが必要である。ここでは、まず、仮説検定の観点から、次に、モデル選択規準の観点から、モデル診断の方法について説明する。

推定されたモデルが適切ならば、そのモデルから計算される残差系列は無相関であると考えられる。サイズ T の標本から ARMA(p, q) モデル $\phi(L)y_t = \theta(L)\varepsilon_t$ を推定した場合の残差は、

$$e_t = y_t - \sum_{j=1}^p \hat{\phi}_j y_{t-j} + \sum_{j=1}^q \hat{\theta}_j e_{t-j} \quad (t = 1, \dots, T) \quad (2.19)$$

で計算される。このような残差系列を使って、ラグ h の残差自己相関を次のように求める。

$$r_h = \frac{\sum_{t=1}^{T-h} e_t e_{t+h}}{\sum_{t=1}^T e_t^2} \quad (h = 1, \dots, T-1) \quad (2.20)$$

ARMA(p, q) モデルを推定したあとの残差系列の場合は、モデルが適切ならば、 $\sqrt{T} r_h$ の周辺分布は、漸近的に $N(0, 1)$ となることが知られている。このことから、各 h ごとに、有意水準 5% で無相関が受容される r_h の範囲は、

$$\left[-1.96 \times \frac{1}{\sqrt{T}}, 1.96 \times \frac{1}{\sqrt{T}} \right]$$

となる。すなわち、0 のまわりのほぼ $\pm 2/\sqrt{T}$ の区間である。

各ラグごとの残差自己相関の有無を判断する以外に、これらを一括して検定する方法も提案されている。整数 m を $p+q$ よりも大きな値とすると、次の 2 つの統計量

$$Q = T \sum_{h=1}^m r_h^2, \quad \tilde{Q} = T(T+2) \sum_{h=1}^m \frac{1}{T-h} r_h^2 \quad (2.21)$$

は、ARMA(p, q) モデルが適切であるという帰無仮説のもとで、漸近的に自由度 $m-p-q$ の χ^2 分布に従う。 Q は、Box-Pierce 統計量と呼ばれる。他方、 \tilde{Q} は、Ljung-Box 統計量と呼ばれ、 Q の標本分布が χ^2 分布よりも小さな値を取る傾向がある点を修正した統計量である。

ところで、適切なモデルを選択する際、候補となるモデルが多数あれば、検定の観点から最終的なモデルを決定するのは、かなり大変な作業となる。また、残差系列の無相関性に関する検定にパスするようなモデルが複数あれば、どのモデルを採用したらよいか、判断に迷うこともありうる。

このような場合、実際のデータが従う真の分布との距離を各モデルごとに推定し、その推定値が最小になるようなモデルを最良のモデルとして選択する方法が提案されている。ここで、真の分布との距離は、情報理論における

Kullback-Leibler 情報量の考え方を援用して測ることができる．その代表的なもの、本稿の冒頭で述べたように、Akaike (1973) により提案された AIC である．平均が未知の ARMA(p, q) モデルに対する AIC は、

$$\text{AIC}(p, q) = -2 \times \text{対数尤度の最大値} + 2(p + q + 1) \quad (2.22)$$

により定義される．この値を最小にする p と q の組合せからなる ARMA(p, q) モデルを最適なものとして選択するのが AIC によるモデル選択規準である．式 (2.22) において、右辺第 1 項は、モデルのデータへの当てはまりのよさを表し、モデルを複雑にすればするほど、単調に減少する．これに対して、第 2 項は、パラメータの個数 $p + q$ の単調増加関数であり、モデルを複雑にすることに対するペナルティーを考慮している．当然ながら、当てはまりが同程度のモデルの中では、単純なモデルをよしとする原理である．なお、AIC 以外にも、その修正版や改良版がいくつか提案されている．代表的なものは、Schwarz (1978) により提案されたもので、

$$\text{SBC}(p, q) = -2 \times \text{対数尤度の最大値} + (p + q + 1) \log T \quad (2.23)$$

を最小にするモデル選択規準である．SBC の第 2 項は AIC の第 2 項よりも大きいので、SBC は推定すべきパラメータの追加に対してペナルティーが厳しくなっている．したがって、SBC 規準は、AIC 規準よりも「ケチの原理」が強く、より単純なモデルを選ぶ傾向がある．

3 ARIMA モデルと ARFIMA モデル

時系列の中には、経済時系列に代表されるように、定常性をみたまない時系列も多い．その場合でも、例えば d 回の差分を取れば定常とみなすことができる時系列がある．そのような非定常時系列は、定常系列を和分したものであると考えられるので、 d 次の和分過程、あるいは $I(d)$ 過程と呼ばれる． I は、Integrated (和分された) の頭文字である． $I(d)$ 過程に対するモデルは、次のように表すことができる．

$$\phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t, \quad \{\varepsilon_t\} \sim \text{i.i.d.}(0, \sigma^2) \quad (3.24)$$

ここで、 d は非負の整数値を取る差分パラメータである．また、 $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$, $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ であり、2 つの特性方程式 $\phi(x) = 0$, $\theta(x) = 0$ は、互いに共通根をもたず、根の絶対値はすべて 1 より大である．このとき、 $\{y_t\}$ は、ARIMA(p, d, q) モデルに従うという．ARIMA の I は、Integrated の頭文字である．

非定常な ARIMA モデルの最も単純な形は、ARIMA(0, 1, 0) であり、

$$y_t = y_{t-1} + \varepsilon_t = \varepsilon_1 + \dots + \varepsilon_t, \quad y_0 = 0 \quad (3.25)$$

と表される．このモデルは，ランダム・ウォークとも呼ばれる．さらに，AR部分の特性方程式の根が1であることから，単位根モデルとも呼ばれる．ランダム・ウォークは，分散が時間に比例して増大するような確率過程であり， $y_t = O_p(\sqrt{t})$ となる．一般のARIMA(p, d, q)モデルの場合には， $y_t = O_p(t^{d-1/2})$ となることが示される．ARIMAモデルの推定は， d が与えられれば，ARMAモデルの場合と同様に行うことができる．

他方，(3.24)のARIMA(p, d, q)モデルにおいて， d を実数の範囲に拡大したモデルを考えることもできる．その場合のモデルを，ARFIMA(p, d, q)モデルという．Fは，Fractionalの頭文字である．特に， $d < 1/2$ ならば定常， $d > -1/2$ ならば反転可能， $0 < d < 1/2$ ならば長期記憶的であることが示される(Hosking (1981))．ここで，定常な確率過程が長期記憶的であるとは，第4節で述べたように，自己共分散の絶対値の総和が発散することである．したがって，ARFIMA(p, d, q)モデルは， $0 < d < 1/2$ ならば，定常，反転可能，長期記憶的となる．

定常なARFIMA(p, d, q)モデルのスペクトラムと自己共分散は，ARMAの場合と同様に，次のように表現される．

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{i\omega})|^2}{|1 - e^{i\omega}|^{2d} |\phi(e^{i\omega})|^2}, \quad \gamma(h) = \int_{-\pi}^{\pi} f(\omega) e^{ih\omega} d\omega \quad (3.26)$$

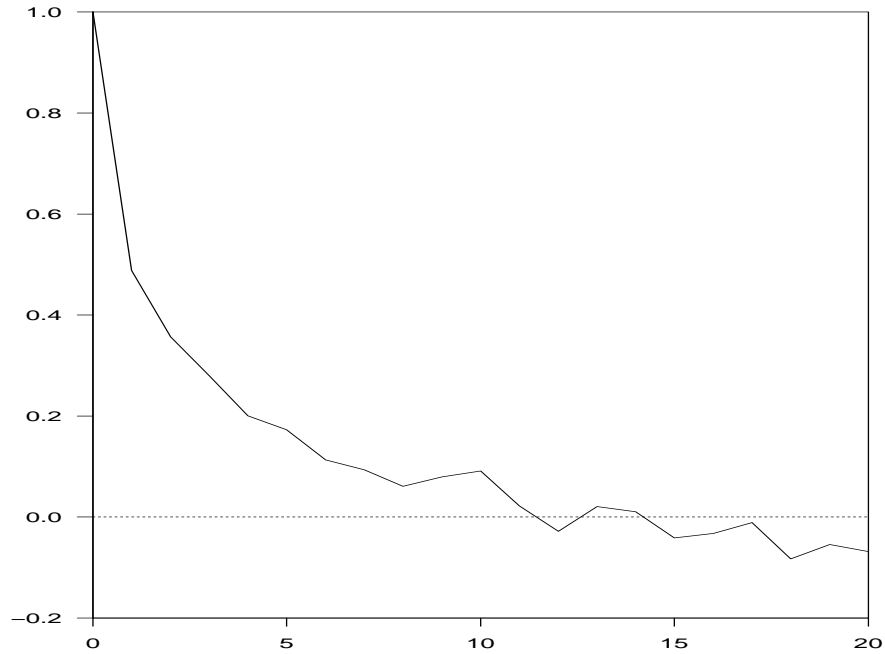
特に，ARFIMA($0, d, 0$)の場合の自己共分散は，次のようになる．

$$\gamma(h) = \sigma^2 \frac{\Gamma(1-2d)\Gamma(h+d)}{\Gamma(d)\Gamma(1-d)\Gamma(h-d+1)} \quad (h > 0) \quad (3.27)$$

式(3.26)から，スペクトラムは， $\omega \rightarrow 0$ のとき， $f(\omega) = O(\omega^{-2d})$ であることがわかる．また，(3.27)から，ARFIMA($0, d, 0$)の自己共分散は， $h \rightarrow \infty$ のとき， $\gamma(h) = O(h^{2d-1})$ であることがわかる．一般の定常なARFIMA(p, d, q)モデルの場合にも，自己共分散は同じオーダーとなることが示される(Hosking (1981))．

第2.1節の図1にある点線の時系列プロットは，ARFIMA($0, 0.45, 0$)モデルからの実現値の例($\sigma^2 = 1$)である．図4は，このデータに対する標本自己相関である．自己相関は，ARMAの場合に比べて，減少のスピードが遅いことが見てとれる．

図 4 長期記憶時系列の標本コレログラム



ARFIMA モデルと ARMA モデルの違いは，次の中心極限定理に関する事実にも見いだすことができる (Hosking (1996)) .

定理 4 定常，反転可能で平均 μ の ARFIMA(p, d, q) モデルからのサイズ T の標本平均 \bar{y} に対して，次のことが成り立つ .

$$T^{1/2-d}(\bar{y} - \mu) = \frac{1}{T^{d+1/2}} \sum_{t=1}^T (y_t - \mu) \Rightarrow N\left(0, \sigma^2(d) \frac{\theta^2(1)}{\phi^2(1)}\right)$$

ここで，

$$\sigma^2(d) = \lim_{T \rightarrow \infty} V\left(\frac{1}{T^{d+1/2}} \sum_{t=1}^T (1-L)^{-d} \varepsilon_t\right) = \frac{\sigma^2 \Gamma(1-2d)}{(1+2d) \Gamma(1+d) \Gamma(1-d)}$$

定常，反転可能な ARMA(p, q) モデルの場合の結果は，この定理において， $d = 0$ とすればよい . 違いは，分布収束のオーダーと，極限の分散の値である .

自己共分散に関しても，漸近的な違いが現れる．特に，ARFIMA モデルの場合は， d の値が $1/4$ を超えると正規分布とは異なる分布に収束することが知られている (Hosking (1996)) ．

ここで説明してきた ARIMA モデルと ARFIMA モデルは，季節性を含むような非定常データを扱うために，季節階差 $1 - L^m$ を適用したモデルに拡張され，それぞれ，SARIMA モデル，SARFIMA モデルとして，提案されている．S は，Seasonal の頭文字である．詳しくは，Box-Jenkins-Reinsel (1994)，Journal of Econometrics (1996)，Vol. 73 の特集号などが参考になるであろう．

4 単位根検定

時系列が単位根をもつ場合には，ランダム・ウォークの表現からわかるように，時系列に固有の過去の確率的なショックが持続することを意味する．他方，単位根をもたないならば，そのようなショックは時間とともに消滅する．このように，両者の違いは，時系列の変動に関して重要な示唆を与えることになる．また，単位根をもつ場合には，階差変換後のデータは定常とみなされるが，複数の単位根時系列を分析する場合には，変換後の定常なデータを分析することは，情報の損失を伴うことがありうる．この点については，第 7 節で説明する．

時系列が単位根をもつかどうかは，検定の立場から判断する必要がある．説明のため，最も単純な次の AR(1) モデル

$$y_t = \rho y_{t-1} + \varepsilon_t \Leftrightarrow \Delta y_t = \delta y_{t-1} + \varepsilon_t \quad (\Delta = 1 - L, \delta = \rho - 1) \quad (4.28)$$

を考えよう．ここで， $y_0 = 0$ ， $\{\varepsilon_t\} \sim \text{i.i.d.}(0, \sigma^2)$ である．単位根検定は，帰無仮説 $H_0 : \rho = 1 (\delta = 0)$ を，対立仮説 $H_1 : \rho < 1 (\delta < 0)$ に対して検定するものであり，直感的には， ρ あるいは δ の LSE の値が小さいときに H_0 を棄却する検定を考えることができる．標本サイズ T から得られる δ の LSE を $\hat{\delta}$ とするとき，局所対立仮説 $H_1 : \rho = 1 - c/T$ (c は正定数) のもとで次のことが成り立つ (Phillips-Perron (1988)) ．

$$T\hat{\delta} = \frac{1}{T\sigma^2} \sum_{t=2}^T y_{t-1} \Delta y_t \Big/ \frac{1}{T^2\sigma^2} \sum_{t=2}^T y_{t-1}^2 \Rightarrow \frac{\int_0^1 Y(t) dY(t)}{\int_0^1 Y^2(t) dt} \quad (4.29)$$

ここで， $\{Y(t)\}$ は $[0,1]$ 上の Ornstein-Uhlenbeck (O-U) 過程であり，

$$dY(t) = -cY(t) dt + dW(t), Y(0) = 0 \Leftrightarrow Y(t) = e^{-ct} \int_0^t e^{cs} dW(s)$$

で定義される． $\{W(t)\}$ は $[0,1]$ 上の標準ブラウン運動である．

式 (4.29) の最後の極限分布は単位根分布と呼ばれ，分布関数は数値計算により求めることができる (Nabeya-Tanaka (1990)) ．単位根検定のための検定統

計量は、この他にも多くのものを考えることができる。また、モデルもより一般的なものに拡張することができる。詳しくは、Fuller (1996), Tanaka (1996) を参照されたい。

単位根検定は、ARMA モデルの特性方程式の根に関する検定であり、上記のように AR 部分に関する単位根検定の他に、MA 部分の単位根検定についても考えることができる。その主たる目的は、原系列に対して、過剰な階差変換をしていないかどうかを調べるものである。以下で見るように、それはまた、AR 部分の単位根検定と解釈することも可能である。

MA 部分の単位根検定のための基本的なモデルは、MA(1) モデル

$$y_t = \varepsilon_t - \alpha \varepsilon_{t-1}, \quad \{\varepsilon_t\} \sim \text{i.i.d.}(0, \sigma^2) \quad (4.30)$$

であり、 y_t は観測値そのものでも階差変換後の値でもよい。検定問題は、 $H_0: \alpha = 1$ vs. $H_1: \alpha < 1$ である。

系列 $\{y_t\}$ が階差系列の場合、すなわち、 $y_t = (1-L)x_t$ ならば、 H_0 のもとでは、 $(1-L)x_t = (1-L)\varepsilon_t$ となり、過剰階差の状況となる。この結果、因数 $1-L$ が消去され、原系列は AR 部分に単位根をもたない。他方、 H_1 のもとでは適切な階差変換であり、原系列は AR 部分に単位根をもつことになる。したがって、階差系列に対する MA 部分の単位根検定は、AR 部分に「単位根なし」という帰無仮説を、「単位根あり」という対立仮説に対して検定するのであり、通常とは逆向きの検定となる。

モデル (4.30) の誤差項 $\{\varepsilon_t\}$ に正規性を仮定するとき、統計量

$$S_T = \frac{1}{T} \frac{\mathbf{y}' \Omega^{-2} \mathbf{y}}{\mathbf{y}' \Omega^{-1} \mathbf{y}} \quad (4.31)$$

が大きいときに H_0 を棄却する検定は、適当な変換群のもとで局所最良不変不偏 (LBIU) 検定となる。ここで、 Ω は、観測ベクトル \mathbf{y} の共分散行列を H_0 のもとで評価したものである。統計量 S_T は、局所対立仮説 $\alpha = 1 - (c/T)$ のもとで、次の漸近的な性質をもっている (詳細は、Tanaka (1996), 田中 (2006) を参照)。

$$S_T \Rightarrow \sum_{n=1}^{\infty} \left[\frac{1}{n^2 \pi^2} + \frac{c^2}{n^4 \pi^4} \right] Z_n^2, \quad \{Z_n\} \sim \text{NID}(0, 1) \quad (4.32)$$

MA 部分の単位根検定についても、より一般的なモデルにおける検定を考えることができる。また、いずれの検定においても、さまざまな検定が考えられるが、それらの検出力の比較結果が報告されている。詳しくは、Elliott-Rothenberg-Stock (1996), 田中 (2006) を参照されたい。

5 構造変化を考慮した時系列モデル

前節で議論した単位根検定の統計量は，モデルの特定化に誤りがあれば，その影響を受けて，想定した分布とは異なる分布をもたらす．その要因の1つとして，説明変数の特定化の誤りによるものが考えられる．すなわち，含めるべき変数をモデルに取り込まない場合である．そのような例としては，構造変化を表す変数をモデルに含めない場合がある．この問題は，Perron (1989) が最初に取り上げたものである．

今，時点 T_B において，レベル・シフトが起きた場合を考えよう．このとき，ダミー変数として，

$$D_t(T_B) = \begin{cases} 0 & (t \leq T_B) \\ 1 & (t > T_B) \end{cases} \quad (5.33)$$

を定義することにより，構造変化を考慮した単位根検定のためのモデルは，次のように表現できる．

$$y_t = \alpha_0 + \alpha_1 D_t(T_B) + \eta_t, \quad \eta_t = \rho \eta_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{i.i.d.}(0, \sigma^2) \quad (5.34)$$

レベル・シフト以外の構造変化を含む一般的な場合についても，新たなダミー変数を定義して，同様に考えることができる．検定方式としては， ρ の推定量に基づく検定が可能である．Perron (1989) は，上記のような構造変化を考慮すれば，単位根の存在が否定される場合があることを実証している．なお，そこでは構造変化が起きる時点が既知であることを前提にしている．未知とした場合については，Zivot-Andrews (1992), Vogelsang-Perron (1998) などで議論されている．

構造変化あるいはレジームの転換を考慮したモデルは，単位根検定とは関係なく，より早期に提案されている．その1つが，Tong (1983) による threshold (閾値) モデルであり，次のような AR モデルの拡張版である．

$$y_t = \begin{cases} \phi_1^{(1)} y_{t-1} + \cdots + \phi_{p_1}^{(1)} + \varepsilon_t^{(1)} & (x_t < a \text{ のとき}) \\ \phi_1^{(2)} y_{t-1} + \cdots + \phi_{p_2}^{(2)} + \varepsilon_t^{(2)} & (x_t \geq a \text{ のとき}) \end{cases}$$

ここで， a が閾値であり，この値を境にして，モデルが異なるとするものである．なお，変数 x_t としては， y_t の過去の値を使ってもよい．

この他には，Hamilton (1989) により提案されたマルコフ・スイッチング・モデルがある．このモデルは，0 か 1 を取るダミー変数 S_t を定義して，それぞれの値を取る確率をマルコフ型の推移確率として与えるものである．すなわち，次のような推移確率を考える．

$$P(S_t = 1 | S_{t-1} = 1) = p, \quad P(S_t = 0 | S_{t-1} = 1) = 1 - p$$

$$P(S_t = 0 | S_{t-1} = 0) = q, \quad P(S_t = 1 | S_{t-1} = 0) = 1 - q$$

このような変数は，一般の構造的な変化，特に，景気の上昇局面と下降局面を表して，時系列モデルに組み込むことができる．

ここで述べたモデルは，大きくは非線形モデルのカテゴリーに入るものであり，通常の線形モデルではとらえられない現象を説明しようとするものである．その意味では，ファイナンスで使われる ARCH や GARCH モデルも非線形モデルであり，さらに，その中に上述のようなレジーム転換を入れたモデルも使われている．推定に関しては，尤度関数が複雑となるので，EM アルゴリズムや MCMC 法など，さまざまな計算上の工夫が提案されている．

6 ウェーブレット解析

従来の時系列分析は，時間領域か周波数領域のいずれかにおいて別々に展開されてきた．それに対して，ウェーブレット解析は，これら 2 つを同時に考慮した領域，すなわち，ウェーブレット領域においてデータを分析するものである．そのために，原系列はウェーブレット変換される．

離散時間確率過程 $\{x_t\}$ からの観測値を列ベクトル $\boldsymbol{x} = (x_1, x_2, \dots, x_T)'$ で表す．ここで，標本サイズ T は $T = 2^J$ (J は自然数) であると仮定する．このとき， \boldsymbol{x} の DWT (Discrete Wavelet Transform: 離散ウェーブレット変換) とは，次の変換

$$\boldsymbol{w} = \mathcal{W} \boldsymbol{x}, \quad \boldsymbol{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_J \\ v_J \end{pmatrix}, \quad \mathcal{W} = \begin{pmatrix} \mathcal{W}_1 \\ \vdots \\ \mathcal{W}_J \\ \mathcal{V}_J \end{pmatrix} \quad (6.35)$$

で，以下で述べる条件をみたすものをいう．

\mathcal{W} はウェーブレット変換行列と呼ばれる直交行列である．その構成部分 \mathcal{W}_j は， $T/2^j$ 個の行からなるレベル j の変換行列である．最後の \mathcal{V}_J は \mathcal{W}_J と同様に行ベクトルであり，すべての成分は $1/\sqrt{T}$ からなっている．

他方， \boldsymbol{w} はウェーブレット係数ベクトルであり，通常は，単にウェーブレットと呼ばれるものである．その構成部分 w_j はレベル j のウェーブレットと呼ばれ， $T/2^j$ 個の成分からなるベクトルである．ウェーブレットは，レベル 1 において最も解像度の高い，高周波の変換をもたらす．そのために，局所的な時点での計算が必要であり，その結果，最も多くの成分 $T/2$ 個を含むことになる．他方，レベル j が大きくなるにつれて，次第に解像度の低い，低周波の変換に移行するので，大局的な時点での計算となり，成分も少なくなる．実際， w_j はスカラーである．また， \boldsymbol{w} の最後の成分である v_J もスカラーであり，

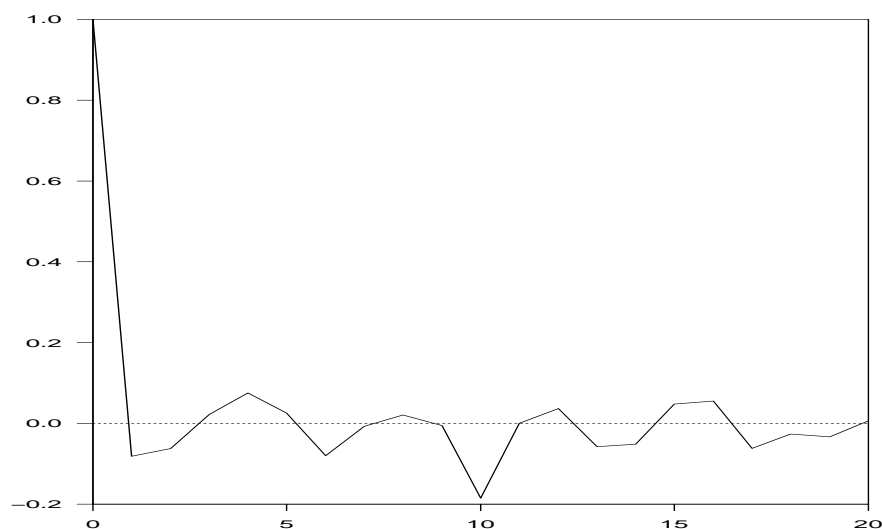
レベル J のスケーリング係数と呼ばれる ν_J の定義から $v_J = \nu_J x = \sqrt{T} \bar{x}$ (\bar{x} は x の平均) となる。

時系列データをウェーブレット変換することの長所としては、大きく次の 3 点を挙げることができる。

- (a) 時間領域と周波数領域における原系列の局所的な挙動に関する情報の抽出能力が優れている。
- (b) 非定常な連続時間確率過程であるフラクショナル・ブラウン運動は、ウェーブレット変換することにより、レベルごとに定常な確率過程となる。
- (c) ARFIMA 過程に従う時系列のウェーブレット変換は、レベルごとに、ほぼ定常無相関、レベル間では無相関な系列となる。

上記 (a) の性質は、文字通り、ウェーブレットの局在性であり、局所的なレベルのシフトやジャンプの検出に有用であることを意味する。(b) は、ウェーブレット変換が、ある意味で階差変換の働きをしていることを示唆している。(c) は、(b) とも関連する性質であるが、ほぼ無相関な系列を実現する点が注目される。実際、図 5 には、図 1 の ARFIMA(0, 0.45, 0) から得られたデータ (ただし、標本サイズは $T = 512$ に拡大) をウェーブレット変換した場合のレベル 1 のコレログラムをプロットしているが、この事実が裏付けられていることがわかる。

図 5 ウェーブレット変換後の標本コレログラム



正規過程の場合には，上記 (c) の性質から，ウェーブレット変換後の系列は，ほぼ独立となる．このことは，非定常な ARFIMA モデルにおいても実現されることが知られており，ウェーブレット変換が ARFIMA モデルの推定を容易にすることが想像できよう．この点も含めて，詳しくは，Percival-Walden (2000) を参照されたい．

7 多変量時系列

複数の時系列間の関係を分析する場合には，多変量時系列モデルを考える必要がある．次の VARMA_m(p, q) モデルは，そのために有用である．

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \cdots + \Phi_p \mathbf{y}_{t-p} + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \cdots - \Theta_q \varepsilon_{t-q} \quad (7.36)$$

ここで， $\{\varepsilon_t\} \sim \text{i.i.d.}(0, \Sigma)$ であり \mathbf{y}_t と ε_t は m 次元ベクトル， Φ_k, Θ_l, Σ は， $m \times m$ 行列である．このモデルは，行列のラグ多項式 $\Phi(L) = I_m - \Phi_1 L - \cdots - \Phi_p L^p$ と $\Theta(L) = I_m - \Theta_1 L - \cdots - \Theta_q L^q$ を定義することにより，次のように表すことができる．

$$\Phi(L) \mathbf{y}_t = \Theta(L) \varepsilon_t \quad (7.37)$$

定常性は，特性方程式 $|\Phi(x)| = 0$ の根の絶対値がすべて 1 より大きい場合に保証される．このとき，一変量の場合と同様に，VMA_m(∞) 表現が可能となる．

多変量の VARMA モデルの推定，特に，MA 部分を含むモデルの推定は，一変量の ARMA モデルに比べれば，格段に複雑になるので，VAR モデルが使われるのが普通である．VAR モデルによる分析は，予測以外にも，さまざまな目的のために使うことができる．例えば，因果性の問題など，さまざまな応用については，山本 (1988), Hamilton (1994) を参照されたい．

以下では，非定常な VAR モデルについて考えてみよう．今， $\{\mathbf{y}_t\}$ の各成分は I(1) であるとする．したがって， $\Delta \mathbf{y}_t$ は定常となり，次のような表現が可能である．

$$\begin{aligned} \Delta \mathbf{y}_t &= \sum_{j=0}^{\infty} C_j \varepsilon_{t-j} = C(L) \varepsilon_t \\ &= [C(1) + (C(L) - C(1))] \varepsilon_t = C(1) \varepsilon_t + \Delta \tilde{C}(L) \varepsilon_t \end{aligned} \quad (7.38)$$

ここで， $\tilde{C}(L)$ は無限次の行列ラグ多項式であり， $\{\tilde{C}(L)\varepsilon_t\}$ は定常な確率過程となる．このとき，(7.37) と (7.38) より，

$$\Phi(L) \Delta \mathbf{y}_t = \Delta \varepsilon_t = \Phi(L) C(L) \varepsilon_t$$

が成り立つから、 $\Phi(L)C(L) = \Delta I_m$ となる。 I_m は、 m 次の単位行列である。したがって、

$$\Phi(1)C(1) = 0 \quad (7.39)$$

を得る。式 (7.38) の最左辺と最右辺の左側から $\Phi(1)$ をかけて、(7.39) を使くと、

$$\Phi(1)\mathbf{y}_t = \Phi(1)\tilde{C}(L)\boldsymbol{\varepsilon}_t$$

が得られる。右辺は定常であるから、このことは、本来は非定常な $\{\mathbf{y}_t\}$ に対して、その成分の線形結合で定常となるものがあることを示している。このとき、 $\{\mathbf{y}_t\}$ の各成分は共和分 (cointegration) の関係にあるという。また、 $\Phi(1)$ は共和分行列、その各行は共和分ベクトルと呼ばれる。

Johansen (1995) は、共和分ランクが r のときの VAR(p) モデルを推定するために、まず、(7.37) の $\Phi(L)$ を次のように表現した。

$$\Phi(L) = \Phi(1)L + \Phi(L) - \Phi(1)L = \Phi(1)L + \Delta\Gamma(L)$$

ここで、

$$\Gamma(L) = I_q - \Gamma_1 L^1 - \dots - \Gamma_{p-1} L^{p-1}, \quad \Gamma_j = - \sum_{i=j+1}^p \Phi_i$$

である。この表現を使って、VAR(p) モデルは次のように変形できる。

$$\Delta\mathbf{y}_t = \gamma\boldsymbol{\alpha}'\mathbf{y}_{t-1} + \Gamma_1\Delta\mathbf{y}_{t-1} + \dots + \Gamma_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\varepsilon}_t \quad (7.40)$$

ここで、 $\gamma\boldsymbol{\alpha}' = -\Phi(1)$ であり、 γ と $\boldsymbol{\alpha}$ は $q \times r$ の行列で、ランクはともに r である。ただし、この分解は一意的ではない。

式 (7.40) のモデルは、ECM (Error Correction Model: 誤差修正モデル) と呼ばれる。右辺第 1 項の $\{\boldsymbol{\alpha}'\mathbf{y}_{t-1}\}$ は I(1) 系列の線形結合からなる r 次元のベクトルである。この項以外は定常であるから、 r が正である限り定常であり、原系列の長期的な均衡関係からの乖離を表している。このように、共和分関係が存在する場合には、階差系列に加えてレベルの変数も混在することになり、通常の ARIMA モデルでは表現不可能な形をしている。別の観点からいえば、共和分関係にあるような多変量時系列においては、階差変換後の定常系列だけを分析すると、情報の損失をもたらすことになる。この点は、一変量時系列の分析と異なる点である。

共和分に関連する統計学的な問題としては、第 4 節で議論した単位根検定が先行する問題としてあり、その上で、共和分関係が存在するかどうかに関する共和分検定、1 次独立な共和分関係の個数を決める共和分ランクの検定などがある。詳しくは、Hamilton (1994), Johansen (1995) を参照されたい。

8 おわりに

本稿では、最初に一変量時系列に関する基礎理論について述べたが、季節性を含むような時系列モデルの説明は割愛した。そのためのモデルは、ARIMA モデルや ARFIMA モデルを拡張したもので、それぞれ、SARIMA モデル、SARFIMA モデルと呼ばれる。S は、Seasonal の頭文字である。SARIMA モデルについては Box-Jenkins-Reinsel (1994)、SARFIMA モデルについては Journal of Econometrics (1996), Vol. 73 の特集号が参考になるであろう。これらのモデルは、予測に使われるとともに、季節調整済み系列を算出するためにも使われる。この点については、国友 (2006) を参照されたい。

応用面のトピックスとしては、単位根検定、構造変化を考慮したモデル、ウェーブレット解析などを説明したが、この他にも、近年はパネル・データの分析が盛んに行われている。時間軸の相関だけでなく、クロス・セクション間の相関を考慮した場合の基本的な文献としては、Arellano-Bond (1991)、Blundell-Bond (1998) がある。

参考文献

- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, Wiley, New York.
- Arellano, M. and Bond, S. (1991). "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies*, **58**, 277-297.
- Beran, J. (1994). *Statistics for Long-Memory Processes*, Chapman & Hall, New York.
- Blundell, R. and Bond, S. (1998). "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, **87**, 115-143.
- Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd Edition, Holden-Day, San Francisco.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 3rd Edition, Springer, New York.
- Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). "Efficient tests for an autoregressive unit root," *Econometrica*, **64**, 813-836.

- Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd Edition, Wiley, New York.
- Hamilton, J. D. (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, **57**, 357-384.
- Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton.
- Hosking, J. R. M. (1981). "Fractional differencing," *Biometrika*, **68**, 165-176.
- Hosking, J. R. M. (1996). "Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series," *Journal of Econometrics*, **73**, 261-284.
- Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.
- 国友直人編 (2006). 「季節調整法 X-12-ARIMA と日本の官庁統計」, 東京大学 CIRJE 研究報告書シリーズ No. R-5 .
- Nabeya, S. and Tanaka, K. (1990). "A general approach to the limiting distribution for estimators in time series regression with nonstable autoregressive errors," *Econometrica*, **58**, 145-163.
- Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge.
- Perron, P. (1989). "The great crash, the oil price shock, and the unit root hypothesis," *Econometrica*, **57**, 1361-1401.
- Phillips, P. C. B. and Perron, P. (1988). "Testing for a unit root in time series regression," *Biometrika*, **75**, 335-346.
- Schwarz, G. (1978). "Estimating the dimension of a model," *Annals of Statistics*, **6**, 461-464.
- Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, Wiley, New York.
- 田中勝人 (2006). 『現代時系列分析』, 岩波書店.
- Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*, Springer, New York.

Tong, H. (1983). *Threshold Models in Non-Linear Time Series Analysis*, Springer, New York.

Vogelsang, T. J. and Perron, P. (1998). “Additional tests for a unit root allowing for a break in the trend function at an unknown time,” *International Economic Review*, **39**, 1073-1100.

山本拓 (1988). 『経済の時系列分析』, 創文社.

Zivot, E. and Andrews, D. W. (1992). “Further evidence on the great crash, the oil price shock, and the unit root hypothesis,” *Journal of Business and Economic Statistics*, **10**, 251-270.

第7章 確率微分方程式の母数推定

内田 雅之¹

(大阪大学大学院基礎工学研究科 教授)

セミマルチンゲールは連続時間確率過程の重要なクラスであり、特に確率微分方程式はその具体的なモデルの一つである。確率微分方程式によって定義される拡散過程の統計解析は、連続パスデータに基づいた統計的漸近推測の基礎的な理論の構築がなされ、最近では離散観測データによる統計的漸近理論が活発に研究されている。本章では、確率微分方程式モデルの母数推定について概観し、離散観測データに基づく小さな拡散をもつ拡散過程の母数推定について解説する。

¹uchida@sigmath.es.osaka-u.ac.jp

1 序

統計モデルとして、次の確率微分方程式 (stochastic differential equation, SDE) によって定義される 1 次元拡散過程 X を考える.

$$dX_t = b(X_t, \alpha)dt + \sigma(X_t, \beta)dw_t, \quad t \in [0, T], \quad X_0 = x_0.$$

ここで、 $b: \mathbf{R} \times \Theta_\alpha \rightarrow \mathbf{R}$, $\sigma: \mathbf{R} \times \Theta_\beta \rightarrow \mathbf{R}$ で w は 1 次元標準 Wiener 過程、 x_0 は非確率的とする. b はドリフト、 σ は拡散係数と呼ばれる. α と β は未知の母数 (パラメータ) で、ドリフト b と拡散係数 σ の関数形は既知とする. さらに、ドリフトのパラメータ空間 Θ_α と拡散係数のパラメータ空間 Θ_β は、それぞれ \mathbf{R} の部分集合とする.

連続時間確率過程である拡散過程のデータとしては、連続パスデータ $\mathbf{X}_T := \{X_t\}_{t \in [0, T]}$ と離散観測データ $\mathbf{X}_n := \{X_{kh_n}\}_{k=0,1,\dots,n}$ 、ただし $nh_n = T$ 、が考えられる. 歴史的には、連続パスデータに基づいた統計的漸近理論の研究が先行し、現在までに数多くの研究がなされている. しかしながら、現実的には連続パスデータを観測することは不可能である. それゆえに、応用上は離散的に観測されたデータに基づいた漸近理論が重要であり、最近活発に研究されている. Yoshida [31] は $h_n \rightarrow 0$, $nh_n \rightarrow \infty$ かつ $nh_n^3 \rightarrow 0$ の下で、近似尤度関数を用いて、ドリフトパラメータ α と拡散係数パラメータ β の推定量を導出し、漸近正規性および漸近有効性を証明した. さらに、ドリフトパラメータ α の推定量の収束率と拡散係数パラメータ β の推定量の収束率が異なることを示した. Kessler [14] は $h_n \rightarrow 0$, $nh_n \rightarrow \infty$ かつ $nh_n^l \rightarrow 0$ ($l \geq 2$) の下で、Ito-Taylor 展開を用いて疑似尤度関数を構成し、ドリフトパラメータ α と拡散係数パラメータ β の同時推定を行い、漸近正規性及び漸近有効性をもつことを証明した.

観測区間 $[0, T]$ を固定した場合、刻み幅 $h_n = T/n$ を十分小さくとっても、一般にはドリフトパラメータ α に対して、一致推定量すら構成できない. しかしながら、拡散項が小さい状況では小さな拡散をもつ拡散過程を考えることにより、この問題を解決することができる. そこで、次の SDE によって定義される 1 次元の小さな拡散をもつ拡散過程を考える.

$$dX_t = b(X_t, \alpha)dt + \varepsilon\sigma(X_t, \beta)dw_t, \quad t \in [0, T], \quad \varepsilon \in (0, 1], \quad X_0 = x_0,$$

ここで ε は微小摂動パラメータと呼ばれる. 離散観測における小さな拡散をもつ拡散過程の母数推定については、Genon-Catalot [4] と Laredo [21] が拡散係数パラメータ β が既知という状況で、ドリフトパラメータ α の漸近正規

性及び漸近有効性をもつ推定量を導出した. Sørensen and Uchida [28] はドリフトパラメータ α と拡散係数パラメータ β の同時推定を行い, 漸近正規性及び漸近有効性について考察した.

本稿では, 確率微分方程式モデルの統計的推測の基礎として, 離散観測データに基づくパラメータ推定の解説を行う. 問題を簡単にするために, 拡散過程は1次元で, ドリフト及び拡散係数のパラメータ空間はそれぞれ1次元で考えることにするが, 多次元化は可能である. 2節では, 離散観測データにおけるエルゴード的拡散過程のパラメータ推定について, オイラー・丸山近似に基づいた擬似対数尤度関数とその擬似最尤推定量の漸近的結果を動機付けを交えながら紹介する. 3節では, 離散観測データにおける小さな拡散をもつ拡散過程のパラメータ推定について基本的な結果を概観する. 4節では, 結論と今後の展望について言及する. 本稿は確率微分方程式を統計モデルとして統計的推測を行っているので, 5節で確率微分方程式及び確率解析の基本的事項を述べておく. 必要に応じて参照していただきたい. 特に, 2節, 3節では, $(\Omega, \mathcal{F}, P; \{\mathcal{F}_t\}_{t \geq 0})$ を通常条件を満たすフィルター付き確率空間とし, その上で定義された1次元 \mathcal{F}_t -標準 Wiener 過程 $\{w_t\}_{t \geq 0}$ が与えられているとする. 通常条件を満たすフィルター付き確率空間については, 5節を参照していただきたい.

2 離散観測における確率微分方程式の母数推定

2.1 統計モデル

統計モデルとして, 次の確率微分方程式によって定義される1次元拡散過程 X を考える.

$$dX_t = b(X_t, \alpha)dt + \sigma(X_t, \beta)dw_t, \quad t \in [0, T], \quad X_0 = x_0, \quad (2.1)$$

ここで, $b: \mathbf{R} \times \Theta_\alpha \rightarrow \mathbf{R}$, $\sigma: \mathbf{R} \times \Theta_\beta \rightarrow \mathbf{R}$, w は1次元標準 Wiener 過程, x_0 は非確率的とする. さらに, $\theta := (\alpha, \beta) \in \Theta := \Theta_\alpha \times \Theta_\beta \subset \mathbf{R}^2$ とし, Θ はコンパクトな矩形集合とする. $\theta_0 = (\alpha_0, \beta_0)$ を SDE モデル (2.1) のパラメータの真値として, $\theta_0 \in \text{Int}(\Theta)$ と仮定する.

例 1. (*Ornstein-Uhlenbeck* 過程). $b(x, \alpha) = -\alpha x$, $\sigma(x, \beta) = \beta$ とすると,

$$dX_t = -\alpha X_t dt + \beta dw_t, \quad t \in [0, T], \quad X_0 = x_0, \quad (2.2)$$

ここで $\alpha > 0$, $\beta > 0$ とする.

例 2. (*The hyperbolic diffusion process*). $b(x, \alpha) = -\alpha \frac{x}{\sqrt{1+x^2}}$, $\sigma(x, \beta) = \beta$ とすると,

$$dX_t = -\alpha \frac{X_t}{\sqrt{1+X_t^2}} dt + \beta dw_t, \quad t \in [0, T], \quad X_0 = x_0, \quad (2.3)$$

ここで $\alpha > 0$, $\beta > 0$ とする.

2.2 離散観測データ

SDE モデル (2.1) において $\theta = \theta_0$ とした真のモデル

$$dX_t = b(X_t, \alpha_0) dt + \sigma(X_t, \beta_0) dw_t, \quad t \in [0, T], \quad X_0 = x_0 \quad (2.4)$$

から各時点 $t_k^n = kh_n$, ($k = 0, 1, \dots, n$) で観測された離散データを考える. ここで, $t_n^n = nh_n = T$ とし, h_n は刻み幅と呼ばれる. すなわち, 離散観測データは $\mathbf{X}_n = \{X_{t_k^n}\}_{k=0,1,\dots,n}$ である. 離散観測データは大きく分けて次の 3 つのタイプがある.

- (i) 観測区間は固定で微小観測幅 ($nh_n = T$ は固定で $h_n \rightarrow 0$).
- (ii) 観測幅は固定で観測区間が増大 ($h_n = \Delta$ で $nh_n = n\Delta = T \rightarrow \infty$).
- (iii) 微小観測幅で, かつ観測区間が増大 ($h_n \rightarrow 0$ かつ $nh_n = T \rightarrow \infty$).

タイプ (i) は高頻度データと呼ばれ, 連続時間確率過程である拡散過程を統計モデルとして採用する強い動機付けとなる. タイプ (ii) は主に時系列解析で取り扱われるもので, 応用上重要である. タイプ (iii) は高頻度データ (タイプ (i)) と長期間観測データ (タイプ (ii)) を併せたもので, 情報量が最も多い. 2 節では, タイプ (iii) の離散観測データを取り扱う. すなわち, $n \rightarrow \infty$ の時, $h_n \rightarrow 0$ かつ $T = nh_n \rightarrow \infty$ の下で考える.

例 3. (1985/1/1~2004/12/31 までの 20 年間を日ごとに観測した場合). $T = 20$, $h_n = T/n = 20/(1461 \times 5) = 20/7305 \approx 0.0027$. $nh_n^2 = Th_n \approx 20 \times 0.0027 = 0.054$.

例 4. (2000/1/1~2004/12/31 までの 60ヶ月間を日ごとに観測した場合). $T = 60$, $h_n = T/n = 60/1827 \approx 0.032$. (i) $nh_n^2 = Th_n \approx 60 \times 0.032 = 1.92 \gg 0$. (ii) $nh_n^3 = Th_n^2 \approx 60 \times 0.032^2 \approx 0.061$.

例 5. (1985/1/1~2004/12/31 までの 20 年間を月ごとに観測した場合). $T = 20$, $h_n = T/n = 20/240 \approx 0.083$. (i) $nh_n^2 = Th_n \approx 20 \times 0.083 = 1.66 \gg 0$. (ii) $nh_n^3 = Th_n^2 \approx 20 \times 0.083^2 \approx 0.13$. (iii) $nh_n^4 = Th_n^3 \approx 20 \times 0.083^3 \approx 0.011$.

仮定している極限 ($T \rightarrow \infty$ かつ $h_n \rightarrow 0$) に近い状況を確保するために、 T の単位をどのようにとるかが重要である。例 4 の場合、 $T = 5$ (年) とすると漸近理論が機能しない可能性がある。さらに、推定量の漸近正規性を示すために、 h_n のオーダーについても注意を払う必要がある。例 3 の場合: $Th_n = nh_n^2 \approx 0$. 例 4 の場合: $Th_n^2 = nh_n^3 \approx 0$. 例 5 の場合: $Th_n^3 = nh_n^4 \approx 0$.

2.3 サンプルパスの発生 (simulation)

SDE モデル (2.1) の解 $X = \{X_t\}_{t \in [0, T]}$ のサンプルパス (標本路) をシミュレーションにより発生する仕方について述べる。ただし、 T は固定しておく。まず、真値 α_0, β_0 を与えておく。 $t > s \geq 0$ に対して、

$$X_t - X_s = \int_s^t b(X_u, \alpha_0) du + \int_s^t \sigma(X_u, \beta_0) dw_u, \quad X_0 = x_0.$$

一般に X_t の明示解を得るのは困難なので、次の差分近似を用いて近似解 $X_t^{(n)}$ を考える。

$$\begin{aligned} X_{t_k}^{(n)} - X_{t_{k-1}}^{(n)} &= b(X_{t_{k-1}}^{(n)}, \alpha_0)(t_k - t_{k-1}) + \sigma(X_{t_{k-1}}^{(n)}, \beta_0)(w_{t_k} - w_{t_{k-1}}), \\ X_{t_0}^{(n)} &= x_0. \end{aligned}$$

これをオイラー・丸山近似という。さらに、 $h_n = t_k^n - t_{k-1}^n$, $\Delta W_{t_{k-1}^n} := w_{t_k^n} - w_{t_{k-1}^n} \sim N(0, h_n)$ に注意すると、 $k = 1, \dots, n$ に対して、逐次的に

$$X_{t_k}^{(n)} = x_{t_{k-1}^n} + b(x_{t_{k-1}^n}, \alpha_0)h_n + \sigma(x_{t_{k-1}^n}, \beta_0)\Delta W_{t_{k-1}^n}, \quad X_{t_{k-1}^n}^{(n)} = x_{t_{k-1}^n}$$

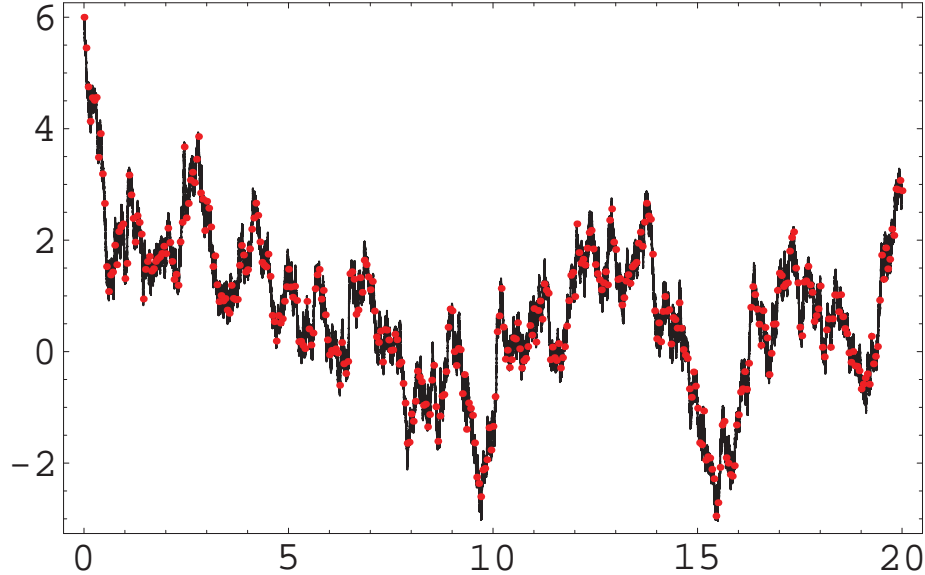
を正規乱数を用いてシミュレーションすることによって、 X の近似 $X^{(n)} = \{X_{t_0}^{(n)}, X_{t_1}^{(n)}, \dots, X_{t_n}^{(n)}\}$ が発生できる。ここで、 $\Delta W_{t_0}^n, \Delta W_{t_1}^n, \dots, \Delta W_{t_{n-1}}^n$ は互いに独立であることに注意する。オイラー・丸山近似で得られた $X_T^{(n)} = X_{t_n}^{(n)}$ は正則条件の下で、次の評価を得る (Kloeden and Platen [15]):ある定数 $c > 0$ が存在して、 $E \left[\left\| X_T - X_T^{(n)} \right\|^2 \right] \leq c\sqrt{h_n}$.

例 6. (Ornstein-Uhlenbeck 過程 のサンプルパスと離散観測データ)

$$dX_t = -\alpha X_t dt + \beta dw_t, \quad X_0 = x_0, \quad t \in [0, T].$$

図 7-1 の実線 (黒) は $x_0 = 6$, $\alpha = 1$, $\beta = 2$, $T = 20$, $h_n = 1/10000$ とした場合のサンプルパスである。赤の点は、実線のデータを刻み幅 $1/20$ でプロットしたものである。

図 7-1: サンプルパスと高頻度データ



実際問題として、 $X = \{X_t\}_{t \in [0, T]}$ の連続なサンプルパスを観測することは困難であるので、離散観測データに基づいた統計的推測が必要である。例 6 の図 7-1 で言えば、実線が連続パスデータに相当するが、実際には観測不可能であるので、連続サンプルパスから離散的に観測されたデータ (赤の点) を用いて統計解析が行われる。

2.4 統計モデルの仮定

SDE モデル (2.1) の解 X のとりうる値の集合 (state space) は $(-\infty, \infty)$ とする。 X が $\theta = \theta_0$ に対してエルゴード性をもつとは、ある不変測度 μ_{θ_0} が存在して、 $\int_{-\infty}^{\infty} |f(x)| \mu_{\theta_0}(dx) < \infty$ となる任意の可測関数 f に対して、 $T \rightarrow \infty$ の時、

$$\frac{1}{T} \int_0^T f(X_t) dt \rightarrow \int_{-\infty}^{\infty} f(x) \mu_{\theta_0}(dx)$$

が確率 1 で成り立つこととする。ここで、 μ_{θ_0} は $\theta = \theta_0$ に対する X の不変測度、すなわち、任意の $t > 0$ と任意の 1 次元ボレル集合 B に対して、 $\mu_{\theta_0}(B) = \int_{-\infty}^{\infty} P_x(X_t \in B) \mu_{\theta_0}(dx)$ を満たす確率測度である。ただし、 P_x は出発点が x である X の法則を表す。今後用いる記号について述べておく。

- X の scale measure $s(x, \theta)dx$ と speed measure $\xi(x, \theta)dx$ を次で定義する. $x \in \mathbf{R}, \theta \in \Theta$ に対して,

$$s(x, \theta) := \exp \left\{ - \int_0^x \frac{2b(y, \alpha)}{\sigma^2(y, \beta)} dy \right\},$$

$$\xi(x, \theta) := \frac{1}{\sigma^2(x, \beta)} \exp \left\{ \int_0^x \frac{2b(y, \alpha)}{\sigma^2(y, \beta)} dy \right\} = \frac{1}{\sigma^2(x, \beta)s(x, \theta)}.$$

- $\partial_x = \partial/\partial x, \delta_\alpha = \partial/\partial \alpha, \delta_\beta = \partial/\partial \beta$ とする. また $\mathbf{R} \times \Theta$ 上で定義される実数値関数 $f(x, \theta)$ に対して, $\delta_\theta f(x, \theta) = (\delta_\alpha f(x, \theta), \delta_\beta f(x, \theta))^*$ とする. ここで $*$ は転置を表す.
- $C_{\uparrow}^{k,l}(\mathbf{R} \times \Theta_\alpha)$ を次の条件を満たす関数 f の空間とする: (i) $f(x, \alpha)$ は $\mathbf{R} \times \Theta_\alpha$ 上で定義された実数値関数で, x について k 回連続微分可能で, $n = 0, 1, \dots, k$ に対して, ある定数 $C > 0$ が存在して, すべての x について, $\sup_\alpha |\partial_x^n f(x, \alpha)| \leq C(1 + |x|)^C$. (ii) $n = 0, 1, \dots, k$ に対して, $\partial_x^n f(x, \alpha)$ は α について l 回連続微分可能で, $\nu = 0, 1, \dots, l$ に対して, 定数 $C > 0$ が存在して, すべての x に対して $\sup_\alpha |\delta_\alpha^\nu \partial_x^n f(x, \alpha)| \leq C(1 + |x|)^C$.
- 2×2 の (Fisher 情報) 行列 $I(\theta_0)$ を

$$I(\theta_0) = \begin{pmatrix} I_b(\theta_0) & 0 \\ 0 & I_\sigma(\theta_0) \end{pmatrix}$$

とする. ただし,

$$I_b(\theta_0) = \int_{-\infty}^{\infty} \left(\frac{\delta_\alpha b(x, \alpha_0)}{\sigma(x, \beta_0)} \right)^2 \mu_{\theta_0}(dx),$$

$$I_\sigma(\theta_0) = 2 \int_{-\infty}^{\infty} \left(\frac{\delta_\beta \sigma(x, \beta_0)}{\sigma(x, \beta_0)} \right)^2 \mu_{\theta_0}(dx).$$

- P_θ を SDE(2.1) の解 X の法則とし, E_θ を P_θ の下での期待値とする. \rightarrow^p および \rightarrow^d はそれぞれ P_{θ_0} の下での確率収束および分布収束を表す.

次の仮定をおく.

A1 すべての $\theta \in \Theta$ に対して,

$$\int_0^\infty s(x, \theta) dx = \infty, \quad \int_{-\infty}^0 s(x, \theta) dx = \infty,$$

$$\Xi(\theta) = \int_{-\infty}^{\infty} \xi(x, \theta) dx < \infty.$$

A2 (i) ある定数 $L > 0$ が存在して, 任意の $x, y \in \mathbf{R}, \theta \in \Theta$ に対し

$$|b(x, \alpha) - b(y, \alpha)| + |\sigma(x, \beta) - \sigma(y, \beta)| \leq L|x - y|.$$

(ii) $b(x, \alpha) \in C_{\uparrow}^{2,3}(\mathbf{R} \times \Theta_{\alpha})$. (iii) $\sigma(x, \beta) \in C_{\uparrow}^{2,3}(\mathbf{R} \times \Theta_{\beta})$ であり, さらに, $\inf_{x, \beta} \sigma^2(x, \beta) > 0$. (v) $p \geq 0$ に対して, $\sup_t E_{\theta_0}[|X_t|^p] < \infty$.

A3 (i) $I(\theta_0)$ は正則である. (ii) すべての x に対して, $b(x, \alpha) = b(x, \alpha_0) \implies \alpha = \alpha_0$. すべての x に対して, $\sigma^2(x, \beta) = \sigma^2(x, \beta_0) \implies \beta = \beta_0$.

エルゴード性について, 次が成り立つ (例えば, Kutoyants [19]).

補題 1. $A1, A2$ -(i) そして, すべての x, β に対して, $\sigma^2(x, \beta) > 0$ を仮定する. この時, X はエルゴード性をもつ. 特に, 不変測度 $\mu_{\theta}(x)$ はルベーグ測度に関する密度 $\xi(x, \theta)/\Xi(\theta)$ をもつ.

仮定 $A2$ -(i) から SDE(2.1) の解の存在と一意性が保証される (5 節の定理 6). 仮定 $A3$ -(ii) は識別可能条件 (identifiability condition) とよばれ, 推定量の一致性を示すときに必要となる.

例 7. (*Ornstein-Uhlenbeck* 過程のエルゴード性と不変測度) *scale measure* $s(x, \theta)dx$ と *speed measure* $\xi(x, \theta)dx$ は

$$\begin{aligned} s(x, \theta) &= \exp \left\{ \frac{\alpha}{\beta^2} x^2 \right\}, \\ \xi(x, \theta) &= \frac{1}{\beta^2} \exp \left\{ -\frac{\alpha}{\beta^2} x^2 \right\} \end{aligned}$$

となり, $\alpha, \beta > 0$ から

$$\begin{aligned} \int_0^{\infty} \exp \left\{ \frac{\alpha}{\beta^2} x^2 \right\} dx &= \infty, \quad \int_{-\infty}^0 \exp \left\{ \frac{\alpha}{\beta^2} x^2 \right\} dx = \infty, \\ \Xi(\theta) &= \int_{-\infty}^{\infty} \frac{1}{\beta^2} \exp \left\{ -\frac{\alpha}{\beta^2} x^2 \right\} dx = \sqrt{\frac{\pi}{\alpha\beta^2}} < \infty \end{aligned}$$

である. ゆえに, 補題 1 より *Ornstein-Uhlenbeck* 過程 (例 1) はエルゴード性を持ち, 不変測度 μ_{θ} は

$$\mu_{\theta}(dx) = \frac{\xi(x, \theta)}{\Xi(\theta)} dx = \frac{1}{\sqrt{\pi\beta^2/\alpha}} \exp \left\{ -\frac{x^2}{\beta^2/\alpha} \right\} dx$$

となり, 平均 0, 分散 $\beta^2/(2\alpha)$ の正規密度を持つ.

例 8. (Ornstein-Uhlenbeck 過程の期待値と分散) SDE (2.2) は具体的に解くことができ,

$$X_t = e^{-\alpha t} \left(X_0 + \int_0^t \beta e^{\alpha s} dw_s \right). \quad (2.5)$$

実際, この X_t が解であることは, 伊藤の公式 (5 節の定理 5) を用いて, 次のように確かめられる. $f(t, v) = e^{-\alpha t} v, V_t = X_0 + \int_0^t \beta e^{\alpha s} dw_s$ として, 伊藤の公式を用いると,

$$\begin{aligned} X_t &= f(t, V_t) = V_0 - \alpha \int_0^t e^{-\alpha s} V_s ds + \int_0^t e^{-\alpha s} \beta e^{\alpha s} dw_s \\ &= X_0 - \alpha \int_0^t X_s ds + \int_0^t \beta dw_s. \end{aligned}$$

特に, X_t の期待値 $E_\theta[X_t]$ と分散 $V_\theta[X_t]$ は (2.5) と 5 節の定理 4 より,

$$\begin{aligned} E_\theta[X_t] &= x_0 e^{-\alpha t}, \\ V_\theta[X_t] &= \beta^2 \frac{1 - e^{-2\alpha t}}{2\alpha}. \end{aligned}$$

X_t の明示解から, $p \geq 0$ に対して $\sup_t E_\theta[|X_t|^p] < \infty$ も示せる.

例 9. (Ornstein-Uhlenbeck 過程に対する Fisher 情報行列) 例 7 より, 不変測度 μ_θ は $N(0, \beta^2/(2\alpha))$ の密度をもつ. よって

$$I(\theta_0) = \begin{pmatrix} \frac{1}{\beta_0^2} \int_{-\infty}^{\infty} x^2 \mu_{\theta_0}(dx) & 0 \\ 0 & \frac{2}{\beta_0^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2\alpha_0} & 0 \\ 0 & \frac{2}{\beta_0^2} \end{pmatrix}.$$

$$\alpha_0, \beta_0 > 0 \text{ より, } \det(I(\theta_0)) = \frac{1}{\alpha_0 \beta_0^2} \neq 0.$$

2.5 尤度関数とマルチンゲール推定関数

統計モデル (2.1) に対して, $X_{t_{k-1}} = x$ が与えられた条件の下での X_{t_k} の推移密度関数を $y \mapsto p(h_n, x, y; \theta)$ と表す. $\mathbf{X}_n = (X_{t_0}^{t_n}, X_{t_1}^{t_n}, \dots, X_{t_n}^{t_n})$ の ($X_{t_0}^{t_n} = x_0$ が与えられた条件の下での) 同時密度関数は

$$f_{X_{t_0}^{t_n}, X_{t_1}^{t_n}, \dots, X_{t_n}^{t_n}}(x_{t_0}, x_{t_1}, \dots, x_{t_n}; \theta) = \prod_{k=1}^n p(h_n, x_{t_{k-1}}, x_{t_k}; \theta)$$

であり, 対数尤度関数は

$$l_n(\theta) = \sum_{k=1}^n \log p(h_n, X_{t_{k-1}^n}, X_{t_k^n}; \theta)$$

である. 最尤推定量 $\hat{\theta}_n^{(ML)}$ を

$$l_n(\hat{\theta}_n^{(ML)}) = \sup_{\theta} l_n(\theta)$$

で定義する. 推移密度関数 p が明示的に求まれば, 尤度解析を用いて, 統計的推測が可能となる. $l_n(\theta)$ が θ について微分可能である場合,

$$S_n(\theta) := \delta_{\theta} l_n(\theta) = \sum_{k=1}^n \frac{\delta_{\theta} p(h_n, X_{t_{k-1}^n}, X_{t_k^n}; \theta)}{p(h_n, X_{t_{k-1}^n}, X_{t_k^n}; \theta)}$$

はスコア関数と呼ばれ, 最尤推定量 $\hat{\theta}_n^{(ML)}$ は $S_n(\hat{\theta}_n^{(ML)}) = 0$ となる. \mathcal{F}_n を $\{X_{t_k^n}\}_{k=0,1,\dots,n}$ で生成される σ -加法族, すなわち, $\mathcal{F}_n = \sigma(X_{t_0^n}, X_{t_1^n}, \dots, X_{t_n^n})$ とすると, 正則条件の下で,

$$\begin{aligned} E_{\theta}[S_n(\theta) - S_{n-1}(\theta) | \mathcal{F}_{n-1}] &= E_{\theta} \left[\frac{\delta_{\theta} p(h_n, X_{t_{n-1}^n}, X_{t_n^n}; \theta)}{p(h_n, X_{t_{n-1}^n}, X_{t_n^n}; \theta)} \middle| X_{t_{n-1}^n} \right] \\ &= \int_{\mathbf{R}} \frac{\delta_{\theta} p(h_n, X_{t_{n-1}^n}, y; \theta)}{p(h_n, X_{t_{n-1}^n}, y; \theta)} p(h_n, X_{t_{n-1}^n}, y; \theta) dy \\ &= \int_{\mathbf{R}} \delta_{\theta} p(h_n, X_{t_{n-1}^n}, y; \theta) dy \\ &= \delta_{\theta} \int_{\mathbf{R}} p(h_n, X_{t_{n-1}^n}, y; \theta) dy = 0. \end{aligned}$$

ゆえに, スコア関数 $S_n(\theta)$ は \mathcal{F}_n -マルチンゲールとなる. よって, マルチンゲール中心極限定理を用いて, 正則条件の下で, $h_n \rightarrow 0$ かつ $nh_n \rightarrow \infty$ の時, 最尤推定量 $\hat{\theta}_n^{(ML)}$ の漸近正規性を示すことができる.

以後, 条件付き期待値を $m(x, \theta) := E_{\theta}[X_{t_k^n} | X_{t_{k-1}^n} = x]$ とし, 条件付き分散を $v(x, \theta) := V_{\theta}[X_{t_k^n} | X_{t_{k-1}^n} = x]$ とする.

例 10. (*Ornstein-Uhlenbeck* 過程の推移密度関数) 例 8 より,

$$X_t = x_0 e^{-\alpha t} + \beta e^{-\alpha t} \int_0^t e^{\alpha s} dw_s.$$

$\int_0^t e^{\alpha s} dw_s$ は Wiener 積分と呼ばれる確率積分の特別な場合であり, 平均 0, 分散 $\int_0^t e^{2\alpha s} ds = \frac{1}{2\alpha}(e^{2\alpha t} - 1)$ の正規分布に従う. よって, 推移密度関数は

$$p(t, x, y; \theta) = \frac{1}{\sqrt{\pi\beta^2(1 - e^{-2\alpha t})/\alpha}} \exp\left\{-\frac{(y - e^{-\alpha t}x)^2}{\beta^2(1 - e^{-2\alpha t})/\alpha}\right\}.$$

つまり, $X_0 = x$ が与えられた下での X_t の条件付き分布 $\mathcal{L}(X_t | X_0 = x)$ は

$$\mathcal{L}(X_t | X_0 = x) \sim N\left(e^{-\alpha t}x, \frac{\beta^2(1 - e^{-2\alpha t})}{2\alpha}\right)$$

であり, 対数尤度関数は

$$l_n(\theta) = -\frac{1}{2} \sum_{k=1}^n \left\{ \log(2\pi) + \log\left(\frac{\beta^2(1 - e^{-2\alpha h_n})}{2\alpha}\right) + \frac{(X_{t_k^n} - e^{-\alpha h_n} X_{t_{k-1}^n})^2}{\beta^2(1 - e^{-2\alpha h_n})/(2\alpha)} \right\} \quad (2.6)$$

である. さらに, 条件付き期待値 $m(x, \theta)$ と条件付き分散 $v(x, \theta)$ は

$$\begin{aligned} m(x, \theta) &= e^{-\alpha h_n} x, \\ v(x, \theta) &= \frac{\beta^2(1 - e^{-2\alpha h_n})}{2\alpha} \end{aligned}$$

となることから,

$$\mathcal{L}(X_{t_k^n} | X_{t_{k-1}^n} = x) \sim N(m(x, \theta), v(x, \theta))$$

であり, 対数尤度関数 (2.6) は条件付き期待値と条件付き分散を用いて, 次式で与えられる.

$$l_n(\theta) = -\frac{1}{2} \sum_{k=1}^n \left\{ \log(2\pi) + \log v(X_{t_{k-1}^n}, \theta) + \frac{(X_{t_k^n} - m(X_{t_{k-1}^n}, \theta))^2}{v(X_{t_{k-1}^n}, \theta)} \right\}.$$

例 10 において, Ornstein-Uhlenbeck 過程の推移密度関数 p が明示的に求められたので, 最尤推定量 $\hat{\theta}_n^{(ML)}$ を求めることができる. 実際に, Ornstein-Uhlenbeck 過程の場合, $h_n \rightarrow 0$ かつ $nh_n \rightarrow \infty$ の下, 最尤推定量 $\hat{\theta}_n^{(ML)}$ の漸近正規性及び漸近有効性を示すことができる.

しかしながら, 一般に SDE (2.1) の解 X の推移密度関数は明示的に導出できない. よって, 離散観測データに基づいた SDE モデルのパラメータ推

定においては、一般に尤度解析を用いることができない。そこで、例 10 を参考にして、条件付き分布 $\mathcal{L}(X_{t_k^n} | X_{t_{k-1}^n} = x)$ を平均 $m(x, \theta)$ 、分散 $v(x, \theta)$ の正規分布で近似した次の近似対数尤度関数を考える。

$$U_n(\theta) = \sum_{k=1}^n U(X_{t_{k-1}^n}, X_{t_k^n}, \theta), \quad (2.7)$$

$$U(x, y, \theta) = -\frac{1}{2} \log(2\pi v(x, \theta)) - \frac{(y - m(x, \theta))^2}{2v(x, \theta)}. \quad (2.8)$$

この近似 $U_n(\theta)$ は局所的に正規分布で近似していることから局所正規近似とよばれ、拡散過程の推移密度関数の近似としては基本的である。本節では、 $U_n(\theta)$ をコントラスト関数とよぶことにする。最大コントラスト推定量 $\hat{\theta}_n^{(M)}$ を

$$U_n(\hat{\theta}_n^{(M)}) = \sup_{\theta} U_n(\theta)$$

で定義する。さらに、コントラスト関数 $U_n(\theta)$ が θ について微分可能な時、最大コントラスト推定量は θ に関する推定方程式

$$M_n(\theta) = \begin{pmatrix} M_n^{(\alpha)}(\theta) \\ M_n^{(\beta)}(\theta) \end{pmatrix} := \begin{pmatrix} \delta_{\alpha} U_n(\theta) \\ \delta_{\beta} U_n(\theta) \end{pmatrix} = 0$$

を満足する M -推定量と見なせる。ここで、

$$\begin{aligned} M_n(\theta) &= \sum_{k=1}^n \frac{\delta_{\theta} m(X_{t_{k-1}^n}, \theta)}{v(X_{t_{k-1}^n}, \theta)} \left[X_{t_k^n} - m(X_{t_{k-1}^n}, \theta) \right] \\ &+ \frac{1}{2} \sum_{k=1}^n \frac{\delta_{\theta} v(X_{t_{k-1}^n}, \theta)}{v^2(X_{t_{k-1}^n}, \theta)} \left[(X_{t_k^n} - m(X_{t_{k-1}^n}, \theta))^2 - v(X_{t_{k-1}^n}, \theta) \right]. \end{aligned}$$

また、

$$\begin{aligned} &E_{\theta}[M_n(\theta) - M_{n-1}(\theta) | \mathcal{F}_{n-1}] \\ &= \frac{\delta_{\theta} m(X_{t_{n-1}^n}, \theta)}{v(X_{t_{n-1}^n}, \theta)} E_{\theta}[(X_{t_n^n} - m(X_{t_{n-1}^n}, \theta)) | X_{t_{n-1}^n}] \\ &+ \frac{1}{2} \frac{\delta_{\theta} v(X_{t_{n-1}^n}, \theta)}{v^2(X_{t_{n-1}^n}, \theta)} E_{\theta}[(X_{t_n^n} - m(X_{t_{n-1}^n}, \theta))^2 - v(X_{t_{n-1}^n}, \theta) | X_{t_{n-1}^n}] \\ &= 0 \end{aligned}$$

が成り立つことから、 $M_n(\theta)$ は \mathcal{F}_n -マルチンゲールになっている。推定関数 $M_n(\theta)$ はマルチンゲール推定関数と呼ばれる。コントラスト関数 $U_n(\theta)$ が対

数尤度関数 $l_n(\theta)$ の近似であるのに対して、マルチンゲール推定関数 $M_n(\theta)$ はスコア関数 $S_n(\theta)$ の近似になっている。 $m(x, \theta)$ と $v(x, \theta)$ が明示的に求められる SDE モデルについては、コントラスト関数 $U_n(\theta)$ 及びマルチンゲール推定関数 $M_n(\theta)$ が明示的に導出できる。先述したように、マルチンゲール推定関数が明示的に求められる場合は、正則条件の下、 M -推定量の漸近正規性が証明できる。しかしながら、条件付き期待値 $m(x, \theta)$ と条件付き分散 $v(x, \theta)$ が明示的に求まることは稀である。例えば、例 2 の SDE モデルに対する条件付き期待値 $m(x, \theta)$ と条件付き分散 $v(x, \theta)$ は明示的に導出できない。SDE モデルの場合は、推移密度関数だけではなく、条件付きモーメント、特に条件付き分散を明示的に導出するのは一般には困難である。そこで、離散観測に基づいた SDE モデルのパラメータ推定においては、次節で述べる疑似尤度解析が重要な役割を果たすことになる。

2.6 疑似対数尤度関数

本節では、前節で構成したコントラスト関数 (2.7)-(2.8) の近似を行う。そのために、2.3 節で用いたオイラー・丸山近似を考える。近似解を $Z_{t_{k-1}^n} := X_{t_{k-1}^n}^{(n)}$ とおくと、

$$Z_{t_k^n} = Z_{t_{k-1}^n} + b(Z_{t_{k-1}^n}, \alpha)h_n + \sigma(Z_{t_{k-1}^n}, \beta)(w_{t_k^n} - w_{t_{k-1}^n}), \quad Z_0 = x_0.$$

さらに、 $w_{t_k^n} - w_{t_{k-1}^n}$ は平均 0、分散 h_n の正規分布に従うことに注意すると、条件付き分布 $\mathcal{L}(Z_{t_k^n} | Z_{t_{k-1}^n} = z)$ は、

$$\mathcal{L}(Z_{t_k^n} | Z_{t_{k-1}^n} = z) \sim N(z + b(z, \alpha)h_n, \sigma^2(z, \beta)h_n).$$

オイラー・丸山近似は局所的にガウスの構造が出てくるため、局所正規近似の一種と考えることができる。実際に、仮定 A2 の下で次が成り立つ。

$$E_\theta[X_{t_k^n} | X_{t_{k-1}^n}] = X_{t_{k-1}^n} + h_n b(X_{t_{k-1}^n}, \alpha) + O_p(h_n^2),$$

$$V_\theta[X_{t_k^n} | X_{t_{k-1}^n}] = h_n \sigma^2(X_{t_{k-1}^n}, \alpha) + O_p(h_n^2).$$

よって、オイラー・丸山近似は、コントラスト関数 (2.7)-(2.8) の条件付き期待値及び条件付き分散を h_n のオーダーまで近似したものと考えられる。

そこで、次の疑似対数尤度関数 g_n を考える。

$$g_n(\theta) = \sum_{k=1}^n g(X_{t_{k-1}^n}, X_{t_k^n}, \theta), \quad (2.9)$$

$$g(x, y, \theta) = -\frac{1}{2} \log(\sigma^2(x, \beta)) - \frac{(y - x - h_n b(x, \alpha))^2}{2h_n \sigma^2(x, \beta)}. \quad (2.10)$$

本来, コントラスト関数 (2.7)-(2.8) を近似するならば, (2.10) の右辺の第1項は $-\frac{1}{2} \log(2\pi h_n \sigma^2(x, \beta))$ とすべきであるが, θ について最大化するので, 定数項を無視した (2.10) を考えることにする.

疑似最尤推定量 $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)^*$ を

$$g_n(\hat{\theta}_n) = \sup_{\theta} g_n(\theta)$$

で定義する. 次の結果は, Kessler [14] の特別な場合である.

定理 1. $A1-A3$ の下, $nh_n^2 \rightarrow 0$ の時,

$$\begin{pmatrix} \sqrt{nh_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \rightarrow^d N(0, I^{-1}(\theta_0)).$$

オイラー・丸山近似に基づいて疑似対数尤度関数 g_n を構成したが, $h_n \rightarrow 0$ かつ $nh_n \rightarrow \infty$ だけでは, 定理 1 は成立せず, $nh_n^2 \rightarrow 0$ が必要である. これは, オイラー・丸山近似がコントラスト関数 (2.7)-(2.8) の条件付き期待値及び条件付き分散を h_n のオーダーまでしか近似していないことに関係する. 定理 1 は例 3 の離散観測データ ($nh_n^2 \rightarrow 0$ に近い状況) に対して適用可能であるが, 例 4-5 の離散観測データではうまく機能しない可能性がある. 条件 $nh_n^2 \rightarrow 0$ を緩めるには, オイラー・丸山近似よりも精度の良い近似を行う必要がある. Kessler [14] は Ito-Taylor 展開を用いて, この条件を緩めたが, 疑似対数尤度関数は当然複雑になる.

定理 1 から, ドリフトパラメータの推定量の収束率と拡散係数パラメータの推定量の収束率は異なることに注意する. 特に, 拡散係数パラメータの推定量はドリフトパラメータの推定量よりも早く収束する. これは, タイプ (iii) の離散観測データに基づいた拡散過程のパラメータ推定における一つの特徴であり, 興味深い点である. 定理 1 により疑似最尤推定量 $\hat{\theta}_n$ は漸近正規性をもつことがわかるが, 漸近有効性まで保証するならば, 条件を強める必要がある. Gobet [8] は, 正則条件の下, すべての $u, v \in \mathbf{R}$ に対して, P_{θ_0} の下で,

$$l_n(\alpha_0 + u/\sqrt{nh_n}, \beta_0 + v/\sqrt{n}) - l_n(\alpha_0, \beta_0) \rightarrow^d \begin{pmatrix} u \\ v \end{pmatrix}^* Z - \frac{1}{2} \begin{pmatrix} u \\ v \end{pmatrix}^* I(\theta_0) \begin{pmatrix} u \\ v \end{pmatrix}$$

が成り立つことを示した. ここで, Z は平均0, 分散共分散行列 $I(\theta_0)$ の2次元正規確率変数である. これは尤度に対する局所漸近正規性 (Local Asymptotic Normality, LAN) とよばれる. $I(\theta_0)$ が非退化の場合, ミニマックス定理から $I(\theta_0)^{-1}$ は正則な推定量の漸近分散に対する下限になる. この事実と定理 1 から疑似最尤推定量 $\hat{\theta}_n$ の漸近有効性が保証される.

3 小さな拡散をもつ拡散過程の母数推定

本節では、次の確率微分方程式によって定義される 1 次元拡散過程を考える。

$$dX_t^\varepsilon = b(X_t^\varepsilon, \alpha)dt + \varepsilon\sigma(X_t^\varepsilon, \beta)dw_t, \quad t \in [0, T], \quad \varepsilon \in (0, 1], \quad X_0^\varepsilon = x_0, \quad (3.1)$$

ここで ε は微小摂動パラメータと呼ばれる既知の定数で、それ以外は SDE モデル (2.1) と同様とする。 ε は微小であるので、SDE (3.1) の解 X_t^ε は常微分方程式

$$dX_t = b(X_t, \alpha)dt, \quad t \in [0, T], \quad X_0 = x_0$$

の解 X_t に小さなノイズが加えられたものと考えられる。SDE (2.1) はランダムな現象をモデリングしているのに対して、SDE (3.1) は非確率的な現象に小さなノイズを加えたモデリングである。拡散項が小さいことから、SDE (3.1) の解 X_t^ε は小さな拡散をもつ拡散過程とよばれる。小さな拡散をもつ拡散過程の応用例としては、オプションの価格付け理論の研究がある (Yoshida [33], Kunitomo and Takahashi [16])。

2 節で用いた記号に加えて、以下を定義しておく。

- X_t^0 は SDE (3.1) において $\varepsilon = 0$ に対応する次の常微分方程式の解とする:

$$dX_t^0 = b(X_t^0, \alpha_0)dt, \quad X_0^0 = x_0.$$

- $C_{\uparrow}^{\infty, l}(\mathbf{R} \times \Theta_\alpha)$ を次の条件を満たす関数 f の空間とする: (i) $f(x, \alpha)$ は $\mathbf{R} \times \Theta_\alpha$ 上で定義された実数値関数で、 x について無限回連続微分可能で、 $n \geq 0$ に対して、ある定数 $C > 0$ が存在して、すべての x について、 $\sup_\alpha |\partial_x^n f(x, \alpha)| \leq C(1 + |x|)^C$. (ii) $n \geq 0$ に対して、 $\partial_x^n f(x, \alpha)$ は α について l 回連続微分可能で、 $\nu = 0, 1, \dots, l$ に対して、定数 $C > 0$ が存在して、すべての x に対して $\sup_\alpha |\delta_\alpha^\nu \partial_x^n f(x, \alpha)| \leq C(1 + |x|)^C$.
- 2×2 の (Fisher 情報) 行列 $\mathcal{I}(\theta_0)$ を

$$\mathcal{I}(\theta_0) = \begin{pmatrix} \mathcal{I}_b(\theta_0) & 0 \\ 0 & \mathcal{I}_\sigma(\theta_0) \end{pmatrix}$$

とする。ここで、

$$\begin{aligned} \mathcal{I}_b(\theta_0) &= \int_0^T \left(\frac{\delta_\alpha b(X_s^0, \alpha_0)}{\sigma(X_s^0, \beta_0)} \right)^2 ds, \\ \mathcal{I}_\sigma(\theta_0) &= 2 \int_0^T \left(\frac{\delta_\beta \sigma(X_s^0, \beta_0)}{\sigma(X_s^0, \beta_0)} \right)^2 ds. \end{aligned}$$

例 11. (小さな拡散をもつ拡散過程のサンプルパス)

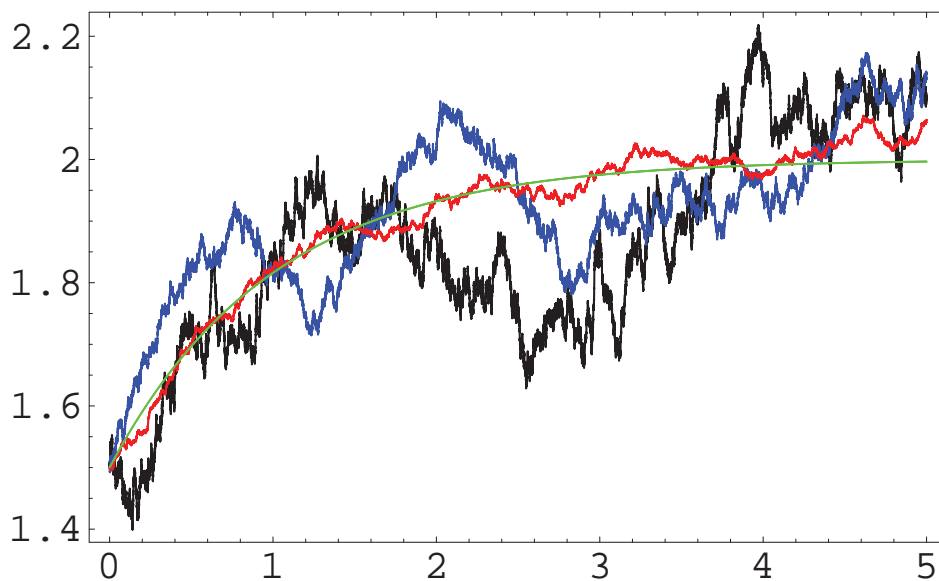
$$dX_t^\varepsilon = -(X_t^\varepsilon - \alpha)dt + \varepsilon \sqrt{\frac{\beta + (X_t^\varepsilon)^2}{1 + (X_t^\varepsilon)^2}} dw_t, \quad t \in [0, T], \quad \varepsilon \in (0, 1], \quad X_0^\varepsilon = x_0$$

で定義される小さな拡散をもつ拡散過程を考える. ただし, $\alpha, \beta > 0$ とする. $\varepsilon = 0$ のとき, 常微分方程式を解くことにより,

$$X_t^0 = x_0 e^{-t} + \alpha(1 - e^{-t})$$

である. 図 7-2 は $x_0 = 1.5, \alpha = 2, \beta = 1, T = 5$ として, 黒の実線は $\varepsilon = 0.25$, 青の実線は $\varepsilon = 0.15$, 赤の実線は $\varepsilon = 0.05$, 緑の実線は $\varepsilon = 0$ で, それぞれ $h_n = 1/10000$ とした場合の X_t^ε のサンプルパスである. ε が小さくなると, X_t^ε のサンプルパスが, X_t^0 (緑の実線) に近づくことがわかる.

図 7-2: 小さな拡散をもつ拡散過程のサンプルパス



本節では, 区間 $[0, T]$ を固定して考えるので, $T = 1$ としても一般性を失わない. よって, 取り扱うデータは区間 $[0, 1]$ 上で等間隔に観測された離散データとする. つまり, $\mathbb{X}_n^\varepsilon = \{X_{t_k}^\varepsilon\}_{k=0,1,\dots,n}, t_k = k/n$ である. 離散観測データ \mathbb{X}_n^ε に基づいて未知パラメータ α と β の推定を行う. 極限は, $\varepsilon \rightarrow 0$ かつ $n \rightarrow \infty$ の下で考える. 2 節ではタイプ (iii) の離散観測データ, つまり $t_k^n = kh_n, h_n \rightarrow 0, t_n^n = nh_n \rightarrow \infty$ を取り扱ったが, 本節はタイプ (i) の離散観測データであることに注意する.

3.1 疑似最尤推定量

SDE モデル (3.1) について, 次の仮定をおく. ただし, $T = 1$ としていることに注意する.

B1 (i) ある定数 $C > 0$ が存在して, すべての $x, y \in \mathbf{R}, \theta \in \Theta$ に対して,

$$|b(x, \alpha) - b(y, \alpha)| + |\sigma(x, \beta) - \sigma(y, \beta)| \leq C|x - y|.$$

(ii) $b(x, \alpha) \in C_{\uparrow}^{\infty,3}(\mathbf{R} \times \Theta_{\alpha})$. (iii) $\sigma(x, \beta) \in C_{\uparrow}^{\infty,3}(\mathbf{R} \times \Theta_{\beta})$ であり, さらに, $\inf_{x,\beta} \sigma^2(x, \beta) > 0$.

B2 (i) $\mathcal{I}(\theta_0)$ は正則である. (ii) $\alpha \neq \alpha_0 \Rightarrow$ ある $t \in [0, 1]$ が存在して, $b(X_t^0, \alpha) \neq b(X_t^0, \alpha_0)$. $\beta \neq \beta_0 \Rightarrow$ ある $t \in [0, 1]$ が存在して, $\sigma^2(X_t^0, \beta) \neq \sigma^2(X_t^0, \beta_0)$.

オイラー・丸山近似を用いて, 次の疑似対数尤度関数を得る.

$$\mathcal{G}_{\varepsilon,n}(\theta) = -\frac{1}{2} \sum_{k=1}^n \left\{ \log \sigma^2(X_{t_{k-1}}^{\varepsilon}, \beta) + \frac{n (X_{t_k}^{\varepsilon} - X_{t_{k-1}}^{\varepsilon} - \frac{1}{n} b(X_{t_{k-1}}^{\varepsilon}, \alpha))^2}{\varepsilon^2 \sigma^2(X_{t_{k-1}}^{\varepsilon}, \beta)} \right\}.$$

ここで 2.6 節の疑似対数尤度関数 (2.9)-(2.10) と同様に定数項を無視していることに注意する.

$\hat{\theta}_{\varepsilon,n} = (\hat{\alpha}_{\varepsilon,n}, \hat{\beta}_{\varepsilon,n})^*$ を

$$\mathcal{G}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) = \sup_{\theta \in \Theta} \mathcal{G}_{\varepsilon,n}(\theta).$$

で定義された疑似最尤推定量とする. この時, 疑似最尤推定量 $\hat{\theta}_{\varepsilon,n}$ に対して次が成り立つ (Sørensen and Uchida [28]).

定理 2. $B1$ - $B2$ の下, $(\varepsilon\sqrt{n})^{-1} = O(1)$ の時,

$$\begin{pmatrix} \varepsilon^{-1}(\hat{\alpha}_{\varepsilon,n} - \alpha_0) \\ \sqrt{n}(\hat{\beta}_{\varepsilon,n} - \beta_0) \end{pmatrix} \rightarrow^d N(0, \mathcal{I}(\theta_0)^{-1}).$$

$(\varepsilon\sqrt{n})^{-1} \rightarrow 0$ の時, 拡散係数パラメータの推定量 $\hat{\beta}_{\varepsilon,n}$ はドリフトパラメータの推定量 $\hat{\alpha}_{\varepsilon,n}$ よりも早く収束する. また, $\varepsilon \rightarrow 0$ かつ $n \rightarrow \infty$ の設定において, ε は n に依存して小さくなる, すなわち, $\varepsilon := \varepsilon_n \rightarrow 0$ ($n \rightarrow \infty$) とする立場があるが, これはデータ数が大きくなるに従って, 拡散項が小さくなることを意味する. もちろん定理 2 はこの場合を含むが, データ数 n を ε に依存させて大きくとる場合, すなわち $n := n_{\varepsilon} \rightarrow \infty$ ($\varepsilon \rightarrow 0$) も考慮している.

4 結論と展望

本稿では、確率微分方程式に対する統計解析の基礎として、離散観測に基づく確率微分方程式の母数推定に関するごく限られた内容の紹介にとどめたが、この分野のレビューとしては Prakasa Rao [24, 25] に研究結果がまとめられている。最近の研究結果について言えば、Genon-Catalot and Jacod [5, 6] はタイプ (i) の設定で拡散係数パラメータ β の推定量の一致性、漸近混合正規性および漸近有効性を証明した。しかしながら、タイプ (i) では、ドリフトパラメータ α については、一致推定量すら構成できないことに注意する。Bibby and Sørensen [1] はタイプ (ii) の離散観測データに対してマルチンゲール推定関数を構成した。その推定関数からドリフトパラメータ α と拡散係数パラメータ β の推定量を導出し、それらが一致性および漸近正規性をもつことを示した。Uchida [29] は小さな拡散をもつ拡散過程に対して、ドリフト及び拡散係数が共有パラメータを持つ状況で、推定量の漸近正規性を示した。Uchida [30] は緩い条件の下で、小さな拡散をもつ拡散過程のドリフトパラメータ α の M -推定量を構成し、漸近有効性を証明した。上述したように、離散観測における確率微分方程式の統計的推測は最近急速に発展しているが、それらは連続パスデータにおける研究結果に基づいていることを軽視してはならない。例えば、Kutoyants [17] は Ibragimov and Has'minskii [10] の理論を小さな拡散をもつ拡散過程に移植し、Küchler and Sørensen [20] は指数型分布族の理論をジャンプ付き拡散過程に一般化した。セミマルチンゲールに対する 1 次の漸近理論の詳細については、Kutoyants [18, 19], Prakasa Rao [25] などを参照。確率微分方程式の解の存在や一意性の十分条件等の基本的事項についても非常に重要であり、拡散過程の統計的モデリングに不可欠である。また、拡散過程の統計解析は漸近理論を多用する。確率微分方程式及び確率解析については、例えば、渡辺 [35], Ikeda and Watanabe [11], Karatzas and Shreve [13], 舟木 [3], 長井 [23] などがあり、小さな拡散をもつ拡散過程については、Freidlin and Wentzell [2], 確率過程の極限定理については、Jacod and Shiryaev [12], シミュレーション等で必要不可欠な確率数値解析については、Kloeden and Platen [15] がある。これらの精緻な道具を駆使し、確率微分方程式の統計的推測理論が展開され、金融工学や数理ファイナンス等への応用 (林・吉田 [9]) が活発に行われている。最近では、拡散過程にとどまらず、Shimizu and Yoshida [27] や Masuda [22] によって、飛躍付き拡散過程に対する漸近理論が研究されている。

高次の漸近理論については、連続パスデータに対して Yoshida [32] がマリアバン解析を使って、小さな拡散をもつ拡散過程における最尤推定量の漸

近展開の正当性を証明して以来, 拡散過程にとどまらず, 一般の確率過程に対して漸近展開の研究が盛んに行われている. 詳細については, 吉田 [34], Sakamoto and Yoshida [26] を参照. 連続パスデータに基づいた確率微分方程式の統計的漸近理論に関する研究結果から, 1 次の漸近理論に対しては, セミマルチンゲールアプローチが有効であり, 高次漸近分布論に対してはマリアバン解析が重要な役割を果たすことがわかった. これにより, 離散観測における確率微分方程式モデルの高次漸近理論の発展が期待される.

5 補遺

本節では, 確率微分方程式の基本的事項について簡単に述べる. 確率微分方程式及び確率解析については多くのテキスト (渡辺 [35], Ikeda and Watanabe [11], Karatzas and Shreve [13], 舟木 [3], 長井 [23]) があるので, 証明については, そちらを参照していただきたい.

5.1 Wiener 過程とマルチンゲール

定義 1. (Ω, \mathcal{F}, P) を確率空間とする.

- (i) 実数値確率変数の族 $\{X_t\}_{t \geq 0}$ を 1 次元確率過程という.
- (ii) 1 次元確率過程 $\{X_t\}_{t \geq 0}$ が連続とは次が成り立つことである.

$$P(\{\omega \in \Omega \mid [0, \infty) \ni t \mapsto X_t(\omega) \in \mathbf{R} \text{ は連続}\}) = 1.$$

定義 2. (Ω, \mathcal{F}, P) を確率空間とする. $\{\mathcal{F}_t\}_{t \geq 0}$ がフィルトレーションとは,

- (i) 各 t に対して, \mathcal{F}_t は σ -加法族.
- (ii) $0 \leq s < t$ に対して, $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$

となることをいう. $(\Omega, \mathcal{F}, P; \{\mathcal{F}_t\}_{t \geq 0})$ をフィルター付き確率空間という.

以下では, フィルター付き確率空間が与えられているものとする.

定義 3. $\{X_t\}_{t \geq 0}$ を 1 次元確率過程とする.

- (i) 各 t に対して, X_t が \mathcal{F}_t -可測であるとき, $\{X_t\}_{t \geq 0}$ が \mathcal{F}_t に適合しているという.

(ii) 各 t に対して, 写像 $(s, \omega) \in [0, t] \times \Omega \mapsto X_s(\omega) \in \mathbf{R}$ が $\mathcal{B}([0, t]) \times \mathcal{F}_t$ に関して可測であるとき, $\{X_t\}_{t \geq 0}$ が \mathcal{F}_t -発展的可測であるという.

定義 4. 1次元確率過程 $\{w_t\}_{t \geq 0}$ が 次の (i)-(iv) を満たすとき, 1次元 \mathcal{F}_t -標準 Wiener 過程であるという.

- (i) $\{w_t\}_{t \geq 0}$ は \mathcal{F}_t に適合している.
- (ii) $P(w_0 = 0) = 1$.
- (iii) 各 $t > s \geq 0$ に対して, $w_t - w_s$ は \mathcal{F}_s と独立であり, 平均 0, 分散 $t - s$ の 1次元正規確率変数である.
- (iv) $\{w_t\}_{t \geq 0}$ は連続確率過程である.

定理 3. 1次元 \mathcal{F}_t -標準 Wiener 過程 $\{w_t\}_{t \geq 0}$ に対して, 以下が成り立つ.

- (i) $p = 1, 2, \dots$ に対して, $E[w_t^{2p}] = (2p - 1)!! t^p$, $E[w_t^{2p-1}] = 0$.
- (ii) a.s. ω に対し, $w_t(\omega)$ は t の関数として, 至るところ微分不可能である.
- (iii) $w_t(\omega)$ は t の関数として 2 次変動をもつ.
- (iv) a.s. ω に対し, $w_t(\omega)$ は t の関数として有界変動でない.

定義 5. $p \geq 1$ とする. 1次元確率過程 $X = \{X_t\}_{t \geq 0}$ が \mathcal{F}_t -連続 L^p マルチンゲールとは次の (i)-(iv) が成り立つことをいう.

- (i) 各 t に対して, $E[|X_t|^p] < \infty$.
- (ii) $X = \{X_t\}_{t \geq 0}$ は \mathcal{F}_t に適合している.
- (iii) 各 $t > s \geq 0$ に対して, $E[X_t | \mathcal{F}_s] = X_s$ a.s.
- (iv) $\{X_t\}_{t \geq 0}$ は連続である.

1次元 \mathcal{F}_t -標準 Wiener 過程 $\{w_t\}_{t \geq 0}$ は \mathcal{F}_t -連続 L^p マルチンゲールである.

5.2 確率積分と伊藤の公式

$(\Omega, \mathcal{F}, P; \{\mathcal{F}_t\}_{t \geq 0})$ が次の (i)-(iii) を満たすとき, “通常条件” を満たすという. (i) (Ω, \mathcal{F}, P) は完備確率空間, つまり, $M \subset N \in \mathcal{F}$ で $P(N) = 0$ ならば $M \in \mathcal{F}$. (ii) \mathcal{F}_t は右連続, つまり, 任意の $t \geq 0$ に対して $\mathcal{F}_t = \bigcap_{s > t} \mathcal{F}_s$. (iii) \mathcal{F}_0 はすべての P -零集合を含む, つまり, $\mathcal{N} := \{A \subset \Omega \mid \text{ある } B \in \mathcal{F} \text{ が存在して } A \subset B, P(B) = 0\} \subset \mathcal{F}_0$.

以下, $(\Omega, \mathcal{F}, P; \{\mathcal{F}_t\}_{t \geq 0})$ を通常条件を満たすフィルター付き確率空間とし, その上に 1次元 \mathcal{F}_t -標準 Wiener 過程 $\{w_t\}_{t \geq 0}$ が与えられているとする.

w_t は a.s. ω に対し, t の関数として有界変動でない (定理 3-(iv)) ので, 積分

$$\int_0^T f_t(\omega) dw_t(\omega) \quad (5.1)$$

を ω ごとに通常 Lebesgue-Stieltjes 積分として定義できない. しかしながら, 標準 Wiener 過程 w の性質及び確率論的手法を用いることによって, $\mathcal{L}_{loc}^2(\mathcal{F}_t)$ に属する被積分関数 f に対して標準 Wiener 過程 w に関する確率積分 (5.1) が定義できる. ここで, $\mathcal{L}_{loc}^2(\mathcal{F}_t) = \{\{f_t\}_{t \geq 0} \mid 1 \text{次元 } \mathcal{F}_t \text{発展的可測過程で, 任意の } T \geq 0 \text{ に対して, } P(\int_0^T f_t^2(\omega) dt < \infty) = 1\}$. 定義の仕方については, 渡辺 [35] や長井 [23] を参照.

$\mathcal{L}^2(\mathcal{F}_t) = \{\{f_t\}_{t \geq 0} \mid 1 \text{次元 } \mathcal{F}_t \text{発展的可測過程で, 任意の } T \geq 0 \text{ に対して, } E[\int_0^T f_t^2(\omega) dt] < \infty\}$ とする. $\mathcal{L}^2(\mathcal{F}_t) \subset \mathcal{L}_{loc}^2(\mathcal{F}_t)$ であることに注意.

定理 4. $f \in \mathcal{L}^2(\mathcal{F}_t)$ とする.

(i) $\left\{ \int_0^t f_s dw_s \right\}_{t \geq 0}$ は \mathcal{F}_t -連続 L^2 マルチンゲールである.

(ii) 任意の $T \geq 0$ に対して,

$$\begin{aligned} E \left[\int_0^T f_t dw_t \right] &= 0, \\ E \left[\left(\int_0^T f_t dw_t \right)^2 \right] &= E \left[\int_0^T f_t^2 dt \right]. \end{aligned}$$

$\mathcal{L}_{loc}^1(\mathcal{F}_t) = \{\{f_t\}_{t \geq 0} \mid 1 \text{次元 } \mathcal{F}_t \text{発展的可測過程で, 任意の } T \geq 0 \text{ に対して, } P(\int_0^T |f_t(\omega)| dt < \infty) = 1\}$ とする.

定理 5 (伊藤の公式). X_0 は \mathcal{F}_0 -可測で, $b \in \mathcal{L}_{loc}^1(\mathcal{F}_t)$, $\sigma \in \mathcal{L}_{loc}^2(\mathcal{F}_t)$ とする. 1次元確率過程 (伊藤過程)

$$X_T = X_0 + \int_0^T b_t dt + \int_0^T \sigma_t dw_t$$

を考える. このとき, $f \in C^2(\mathbf{R})$ に対して,

$$\begin{aligned} f(X_T) &= f(X_0) + \int_0^T \frac{\partial f}{\partial x}(X_t) b_t dt + \frac{1}{2} \int_0^T \frac{\partial^2 f}{\partial x^2}(X_t) \sigma_t^2 dt \\ &\quad + \int_0^T \frac{\partial f}{\partial x}(X_t) \sigma_t dw_t \end{aligned}$$

が成り立つ. また, $(t, x) \rightarrow f(t, x)$ が x について 2 回微分可能で, t について 1 回微分可能な関数で, これらの偏導関数が (t, x) について連続であるとき,

$$\begin{aligned} f(T, X_T) &= f(0, X_0) + \int_0^T \frac{\partial f}{\partial t}(t, X_t) dt \\ &\quad + \int_0^T \frac{\partial f}{\partial x}(t, X_t) b_t dt + \frac{1}{2} \int_0^T \frac{\partial^2 f}{\partial x^2}(t, X_t) \sigma_t^2 dt \\ &\quad + \int_0^T \frac{\partial f}{\partial x}(t, X_t) \sigma_t dw_t \end{aligned}$$

が成り立つ.

5.3 確率微分方程式

$(\Omega, \mathcal{F}, P; \{\mathcal{F}_t\}_{t \geq 0})$ を通常条件を満たすフィルター付き確率空間とし, その上で定義された 1 次元 \mathcal{F}_t -標準 Wiener 過程 $\{w_t\}_{t \geq 0}$ が与えられているとする. b, σ を \mathbf{R} から \mathbf{R} への Borel 可測関数とする.

定義 6. $X_t(\omega) : [0, \infty) \times \Omega \rightarrow \mathbf{R}$ が出発点 x_0 をもつ確率微分方程式

$$dX_t = b(X_t)dt + \sigma(X_t)dw_t, \quad X_0 = x_0 \quad (5.2)$$

の解であるとは, 次が満たされることをいう.

(i) $X = \{X_t\}_{t \geq 0}$ は \mathcal{F}_t に適合した連続確率過程で, 任意の $T \geq 0$ に対して

$$P \left(\int_0^T \{ |b(X_t)| + \sigma^2(X_t) \} dt < \infty \right) = 1.$$

(ii) $P(X_0 = x_0) = 1,$

(iii) $X_t = X_0 + \int_0^t b(X_s)dt + \int_0^t \sigma(X_s)dw_t, \forall t \geq 0$ が確率 1 で成り立つ.

定理 6. リプシッツ (*Lipschitz*) 条件 : ある定数 $L > 0$ が存在して, 任意の $x, y \in \mathbf{R}$ に対し

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq L|x - y|$$

が成り立つとする. このとき, 確率微分方程式 (5.2) は一意解 $X = \{X_t\}_{t \geq 0}$ をもつ. ここで, 一意性は次の意味である: X と Y がともに (5.2) の解ならば, $P(X_t = Y_t, \forall t \geq 0) = 1.$

参考文献

- [1] Bibby, B. M. and Sørensen, M. (1995). Martingale estimating functions for discretely observed diffusion processes. *Bernoulli* **1**, 17-39.
- [2] Freidlin M. I. and Wentzell A. D. (1998). *Random perturbations of dynamical systems, second edition.* Springer-Verlag, New York.
- [3] 舟木直久 (1997). 確率微分方程式. 岩波書店.
- [4] Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. *Statistics* **21**, 99-116.
- [5] Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Ann. Inst. Henri Poincaré Probab. Statist.* **29**, 119-151.
- [6] Genon-Catalot, V. and Jacod, J. (1994). Estimation of the diffusion coefficient for diffusion processes: random sampling. *Scand. J. Statist.* **21**, 193-221.
- [7] Gobet, E. (2001). Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli* **7**, 899-912.
- [8] Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.* **38**, 711-737.
- [9] 林高樹・吉田朋広 (2008). 高頻度金融データと統計科学. 21 世紀の統計科学 I: 社会・経済の統計科学, 第 10 章, 東京大学出版会.

- [10] Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical estimation*. Springer Verlag, New York.
- [11] Ikeda, N. and Watanabe, S. (1989). *Stochastic differential equations and diffusion processes, second edition*. North-Holland/Kodansha, Tokyo.
- [12] Jacod, J. and Shiryaev, A. N. (1987). *Limit theorems for stochastic processes*. Springer, Heidelberg.
- [13] Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus, second edition*. Springer-Verlag, New York.
- [14] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211-229.
- [15] Kloeden, P. E. and Platen, E. (1992). *Numerical solution of stochastic differential equations*. Springer-Verlag, New York.
- [16] Kunitomo, N. and Takahashi, A. (2001). The asymptotic expansion approach to the valuation of interest rate contingent claims. *Mathematical Finance*, **11**, 117-151. (2001)
- [17] Kutoyants, Yu. A. (1984). *Parameter estimation for stochastic processes*. Prakasa Rao, B.L.S. (ed.) Heldermann, Berlin.
- [18] Kutoyants, Yu. A. (1994). *Identification of dynamical systems with small noise*. Kluwer, Dordrecht.
- [19] Kutoyants, Yu. A. (2004). *Statistical inference for ergodic diffusion processes*. Springer-Verlag, London.
- [20] Küchler, U. and Sørensen, M. (1997). *Exponential families of stochastic processes*. Springer, New York.
- [21] Laredo, C. F. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. *Ann. Statist.* **18**, 1158-1171.
- [22] Masuda, H. (2007). Ergodicity and exponential β -mixing bound for multidimensional diffusions with jumps. *Stochastic Processes Appl.* **117**, 35-56.
- [23] 長井英生 (1999). 確率微分方程式. 共立出版.

- [24] Prakasa Rao, B. L. S. (1999a) . *Statistical inference for diffusion type processes*. London, Arnold.
- [25] Prakasa Rao, B. L. S. (1999b). *Semimartingales and their statistical inference*. Chapman & Hall/CRC.
- [26] Sakamoto, Y. and Yoshida, N. (2004). Asymptotic expansion formulas for functionals of ϵ -Markov processes with a mixing property. *Ann. Inst. Statist. Math.* **56**, 545-597.
- [27] Shimizu, Y. and Yoshida, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations. *Statist. Infer. Stochast. Process.*, **9**, 227 - 277.
- [28] Sørensen, M. and Uchida, M. (2003). Small diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli* **9**, 1051-1069.
- [29] Uchida, M. (2003). Estimation for dynamical systems with small noise from discrete observations. *J. Japan Statist. Soc.*, **33**, 157-167.
- [30] Uchida, M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions. *Scand. J. Statist.* **31**, 553-566.
- [31] Yoshida, N. (1992a). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.* **41**, 220–242.
- [32] Yoshida, N. (1992b). Asymptotic expansion of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe. *Probab. Theory Relat. Fields* **92**, 275-311.
- [33] Yoshida, N. (1992c). Asymptotic expansion for statistics related to small diffusions. *J. Japan Statist. Soc.* **22**, 139-159.
- [34] 吉田朋広 (2003). Malliavin 解析と数理統計, 数学, **55**, 225-244.
- [35] 渡辺信三 (1975). 確率微分方程式. 産業図書.

21世紀の統計科学3 数理・計算の統計科学

日本統計学会HP版 2011年10月

第III部 統計計算の展開と統計科学

第8章 ブートストラップ

下平英寿

東京工業大学情報理工学研究科

shimo@is.titech.ac.jp

どのような予測や推測であっても必ず不確実性があるから、そのバラツキの程度をコンピュータ・シミュレーションによって見積もることがある。そのような手法のひとつにブートストラップがあり、従来の複雑な数式を単純な反復計算に置き換えて、データのランダムネスを測定する。本章ではバイアス補正した信頼区間の計算法や、最新のマルチスケール・ブートストラップによる信頼度の計算法を紹介する。

1 まえがき

ブートストラップは、データ解析の確からしさを評価するための統計手法のひとつであり、Efron (1979) によって提唱された。誤差推定、信頼区間の構成、仮説検定などに用いられる。従来の複雑な数式に基づく理論を莫大な数値計算による単純なシミュレーションでおきかえる。データ解析の前提となる仮定を緩和し、理論的な解析が困難な問題で広く応用されている。ブートストラップに伴う反復計算は並列化が容易であるから、今後の計算環境を考慮すればその重要性はさらに増すだろう。

理論を計算でおきかえるブートストラップには、やはり理論が必要である。ブートストラップの理論研究は 1980 年代から 1990 年代前半に集中的に行われ、とくに信頼区間の計算法について理論が発展した。この話題を、本稿の 2 節から 7 節であつかう。標準的な教科書には Efron and Tibshirani (1993), Davison and Hinkley (1997), Hall (1992) がある。和書では汪・田栗 (2003) によくまとめられており、その文献案内にあげられた論文リストは解説付きで参考になる。一方、8 節は 7 節を発展させた内容である。これは筆者の研究紹介であり、下平 (2004) の 3 章も参考になる。

2 リサンプリング

まず以下の状況を設定して議論を進める。データを \mathcal{X} で表し、データを構成する要素をベクトル \boldsymbol{x} で表す。データは

$$\mathcal{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$$

のように n 個の要素 $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ で構成されている。データのことをサンプルまたは標本ともいうので、この n はサンプルサイズ (sample size) とよばれる。データを入力すると実数を出力する装置を考える。たとえば、各種の気象データから 100 年後の東京の年平均気温を予測するソフトウェアを想像すればよい。出力値を $\hat{\theta}$ で表し、入出力の関係を $\hat{\theta} = \hat{\theta}(\mathcal{X})$ と書く。

具体的には、次のデータを例題に方法の説明を行う。

$$\mathbf{x}_1 = \begin{bmatrix} 5.20 \\ 12.17 \\ -12.30 \\ 9.85 \\ 6.02 \\ 28.67 \\ 12.21 \\ -1.83 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 30.06 \\ -3.35 \\ 11.72 \\ -10.33 \\ 28.36 \\ -24.15 \\ -4.90 \\ -3.44 \end{bmatrix}, \dots, \mathbf{x}_{100} = \begin{bmatrix} -5.27 \\ 2.29 \\ -15.16 \\ -35.44 \\ -3.47 \\ 19.36 \\ -5.52 \\ -28.85 \end{bmatrix}$$

サンプルサイズは $n = 100$ である。このデータに対して $\hat{\theta} = 14.45$ が出力されたとしておく。この数値例はあくまで説明のため多変量正規分布から人工的に作成した。この詳細はずっと後で述べるが、そのような確率分布のことをあまり気にせず数値例を体感していただきたい。 \mathcal{X} や $\hat{\theta}(\mathcal{X})$ の詳細を知らなくてもそれなりに機能するのがブートストラップ法の良いところである。

ここで $\hat{\theta}$ が推定（または予測）しようとする量を θ で表し、これはデータ生成メカニズムを特徴付けるパラメータのひとつであると考え、そして、 $\hat{\theta}(\mathcal{X})$ は θ の十分に良い推定量であることを仮定する。それにもかかわらず、 \mathcal{X} にひそむランダムネスの影響によって $\hat{\theta}(\mathcal{X})$ も変動する。そこで、 $\hat{\theta}$ の値をどれほど信用してよいのか、その程度を定量的に評価するために考案された方法がブートストラップである。

その意味を理解するために、まず計算機シミュレーションによってデータを人工的に B 個生成したものを

$$\mathcal{X}^1, \dots, \mathcal{X}^B$$

と書く。これから $\hat{\theta}^b = \hat{\theta}(\mathcal{X}^b)$, $b = 1, \dots, B$ を計算したときの結果を図 1 (左) に示す。ただし、反復回数は $B = 10000$ 、真値は仮に $\theta = 10$ とした。真値 θ のまわりに $\hat{\theta}$ が分布しているようすが見える。実際の観測値 $\hat{\theta}$ はこのような分布から偶然に得られたひとつの結果に過ぎない。

この数値例は人工的に作成したデータだったから、 $\hat{\theta}$ のバラツキを計算機シミュレーションで見せることができた。ところが現実には θ の値もデータ生成のメカニズムもよくわからないことが多い。このような状況で偶然に観測した \mathcal{X} だけをつかって $\hat{\theta}$ のバラツキを調べたい。この目的のために、ブー

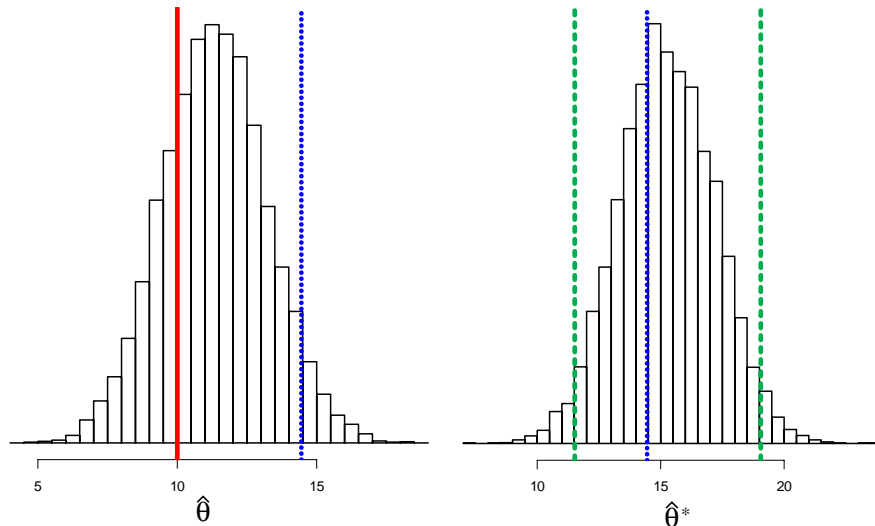


図 1 (左) 計算機シミュレーションによって生成した $\hat{\theta}(x^1), \dots, \hat{\theta}(x^{10000})$ のヒストグラム. 真値は $\theta = 10$ (実線). (右) ブートストラップによって生成した $\hat{\theta}(x^{*1}), \dots, \hat{\theta}(x^{*10000})$ のヒストグラム. 観測値は $\hat{\theta} = 14.45$ (点線). 2本の破線はアルゴリズム 3 によって計算された 95% 信頼区間を表す.

トストラップ法では以下に示した手順を実行する.

アルゴリズム 1 (ノンパラメトリック・ブートストラップ法)

1. 整数 n' をひとつ定める. とくに断らない限り $n' = n$ であるが, 後ほど $n' \neq n$ とする場合も扱う. また十分に大きな反復回数 B を定める. 以下の数値例では $B = 10000$ である.
2. $\{1, 2, \dots, n\}$ から等確率 (すなわち $1/n$) でランダムに整数を選ぶ. これを n' 回繰り返して得られた整数列を $i_1, i_2, \dots, i_{n'}$ とする. 同じ整数が複数回選ばれていてもよい.
3. 得られた整数を添え字とする要素をデータ \mathcal{X} から取り出して

$$\mathbf{x}_1^* = \mathbf{x}_{i_1}, \mathbf{x}_2^* = \mathbf{x}_{i_2}, \dots, \mathbf{x}_{n'}^* = \mathbf{x}_{i_{n'}}$$

とおき, データ \mathcal{X}^* を次式で与える.

$$\mathcal{X}^* = \{\mathbf{x}_1^*, \dots, \mathbf{x}_{n'}^*\}$$

4. 上記 2 と 3 を B 回繰り返して得られた B 個のデータを

$$\mathcal{X}^{*1}, \dots, \mathcal{X}^{*B}$$

とおく. これから $\hat{\theta}^{*b} = \hat{\theta}(\mathcal{X}^{*b})$, $b = 1, \dots, B$ を計算する.

この手順をさきほどのデータへ適用した結果を図 1 (右) に示す. 実際の観測値 $\hat{\theta}$ のまわりに $\hat{\theta}^* = \hat{\theta}(\mathcal{X}^*)$ が分布しているようすが見える. その分散を

$$\hat{\sigma}^2 = \frac{1}{B-1} \sum_{b=1}^B \left\{ \hat{\theta}(\mathcal{X}^{*b}) - \frac{1}{B} \sum_{b'=1}^B \hat{\theta}(\mathcal{X}^{*b'}) \right\}^2 \quad (1)$$

で計算して平方根をとると, $\hat{\theta}^*$ の標準偏差は $\hat{\sigma} = 1.94$ である. そして左右の図を比べると, 分布の形状やバラツキの幅に大差ないことがわかる. 実際に左図へ (1) を適用して標準偏差を計算すれば $\hat{\sigma} = 1.87$ である. したがって, $\hat{\theta}^*$ のバラツキを調べれば, $\hat{\theta}$ のバラツキもわかる. 観測値 $\hat{\theta} = 14.45$ に標準偏差 1.94 を添えて表記すれば, $\hat{\theta}$ をどれほど信用してよいかの目安になる. きわめて単純であるが強力な方法である.

アルゴリズム 1 のことを, ふつうは単に「ブートストラップ」(bootstrap) とよぶ. ノンパラメトリックというのはデータ生成のメカニズムを仮定しないことを意味していて, 後ほど説明するパラメトリック・ブートストラップ (parametric bootstrap) と区別するときにはノンパラメトリック・ブートストラップ (nonparametric bootstrap) とよぶ. \mathcal{X}^* はブートストラップ標本 (bootstrap sample), $\hat{\theta}^*$ はブートストラップ複製 (bootstrap replicate) とよばれることもある. $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B}$ の度数分布, すなわち図 1 (右) はブートストラップ分布 (bootstrap distribution) という.

背景となる母集団または確率分布からデータを得ることを標本抽出またはサンプリング (sampling) というが, ブートストラップはデータから再度データを得るという意味の「リサンプリング」(resampling) の一種である. 同じ要素を複数回選べるので, 復元抽出 (sampling with replacement) ともいう. これに対して, もしアルゴリズム 1 のステップ 2 で同じ添え字は 1 回だけしか選べない (ただし $n' = n$) とすれば, i_1, i_2, \dots, i_n はもとの添え字を並べ替えただけになり, 並べ替え検定で用いられる非復元抽出 (sampling without replacement) というリサンプリングになる.

ブートストラップの基本的なアイデアは、ブートストラップ標本のバラツキを調べることによりデータのバラツキが推測できることである。ブートストラップという B. Efron の命名は、pulling yourself up by own bootstraps (独力でやり遂げる) という言い回しに由来しており、たった1組の標本 \mathcal{X} 自身から多数の標本 $\mathcal{X}^{*1}, \dots, \mathcal{X}^{*B}$ を生成することを反映している。

ブートストラップの反復回数 B は大きいほどブートストラップ分布の近似は良くなり、リサンプリングに伴う誤差は $B^{-1/2}$ に比例して小さくなる。一方、ブートストラップの計算量は B に比例して増える。どの程度の B を用いればよいかは個別の問題に依存するので一概には決められないが、分散推定には $B = 10^2$ 、信頼区間には $B = 10^3$ 程度は必要とされる。多くの場合、 $B = 10^4$ とすれば十分であろう。複雑な応用では、 $\hat{\theta}(\mathcal{X}^*)$ の計算時間が大半を占め、リサンプリングによって \mathcal{X}^* を生成する計算時間は相対的に小さい。 $\hat{\theta}(\mathcal{X}^*)$ を B 回繰り返す計算は並列に実行できるから、計算コストは今後さらに低下するだろう。

3 パラメータの信頼区間

さきほど (1) でブートストラップ複製 $\hat{\theta}^*$ の分散を計算して $\hat{\theta}$ のバラツキの目安とした。これだけでも十分に役立つが、95% の確率でパラメータ θ の真値がとりうる範囲、つまり信頼区間 (confidence interval) が知りたい、というような要求がなされることがある。この意味を厳密に考えるのは後回しにして、とりあえず計算手順を与えてさきほどのデータに適用してみる。信頼区間が θ の真値を含まない確率を $0 < \alpha < 1$ 、したがって信頼度または信頼水準 (confidence level) を $1 - \alpha$ とする。以下の例題では $\alpha = 0.05$ とおき、信頼度 95% の信頼区間を求める。

アルゴリズム 1 で生成したブートストラップ複製 $\hat{\theta}^*$ の標準偏差は $\hat{\sigma} = 1.94$ であった。これを利用して、 θ の信頼区間を次のように計算できる。

アルゴリズム 2 (標準信頼区間)

1. $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B}$ から (1) によって $\hat{\sigma}^2$ を計算する。 $\hat{\sigma}$ は $\hat{\theta}^*$ の標準偏差である。

2. 平均 0, 分散 1 の正規分布の下側 $100p\%$ 点 ($0 < p < 1$) を $z^{(p)}$ と表し, この値を $p = 1 - \alpha/2$ に対して調べておく. とくに $\alpha = 0.05$ では $p = 1 - \alpha/2 = 0.975$ に対して $z^{(0.975)} = 1.96$ である. なお, その正規分布に従う確率変数を X で表し, $P(A)$ が事象 A の確率を表すとすれば, $z^{(p)}$ は

$$P(X \leq z^{(p)}) = p$$

によって定義される定数である.

3. θ の $100(1 - \alpha)\%$ 両側信頼区間を

$$[\hat{\theta} - z^{(1-\alpha/2)}\hat{\sigma}, \hat{\theta} + z^{(1-\alpha/2)}\hat{\sigma}] \quad (2)$$

で与える. とくに θ の 95% 信頼区間は

$$[\hat{\theta} - 1.96 \times \hat{\sigma}, \hat{\theta} + 1.96 \times \hat{\sigma}]$$

である.

例題に (2) を適用すると, θ の 95% 信頼区間は $[14.45 - 1.96 \times 1.94, 14.45 + 1.96 \times 1.94] = [14.45 - 3.80, 14.45 + 3.80] = [10.65, 18.25]$ となる. このように計算した範囲に θ の真値が含まれる確率 (これを信頼区間の被覆確率という) が 95% であると解釈してよい. 図 1 (左) のシミュレーションでは $\theta = 10$ であったが, 仮にこれが真値であるとすれば,

$$P\{\hat{\theta} - 1.96 \times \hat{\sigma} \leq 10 \leq \hat{\theta} + 1.96 \times \hat{\sigma}\} = 0.95 \quad (3)$$

という意味である. 単純に $\hat{\theta} = 14.45$ であるというだけでは, この観測値をどれほど信用してよいか分からない. 信頼区間を示すことによって, $\hat{\theta}$ のバラツキの程度を把握できる.

アルゴリズム 2 は正規分布のパーセント点 $z^{(p)}$ を利用していた. パーセント点はパーセンタイル (percentile) ともいう. 正規分布がここで出てくるのは, (2) を導出するときには実は $\hat{\theta}$ が平均 θ , 分散 σ^2 の正規分布

$$\hat{\theta} \sim N(\theta, \sigma^2) \quad (4)$$

に従うと仮定 (もしくは近似) しているからである. ところが図 1 (左) を良く見ると, 分布は左右対称ではないし, $\hat{\theta}$ の期待値 $E(\hat{\theta}) = 11.35$ は $\theta = 10$

よりずいぶん大きいから、(4)が良い近似とは言えない。そこで正規分布を仮定せず、ブートストラップ法によって直接に信頼区間を得る方法として次のものが知られている。

アルゴリズム 3 (パーセンタイル信頼区間)

1. $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B}$ を小さい順にソートする。 $0 < p < 1$ に対して pB 番目に小さい数値を $\hat{\theta}^{*(p)}$ と表す。 pB が整数値でなければ、となりあう値を線形補間する。
2. $p = \alpha/2$ および $p = 1 - \alpha/2$ とおいて、 $100(1 - \alpha)\%$ 両側信頼区間を次式で与える。

$$[\hat{\theta}^{*(\alpha/2)}, \hat{\theta}^{*(1-\alpha/2)}] \quad (5)$$

とくに θ の 95% 信頼区間は

$$[\hat{\theta}^{*(0.025)}, \hat{\theta}^{*(0.975)}]$$

である。

なお、 $\hat{\theta}^{*(p)}$ はブートストラップ分布の下側 $100p\%$ 点であり、上記の定義のほかに数種類あるが、十分大きな B を用いれば大差ない。また全体のソートをしなくて効率よく $\hat{\theta}^{*(p)}$ の計算が可能である。

例題に (5) を適用すると、 θ の 95% 信頼区間は $[\hat{\theta}^{*(0.025)}, \hat{\theta}^{*(0.975)}] = [11.52, 19.05]$ となる。この信頼区間は図 1 (右) に 2 本の破線で示されている。左側の破線より小さなブートストラップ複製は全体の 2.5%、右側の破線より大きなブートストラップ複製も全体の 2.5% であるから、この 2 本の破線で囲まれた区間には全体の 95% が含まれる。

ここで示した二つの方法から得られる信頼区間はどちらも数値的に安定で十分に実用的である。信頼区間の精度は必ずしもよくないが、多くの場合許容範囲である。(3) で言えば左辺で表された被覆確率とあらかじめ想定した信頼水準 0.95 のズレが誤差であり、これが小さいほど精度が高いという。この被覆確率の誤差のことをバイアスということもあるが、いわゆるパラメータ推定におけるバイアスとは意味が異なるので注意する。アルゴリズム 2 では $\hat{\theta}$ の分布を正規分布で近似してしまったので被覆確率に誤差が入るのはし

かたがないが、ブートストラップ複製 $\hat{\theta}^*$ の分布がだいたい正規分布に近ければ、実用的に問題ないだろう。

アルゴリズム 3 では正規分布の仮定をせずに信頼区間を計算するので精度が改善するように思えるかもしれない。ところが実際には、アルゴリズム 2 のほうがアルゴリズム 3 より誤差が小さい傾向がある。この原因は $\hat{\theta}$ の分布が左右非対称なことである。図 1 (左) では全体の 5.06% が $\hat{\theta} \geq 14.45$ であるが、図 1 (右) では全体の 0.28% が $\hat{\theta}^* \leq 10$ である。このズレが誤差につながるのであるが、その意味および改善した信頼区間のアルゴリズムの説明は後ほど 6 節から 8 節で行う。

4 プラグイン推定量

ブートストラップによって推定した値が本当にその真値のよい近似になっているのか、ブートストラップが正しい方法といえるのかを調べておく。図 1 で言えば、右図の分散 $\hat{\sigma}^2$ が左図の分散 σ^2 を近似することを説明する。本節は数学的にやや高度な表現を含むが、これを読み飛ばしても本稿全体の理解にさほど支障はない。

アルゴリズム 1 では、データの要素 $\mathbf{x}_1, \dots, \mathbf{x}_n$ をどれも対等に扱っていた。この方法を正当化するには、各要素が同じ確率分布に従う確率変数の実現値であることを仮定する。その分布の確率密度関数を $f(\mathbf{x})$ と表し、各要素が互いに独立に $f(\mathbf{x})$ に従うことを

$$\mathbf{x}_1, \dots, \mathbf{x}_n \sim f(\mathbf{x}) \quad (6)$$

と表す。ただし確率変数とその実現値の記号を区別しないことにする。

データ $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ が与えられたとき、

$$\hat{f}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \delta(\mathbf{x} - \mathbf{x}_i)$$

によって確率分布 $\hat{f}(\mathbf{x})$ を定義する。ただし $\delta(\mathbf{x})$ はディラックのデルタ関数とよばれ、簡単のため x が整数値をとる離散分布で考えると、 $x = x_i$ のとき $\delta(x - x_i) = 1$ 、 $x \neq x_i$ のとき $\delta(x - x_i) = 0$ である。一般に $\hat{f}(\mathbf{x})$ は $\mathbf{x}_1, \dots, \mathbf{x}_n$ の各要素に確率 $1/n$ を割り当てた確率分布であり、経験分布

(empirical distribution) とよばれる。そして、アルゴリズム 1 のステップ 2 と 3 で得られるブートストラップ標本 $\mathcal{X}^* = \{\mathbf{x}_1^*, \dots, \mathbf{x}_{n'}^*\}$ の各要素は、この経験分布に従う確率変数の実現値とみなせる。これを (6) と対比させて

$$\mathbf{x}_1^*, \dots, \mathbf{x}_{n'}^* \sim \hat{f}(\mathbf{x}) \quad (7)$$

と表す。 \hat{f} は n と \mathcal{X} に依存するが、簡単のため記号では明示しない。

十分にサンプルサイズ n が大きくなると $\hat{f}(\mathbf{x})$ は $f(\mathbf{x})$ の良い近似になることがアルゴリズム 1 を正当化するポイントである。以下で説明するように $n \rightarrow \infty$ の極限で分布の収束 $\hat{f} \rightarrow f$ がいえるから、データ \mathcal{X} のバラツキをブートストラップ標本 \mathcal{X}^* のバラツキから推定できる。

簡単のため、 \mathbf{x} がスカラー x の場合を考える。とくに x が整数値をとる離散分布ならば、 $f(x)$ は密度関数ではなく x の確率を与える確率関数としてよい。そして $\hat{f}(x) = (\text{データ } x_1, \dots, x_n \text{ のうち } x \text{ に一致した個数})/n$ であるから、 $n \rightarrow \infty$ で $\hat{f}(x) \rightarrow f(x)$ である。より一般に x が実数の場合、 $f(x)$ と $\hat{f}(x)$ の累積分布関数を $F(x)$ と $\hat{F}(x)$ で表す。定義より $F(x) = \int_{-\infty}^x f(x') dx'$ 、 $\hat{F}(x) = \int_{-\infty}^x \hat{f}(x') dx'$ であるが、デルタ関数の性質 ($x = 0$ で $\delta(x) = \infty$ 、 $x \neq 0$ で $\delta(x) = 0$ 、 $\int_{-\infty}^{\infty} \delta(x) dx = 1$) より、 $\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x)$ である。ただし $I(\cdot)$ は指示関数を表し、事象 A が真のとき $I(A) = 1$ 、偽のとき $I(A) = 0$ とする。 x の値をひとつ固定して考えると、 $I(x_i \leq x)$ 、 $i = 1, \dots, n$ は成功確率 $F(x)$ のベルヌーイ試行になるから、大数の法則より $\hat{F}(x)$ は $F(x)$ に確率収束する。このような x の各点での確率収束よりもっと強い結果が Glivenko-Cantelli の定理

$$P \left(\lim_{n \rightarrow \infty} \sup_x |\hat{F}(x) - F(x)| = 0 \right) = 1$$

である。すなわち $n \rightarrow \infty$ で一様収束 $\hat{F} \rightarrow F$ を概収束の意味で言える。これがさきほどのべた分布の収束 $\hat{f} \rightarrow f$ の意味である。

アルゴリズム 1 では $\mathcal{X}^{*1}, \dots, \mathcal{X}^{*B}$ を生成してから $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B}$ を計算し、これから分散 $\hat{\sigma}^2$ を (1) で計算した。このようにして得られた $\hat{\sigma}^2$ が、 $\hat{\theta} = \hat{\theta}(\mathcal{X})$ の分散

$$\sigma^2 = E_F \left\{ \left(\hat{\theta} - E_F(\hat{\theta}) \right)^2 \right\} \quad (8)$$

を近似することを以下で説明する。ただし $E_F(\cdot)$ は単に $E(\cdot)$ と書いていたものであり、分布 F に関する期待値を表す。

まず、(1) では B 個のブートストラップ標本から分散 $\hat{\sigma}^2$ を計算しているが、この記号を $\hat{\sigma}_B^2$ に変更する。同じ \mathcal{X} を入力しても乱数を変えてアルゴリズム 1 を実行すれば、そのたびに $\hat{\sigma}_B^2$ の値は異なる。しかし計算量を増やすことの引き換えに B をいくらでも大きくできるので、 $B \rightarrow \infty$ とした極限を考える。すると $\hat{\sigma}_B^2$ の標準偏差は $B^{-1/2}$ に比例して小さくなり、 $\hat{\sigma}_B^2$ はある値に収束する。その収束値を $\hat{\sigma}^2$ と書くと、 B が十分に大きければ $\hat{\sigma}_B^2$ は $\hat{\sigma}^2$ の良い近似である。

問題は、(上で定義した意味の) $\hat{\sigma}^2$ が σ^2 を近似するかどうかである。 $\hat{\sigma}_B^2$ は標本分散の式であるから、 $\hat{\sigma}^2$ は \mathcal{X} を与えたときの $\hat{\theta}^* = \hat{\theta}(\mathcal{X}^*)$ の条件付分散である。ところで \mathcal{X} を与えたときの \mathcal{X}^* の要素の条件付分布は (7) であり、一方 \mathcal{X} の要素の分布は (6) であることを思い出すと、 $\hat{\sigma}^2$ と σ^2 の違いは分布が \hat{F} か F かという点だけである。したがって、(8) の F を \hat{F} で置き換えて、

$$\hat{\sigma}^2 = E_{\hat{F}} \left\{ \left(\hat{\theta}^* - E_{\hat{F}}(\hat{\theta}^*) \right)^2 \right\} \quad (9)$$

が得られる。さきほど述べたように $n \rightarrow \infty$ で $\hat{F} \rightarrow F$ であるから、 $\hat{\sigma}^2 \rightarrow \sigma^2$ が示されたことになる。十分に n が大きければ、 $\hat{\sigma}^2$ は σ^2 の良い近似である。

この考え方は σ^2 に限らず一般化できる。分布 F を入力、スカラー σ^2 を出力とした滑らかな汎関数 $\sigma^2 = H(F)$ として (8) を表すと、(9) は $\hat{\sigma}^2 = H(\hat{F})$ と表される。 $n \rightarrow \infty$ で $\hat{F} \rightarrow F$ から $H(\hat{F}) \rightarrow H(F)$ がいえるが、これは他の汎関数 $H(\cdot)$ に対しても成り立つ。たとえば、パラメータ推定における $\hat{\theta}$ のバイアス $H(F) = E_F(\hat{\theta}) - \theta$ や、 t をひとつ固定したときの $\hat{\theta}$ の分布関数の値 $H(F) = P(\hat{\theta} \leq t) = E_F(I(\hat{\theta} \leq t))$ に対しても収束がいえる。とくに最後の例は任意の t でいえるから、 $\hat{\theta}^*$ の分布は $\hat{\theta}$ の分布の近似とみなしてもよい。

(9) の $\hat{\sigma}^2$ のように $H(F)$ の F を \hat{F} でおきかえて近似したものは一般にプラグイン推定量 (plug-in estimator) とよばれる。さらに $\hat{\sigma}_B^2$ のように \hat{F} を多数のブートストラップ標本でおきかえて、数値的に $H(\hat{F})$ を計算したものがブートストラップ推定量 (bootstrap estimator) である。ブートストラップの理論では、 B が十分に大きいと仮定して、はじめから $H(\hat{F})$ をブートスト

ラップ推定量とよぶこともある。

ここで見てきたように、 n と B が十分に大きいと仮定してブートストラップ法は正当化される。必要に応じて B は大きくできるが、 n を増やすことは困難な場合が多い。もし n が小さい場合 ($n \leq 30$ など) の応用では結果の妥当性を別途シミュレーション等で確認するなどの注意が必要である。

5 構造のあるデータ

データ \mathcal{X} の従う同時密度関数を $g(\mathcal{X})$ で表す。アルゴリズム 1 は (6) を仮定していたので、 $g(\mathcal{X})$ はデータ要素の密度関数の積

$$g(\mathcal{X}) = f(\mathbf{x}_1) \cdots f(\mathbf{x}_n)$$

であった。計算機シミュレーションではこれから

$$\mathcal{X}^1, \dots, \mathcal{X}^B \sim g(\mathcal{X})$$

を生成していたことになる。ここでは回帰分析や時系列解析などのように、データに何らかの構造がある状況へ適用可能な方法について簡単に説明する。

まずデータ生成のメカニズムに関する事前知識を総合してパラメトリックモデル (parametric model) $g(\mathcal{X}; \boldsymbol{\xi})$ を与えた場合を考える。パラメータベクトル $\boldsymbol{\xi}$ を適切な値にすれば \mathcal{X} の確率密度関数を表すと仮定する。原理的にはいくらでも複雑なモデルを工夫して与えることにより、どのような構造のデータ生成メカニズムも表現できるはずである。(6) の状況ならば \mathbf{x} のパラメトリックモデル $f(\mathbf{x}; \boldsymbol{\xi})$ を与えて $g(\mathcal{X}; \boldsymbol{\xi}) = f(\mathbf{x}_1; \boldsymbol{\xi}) \cdots f(\mathbf{x}_n; \boldsymbol{\xi})$ とすれば良いし、より複雑な状況では直接 $g(\mathcal{X}; \boldsymbol{\xi})$ を指定すれば良い。このとき、次の方法が利用できる。

アルゴリズム 4 (パラメトリック・ブートストラップ法)

1. パラメータ $\boldsymbol{\xi}$ の推定値 $\hat{\boldsymbol{\xi}}$ を計算する。たとえば最尤法を適用して最尤推定量 $\hat{\boldsymbol{\xi}}$ を計算するには、データにモデルが最も適合するように次式によって $\boldsymbol{\xi}$ を調節する。

$$\hat{\boldsymbol{\xi}} = \arg \max_{\boldsymbol{\xi}} g(\mathcal{X}; \boldsymbol{\xi})$$

2. 推定したパラメータ値 $\hat{\xi}$ をモデルに代入して、計算機シミュレーションによってデータを生成する。

$$\mathcal{X}^{*1}, \dots, \mathcal{X}^{*B} \sim g(\mathcal{X}; \hat{\xi})$$

アルゴリズム 1 に比べると (6) の制約がなく、また (6) がいえる場合は 4 節の意味での収束が早くなる点でアルゴリズム 4 のほうが優れている。しかし、一般に $g(\mathcal{X}; \xi)$ を正しく指定することは困難であり、現実にはどのようなパラメータ値 ξ を与えても真実の確率密度関数の近似にすぎない。したがってアルゴリズム 4 で得られるブートストラップ分布がどの程度信頼できるかはモデルのよさ次第である。むしろ、このような問題点を克服するアルゴリズム 1 のほうが優れているともいえる。

アルゴリズム 1 とアルゴリズム 4 は一長一短があるので、それらの中間的な方法も考えられる。この方法を回帰分析を例にして具体的に説明するが、自己回帰モデルなど時系列解析にも適用可能である。まず次のような線形回帰モデルを考える。

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, n$$

ここで y_i は目的変数、 x_i は説明変数、 ϵ_i は誤差、 $\beta = (\beta_0, \beta_1)$ は回帰係数である。データの要素を $\mathbf{x}_i = (x_i, y_i)$ とおいて

$$\mathcal{X} = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

とすればアルゴリズム 1 を適用可能であるが、もし説明変数が定数（あらかじめ設定した投薬量など）ならば (6) の想定は不適切である。一方、誤差 ϵ_i に注目して、

$$\mathcal{X} = \{\epsilon_1, \dots, \epsilon_n\}$$

とおけば (6) の想定が妥当になりアルゴリズム 1 を適用できるが、実際には誤差を直接観測できない。そこで、回帰モデルを利用して誤差を推定する方法によってブートストラップ標本を生成する。

アルゴリズム 5 (残差のリサンプリング)

1. 回帰係数の推定値 $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ を計算する。最小 2 乗法を適用すれ

ば, $\bar{x} = \sum_{i=1}^n x_i/n$, $\bar{y} = \sum_{i=1}^n y_i/n$ において

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

2. 残差 (residuals) $e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$, $i = 1, \dots, n$ を計算する. 残差 e_i は誤差 ϵ_i の推定値と考えてもよい.
3. 残差 $\{e_1, \dots, e_n\}$ から復元抽出によって残差のブートストラップ標本 $\{e_1^*, \dots, e_n^*\}$ を生成する. すなわちアルゴリズム 1 のステップ 2 で得られた添え字 (ただし $n' = n$) をつかって $e_1^* = e_{i_1}, e_2^* = e_{i_2}, \dots, e_n^* = e_{i_n}$ である.
4. 回帰係数の推定値 $\hat{\beta}$ と残差のブートストラップ標本 $\{e_1^*, \dots, e_n^*\}$ から目的変数のブートストラップ標本を計算する.

$$y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i^*, \quad i = 1, \dots, n$$

これからデータ \mathcal{X}^* を次式で与える.

$$\mathcal{X}^* = \{(x_1, y_1^*), \dots, (x_n, y_n^*)\}$$

5. 上記 3 と 4 を B 回繰り返して $\mathcal{X}^{*1}, \dots, \mathcal{X}^{*B}$ を得る.

ここでは単回帰モデルで説明したが, 説明変数 x_i をベクトルでおきかえた重回帰モデルでも同様である. 標準的な記法で結論を書くと, ステップ 1 で回帰係数ベクトルを $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ で計算して, ステップ 2 で残差 e_1, \dots, e_n をならべたベクトル $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}$ を計算する. ステップ 3 でベクトル \mathbf{e}^* を得て, ステップ 4 で \mathbf{y} のブートストラップ標本 $\mathbf{y}^* = \mathbf{X}\hat{\beta} + \mathbf{e}^*$ を計算する.

上記のステップ 3 を改良して残差 e_i の代わりに修正残差 $r_i = e_i/\sqrt{1-h_i}$ が用いられることもある. h_i (てこ比) はハット行列 $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ の対角成分である. このような修正をする理由は e_i の分散が ϵ_i の分散の $1-h_i$ 倍になってしまうからである. 平均値 $\bar{r} = \sum_{i=1}^n r_i/n$ の調整も行って, $r_1 - \bar{r}, \dots, r_n - \bar{r}$ から復元抽出したものを e_1^*, \dots, e_n^* の代わりに用いる.

時系列データ $\{y_1, y_2, \dots, y_n\}$ にアルゴリズム 5 を適用するには, 時系列の生成メカニズムをあらわすモデルを指定する. たとえば次数 1 の自己回帰

モデル $y_i = \beta_1 y_{i-1} + \epsilon_i$ を指定すれば、形式的に目的変数 y_i 、説明変数 y_{i-1} の回帰モデルになり、さきほどの線形回帰モデルと同様に扱える。パラメータ β_1 を推定して残差 $e_i = y_i - \hat{\beta}_1 y_{i-1}$, $i = 2, \dots, n$ を計算し、復元抽出で得た e_i^* , $i = -k, \dots, n$ から $y_i^* = \hat{\beta}_1 y_{i-1}^* + e_i^*$, $i = -k, \dots, n$ とする。 k は十分に大きい数で $y_{-k-1}^* = 0$ としておき、 y_{-k}^*, \dots, y_0^* は捨てる。

これを次数 2 の自己回帰モデル $y_i = \beta_1 y_{i-1} + \beta_2 y_{i-2} + \epsilon_i$ におきかえてもよいが、得られるブートストラップ分布は異なるであろう。このようにアルゴリズム 5 から得られるブートストラップ分布は指定したモデルに依存する。それでも ϵ_i の従う確率密度関数を指定する必要がないからアルゴリズム 4 に比べれば改善されているが、(6) の状況におけるアルゴリズム 1 のようにモデルにまったく依存しない方法が望ましい。そこで、モデルを指定せずアルゴリズム 1 を時系列データへ適用するための工夫が次の方法である。

アルゴリズム 6 (ブロック・ブートストラップ)

1. 時系列データ $\mathcal{X} = \{x_1, x_2, \dots\}$ へ前処理を行ってトレンドなどの傾向を取り除き定常化する。たとえば 1 階差分 $y_i = x_{i+1} - x_i$ が一般的である。得られた $\{y_1, \dots, y_n\}$ は定常な時系列の標本と仮定する。
2. あらかじめ定めたブロック長 m にデータを(重複を許して)分割する。データ要素を $z_i = (y_{i-m+1}, y_{i-m+2}, \dots, y_i)$, $i = m, m+1, \dots, n$ とおく。簡単のため $k = n/m$ が整数と仮定する。
3. $\{z_m, \dots, z_n\}$ から復元抽出によって k 個の要素をとりだしてブートストラップ標本 $\{z_1^*, \dots, z_k^*\}$ を生成する。
4. $\{z_1^*, \dots, z_k^*\}$ の要素をつないで (y_1^*, \dots, y_n^*) とし、これに前処理と逆の操作を行って \mathcal{X}^* とする。
5. 上記 3 と 4 を B 回繰り返して $\mathcal{X}^{*1}, \dots, \mathcal{X}^{*B}$ を得る。

ブロック長 m を $m = 1$ とおけばアルゴリズム 1 に帰着される。実際にはデータの特徴を失わない程度の大きさの m を指定する。結局、自己回帰モデルにおける次数選択と同様に m の選択を考慮する必要があるのはアルゴリズム 5 の問題点と同じである。構造のあるデータにおけるブートストラップはまだ発展段階にあり、いまのところデータに応じて十分な検討が必要である。

6 精度の高い信頼区間

パラメータ θ の信頼区間を 3 節で簡単に紹介してアルゴリズム 2 とアルゴリズム 3 を示した。その二つの方法は十分実用的であるが必ずしも精度が高くない。ここでは信頼区間が満たすべき性質を説明して、より精度の高い二つの方法を説明する。

信頼水準 $1 - \alpha$ の上側信頼区間を一般に $\theta \in [\hat{\theta}_\alpha, \infty)$ と書くことにする。これから両側信頼区間を得るのは容易で $\theta \in [\hat{\theta}_{\alpha/2}, \hat{\theta}_{1-\alpha/2}]$ とすればよいので、以下では $\hat{\theta}_\alpha$ の決め方について考える。 $\hat{\theta}_\alpha$ は \mathcal{X} から計算する量であり、上側信頼区間の被覆確率が信頼水準に一致する、すなわち

$$P_\theta(\hat{\theta}_\alpha \leq \theta) = 1 - \alpha \quad (10)$$

を満たすようにしたい。ただし記号 $P_\theta(\cdot)$ は確率が θ の値に依存することを明示するために用いる。(10) から直ちに

$$\begin{aligned} P_\theta(\hat{\theta}_{\alpha/2} \leq \theta \leq \hat{\theta}_{1-\alpha/2}) &= 1 - \left\{ P_\theta(\theta < \hat{\theta}_{\alpha/2}) + P_\theta(\hat{\theta}_{1-\alpha/2} < \theta) \right\} \\ &= 1 - \left(\frac{\alpha}{2} + \frac{\alpha}{2} \right) = 1 - \alpha \end{aligned}$$

となり、さきほどの両側信頼区間の被覆確率が信頼水準に一致することが確かめられる。

近似的な信頼区間を得るためにアルゴリズム 2 では

$$\hat{\theta}_\alpha = \hat{\theta} - z^{(1-\alpha)} \hat{\sigma} \quad (11)$$

によって $\hat{\theta}_\alpha$ を計算し、アルゴリズム 3 は

$$\hat{\theta}_\alpha = \hat{\theta}^{*(\alpha)} \quad (12)$$

で計算する。このように $\hat{\theta}_\alpha$ の計算法を与えたとき、(10) の左辺と右辺の差がその信頼区間の誤差 (またはバイアス) である。この誤差が $n \rightarrow \infty$ で $n^{-k/2}$ に比例して小さくなる時、信頼区間は k 次の精度 (k -th order accuracy) をもつという。(11) と (12) は一般に 1 次の精度である。

アルゴリズム 2 は (4) を仮定して導かれると 3 節で述べたが、これを確認しておく。まず $\hat{\theta}$ を標準化 (平均 0, 分散 1 に変換) して

$$v = \frac{\hat{\theta} - \theta}{\sigma}$$

表 1 95% 両側信頼区間の数値例 (図 1 のデータ)

手法	計算式	$\hat{\theta}_{0.025}$	$\hat{\theta}_{0.975}$
アルゴリズム 2 (標準信頼区間)	(11)	10.65	18.25
アルゴリズム 3 (パーセンタイル信頼区間)	(12)	11.52	19.05
アルゴリズム 7 (ブートストラップ- t 法)	(15)	9.85	17.38
アルゴリズム 8 (BCa 法)	(17)	9.98	17.48
厳密値	(18)	9.33	17.49

と定義すると $v \sim N(0, 1)$ である。パーセント点の定義から

$$P_{\theta} \left(v \leq z^{(1-\alpha)} \right) = 1 - \alpha$$

であり、これを変形すると次式が得られる。

$$P_{\theta} \left(\hat{\theta} - z^{(1-\alpha)} \sigma \leq \theta \right) = 1 - \alpha \quad (13)$$

ここで σ が既知と仮定すれば (11) によって定義される $\hat{\theta}_{\alpha}$ が (10) を厳密に満たすことが分かる。実際には σ をデータから推定して $\hat{\sigma}$ を用いるので、近似的に (10) がいえる。

一般には (4) が正しいとは言えないので、これを仮定しないで信頼区間を導いてみる。 v の下側 $100p\%$ 点を $v^{(p)}$ とすれば (13) は

$$P_{\theta} \left(\hat{\theta} - v^{(1-\alpha)} \sigma \leq \theta \right) = 1 - \alpha$$

と書き換えられるから、(11) を再定義して

$$\hat{\theta}_{\alpha} = \hat{\theta} - v^{(1-\alpha)} \sigma \quad (14)$$

とおく。 v のブートストラップ複製

$$v^* = \frac{\hat{\theta}^* - \hat{\theta}}{\sigma}$$

のブートストラップ分布と σ の値が θ に依存しないと仮定すれば

$$v^{(p)} = \frac{\hat{\theta}^{*(p)} - \hat{\theta}}{\sigma}$$

である。これを (14) に代入すれば、

$$\hat{\theta}_\alpha = \hat{\theta} - (\hat{\theta}^{*(1-\alpha)} - \hat{\theta}) = 2\hat{\theta} - \hat{\theta}^{*(1-\alpha)} \quad (15)$$

が得られる。これは仮定の下で (10) を満たす。この方法（一般的には以下で述べるアルゴリズム 7）を 2 節の数値例に適用して得た両側信頼区間を表 1 に示した。

もし $\hat{\theta}^*$ のブートストラップ分布が $\hat{\theta}$ を中心にして左右対称ならば $\hat{\theta}^{*(1-\alpha)} - \hat{\theta} = \hat{\theta} - \hat{\theta}^{*(\alpha)}$ であるから、これを (15) に代入すると

$$\hat{\theta}_\alpha = \hat{\theta} - (\hat{\theta} - \hat{\theta}^{*(\alpha)}) = \hat{\theta}^{*(\alpha)}$$

すなわちアルゴリズム 3 の (12) が得られる。しかし対称性からのズレが大きければ (12) ではなく (15) を用いるべきである。

一般的には σ が θ に依存するので、データから推定した $\hat{\sigma}$ を用いる。ブートストラップ標本 \mathcal{X}^* から計算した $\hat{\sigma}$ を $\hat{\sigma}^*$ と書いて (15) を定義しなおすと、次の方法になる。

アルゴリズム 7 (ブートストラップ- t 法)

1. $\hat{\theta}^*$ を標準化した

$$t^* = \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*}$$

のブートストラップ複製

$$t^{*1}, \dots, t^{*B}$$

を小さい順にソートする。

2. t^* のブートストラップ分布の下側 $100(1-\alpha)\%$ 点、すなわち $t^{*(1-\alpha)}$ を計算する。
3. 信頼水準 $1-\alpha$ の上側信頼区間の下限を次式で与える。

$$\hat{\theta}_\alpha = \hat{\theta} - t^{*(1-\alpha)} \hat{\sigma} \quad (16)$$

なお、ステップ 1 とステップ 3 で形式的に $\hat{\sigma}^* = \hat{\sigma} = 1$ とおけば、(16) は (15) に一致する。

アルゴリズム 7 は一般に 2 次の精度の信頼区間を与える。十分に n が大き

ければ、精度の次数が高いほど誤差が少ないので理論的に良い方法とみなされる。もちろん現実の n では次数の高い方法が必ずしも誤差を小さくするとは限らないのだが、表 1 を見るとアルゴリズム 7 のほうがアルゴリズム 2 とアルゴリズム 3 より厳密値に近いことがわかる。

あらかじめ興味のあるパラメータを非線形単調変換して θ が定義されると仮定する。その変換を工夫して t^* の分布の θ への依存性を少しでも小さくしてからアルゴリズム 7 を適用して信頼区間を計算し、その結果を逆変換すれば興味のあるパラメータの信頼区間の誤差は小さくできる。一方でアルゴリズム 3 は、そのような変換に計算結果が依存せず、いわば常にベストの変換を用いたときの信頼区間が自動的に得られる。このような性質を変換保存 (transformation-respecting) とよび、アルゴリズム 3 を改良した以下の BCa (bias-corrected and accelerated) 法にも共通する性質である。

アルゴリズム 8 (BCa 法)

1. 事象が成立した回数を $\#\{\cdot\}$ と書いて、

$$\hat{\alpha}_0 = \frac{\#\{\hat{\theta}^{*b} \leq \hat{\theta}, b = 1, \dots, B\}}{B}$$

および $\hat{z}_0 = z^{(\hat{\alpha}_0)}$ を計算する。十分に B が大きければ、 $\hat{\alpha}_0 = P_{\hat{\theta}}(\hat{\theta}^* \leq \hat{\theta})$ である。 $\hat{\alpha}_0 < 1/2$ (すなわち $\hat{z}_0 < 0$) なら $\hat{\theta}^* - \hat{\theta}$ の分布は正に偏り、逆に $\hat{\alpha}_0 > 1/2$ (すなわち $\hat{z}_0 > 0$) なら分布は負に偏る。

2. X から要素 x_i を取り除いて計算した $\hat{\theta}$ を $\hat{\theta}_{(i)}$ と書いて、加速定数とよばれる量を次式で計算する。

$$\hat{a} = - \frac{\sum_{i=1}^n (\hat{\theta}_{(i)} - \frac{1}{n} \sum_{j=1}^n \hat{\theta}_{(j)})^3}{6 \left\{ \sum_{i=1}^n (\hat{\theta}_{(i)} - \frac{1}{n} \sum_{j=1}^n \hat{\theta}_{(j)})^2 \right\}^{3/2}}$$

3. $X \sim N(0, 1)$ の累積分布関数を $\Phi(x)$ と書く (つまり $\Phi(z^{(p)}) = p$ である)。次の量を計算する。

$$\hat{\alpha} = \Phi \left\{ \hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})} \right\}$$

4. 信頼水準 $1 - \alpha$ の上側信頼区間の下限を次式で与える.

$$\hat{\theta}_\alpha = \hat{\theta}^*(\hat{\alpha}) \quad (17)$$

なお, σ が θ に依存しないと仮定 (または近似) すればステップ 2 で $\hat{\alpha} = 0$ とおいてよい.

アルゴリズム 8 はアルゴリズム 7 と同様に一般に 2 次の精度である. 表 1 に示したアルゴリズム 8 の結果は $\hat{\alpha} = 0$ とおいて計算したものであり, やはり厳密値に近い.

表 1 の最後の行に示した厳密値は次のように得られている. 実は 2 節のデータは多変量正規分布 $\mathbf{x}_1, \dots, \mathbf{x}_{100} \sim N_8(\mathbf{a}, b^2 \mathbf{I})$ にしたがって生成されていた. パラメータは $a_1 = 10, a_2 = \dots = a_8 = 0, b = 20, n = 100$ であった. 興味のあるパラメータは $\theta = \|\mathbf{a}\| = 10$ と定義され, その推定量は $\hat{\theta} = \|\bar{\mathbf{x}}\| = \|\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i\|$ であった. このようにデータの生成メカニズムが明らかならば, 信頼区間の厳密値が計算できる. $\mathbf{y} = \sqrt{n}\bar{\mathbf{x}}/b, \boldsymbol{\mu} = \sqrt{n}\mathbf{a}/b$ とおけば, $\mathbf{y} \sim N_8(\boldsymbol{\mu}, \mathbf{I})$ である. $\|\mathbf{y}\|^2$ は自由度 8, 非心度 $\|\boldsymbol{\mu}\|^2 = n\theta^2/b^2$ のカイ二乗分布に従う. この分布の下側 $100(1 - \alpha)\%$ 点がちょうど $n\hat{\theta}^2/b^2$ に一致するような θ の値を $\hat{\theta}_\alpha$ と定義する. すなわち,

$$P_{\hat{\theta}_\alpha}(\hat{\theta}^* \leq \hat{\theta}) = 1 - \alpha \quad (18)$$

とおいて $\hat{\theta}_\alpha$ を定義すれば (10) を厳密に満たすことは以下のように容易に示せる. ここで $F_\theta(x) = P_\theta(\hat{\theta} \leq x)$ とおけば $P_\theta(\hat{\theta}_\alpha \leq \theta) = P_\theta(F_{\hat{\theta}_\alpha}(\hat{\theta}) \geq F_\theta(\hat{\theta})) = P_\theta(1 - \alpha \geq U) = 1 - \alpha$, ただし U は区間 $(0,1)$ の一様分布に従う確率変数である.

7 ブートストラップ確率

パラメータ θ に関する次のような仮説を考える.

$$\theta \leq \theta_0 \quad (19)$$

ただし θ_0 は定数である. たとえば $\hat{\theta}$ が 100 年後の予測気温で $\theta_0 = 10$ とすれば, $\theta \leq 10$ は気温が 10 度以下であるという仮説を表す.

このような仮説の真偽をデータから判断するには、6節の上側信頼区間を利用すればよい。もし $\hat{\theta}_\alpha \leq \theta_0$ ならば (19) を受容 (accept) して仮説が真である可能性があるとして判断し (ただし仮説が真実であると判断したわけではないことに注意)、逆に $\hat{\theta}_\alpha > \theta_0$ ならば (19) を棄却 (reject) して仮説が偽である、すなわち $\theta > \theta_0$ と判断する。この手続きを仮説検定という。(10) を厳密に満たす $\hat{\theta}_\alpha$ を使えば、仮説の境界 $\theta = \theta_0$ が真値であるとき誤って仮説を棄却してしまう確率は

$$P_{\theta_0}(\hat{\theta}_\alpha > \theta_0) = \alpha \quad (20)$$

である。 α の値は有意水準とよばれ、 $\alpha = 0.05$ とするのが一般的であるが、より慎重な判断が要求される場合は $\alpha = 0.01$ やもっと小さな値を用いる。仮に誤った判断をしたときの深刻さに応じて、判断を下す前に α の値をきめておく。

簡単のため $\hat{\theta}_\alpha$ として (11) や (12) を考えて図 1 をみると、 α を小さくすれば $\hat{\theta}_\alpha$ は左に移動して仮説は棄却されにくくなり、逆に α を大きくすれば $\hat{\theta}_\alpha$ は右に移動して仮説は棄却されやすくなる。棄却するかしないかの判断が切り替わるときの α の値を p で表す。 α や θ_0 は定数であるが p は \mathcal{X} に依存することに注意。この p を仮説検定の p -値といい、仮説 (19) の信頼度と解釈してもよい。つまり p が大きいほどデータが仮説を支持しており、逆に p が小さいほど仮説が支持されていない。

任意の $0 < \alpha < 1$ で (20) が成り立つとしておく。 p の定義は

$$\hat{\theta}_p = \theta_0 \quad (21)$$

であるから $P_{\theta_0}(p < \alpha) = \alpha$ となり、したがって $\theta = \theta_0$ のとき p は区間 $(0, 1)$ の一様分布に従う。検定の結果を p -値で表しておけば、判断を下す人があとから α を与えても $p \geq \alpha$ なら受容、 $p < \alpha$ なら棄却とすぐに分かり便利である。

とくにパーセントイル信頼区間、すなわち (12) 式から得られる p -値の近似値はブートストラップ分布から容易に計算できる。これをブートストラップ確率 (bootstrap probability) という。

アルゴリズム 9 (ブートストラップ確率)

1. ブートストラップ標本 $\mathcal{X}^{*1}, \dots, \mathcal{X}^{*B}$ において $\hat{\theta}(\mathcal{X}^{*b}) \leq \theta_0$ となる

回数 C を数える.

$$C = \#\{\hat{\theta}(\mathcal{X}^{*b}) \leq \theta_0, \quad b = 1, \dots, B\}$$

2. p -値を次式で計算する.

$$p = \frac{C}{B}$$

ちなみにアルゴリズム 8 の $\hat{\alpha}_0$ は $\theta_0 = \hat{\theta}$ としたときのブートストラップ確率である.

図 1 (右) において $\theta_0 = 10$ としてブートストラップ確率を計算すると, $C = 28$, $B = 10000$ より $p = 0.0028$ であった. C は 2 項分布に従う確率変数であることを考慮すると, リサンプリングによる p の標準誤差は $\sqrt{p(1-p)/B} = 0.0005$ と計算される. $B \rightarrow \infty$ の極限では

$$p = P_{\hat{\theta}}(\hat{\theta}(\mathcal{X}^*) \leq \theta_0) \quad (22)$$

となるので, 十分に B が大きければ (22) をブートストラップ確率とみなしてよい.

一方, (18) と (21) から容易に示されるように厳密な p -値は

$$p = P_{\theta_0}(\hat{\theta}(\mathcal{X}) \geq \hat{\theta}) \quad (23)$$

と表される (ただし \mathcal{X} は確率変数, $\hat{\theta}$ は定数とみなす). すなわち, 実際に観測した $\hat{\theta}$ の値より大きな $\hat{\theta}(\mathcal{X})$ が得られる確率が p -値と定義される. 図 1 (左) において $\hat{\theta}(\mathcal{X}^b) \geq 14.45$ となった回数は 506 回であるから, $p = 0.0506$, 標準誤差 0.0022 である. このデータの生成モデルが 6 節の最後で説明したものであることを利用すると, $B \rightarrow \infty$ の極限における (23) の厳密値は $p = 0.05$ と計算される. こうなるように, はじめからこの例題をつくってあった.

この数値例ではブートストラップ確率 $p = 0.0028$ と厳密な $p = 0.05$ がかなりズレている. この原因は以前に述べたように, ブートストラップ分布が左右非対称なこと (3 節), もしくはパーセンタイル信頼区間における $\hat{\theta}_\alpha$ がブートストラップ- t 法と逆向きにブートストラップ分布を利用していること (6 節) である. この問題点を解決するのが次節の主題であり, そこで与えられる方法を適用すると $p = 0.051$ が得られる.

ブートストラップ確率である。これはブートストラップ標本に階層型クラスタリングを適用して同じクラスが観測された回数を数えて計算したものである。ここでは (19) を逆向きに考えた仮説 $\theta \geq \theta_0$ の p -値が $1 - p$ で与えられることを考慮して、 $p > 0.95$ に注目している。この手続きにおいて $\hat{\theta}(\mathcal{X})$ を明示的に定義しなくても (22) を計算したことになる。一方、各枝の左側にある数値は次節の方法で計算したものであり、(23) に相当する量が自動的に得られている。図 2 では両者の値がかなりズれているので、ブートストラップ確率よりも次節の方法を利用したほうよい。

8 マルチスケール・ブートストラップ

これまでアルゴリズム 1 において $n' = n$ としていたが、これ以外の値を利用することを考える。以下では $\sigma = \sqrt{n/n'}$ とおきスケールとよぶ。アルゴリズム 9 のステップ 1 で得られる回数を C_{σ^2} 、ブートストラップ確率を $\alpha_{\sigma^2} = C_{\sigma^2}/B$ で表すことにする。このとき、Shimodaira (2002, 2004) で提案された次の方法によって精度の高い p -値が計算できる。

アルゴリズム 10 (マルチスケール・ブートストラップ)

1. あらかじめ $M \geq 2$ 個の n' の値を定めておき、 n'_1, \dots, n'_M とする。スケールを $\sigma_i = \sqrt{n/n'_i}$, $i = 1, \dots, M$ とおく。本節の数値例では $M = 13$ として、 $\sigma^2 = 1/9$ と $\sigma^2 = 9$ の間を $\log \sigma^2$ に関して等間隔に分割した。
2. $i = 1, \dots, M$ に対してアルゴリズム 1 を $n' = n'_i$ で実行して事象の回数 $C_{\sigma_i^2}$ を数え、ブートストラップ確率 $\alpha_{\sigma_i^2} = C_{\sigma_i^2}/B$ を計算する。
3. 上記で求めたブートストラップ確率の測定値に次式のモデルを当てはめてパラメータベクトル $\beta = (v, c)$ を推定する。

$$\alpha_{\sigma_i^2} = 1 - \Phi\left(\frac{v}{\sigma_i} + c\sigma_i\right), \quad i = 1, \dots, M \quad (24)$$

4. 推定したパラメータ値 $\hat{\beta} = (\hat{v}, \hat{c})$ を用いて、次式の p -値を計算する。

$$p = 1 - \Phi(\hat{v} - \hat{c}) \quad (25)$$

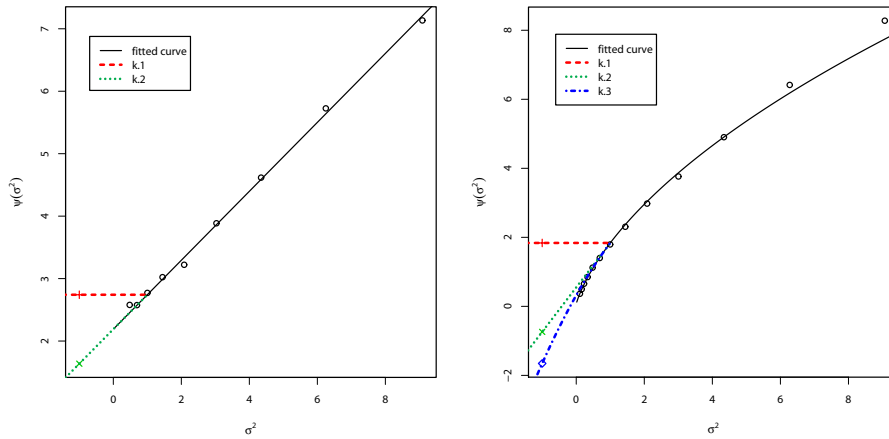


図3 正規化ブートストラップ z -値 (σz_{σ^2}) の変化. (左) 図1 のデータで仮説 $\theta \leq 10$ としたものと, (右) 図2 のデータで67番のクラスタ (左から3番目の赤い箱で示された22個の被験者からなる集合) に注目したもの. ブートストラップ確率の実測値 $\alpha_{\sigma_i^2}$, $i = 1, \dots, M$ から計算した $\sigma_i z_{\sigma_i^2}$ をプロットして (ただし $\alpha_{\sigma_i^2} = 0$ または 1 は取り除いた), モデル式から推定した曲線と $\sigma^2 = -1$ への外挿を示した.

このアルゴリズムを図1のデータに適用する. $n = 100$ であるが, $n'_i = 900, 624, 433, 300, 208, 144, 100, 69, 48, 33, 23, 16, 11$ とおいて $\sigma_i^2 \approx 1/9, \dots, 9$ とした. ブートストラップ複製 ($B = 10000$) で $\hat{\theta}^* \leq 10$ となった回数は $C_{\sigma_i^2} = 0, 0, 0, 0, 1, 10, 28, 60, 128, 128, 134, 110, 90$ であり, とくに $n'_i = n$ のときのブートストラップ確率は $\alpha_1 = 28/10000 = 0.0028$ であった.

ステップ3を図示するために, ブートストラップ確率 α_{σ^2} から変換 $z_{\sigma^2} = z^{(1-\alpha_{\sigma^2})}$ によってブートストラップ z -値を定義する. 逆に $\alpha_{\sigma^2} = 1 - \Phi(z_{\sigma^2})$ としてもよい. そして σz_{σ^2} を正規化ブートストラップ z -値 (normalized bootstrap z -value) とよぶことにする. これを縦軸にプロットしたのが図3 (左) である. (24) を変形すると, σz_{σ^2} のスケール変換のモデル式

$$\sigma_i z_{\sigma_i^2} = v + c\sigma_i^2, \quad i = 1, \dots, M \quad (26)$$

が得られ, 図3 (左) の直線の y 切片が $\hat{v} = 2.190$, 傾きが $\hat{c} = 0.552$ と推定される.

ステップ4の(25)を計算すると, $p = 1 - \Phi(2.190 - 0.552) = 1 -$

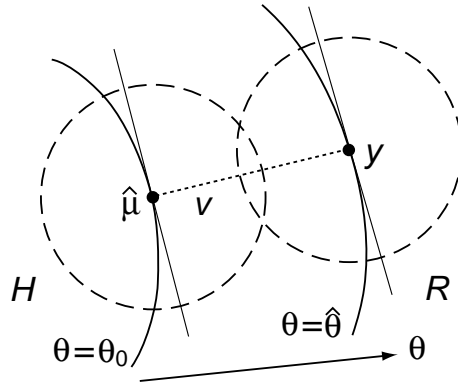


図4 ブートストラップと p -値の関係. 「 $\theta = \text{一定}$ 」の曲面は θ が大きくなると右へ移動する. $n' = n$ の場合のブートストラップ標本 \mathbf{y}^* の分布は \mathbf{y} を中心にした半径 $\sigma = 1$ の円で表現されている. p -値を得るための \mathbf{y}^* の分布は $\hat{\boldsymbol{\mu}}$ を中心にした半径 1 の円で表現されている.

$\Phi(1.638) = 0.051$ となる. ブートストラップ確率 $\alpha_1 = 0.0028$ であったことを考慮すれば, アルゴリズム 10 で得られる p -値が 7 節で述べた厳密値 $p = 0.05$ にきわめて近い数値であることがわかる. (26) の右辺で $\sigma_i^2 = -1$ とおけば $v - c$ となることに注意すると, 図 3 (左) の直線を $\sigma^2 = -1$ へ外挿した値が $\hat{v} - \hat{c}$ であり, これから (25) を計算すると解釈してもよい. 標語的には, ブートストラップ確率を $n' = -n$ へ外挿したものが p -値になる.

アルゴリズム 10 の原理を導出するためのモデルを Efron and Tibshirani (1998) にならって説明する. まずデータ \mathcal{X} を適当に非線形変換したベクトル \mathbf{y} が多変量正規分布

$$\mathbf{y} \sim N_K(\boldsymbol{\mu}, \mathbf{I}) \quad (27)$$

に従うと仮定 (または近似) する. 平均ベクトル $\boldsymbol{\mu}$ は未知パラメータ, 分散共分散行列は単位行列 \mathbf{I} である. これはアルゴリズム導出のためのモデルであり, 実際にその非線形変換を求める必要はないし, 次元 K も任意である.

\mathcal{X} から $\hat{\theta}$ への関係 $\hat{\theta}(\mathcal{X})$ は, \mathcal{X} の代わりに \mathbf{y} を使って $\hat{\theta} = \hat{\theta}(\mathbf{y})$ と書かれる. この \mathbf{y} を $\boldsymbol{\mu}$ でおきかえたものが真のパラメータ値 $\theta = \hat{\theta}(\boldsymbol{\mu})$ であるから, (19) の仮説 $\theta \leq \theta_0$ は K 次元空間の領域 $\mathcal{H} = \{\boldsymbol{\mu} | \hat{\theta}(\boldsymbol{\mu}) \leq \theta_0\}$ と表される (図 4). すなわち, $\boldsymbol{\mu} \in \mathcal{H}$ が仮説である.

ブートストラップ標本 \mathcal{X}^* を変換した \mathbf{y}^* は (27) のパラメトリック・ブー

トストラップを形式的に考えて

$$\mathbf{y}^* \sim N_K(\mathbf{y}, \sigma^2 \mathbf{I}) \quad (28)$$

に従うと近似しておく.

このとき, (22) のブートストラップ確率は $\alpha_{\sigma^2} = P_{\mathbf{y}, \sigma^2}(\mathbf{y}^* \in \mathcal{H})$ と表される. ただし $P_{\mathbf{y}, \sigma^2}(\cdot)$ は (28) に関する確率である. 一方, 領域 $\mathcal{R} = \{\mathbf{y}^* | \hat{\theta}(\mathbf{y}^*) \geq \hat{\theta}\}$ を定義すると, (23) の p -値は $p = P_{\hat{\mu}, 1}(\mathbf{y}^* \in \mathcal{R})$ である. ただし $\hat{\mu}$ は領域 \mathcal{H} の境界上で \mathbf{y} への距離が最小になる点を表す. α_1 と p のズレの原因を図 4 で解釈すると, \mathbf{y} から見た \mathcal{H} と $\hat{\mu}$ から見た \mathcal{R} の表裏が入れ替わっていることが原因になっている.

このことを数式で表現するために, 符号付距離を $v = \pm \|\mathbf{y} - \hat{\mu}\|$ で定義し ($\mathbf{y} \in \mathcal{H}$ なら負, $\mathbf{y} \notin \mathcal{H}$ なら正), $\hat{\mu}$ における \mathcal{H} の境界の曲率を c で表す (符号は図 4 のとき正, 逆向きに曲がっているとき負とする). もし \mathcal{H} の境界が平坦で $c = 0$ ならば v は分散 1 の正規分布に従うから, $\alpha_1 = p = 1 - \Phi(v)$ である. 実際には曲がっていて $c \neq 0$ の場合, c が大きくなる分だけ α_1 は減り, 逆に p が増える. それを厳密に評価すると

$$\alpha_1 = 1 - \Phi(v + c), \quad p = 1 - \Phi(v - c) \quad (29)$$

で近似されることが知られていて, この後者が (25) である. アルゴリズム 10 の v と c にはこのような幾何的な意味があったのである. 一方, スケールを $\sigma = 1$ から変化させることは v, c をそれぞれ $v/\sigma, c\sigma$ でおきかえることに等価であるから, (29) の前者から (24) が得られる. 以上がアルゴリズム 10 の原理である.

図 3 (左) ではアルゴリズム 10 がうまく機能したが, 図 3 (右) をみると (26) の直線が当てはまらない. これまでの議論は \mathcal{H} の境界が滑らかな曲面であることを仮定したが, たとえばもし \mathcal{H} が錐ならばスケール変換で形が不変なので c (に相当する量) が変化せず, $\alpha_{\sigma^2} = 1 - \Phi(v/\sigma + c)$ のほうがよいモデルになる. これをすこし一般化してモデル式 $\sigma z_{\sigma^2} = v + c\sigma^2 / (1 + s(\sigma - 1))$ を当てはめて $\beta = (v, c, s)$ を推定した結果が図 3 (右) の曲線である. $s = 0$ が (26), $s = 1$ が錐に相当する. そして Shimodaira (2008) はアルゴリズム

10 のステップ 4 を次式でおきかえた.

$$p_k = 1 - \Phi \left(\sum_{j=0}^{k-1} \frac{(-1 - \sigma_0^2)^j}{j!} \frac{\partial^j (\sigma z_{\sigma^2})}{\partial (\sigma^2)^j} \Big|_{\sigma_0^2} \right) \quad (30)$$

この式は σz_{σ^2} を σ^2 の関数とみなして σ_0^2 においてテイラー展開 (k 項で打ち切り) して $\sigma^2 = -1$ へ外挿したものである. ここでは $\sigma_0^2 = 1$ としておく.

とくに, $k = 1$ はブートストラップ確率 $\alpha_1 = p_1$, $k = 2$ は (25) の $p = p_2$ に一致し, さらに k を大きくした p_3 は誤差が小さいと考えられる. 図 3 (右) では, $p_1 = 0.033$, $p_2 = 0.770$, $p_3 = 0.951$ と計算され, この p_3 を各クラスタで計算したものが図 2 の各枝の左の数値である. 一般に k を増やすと誤差は減少するが数値的に不安定になるので, $k = 3$ 程度が妥当である. この不安定になる現象は Perlman and Wu (1999) で扱われた推測原理の本質的な問題の一例であり, さらに議論を深める必要がある.

参考文献

- [1] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, Cambridge, 1997.
- [2] B. Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, Vol. 7, pp. 1–26, 1979.
- [3] B. Efron and R. Tibshirani. The problem of regions. *Annals of Statistics*, Vol. 26, pp. 1687–1718, 1998.
- [4] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.
- [5] M. E. Garber, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 98, pp. 13784–13789, 2001.
- [6] P. Hall. *The bootstrap and Edgeworth expansion*. Springer-Verlag, New York, 1992.
- [7] M. D. Perlman and L. Wu. The emperor’s new tests. *Statistical Science*, Vol. 14, pp. 355–381, 1999.

- [8] H. Shimodaira. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, Vol. 51, pp. 492–508, 2002.
- [9] H. Shimodaira. Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *Annals of Statistics*, Vol. 32, pp. 2616–2641, 2004.
- [10] 下平英寿. モデル選択 — 予測・検定・推定の交差点, 統計科学のフロンティア 3, 情報量規準によるモデル選択とその信頼性評価, pp. 1–76. 岩波書店, 2004.
- [11] H. Shimodaira. scaleboot: Approximately unbiased p-values via multi-scale bootstrap, 2006. R package is available from CRAN.
- [12] H. Shimodaira. Testing regions with nonsmooth boundaries via multi-scale bootstrap. *Journal of Statistical Planning and Inference*, Vol. 138, pp. 1227–1241, 2008.
- [13] R. Suzuki and H. Shimodaira. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, Vol. 22, pp. 1540–1542, 2006.
- [14] 汪金芳, 田栗正章. 計算統計 I — 確率計算の新しい手法, 統計科学のフロンティア 11, ブートストラップ法入門, pp. 1–64. 岩波書店, 2003.

第9章 EMアルゴリズム

渡辺美智子¹

EMアルゴリズムは、不完全データに基づく統計モデル一般に適用される最尤推定導出のためのアルゴリズムである。不完全データを広義にとらえることで、欠損データの問題に限らず、切断や打ち切りデータ、有限混合分布モデル、ロバスト分布モデルや潜在変数モデル、ベイズモデルと、その適用範囲が広いことが特徴である。本章ではEMアルゴリズムの基本的な考え方や理論及び代表的なモデルに対する具体例を解説し、ECMやECMEなどの拡張EMについても概観する。

¹東洋大学経済学部

Contents

1	EM アルゴリズム	1
1.1	はじめに	3
1.2	EM アルゴリズムの考え方	4
1.3	EM アルゴリズムの理論と一般形	6
1.4	EM アルゴリズムと適用例	8
1.4.1	血液型遺伝子に関する発生確率の推定	8
1.4.2	多変量データにおける欠測：多変量正規モデル	9
1.4.3	重回帰モデル：目的変数に欠測がある場合	12
1.4.4	中途打ち切りデータに基づく例	13
1.4.5	混合分布モデル	16
1.5	EM を利用した漸近分散共分散行列の評価	20
1.5.1	Louis の方法	20
1.5.2	Oakes の方法	22
1.5.3	SEM(Supplemented EM)	23
1.6	EM の特性	23
1.7	GEM アルゴリズムとその他の拡張型 EM	25
1.7.1	GEM(Generalized EM)	25
1.7.2	ECM アルゴリズム	26
1.7.3	ECME アルゴリズム – Contaminated Normal –	27

CONTENTS

1.7.4 加速化を意識した完全データの探索	
— optimal EM アルゴリズム—	29

1.1 はじめに

EM アルゴリズムは、不完全データから最尤推定値を導くための統一的なアルゴリズムで、Dempster, Laird and Rubin(1977) が最初に、EM という一般的なフレームとその背景となる理論および諸種の統計モデルへの広い適用性を示した。当初、EM アルゴリズムは、最適化アルゴリズムとして既に普及していた Newton-Raphson 法やスコアリング法などと比較し、収束に至るまでの反復回数が多いことなどの問題点が指摘されていたが、アルゴリズムのフレームの単純性と収束の安定性、適用モデルの汎用性などの利点が勝り、今日に至るまでに、工学・医学・社会学・経営学など統計モデルが必要とされるほぼすべての分野で、EM アルゴリズムを推測に利用するモデル構築が数多く発表され続け、尤度に基づく統計分析法を構築する上での汎用的なツールになっている。

EM アルゴリズムの利点は、先述したアルゴリズムのフレームの単純性と収束の安定性、適用モデルの汎用性にある。具体的には、フレームが単純であるため、尤度の定式化から具体的なアルゴリズムの導出およびプログラミングの作成に至るまでの作業効率が比較的に高く、また、尤度の形状にも拠るが、Newton-Raphson 法などと比較した場合に初期値によらず局所的な最適解が得られる安定性がある。そして、EM アルゴリズムの適用事例が広範にわたる最大の理由として、対象とするデータの不完全性の解釈に柔軟性があり、かなり自在に適用モデルを拡張できる汎用性が挙げられる。

不完全データ本来の意味である欠測値問題はもとより、切断分布や打ち切り分布、混合分布、分散分析における変量効果モデル・混合効果モデル、ロバスト分析、潜在変数モデル、ベイズモデルと主要な統計モデルがその範疇に入る。また、範疇に入ったモデル同士を組み合わせさらに複雑なモデルに最尤解を与えることも可能となる。この意味で、EM アルゴリズムは単なる

CONTENTS

計算手段というより、様々な現実課題の様相により適合し説明力のある複雑な統計モデルを一挙に実用段階に押し上げ、さらにモデルの拡張性へのヒントを与え得るアルゴリズムとも言える。

1.2 EM アルゴリズムの考え方

EM アルゴリズムは、不完全データの問題を完全データのフレームワークで逐次的に解決する方法論の一つと位置づけられる。基底をなす考え方は単純で、次のステップでまとめることができる。

1. 手元にあるデータ \mathbf{y}_{obs} だけでは解を導くことがすぐには困難な問題に対して、解決が容易であると考えられるレベルまでデータを完全化し(欠けているデータ部分 \mathbf{y}_{mis} が存在するとして)問題を定式化する。
2. 例えば、パラメータ θ の推定値 $\hat{\theta}$ を導くことが当面の目的であれば、 \mathbf{y}_{mis} に暫定的な値を埋め込み $\hat{\theta}$ を求める。
3. $\hat{\theta}$ を用いて \mathbf{y}_{mis} を改良し、その値を埋め込む。
4. $\hat{\theta}$ の値が収束するまで、前の2つのステップを繰り返す。

例として、血液遺伝子(A,B,O)に関する発生確率 $\theta = (p_A, p_B, p_O)$, ($p_O = 1 - p_A - p_B$) の推定の問題を考えてみる。実際に観測されるデータは、表 1.1 のように血液型に関する頻度データのみで、このデータに基づく直接的な最尤解は数式で明示的に表現することはできず、何等かの反復計算が必要となる。

Table 1.1: 観測データ

CONTENTS

血液型	セル確率	観測データ
O	p_O^2	n_O
A	$p_A^2 + 2p_{AP_O}$	n_A
B	$p_B^2 + 2p_{BP_O}$	n_B
AB	$2p_{AP_B}$	n_{AB}
	合計	n

そこで、観測された頻度データは、目的とする血液遺伝子発生確率を推定するためには不十分な不完全データとみなし、表 1.2 のような血液遺伝子の発生確率と直接対応する頻度データの存在を想定してみると、この完全データの下では、目的パラメータの推定問題ははるかに容易になる。

Table 1.2: 観測不可能な完全データ

遺伝子型	セル確率	完全データ
O	p_O^2	n_O
AA	p_A^2	n_{AA}
AO	$2p_{AP_O}$	n_{AO}
BB	p_B^2	n_{BB}
BO	$2p_{BP_O}$	n_{BO}
AB	$2p_{AP_B}$	n_{AB}
	合計	n

もちろん、想定した完全データ (表 1.2) は実際には観測されないので、これをどのように推定して置き換えるかの問題になる。EM アルゴリズムでは、観測される不完全データが与えられた下での完全データの対数尤度の条件付き期待値を介してこの置き換えを行う。具体的な計算例は後述するが、EM アルゴリズムの他にも、このようなロジックを背景にした方法論として、(a) multiple imputation (Rubin 1987a), (b) data augmentation (Tanner and Wong 1987), (c) Gibbs sampler (Geman and Geman 1984), (d)

CONTENTS

Sampling/Importance Resampling Algorithm (Rubin 1987b) などがある。Rubin(1991) では、これらを拡張型 EM のフレームワークの下で解説しているが、その区別は一般に、ステップ 2 における \mathbf{y}_{mis} への値の埋め込み方法の違いで説明される。

1.3 EM アルゴリズムの理論と一般形

いま、完全データに対応する変数を \mathbf{Y} とし、 \mathbf{Y} は観測データ（不完全データ）に対する変数 \mathbf{Y}_{obs} と観測されていないデータ（欠測データ）に対応する変数 \mathbf{Y}_{mis} で構成されるとする。ここでは、2つの標本空間： $\Omega_{\mathbf{Y}}$ （完全データの空間）と $\Omega_{\mathbf{Y}_{obs}}$ （不完全データの空間）があり、これらは、 $\Omega_{\mathbf{Y}}$ から $\Omega_{\mathbf{Y}_{obs}}$ への多対 1 写像で関連づけられているとする。

このとき、 \mathbf{Y} の密度関数を $f(\mathbf{y} | \boldsymbol{\theta})$ とすると、観測変数 \mathbf{Y}_{obs} の密度関数 $g(\mathbf{y}_{obs} | \boldsymbol{\theta})$ は

$$g(\mathbf{y}_{obs} | \boldsymbol{\theta}) = \int_{\mathbf{y}_{mis}} f(\mathbf{y} | \boldsymbol{\theta}) d\mathbf{y}_{mis} \quad (1.1)$$

となる。

不完全データの発生の過程に関しては、欠測がランダムである (missing at random), 即ち、欠測メカニズムを無視した下で、観測データ \mathbf{y}_{obs} の尤度に基づいてパラメータ $\boldsymbol{\theta}$ の推定を行うことが妥当であることを仮定しておく (Rubin(1976))。EM アルゴリズムの目的は、パラメータ $\boldsymbol{\theta}$ を最尤推定することである。

パラメータ $\boldsymbol{\theta}$ の定義域に対して完全データに基づく対数尤度が定義でき、その関数を $LL_c(\boldsymbol{\theta}) = \log f(\mathbf{y} | \boldsymbol{\theta})$, 不完全データに基づく対数尤度関数を $LL(\boldsymbol{\theta}) = \log g(\mathbf{y}_{obs} | \boldsymbol{\theta})$ とすれば、適当な条件の下で、 $LL(\boldsymbol{\theta})$ の最大化は $LL_c(\boldsymbol{\theta})$ の条件付き期待値の最大化を通して達成できる (Dempster, Laird and Rubin (1977)). つまり、一般に最適化が複雑になる不完全データ \mathbf{y}_{obs} の対数尤度 $LL(\boldsymbol{\theta})$ の直接的な導出を回避して、観測データ \mathbf{y}_{obs} と k 回目の $\boldsymbol{\theta}$ の暫定値 $\boldsymbol{\theta}^{(k)}$ が与えられた下で、完全データ \mathbf{y} の対数尤度 $LL_c(\boldsymbol{\theta})$ の条件付

CONTENTS

き期待値を最大化する値としてパラメータ θ の更新値を得る以下のプロセス

$$\theta^{(k+1)} = \max_{\theta \in \Theta} E[LL_c(\theta) | \mathbf{y}, \theta^{(k)}] \quad (1.2)$$

を繰り返すことによって、目的である $LL(\theta)$ の最大化を達成する θ の最尤推定値を導くことができる。

(1.2) 式は、具体的に、E-step(Expectation Step) と M-step(Maximization Step) の2つのステップに分解され、この2つのステップを反復することから EM アルゴリズムと呼ばれている。

E-step: 観測データ \mathbf{y}_{obs} と k 回目の θ の暫定値が与えられた下で、完全データの対数尤度の条件付期待値 (Q 関数) を計算するステップ

$$Q(\theta; \theta^{(k)}) = E[LL_c(\theta) | \mathbf{y}_{obs}, \theta^{(k)}] \quad (1.3)$$

M-step: E-step で求めた $Q(\theta; \theta^{(k)})$ を最大化する $\theta^{(k+1)}$ を求めるステップ

$$\theta^{(k+1)} = \max_{\theta \in \Theta} Q(\theta; \theta^{(k)}), \quad (1.4)$$

具体的なモデルの下で、推定値の系列 $\{\theta^{(k)}\}$ の変化率および不完全データの対数尤度の系列 $LL(\theta^{(k)})$ の変化量が十分に微少になるまで、上記の E-step と M-step を交互に繰り返すことで、一般に $LL(\theta^{(k)})$ の最尤推定値 $\hat{\theta}$ を導くことができる。Dempster, Laird and Rubin (1977) と Wu(1983) には、EM アルゴリズムが $LL(\theta)$ の定常点に収束するための条件が与えられている。

もし、 $f(\mathbf{y} | \theta)$ が、多項分布、ポアソン分布、正規分布などの指数分布族の範疇に入るモデルであれば、E-step と M-step は θ の十分統計量 $\mathbf{t} = \mathbf{t}(\mathbf{x})$ を媒介にしたステップに置き換えることができ、更に分かり易くなる。

E-step: 観測データ \mathbf{y}_{obs} と k 回目の θ の暫定値 $\theta^{(k)}$ が与えられた下で、十分統計量 \mathbf{t} の条件付期待値を計算するステップ

$$\mathbf{t}^{(k+1)} = E[\mathbf{t}(\mathbf{x}) | \mathbf{y}_{obs}, \theta^{(k)}] \quad (1.5)$$

CONTENTS

M-step: 次式をみたま $\theta^{(k+1)}$ を求めるステップ, すなわち, E-step で得られた $t^{(k+1)}$ を完全データから観測された値と見做して, 通常の指数分布族における良く知られた最尤解を与える式で $\theta^{(k+1)}$ を推定するステップ

$$E[t(\boldsymbol{x}) | \boldsymbol{\theta}] = \boldsymbol{t}^{(k+1)} \quad (1.6)$$

E-step を単に個々の欠測データではなく十分統計量を条件付き期待値で代挿する, と解釈すれば, EM は前節の Rubin(1991) のフレームワークに入る. とくに, 十分統計量が個々のデータ値の線形結合で表現できる場合は, 欠測データの値自身を条件付き期待値で代挿する, と考えても差し支えない.

1.4 EM アルゴリズムと適用例

1.4.1 血液型遺伝子に関する発生確率の推定

前節で示した血液遺伝子 (A,B,O) に関する発生確率 $\boldsymbol{\theta} = (p_A, p_B, p_O)$ の推定問題の場合, 表 2.1 の観測データ (不完全データ) \boldsymbol{y}_{obs} に基づく対数尤度は次式で与えられる.

$$LL(\boldsymbol{\theta}) = 2n_O \log p_O + n_A \log(p_A^2 + 2p_A p_O) + n_B \log(p_B^2 + 2p_B p_O) + n_{AB} \log(2p_A p_B) \quad (1.7)$$

(1.7) 式を $\boldsymbol{\theta} = (p_A, p_B, p_O)$ に関して最大化するための尤度方程式は, 明示的な式で解くことはできないが, 表 2.2 の遺伝子型に対応する頻度データに基づく以下の完全データ \boldsymbol{y} の対数尤度式は, 容易に解くことができる.

$$LL_c(\boldsymbol{\theta}) = 2n_A^+ \log p_A + 2n_B^+ \log p_B + 2n_O^+ \log p_O \quad (1.8)$$

ここに, n_A^+, n_B^+, n_O^+ は, 以下で与えられる頻度である.

$$n_A^+ = n_{AA} + \frac{1}{2}n_{AO} + \frac{1}{2}n_{AB} \quad (1.9)$$

$$n_B^+ = n_{BB} + \frac{1}{2}n_{BO} + \frac{1}{2}n_{AB} \quad (1.10)$$

$$n_O^+ = n_O + \frac{1}{2}n_{AO} + \frac{1}{2}n_{BO}. \quad (1.11)$$

CONTENTS

EM アルゴリズムをこの推定問題で具体的に構築する. E-step は, (1.3) 式の $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) = E[LL_c(\boldsymbol{\theta}) | \mathbf{y}_{obs}, \boldsymbol{\theta}^{(k)}]$ を具体的に評価していると考えられることもできるし, (2.7) 式が多項分布であることから, 十分統計量 $\{n_A^+, n_B^+, n_O^+\}$ を評価していると考えても良い.

E-step: 観測データとパラメータの k 回目の推定値が与えられた下で, 完全データの条件付き期待値を計算する.

$$n_{AA}^{(k)} \equiv E[n_{AA} | n_A; p_A^{(k)}, p_O^{(k)}] = \frac{n_A p_A^{(k)^2}}{p_A^{(k)^2} + 2p_A^{(k)} p_O^{(k)}} \quad (1.12)$$

$n_{AO}, n_{BB}, n_{BO}, n_{AB}$ の条件付き期待値も同様の式で計算する.

M-step: E-step で求めた値 (完全データの推定値) を使って, $LL_c(\boldsymbol{\theta})$ を最大化する $\boldsymbol{\theta}^{(k+1)}$ を求め, パラメータの推定値を更新する.

$$p_A^{(k+1)} = \frac{n_A^{+(k)}}{n} = \frac{n_{AA}^{(k)} + \frac{1}{2}n_{AO}^{(k)} + \frac{1}{2}n_{AB}^{(k)}}{n} \quad (1.13)$$

$$p_B^{(k+1)} = \frac{n_B^{+(k)}}{n}, \quad p_O^{(k+1)} = \frac{n_O^{+(k)}}{n}. \quad (1.14)$$

適当な初期値から始めて上記の E-step と M-step を反復することにより, とくに制約を設けることなく, 確率の定義域内で目的である血液遺伝子の発生確率を求めることができる. 例えば, $n_A = 20, n_B = 10, n_{AB} = 5, n_O = 10$ が観測されたとして, 初期値 $p_A = 0.3, p_B = 0.3, p_O = 0.4$ の下に EM アルゴリズムをスタートすると, 表 (1.3) のように値が収束していく.

1.4.2 多変量データにおける欠測：多変量正規モデル

この節では, p 次の多変量正規確率変数ベクトル $\mathbf{X} = (X_1, \dots, X_p)'$ の平均ベクトルが $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$, 分散共分散行列が $\boldsymbol{\Sigma} = (\sigma_{jk})$ のとき, 欠測値を含む大きさ n の標本から $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ の最尤推定値を求める問題を考える.

CONTENTS

Table 1.3: 収束の様子

反復回数	p_A	p_B	$p_O = 1 - p_A - p_B$
初期値	0.3000	0.3000	0.4000
1	0.3384	0.1970	0.4646
2	0.3371	0.1861	0.4768
3	0.3358	0.1848	0.4794
4	0.3354	0.1846	0.4800
5	0.3353	0.1846	0.4801
6	0.3353	0.1846	0.4801

第 i 番目のケースの観測値 \mathbf{x}_i は、欠測に応じて以下のように分解されているとする.

$$\mathbf{x}_i = \begin{bmatrix} \mathbf{x}_i^0 \\ \mathbf{x}_i^1 \end{bmatrix} \equiv \begin{bmatrix} \mathbf{x}_{i,mis} \\ \mathbf{x}_{i,obs} \end{bmatrix}$$

この欠測パターンに応じたパラメータ $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ の分解を

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_0 \\ \boldsymbol{\mu}_1 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} \\ \boldsymbol{\Sigma}_{10} & \boldsymbol{\Sigma}_{11} \end{bmatrix}$$

とする. このとき, $\{\mathbf{x}_{i,obs}\}, i = 1, 2, \dots, n$ に基づくパラメータ $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ の最尤推定値を導く EM アルゴリズムは, 以下のように具体化される.

E-step : 十分統計量 $(\mathbf{x}_i, \mathbf{x}_i \mathbf{x}_i^T)$ の条件付期待値を計算するステップ

$$\begin{aligned} E[\mathbf{x}_{i,mis} \mid \mathbf{x}_{i,obs}; \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}] &= \boldsymbol{\mu}_0^{(k)} + \boldsymbol{\Sigma}_{01}^{(k)} \boldsymbol{\Sigma}_{11}^{(k)-1} (\mathbf{x}_{i,obs} - \boldsymbol{\mu}_1^{(k)}) \equiv \mathbf{x}_{i,mis}^{(k)} \\ E[\mathbf{x}_{i,mis} \mathbf{x}_{i,obs}^T \mid \mathbf{x}_{i,obs}; \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}] &= \mathbf{x}_{i,mis}^{(k)} \mathbf{x}_{i,obs}^T \\ E[\mathbf{x}_{i,mis} \mathbf{x}_{i,mis}^T \mid \mathbf{x}_{i,obs}; \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}] &= \mathbf{x}_{i,mis}^{(k)} \mathbf{x}_{i,mis}^{(k)T} + \boldsymbol{\Sigma}_{00}^{(k)} - \boldsymbol{\Sigma}_{01}^{(k)} \boldsymbol{\Sigma}_{11}^{(k)-1} \boldsymbol{\Sigma}_{10}^{(k)} \\ &\equiv [\mathbf{x}_{i,mis} \mathbf{x}_{i,mis}^T]^{(k)} \end{aligned} \quad (1.15)$$

CONTENTS

上記の結果を利用した十分統計量 $(\mathbf{x}_i, \mathbf{x}_i \mathbf{x}_i^T)$ の k 回目の反復値

$$\mathbf{x}_i^{(k)} \equiv \begin{bmatrix} \mathbf{x}_{i,mis}^{(k)} \\ \mathbf{x}_{i,obs} \end{bmatrix}, \quad [\mathbf{x}_i \mathbf{x}_i^T]^{(k)} \equiv \begin{bmatrix} [\mathbf{x}_{i,mis} \mathbf{x}_{i,mis}^T]^{(k)} & \mathbf{x}_{i,mis}^{(k)} \mathbf{x}_{i,obs}^T \\ \mathbf{x}_{i,obs} \mathbf{x}_{i,mis}^{(k)T} & \mathbf{x}_{i,obs} \mathbf{x}_{i,obs}^T \end{bmatrix}$$

に対して、次の M-step の計算が適用される。

M-step : E-step で求めた十分統計量の期待値を完全データからの観測結果と見做して、 $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ の通常の最尤解を計算するステップ

$$\boldsymbol{\mu}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^{(k)} \quad (1.16)$$

$$\boldsymbol{\Sigma}^{(k+1)} = \frac{1}{N} \sum_{i=1}^n [\mathbf{x}_i \mathbf{x}_i^T]^{(k)} - \boldsymbol{\mu}^{(k+1)} \boldsymbol{\mu}^{(k+1)T} \quad (1.17)$$

E-step と M-step の反復により、とくに制約を設けることなく、 $\boldsymbol{\Sigma}$ に関して正定値性が保証された下で、最尤解を求めることができる。

表 1.4 を 2 変量正規分布からの標本と仮定し、EM アルゴリズムによって平均と分散、共分散の最尤推定値を求めてみる。ここで、表中の * は欠測を意味し、欠測は Rubin(1976) の意味でランダムであるとする。初期値として、データが完全に得られている 3 つのケース 1,2,6 から計算した値を使って E-step と M-step を反復させた収束の様子が、表 1.5 である。

Table 1.4: 欠測値を含む 2 変量データ

No.	1	2	3	4	5	6	7
X_1	5	8	*	*	4	10	6
X_2	10	20	15	40	*	60	*

Table 1.5: 収束の様子

反復回数	μ_1	μ_2	σ_1	σ_2	σ_{12}
初期値	7.667	30.000	6.333	700.000	40.000
1	6.864	24.474	4.791	436.150	29.996
2	6.735	24.372	4.350	381.408	30.286
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
34	6.767	22.861	3.995	385.525	34.695
35	6.767	22.861	3.995	385.525	34.695

1.4.3 重回帰モデル：目的変数に欠測がある場合

重回帰モデル $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I})$ で、目的変数 \mathbf{y} に欠測データが含まれる場合のパラメータの最尤推定値を求める問題を考える。この問題は、 \mathbf{X} が計画行列の場合には、意図された不完備実験計画および予期せぬ欠測などの理由でバランスの崩れた実験データの推測にも適用できる。

観測と欠測に応じて、変数 \mathbf{y} を観測変数 \mathbf{y}_{obs} , 欠測変数 \mathbf{y}_{mis} に分解する。ここでは、欠測はランダムであると仮定した下で、 \mathbf{y}_{obs} に基づく回帰パラメータ $\boldsymbol{\beta}$ と誤差分散 σ^2 の推定を EM アルゴリズムで以下に具体化する。ここで、 \mathbf{x}_i^T は説明変数行列 \mathbf{X} の第 i 行ベクトルとする。

E-step : 観測されている説明変数 \mathbf{x}_i の値と回帰パラメータ $\boldsymbol{\beta}$, 誤差分散 σ^2 の k 回目の推定値が与えられた下で、目的変数が欠測したケースに対して、十分統計量 (y_i, y_i^2) の期待値を計算する。

$$y_{i,mis}^{(k)} \equiv E[y_i \mid \mathbf{x}_i, \boldsymbol{\beta}^{(k)}, \sigma^{2(k)}] = \mathbf{x}_i^T \boldsymbol{\beta}^{(k)} \quad (1.18)$$

$$y_{i,mis}^{2(k)} \equiv E[y_i^2 \mid \mathbf{x}_i, \boldsymbol{\beta}^{(k)}, \sigma^{2(k)}] = (\mathbf{x}_i^T \boldsymbol{\beta}^{(k)})^2 + \sigma^{2(k)} \quad (1.19)$$

すべてのケースに対して十分統計量 (y_i, y_i^2) の k 回目の反復値を以下で定義

CONTENTS

する.

$$y_i^{(k)} \equiv \begin{cases} y_{i,obs}, & \text{if } y_i \text{が観測,} \\ y_{i,miss}^{(k)}, & \text{if } y_i \text{が欠測,} \end{cases} \quad (1.20)$$

$$y_i^{2(k)} \equiv \begin{cases} y_{i,obs}^2, & \text{if } y_i \text{が観測,} \\ y_{i,miss}^{2(k)}, & \text{if } y_i \text{が欠測.} \end{cases} \quad (1.21)$$

M-step : E-step で求めた値を使って, 通常の方法で目的パラメータの推定を行う.

$$\boldsymbol{\beta}^{(k+1)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X} \mathbf{y}^{(k)} \quad (1.22)$$

$$\sigma^{2(k+1)} = \frac{\sum_{i=1}^n [y_i^{2(k)} - (\boldsymbol{\beta}^{(k)} \mathbf{x}_i)^2]}{n} \quad (1.23)$$

この場合も, σ^2 は, 正の値の範囲で推定される.

1.4.4 中途打ち切りデータに基づく例

$(n + m)$ 個の個体があり, n 個の個体はその生存時間 y_1, \dots, y_n が観測され, 残り m 個の個体については t 時間の観察の後, 生存か死亡のいずれかの結果 s_1, \dots, s_m だけが観測されている状況を考える. ここで, $s_i = 1$ は生存, $s_i = 0$ は死亡を意味する. また, ℓ は, 生存時間が観測されなかった m 個の個体のうち, 実験打ち切り時間 t において生存していた個体数, すなわち $\ell = \sum_{i=1}^m s_i$ とする.

Flury and Zoppe (2000) は, この設定で EM アルゴリズムが適用できる場合と適用できない場合の 2 つの例を与えている.

指数分布を仮定する場合 : EM が適用できる例

対象としている個体の生存時間が平均 θ を持つ指数分布に従っていると仮定すると, いま観測されている不完全データの対数尤度は,

$$LL(\theta) = -n(\log \theta + \bar{y}/\theta) - \ell t/\theta + (m - \ell) \log(1 - e^{-t/\theta}) \quad (1.24)$$

CONTENTS

となる. ここで, \bar{y} は y_1, \dots, y_n の平均を表す. $\ell = m$ のとき, (1.24) 式を最大化する θ の解は, 代数的に求めることができ, 最尤推定量は, $\hat{\theta} = (n\bar{y} + mt)/n$ となる. また, $\ell \neq m$ のときでも, (1.24) 式の最大化を通常最適化アルゴリズムで達成することは難しくはない. しかし, ここでは, この問題を EM アルゴリズムで考えてみる.

いま生存時間が正確に観測されなかった m 個の個体に対して, 本来観測されたであろう生存時間 (欠測データ) を x_1, \dots, x_m とし, 完全データを $y_1, \dots, y_n, x_1, \dots, x_m$ と想定すると, 完全データの対数尤度は

$$LL_c(\theta) = -n(\log \theta + \bar{y}/\theta) - \sum_{i=1}^m (\log \theta + x_i/\theta). \quad (1.25)$$

となる. (1.25) 式は x_i の線形関数なので, EM アルゴリズムは, 欠測データ x_i 自身の条件付期待値を介して構築される.

E-step: 不完全データとパラメータの k 回目の推定値 $\theta^{(k)}$ が与えられた下, 欠測データ x_i の条件付期待値を計算する.

$$x_i^{(k)} \equiv E[x_i | s_i, \theta^{(k)}] = \begin{cases} t + \theta^{(k)}, & \text{if } s_i = 1, \\ \theta^{(k)} - (te^{-t/\theta^{(k)}})/(1 - e^{-t/\theta^{(k)}}), & \text{if } s_i = 0. \end{cases} \quad (1.26)$$

(1.25) 式の x_i に (1.26) の $x_i^{(k)}$ を代入して,

$$\begin{aligned} Q(\theta; \theta^{(k)}) &= E[LL_c(\theta) | y_i, s_i, \theta^{(k)}] \\ &= -(n+m) \log \theta - \frac{1}{\theta} \{n\bar{y} + \ell(t + \theta^{(k)}) + (m-\ell)(\theta^{(k)} - tp^{(k)})\} \end{aligned} \quad (1.27)$$

を得る. $p^{(k)} = e^{-t/\theta^{(k)}}/(1 - e^{-t/\theta^{(k)}})$.

M-step: (1.27) 式を最大化する θ を求める.

$$\theta^{(k+1)} = \frac{n\bar{y} + \ell(t + \theta^{(k)}) + (m-\ell)(\theta^{(k)} - tp^{(k)})}{n+m} \quad (1.28)$$

適当な正の値を初期値とし, E-step と M-step を繰り返すことで, θ の最尤推定値を得ることができる. とくに, $\ell = m$ のとき, (1.28) 式の θ に関する自己一致解は, (1.24) 式から導いた最尤解 $\hat{\theta} = (n\bar{y} + mt)/n$ と一致する.

CONTENTS

(0, \theta] 上の一様分布を仮定する場合：EM が適用できない例

前節と同様な生存時間の観測状況に対して，生存時間が (0, \theta] 上の一様分布に従う場合のパラメータ \theta の最尤推定を考える。

仮に完全データ x_i が得られたとした場合の最尤推定値は $\max\{x_{max}, y_{max}\}$ である。 x_{max}, y_{max} は，それぞれ $\{x_i\}_{1 \leq i \leq m}, \{y_i\}_{1 \leq i \leq n}$ の最大値。

ここで，このまま先述の指数分布の場合のように， $E[x_i | s_i, \theta^{(k)}]$ を介した EM アルゴリズムを構築すると， $\ell \geq 1$ の場合は， $\theta \geq t$ で， E-step は，

$$E[x_i | s_i, \theta^{(k)}] = \begin{cases} (t + \theta^{(k)})/2, & \text{if } s_i = 1, \\ t/2, & \text{if } s_i = 0. \end{cases} \quad (1.29)$$

M-step は，

$$\theta^{(k+1)} = \max\{y_{max}, (t + \theta^{(k)})/2\} \quad (1.30)$$

となり，この E-step と M-step の反復結果は， $\max\{y_{max}, t\}$ に到達する。しかしこれは，このモデルの下での最尤推定値とはならない。

この場合の観測データ（不完全データ）に基づく尤度は，

$$L(\theta) = \theta^{-N} I_{[y_{max}, \infty)}(\theta) (t / \max(t, \theta))^{m-\ell} (1 - t / \max(t, \theta))^\ell \quad (1.31)$$

で，とくに $\ell = 0$ の場合，

$$L(\theta) = \theta^{-n} I_{[y_{max}, \infty)}(\theta) (t / \max(t, \theta))^m \quad (1.32)$$

となり， $\theta \geq y_{max}$ の範囲で減少関数なので，最尤推定値は y_{max} となる。

$\ell \geq 1$ の場合は， $\theta \geq t$ で， $\theta \geq t$ のとき，関数 $h(\theta) = \theta^{-(n+m)}(\theta - t)^\ell$ は， $\dot{\theta} = \frac{n+m}{n+m-\ell}t$ で最大値をとり， $\theta > \dot{\theta}$ の範囲で単調減少関数なので，(1.31) 式の尤度は， $\dot{\theta} > y_{max}$ のとき $\dot{\theta}$ で，また， $\dot{\theta} < y_{max}$ のとき y_{max} で最大となる。したがって，一様分布の下での θ の最尤推定値 $\hat{\theta}$ は，

$$\hat{\theta} = \begin{cases} \dot{\theta}, & \text{if } \dot{\theta} > y_{max} \text{ and } \ell \geq 1, \\ y_{max}, & \text{その他,} \end{cases} \quad (1.33)$$

となる。

CONTENTS

ここで、前述の (1.29) 式と (1.30) 式による EM アルゴリズムが正しく最尤推定値を導かなかった理由は、すべての $\theta > 0$ について対数尤度関数が存在するわけではないからである。例えば、生存時間が観測されなかった 1 つの個体の生存時間 x_i が t を超えている場合、この x_i の確率密度は

$$f_x(x_i; \theta) = \begin{cases} 1/\theta, & \text{if } 0 \leq x_i \leq \theta, \\ 0, & \text{その他,} \end{cases} \quad (1.34)$$

となる。ここで、 $x_i > t$ の範囲での条件付分布が $[t, \theta^{(k)}]$ の一様分布となるため、 $\theta < \theta^{(k)}$ となる θ においては、密度関数の値は 0 となる。したがって、 $\theta^{(k)}$ が与えられたときの x_i の条件付期待値は存在しないことになる。

この例のように、ある範囲において尤度の値が 0 となる場合、対数尤度の期待値を求めることができず、EM アルゴリズムが適用できない。EM アルゴリズムにおける E-step は、不完全データの補完ではなく、あくまで対数尤度の期待値を求めるステップであることに注意しなければならない。

1.4.5 混合分布モデル

潜在クラスモデルをはじめとする有限混合分布モデルでは、最尤推定に EM アルゴリズムが利用されることが多い。いま、一般に p 個のコンポーネントに対して、各コンポーネントの分布が $f_j(x; \theta_j)$ 、構成率が π_j の混合分布モデル

$$\sum_{j=1}^p \pi_j f_j(x; \theta_j), \quad \text{ここに, } \sum_{j=1}^p \pi_j = 1 \quad (1.35)$$

を想定する。

実際に n 個の対象に対して観測されるデータは x_1, \dots, x_n であるが、更に各対象のコンポーネントへの所属を示す 0-1 型のデータ $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})$ の存在を仮定し、 $(x_1, \mathbf{z}_1), \dots, (x_n, \mathbf{z}_n)$ を観測されるべき完全データと見做す。ここで、 z_{ij} は、第 i 個体が第 j コンポーネントに所属する場合に 1 をとり、それ以外には 0 をとる。ただし、すべての \mathbf{z}_i は欠測して観測されない、いわゆる潜在データと考える。

CONTENTS

すなわち, (1.35) 式のモデルは, 潜在データ z_{ij} を導入することで,

$$x_i | (z_{ij} = 1) \sim f_j(x_i; \boldsymbol{\theta}_j), \quad \text{ただし, } Pr(z_{ij} = 1) = \pi_j \quad (1.36)$$

と書き直すことができる.

ここで, 完全データ $(x_i, \mathbf{z}_i), i = 1, \dots, n$ に基づく対数尤度は,

$$LL_c = \sum_{i=1}^n \sum_{j=1}^p z_{ij} \log \pi_j + \sum_{i=1}^n \sum_{j=1}^p z_{ij} \log f_j(x_i; \boldsymbol{\theta}_j) \quad (1.37)$$

となる. E-step では対数尤度 (1.37) の条件付期待値を求めることになるが, z_{ij} の線形関数であるため, z_{ij} の条件付期待値だけを求めればよいことが分かる. ベイズの定理による事後確率の評価から, 具体的に, E-step の一般形は以下となる.

E-step

$$\begin{aligned} w_{ij}^{(k)} &\equiv E[z_{ij} | x_i; \pi_1^{(k)}, \dots, \pi_p^{(k)}, \boldsymbol{\theta}^{(k)}] \\ &= Pr(z_{ij} = 1 | x_i; \pi_1^{(k)}, \dots, \pi_p^{(k)}, \boldsymbol{\theta}^{(k)}) \\ &= \pi_j f_j(x_i; \boldsymbol{\theta}_j) / \left\{ \sum_{\ell=1}^p \pi_\ell f_\ell(x_i; \boldsymbol{\theta}_\ell) \right\} \end{aligned} \quad (1.38)$$

M-step では, (1.37) 式の z_{ij} を E-step で求めた条件付期待値 $w_{ij}^{(k)}$ に置き換え最大化を達成するパラメータの値を求めるが, 混合割合 π_j を示すパラメータと第 j コンポーネントの分布を規定するパラメータ $\boldsymbol{\theta}_j$ は, それぞれ (1.37) 式右辺の第 1 項と第 2 項に分かれているので, それぞれ別々に最大化を行えばよいことになる.

M-step

$$\pi_j^{(k+1)} = \sum_{i=1}^n w_{ij}^{(k)} / n \quad (1.39)$$

パラメータ $\boldsymbol{\theta}_j$ に関する最大化は, コンポーネントごとに, 分布 $f_j(X; \boldsymbol{\theta}_j)$ のパラメータ $\boldsymbol{\theta}_j$ の $w_{ij}^{(k)}$ を重みとした重み付き推定に帰着する.

CONTENTS

混合正規分布のパラメータの推定

$$x \sim \sum_{j=1}^p \pi_j N(\mu_j, \sigma_j^2) \quad (1.40)$$

に対して、 μ_j, σ_j^2 の最尤推定値を求めるための EM アルゴリズムの具体的計算は以下ようになる：

E-step

$$w_{ij}^{(k)} = \frac{\pi_j^{(k)} \phi(x_i; \mu_j^{(k)}, \sigma_j^{(k)2})}{\sum_{\ell=1}^p \pi_{\ell}^{(k)} \phi(x_i; \mu_{\ell}^{(k)}, \sigma_{\ell}^{(k)2})}, \quad \phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (1.41)$$

M-step

$$\mu_j^{(k+1)} = \frac{\sum_{i=1}^n w_{ij}^{(k)} x_i}{\sum_{i=1}^n w_{ji}^{(k)}}, \quad (1.42)$$

$$\sigma_j^{2(k+1)} = \frac{\sum_{i=1}^n w_{ij}^{(k)} (x_i - \mu_j^{(k)})^2}{\sum_{i=1}^n w_{ji}^{(k)}}. \quad (1.43)$$

E-step と M-step の反復により、目的とするパラメータの収束値を得る。

混合正規分布の場合、平均 μ を観測されたデータの中のある 1 つの値に固定し、分散 σ^2 を 0 に近づけると、尤度が発散するという問題の存在が知られているが、これは、分散を一定以上の値にしておけば、回避ができる。上記の EM アルゴリズムの場合、分散が 0 より大きいという制約が自然と導入されており、尤度が発散するという状況は避けられている。

表 1.6 は、2005 年のシーズン中に松坂 (当時西武ライオンズ) が投げたスライダーに分類された球の球速の度数分布である。図 1.4.5 は、2 つの正規分布の混合モデルをあてはめた結果を表している。最尤推定の結果、全体の球速の分布は、93% が平均が時速 128.75km、分散 3.31、残りの 7% は平均が時速 136.86km、分散 1.74 の 2 つの正規分布の混合に従うと考えられる。

CONTENTS

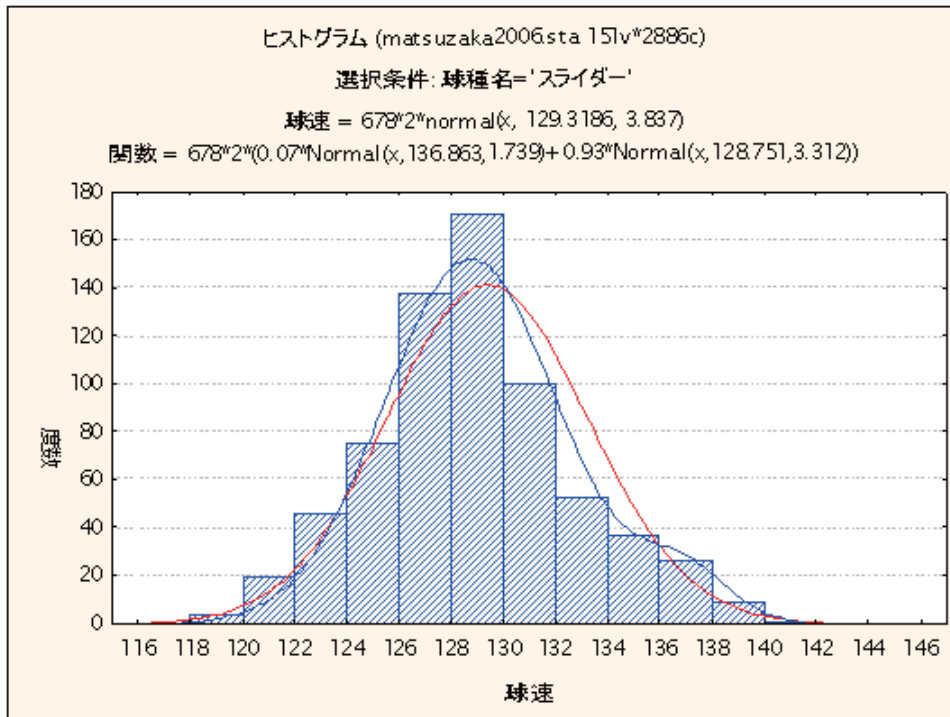


Figure 1.1: 混合正規分布のあてはめ

CONTENTS

Table 1.6: 松坂投手の球速分布 (2005年, 球種:スライダー)

球速 (km/h)	球数	球速 (km/h)	球数	球速 (km/h)	球数
119	2	126	55	133	34
120	1	127	95	134	25
121	6	128	101	135	30
122	8	129	89	136	12
123	15	130	71	137	22
124	42	131	71	138	14
125	64	132	51	139	6
合計					814

1.5 EM を利用した漸近分散共分散行列の評価

EM アルゴリズムでは Newton-Raphson 法と異なりアルゴリズム自身の副産物として推定値の漸近分散共分散行列の推定値を与えない。したがって、直接的には、不完全データの対数尤度 $LL(\boldsymbol{\theta} | \mathbf{y})$ から導かれる観測情報行列の逆行列を数値的に評価する方法で、漸近分散共分散行列を推定しなければならないが、これはモデルによっては導出が困難になる。

しかし、不完全データの対数尤度 $LL(\boldsymbol{\theta} | \mathbf{y})$ を直接意識しないという EM アルゴリズムのメリットを保持したもとの漸近分散共分散行列を近似する手法もいくつか提唱されている。

1.5.1 Louis の方法

Louis(1982) は、以下のような不完全データの観測情報行列 $\mathbf{I}(\hat{\boldsymbol{\theta}}; \mathbf{y})$ を完全データのフレームで評価する式：

$$\mathbf{I}(\hat{\boldsymbol{\theta}}; \mathbf{y}) = E[\mathbf{B}_c(\mathbf{x}; \boldsymbol{\theta}) | \mathbf{y}; \boldsymbol{\theta}]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} - E[\mathbf{S}_c(\mathbf{x} | \boldsymbol{\theta}) \mathbf{S}_c^T(\mathbf{x} | \boldsymbol{\theta}) | \mathbf{y}; \boldsymbol{\theta}]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (1.44)$$

CONTENTS

を示した. ここに, $S_c(x; \theta)$ は完全データ \mathbf{x} の対数尤度の1次導関数ベクトル, $B_c(x; \theta)$ は2次導関数行列の負値である.

(1.44) 式は, いわゆる Orchard and Woodbury (1972) による missing information principle に対応するもので,

Observed information = Complete information - Missing information
を表している. EM アルゴリズムにより $\hat{\theta}$ を得た後, (1.44) を評価すれば観測情報量を得ることができ, 漸近分散共分散行列はその逆行列の計算により導くことができる.

数値例として, Dempster, Laird and Rubin(1977) が取り上げている遺伝連鎖モデル (Genetic Linkage Model) を取り上げる. パラメータ θ を含むセル確率

$$\left(\frac{1}{2} + \frac{\theta}{4}, \frac{1}{4}(1 - \theta), \frac{1}{4}(1 - \theta), \frac{\theta}{4}\right)$$

をもつ4つのカテゴリに対して, 度数が $\mathbf{y} = (y_1, y_2, y_3, y_4)$ のように観測される多項モデルがある. これを不完全データと見做し, 実際は, セル確率

$$\left(\frac{1}{2}, \frac{\theta}{4}, \frac{1}{4}(1 - \theta), \frac{1}{4}(1 - \theta), \frac{\theta}{4}\right)$$

を有する5つのカテゴリがあるとして, これに対応する度数 $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$ を完全データとして想定する. ここで, $x_1 + x_2 = y_1, x_3 = y_2, x_4 = y_3, x_5 = y_4$ と考えると, θ の推定は容易に EM アルゴリズムのフレームで構築できる.

E-step: 完全データの対数尤度の条件付期待値 $E[LL_c(\theta) | \mathbf{y}, \theta^{(k)}]$ の導出

$$\begin{aligned} Q(\theta; \theta^{(k)}) &= E[LL_c(\theta) | \mathbf{y}, \theta^{(k)}] = E[(x_2 + x_5) \log(\theta) + (x_3 + x_4) \log(1 - \theta) | \mathbf{y}, \theta^{(k)}] \\ &= \{E(x_2 | y_1, \theta^{(k)}) + y_4\} \log(\theta) + (y_2 + y_3) \log(1 - \theta) \end{aligned} \quad (1.45)$$

M-step: E-step で求めた $Q(\theta; \theta^{(k)})$ を最大化する $\theta^{(k+1)}$ を求める

$$\theta^{(k+1)} = \frac{E[x_2 | y_1, \theta^{(k)}] + y_4}{E[x_2 | y_1, \theta^{(k)}] + y_2 + y_3 + y_4} \quad (1.46)$$

Rao(1965) による観測データ $\mathbf{y} = (y_1, y_2, y_3, y_4) = (125, 18, 20, 34)$ に対して, 上記の E-step と M-step の反復により収束値 0.6268 が最尤推定値 $\hat{\theta}$ として求められる (初期値 0.5).

CONTENTS

このとき, (1.44) 式の右辺の第 1 項と第 2 項はそれぞれ

$$\begin{aligned} E[\mathbf{B}_c(\mathbf{x}; \boldsymbol{\theta}) | \mathbf{y}; \boldsymbol{\theta}]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{E[x_2 | y_1, \hat{\theta}] + y_4}{\hat{\theta}^2} + \frac{y_2 + y_3}{(1 - \hat{\theta})^2} \\ &= \frac{(29.83 + 34)}{0.6268^2} + \frac{38}{(1 - 0.6268)^2} = 435.3, \end{aligned} \quad (1.47)$$

$$\begin{aligned} E[\mathbf{S}_c(\mathbf{x} | \boldsymbol{\theta}) \mathbf{S}_c^T(\mathbf{x} | \boldsymbol{\theta}) | \mathbf{y}; \boldsymbol{\theta}]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \text{Var}[\mathbf{S}_c(\mathbf{x} | \boldsymbol{\theta}) | \mathbf{y}, \boldsymbol{\theta}]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ &= \text{Var}\left[\left(\frac{E[x_2 | y_1, \boldsymbol{\theta}] + y_4}{\boldsymbol{\theta}} - \frac{y_2 + y_3}{1 - \boldsymbol{\theta}}\right) | \mathbf{y}, \boldsymbol{\theta}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ &= \frac{\text{Var}[E[x_2 | y_1, \hat{\theta}]]}{\hat{\theta}^2} \\ &= \frac{y_1 \frac{\hat{\theta}}{(\frac{1}{2} + \frac{\hat{\theta}}{4})^2}}{\hat{\theta}^2} = 57.8 \end{aligned} \quad (1.48)$$

となり, $\mathbf{I}(\hat{\boldsymbol{\theta}}; \mathbf{y}) = 435.3 - 57.8 = 377.5$ を得る. したがって, $\hat{\theta}$ の標準誤差は, $\sqrt{\frac{1}{377.5}} \approx 0.05$ となる.

1.5.2 Oakes の方法

Oakes(1999) は, 不完全データに基づく観測情報行列 $\mathbf{I}(\boldsymbol{\theta}; \mathbf{y})$ を $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})$ の 2 次導関数で表現する以下の式を与えた.

$$\mathbf{I}(\boldsymbol{\theta}^{(k)}; \mathbf{y}) = - \left[\left(\frac{\partial^2 Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} + \frac{\partial^2 Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{(k)T}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}} \right] \quad (1.49)$$

ここで右辺の第 2 項は, k 回目の $\boldsymbol{\theta}$ の推定値 $\boldsymbol{\theta}^{(k)}$ で評価した missing information に対応している.

先の遺伝連鎖モデルの例では,

$$\frac{\partial^2 Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{(k)T}} = \frac{2y_1}{\theta (2 + \theta^{(k)})^2} \quad (1.50)$$

かつ, (1.47) 式より,

$$\frac{\partial^2 Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \frac{y_1 \frac{\theta}{2+\theta} + y_4}{\theta^2} + \frac{y_2 + y_3}{(1 - \theta)^2} \quad (1.51)$$

CONTENTS

となり, これらを $\theta^{(k)}, \theta$ とも $\hat{\theta} = 0.6268$ で評価すると, 前節でも求めた $I(\hat{\theta}; \mathbf{y}) = 435.3 - 57.8 = 377.5$ を得る.

1.5.3 SEM(Supplemented EM)

Meng and Rubin(1991) は, 不完全データの漸近分散共分散行列 $I^{-1}(\hat{\theta}; \mathbf{y})$ が完全データの分散共分散行列に欠測データが存在することによる増分を加えることで評価できるとして次式を導いた.

$$I^{-1}(\hat{\theta}; \mathbf{y}) = I_c^{-1}(\hat{\theta}; \mathbf{y}) + \Delta V, \quad (1.52)$$

ここに, $I_c(\hat{\theta}; \mathbf{y})$ は完全データの情報行列の条件付き期待値 $E[B_c(\mathbf{x}; \theta) | \mathbf{y}; \theta]$ を表し,

$$\Delta V = [I - J(\hat{\theta})]^{-1} J(\hat{\theta})(\theta; \mathbf{y}) I_c^{-1}(\hat{\theta}; \mathbf{y}), \quad (1.53)$$

ここで, $J(\hat{\theta}) = I_c(\hat{\theta}; \mathbf{y}) I_m^{-1}(\hat{\theta}; \mathbf{y})$ である.

(1.52) 式は, $I_c(\hat{\theta}; \mathbf{y})$ とヤコビアン $J(\hat{\theta})$ を計算することで求められる値である. $I_c(\hat{\theta}; \mathbf{y})$ は先述のように, 完全データにに基づく尤度式から比較的容易に求めることができる. また, ヤコビアン行列 $J(\hat{\theta})$ はパラメータの成分毎の収束率を各要素としているので, EM アルゴリズム自身の出力コードを収束後, 再評価することでこれも容易に計算できる量である.

1.6 EM の特性

ここで, EM アルゴリズムの特徴や収束に関しこれまでに得られている結果をまとめておく.

1. 反復の各過程において $LL(\theta | y)$ の値は増加 (正確には, 非減少) する. これは実際にプログラミングを行う場合, デバッキングに有効である. また, 初期値に関して頑健である.

CONTENTS

2. Newton-Raphson 法が 2 次収束, 準 Newton 法が超 1 次収束するのに対して, EM アルゴリズムは 1 次収束であり, 収束率 γ は, 完全データの情報行列に対する不完全データの情報行列の比 $\tilde{I}_c^{-1}(\boldsymbol{\theta}; \mathbf{y})\tilde{I}_m(\boldsymbol{\theta}; \mathbf{y})$ の最大固有値 (完全データの情報量に占める不完全データの情報量の割合) となる. このとき, 収束のスピード (global speed of convergence) は, $1 - \gamma$ に比例する.
3. EM アルゴリズムでは, とくに完全データが指数分布族の範疇であれば, 各ステップにおける出力が統計的に意味のあるものになっている. すなわち, 自然にパラメータの定義域や制約条件が充たされ, かつ容易に制約条件を組み込んで解を得ることができる. いわゆる不適解が出ない. 例えば, 欠測値のある多次元データ行列から分散共分散行列を推定する場合も初期値を正定値行列から出発すれば各ステップで得られる $\hat{\Sigma}$ も正定値であることが保証される. また, 生起確率の推定では自然に確率の定義域に納まる解が得られる.
4. $LL(\boldsymbol{\theta} | y)$ が有界であれば, 反復の過程での対数尤度の値 $\{LL(\boldsymbol{\theta}^{(k)} | y)\}$ は $LL(\boldsymbol{\theta} | y)$ の定常点 (stationary value) に収束する.
5. かなり一般的な条件の下で, $\boldsymbol{\theta}^{(k)}$ が収束すればその収束値は $LL(\boldsymbol{\theta} | y)$ の局所的最大値 (local maximum) か鞍点 (saddle point) であることが証明できる (Boyle 1983, Wu 1983). したがって, 尤度関数が単峰で, Q 関数の 1 次導関数が $\boldsymbol{\theta}^{(k)}$ と $\boldsymbol{\theta}$ に関して連続であれば, EM は唯一の極大値 (最大値) に収束する. しかし, 一般に不完全データの尤度関数が単峰であるとは限らない. したがって, 多くの初期値を試みる必要がある.
6. EM の収束が極端に遅くなる場合は, その方向に尤度関数がフラットになっていることを示唆している.
7. EM アルゴリズムでは, 一般に計算量が大きい不完全データの対数尤度の 1 次および 2 次の導関数の評価が不要である. このことが多くの場

CONTENTS

合、他の最大化アルゴリズムと比較して各反復における CPU タイムの節約につながる。即ち、EM アルゴリズムの欠点である収束の遅さは、収束に要する反復回数を基準として見ているので、全体的な CPU-time を基準に採れば、各反復における高次元の逆行列演算の回避を考慮するとき、Newton-Raphson 法等との実際的な収束の速さに関する優劣は一概には定まらない。準 Newton 法も、不完全データの情報行列の計算を含まないので、EM アルゴリズム同様 1 回の反復における計算時間は、Newton-Raphson 法より短く、かつ、収束に要する反復回数は、超 1 次収束する準 Newton 法の方が 1 次収束する EM アルゴリズムよりはるかに少なく済む。従って問題によっては、反復の初期の段階は EM アルゴリズム、その後、準 Newton 法に切り替える等の使い方も有効である。

1.7 GEM アルゴリズムとその他の拡張型 EM

1.7.1 GEM(Generalized EM)

EM アルゴリズムの利点は、適用モデルの汎用性にあるが、対象とするモデルの構造が複雑になれば、E-step でモデルから発生させた乱数による評価が必要になったり、M-step の中に更に Newton 法などの反復アルゴリズムを含まざるを得なくなる場合も出てくる。その際、EM を拡張した GEM アルゴリズムやその他の EM の拡張型が有効になることがある。

GEM アルゴリズムとは、M-step を、次式をみたす $\theta^{(k+1)}$ を見つける作業に置き換えたアルゴリズムである：

$$Q(\theta^{(k+1)} | \theta^{(k)}) \geq Q(\theta^{(k)} | \theta^{(k)}) \quad (1.54)$$

これは、M-step において、必ずしも Q 関数の最大化を達成する必要がなく、より大きな値に更新する $\theta^{(k+1)}$ を見つければ良いことを示している。したがって、M-step で代数的な解が得られず Newton-Raphson 法など利用して数値的

CONTENTS

に最大化を図らなければならないような場合でも、1回の反復だけで止めることも可能となる。Lange(1995)では、a gradient algorithmとしてこの方法を提唱している。

1.7.2 ECM アルゴリズム

ECM アルゴリズムは、E-step と CM(Conditional Maximization)-step からなり、名前の通り、従来の M-step を条件付の最大化ステップ (CM-step) に置き換えたものである (Meng and Rubin 1993)。ここで、条件付の最大化ステップでは、最大化する際に動かすパラメータの次元は、従来の M-step に比べ小さくなる。このため、モデルに含まれるパラメータの次元が大きな場合、すべてのパラメータに対して最大化を同時に行う M-step に場合に比べ、CM-step の方が、単純で速く、かつ安定している場合が多い。

具体的な ECM アルゴリズムでは、従来の M-step を、計算が簡素化された複数の (S 個の) CM-step で置き換える。 $(k+1)$ 回目の反復における s 番目の CM-step では、 $c_s(\boldsymbol{\theta}) = c_s(\boldsymbol{\theta}^{(k+(s-1)/S)})$ の制約の下で $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})$ を最大化する $\boldsymbol{\theta}^{(k+(s-1)/S)}$ を求める。ここで、 $c_s(\boldsymbol{\theta}), s = 1, \dots, S$ はあらかじめ設定しておいた既知の関数である。

CM-step を進めるに従い、

$$\begin{aligned} Q(\boldsymbol{\theta}^{(k)}; \boldsymbol{\theta}^{(k)}) &\leq Q(\boldsymbol{\theta}^{(k+1/S)}; \boldsymbol{\theta}^{(k)}) \\ &\leq Q(\boldsymbol{\theta}^{(k+2/S)}; \boldsymbol{\theta}^{(k)}) \\ &\dots \\ &\leq Q(\boldsymbol{\theta}^{(k+(S-1)/S)}; \boldsymbol{\theta}^{(k)}) \\ &\leq Q(\boldsymbol{\theta}^{(k+1)}; \boldsymbol{\theta}^{(k)}) \end{aligned}$$

となるようなパラメータの更新が行われ、結果として一回の CM-step で必ず完全データの尤度の期待値が上昇 (厳密には非減少) する。すなわち、ECM アルゴリズムは GEM アルゴリズムになっている。

また、ECM アルゴリズムにおいて、各 CM-step または一部の CM-step の

CONTENTS

合間に E-step を挿入する場合がある。このアルゴリズムを Multicycle ECM アルゴリズムと呼んでいる。Multicycle ECM アルゴリズムでは、E-step を挿入するため、一回の反復に要する計算量は増大するが、一回の反復でより大きな尤度の改善が期待できる。Multicycle ECM アルゴリズムは、GEM アルゴリズムではないが、各反復における観測データ (不完全データ) の尤度の単調非減少性は成立する (Meng and Rubin(1993))。

1.7.3 ECME アルゴリズム – Contaminated Normal –

ECME アルゴリズムは ECM アルゴリズムの拡張として、Liu and Rubin(1994, 1995) により提唱された。ECME は Expectation-Conditional Maximization Either のことで、最後の "Either" は、CM-step において、場合によっては、不完全データに基づく尤度の最大化を行うことを意味している。

確率変数 \mathbf{y} が p 次の混淆正規分布 (contaminated normal)

$$(1 - \delta) \times N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + \delta \times N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\lambda) \quad (1.55)$$

に従うとする。このとき、 \mathbf{y} とは独立な正の、1 か $\lambda (> 0)$ の値をとる 2 値型の確率変数 q を想定し、 q の確率関数 $M(q)$ を

$$M(q) = \begin{cases} 1 - \delta & \text{if } q = 1, \\ \delta & \text{if } q = \lambda \quad (0 < \delta < 1, \lambda > 0), \\ 0 & \text{その他,} \end{cases} \quad (1.56)$$

とすると、 q が与えられた下での \mathbf{y} の条件付分布は、

$$\mathbf{y} | q \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/q) \quad (1.57)$$

となる。

ここで、各個体に対して、 (\mathbf{y}_i, q_i) , $i = 1, \dots, n$ を観測されるべき完全データと考え (実際は、 q_i はすべて欠測データとなる)、ECME アルゴリズムの枠組みで、(1.55) 式の混淆正規分布に含まれるパラメータ $\lambda, \delta, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ の最尤推定値を導く方法を与える (Yamaguchi(1990))。

CONTENTS

E-step:

$$\begin{aligned} w_i^{(k+1)} &= E(q_i | \mathbf{y}_i; \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}, \delta^{(k)}, \lambda^{(k)}) \\ &= \frac{1 - \delta^{(k)} + \delta^{(k)} \lambda^{(k)1+p/2} \exp\{(1 - \lambda^{(k)})d_i^{(k)2}/2\}}{1 - \delta^{(k)} + \delta^{(k)} \lambda^{(k)p/2} \exp\{(1 - \lambda^{(k)})d_i^{(k)2}/2\}} \end{aligned}$$

$$v_i^{(k+1)} = Pr(q_i = \lambda^{(k)} | \mathbf{y}_i; \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}, \delta^{(k)}) \quad (1.58)$$

$$= \frac{\delta^{(k)} \lambda^{(k)p/2} \exp\{(1 - \lambda^{(k)})d_i^{(k)2}/2\}}{1 - \delta^{(k)} + \delta^{(k)} \lambda^{(k)p/2} \exp\{(1 - \lambda^{(k)})d_i^{(k)2}/2\}} \quad (1.59)$$

ただし, $d_i^{(k)2} = (\mathbf{y}_i - \boldsymbol{\mu}^{(k)})' \boldsymbol{\Sigma}^{(k)-1} (\mathbf{y}_i - \boldsymbol{\mu}^{(k)})$.

CM1-step:

$$\boldsymbol{\mu}^{(k+1)} = \frac{\sum_{i=1}^n w_i^{(k+1)} \mathbf{y}_i}{\sum_{i=1}^n w_i^{(k+1)}}, \quad (1.60)$$

$$\boldsymbol{\Sigma}^{(k+1)} = \frac{\sum_{i=1}^n w_i^{(k+1)} (\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)}) (\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)})'}{n}, \quad (1.61)$$

$$\delta^{(k+1)} = \frac{\sum_{i=1}^n v_i^{(k+1)}}{n} \quad (1.62)$$

CM2-step: λ に関して, Q 関数ではなく, 不完全データ \mathbf{y} に基づく対数尤度 LL に関する最大化を行い, 値を更新する:

$$\begin{aligned} LL = & - \frac{n}{2} \log |\boldsymbol{\Sigma}^{(k+1)}| - \frac{1}{2} \sum_{i=1}^n d_i^{(k+1)2} \\ & + \sum_{i=1}^n \log [1 - \delta^{(k+1)} + \delta^{(k+1)} \lambda^{p/2} \exp\{(1 - \lambda)d_i^{(k+1)2}/2\}] \end{aligned} \quad (1.63)$$

具体的には, 以下の解を求めることになる.

$$\lambda = \frac{p \sum Pr(q_i = \lambda | \mathbf{y}_i; \boldsymbol{\mu}^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}, \delta^{(k+1)})}{\sum d_i^{(k+1)2} Pr(q_i = \lambda | \mathbf{y}_i; \boldsymbol{\mu}^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}, \delta^{(k+1)})} \quad (1.64)$$

ここで, $Pr(q_i = \lambda | \mathbf{y}_i; \boldsymbol{\mu}^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}, \delta^{(k+1)})$, また d_i^2 の値は, E-step の結果を使用する.

CONTENTS

ECME アルゴリズムと ECM アルゴリズムの双方が適用できる場合は、ECME アルゴリズムの方が、観測データ (不完全データ) に対する対数尤度 LL を直接的に最大化するので、必ずしもではないが、多くの場合、収束のスピードが速くなる。これらのアルゴリズムのモデルに応じた比較研究については、Liu and Rubin (1994) が参考になる。

E-step が閉じた式展開で単純に解けない場合の取り扱いに関しては、近似分布で置き換える (Laird, 1978) やラプラス展開を用いる (Steele, 1996) などの工夫がある。Tanner and Wong (1987) は EM のフレームワークの拡張的解釈に基づいて、欠測データがある場合の θ の事後密度の計算にモンテカルロ法や data augmentation 法を利用している。Wei and Tanner (1990) は、E-step でシュミレーションを行うモンテカルロ EM を提唱している。

1.7.4 加速化を意識した完全データの探索

－ optimal EM アルゴリズム－

EM アルゴリズムでは、パラメータの推定が容易になる程度までデータを拡張したものを完全データとするが、Meng and van Dyke (1997) は、加速化も意識した最適なデータ拡張 (Data Augmentation) という考え方を示した。具体的には、EM アルゴリズムの収束スピードは完全データの情報量に依存するので、完全データを変化させるワーキングパラメータ α を導入してその情報量を α の関数で表現し、それを最小化をする α によって最適な完全データを規定する。

加速化については、種々の研究がなされているが (Jamshidian and Jennrich, 1997), その多くは EM アルゴリズムが持つ単純性と安定性のある程度犠牲にするものである。しかし、Optimal EM アルゴリズムでは、本来の単純性と安定性は損なわれていない。

ここでは、自由度 ν を既知とした多変量 t 分布モデルを例に、Optimal EM アルゴリズムを示す。多変量 t 分布は前節の混淆正規分布同様、正規分布の尺度混合分布の一つで、EM アルゴリズムを適用することで比較的簡単に最

CONTENTS

尤推定値を求めることができる。

いま p 次の確率ベクトル \mathbf{y} と \mathbf{y} と独立な正の確率変数 q があり, q が与えられた下での \mathbf{y} の条件付分布は, $\mathbf{y} | q \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/q)$ であつ, $q \sim \chi_\nu^2/\nu$ とすると, \mathbf{y} の周辺分布は自由度 ν の多変量 t 分布となる. 各個体に対して, (\mathbf{y}_i, q_i) , $i = 1, \dots, n$ を完全データ, q_i はすべて観測されない欠測データとすると, E-step と M-step は次のようになる;

E-step:

$$w_i^{(k)} = \frac{\nu + p}{\nu + d_i^{2(k)}}, \quad d_i^{2(k)} = (\mathbf{y}_i - \boldsymbol{\mu}^{(k)})^T \boldsymbol{\Sigma}^{(k)-1} (\mathbf{y}_i - \boldsymbol{\mu}^{(k)}),$$

M-step:

$$\boldsymbol{\mu}^{(k+1)} = \sum_{i=1}^n w_i^{(k)} \mathbf{y}_i / \sum_{i=1}^n w_i^{(k)}, \quad (1.65)$$

$$\boldsymbol{\Sigma}^{(k+1)} = \sum_{i=1}^n w_i^{(k)} (\mathbf{y}_i - \boldsymbol{\mu}^{(k)}) (\mathbf{y}_i - \boldsymbol{\mu}^{(k)})^T / n. \quad (1.66)$$

optimal-EM アルゴリズムでは, モデル $\mathbf{y}|q \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/q)$ を $\mathbf{y} = \boldsymbol{\mu} + \mathbf{z}/q^{1/2}$, ただし $\mathbf{z} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ と表現し直し, 更に α (working parameter) を導入して

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\mu} + |\boldsymbol{\Sigma}|^{-\alpha/2} \mathbf{z} / (|\boldsymbol{\Sigma}|^{-\alpha} q)^{1/2} \\ &= \boldsymbol{\mu} + |\boldsymbol{\Sigma}|^{-\alpha/2} \mathbf{z} / q(\alpha)^{1/2}, \end{aligned}$$

$q(\alpha) = |\boldsymbol{\Sigma}|^{-\alpha} q$ と一般化する. $\alpha = 0$ のとき, $q(0) = q$ でもとのモデルとなるが, ここで, α の関数として $I_{obs} I_{aug}^{-1}(\alpha)$ の最小固有値 $\gamma = \gamma(\alpha)$ を評価すると, $\alpha = 1/(\nu + p) \equiv \alpha_{opt}$ のとき, γ が最大となることがわかる. つまり, $q(\alpha_{opt})$ でデータを完全化 (拡大) すると, 収束のスピードが最適化されることになる.

$q(\alpha_{opt})$ を使って EM アルゴリズムを再構築すると, M-step の (1.66) 式が次のようになる;

$$\boldsymbol{\Sigma}^{(k+1)} = \sum_{i=1}^n w_i^{(k)} (\mathbf{y}_i - \boldsymbol{\mu}^{(k)}) (\mathbf{y}_i - \boldsymbol{\mu}^{(k)})^T / \sum_{i=1}^n w_i^{(k)}. \quad (1.67)$$

CONTENTS

これは、(1.66) 式の分母の n を重みの和でおきかえただけの修正である。

Meng and van Dyke(1977)の論文は、EMアルゴリズムの提唱からちょうど20年が経過したことを記念する総括的な論文でもある。他にも1990年代以降、拡張的なEMを系統化する論文や成書が刊行されている。Rubin(1991)は、シミュレーションをベースにした4つの代表的なアルゴリズム: Multiple Imputation, Data Augmentation Algorithm, Gibbs Sampler, Sampling/Importance Resampling Algorithm を乱数機構をもつ拡張型EMのフレームワークの下に解説している。McLachlan and Krishnan(1997)は、拡張型のEMまでひろくかつ具体的に解説したEMアルゴリズムの成書である。雑誌 *Statistica Sinica* では1995年に、続く1997年には、*Statistical Methods in Medical Research* で、EMアルゴリズムの特集が組まれている。Meng(1997)やAmari(1996)もそれぞれ、医学系や情報系の雑誌で組まれたEMアルゴリズム招待論文である。Tanner(1996), Schafer(1997), Sorensen and Gianola(2002), Watanabe and Yamaguchi(2004)も、EMアルゴリズムおよびMCMCを含めその他のアルゴリズムが解説され参考になる。

Bibliography

- [1] Amari, S.(1995). Information geometry of the EM and em algorithms for neural networks, *Neural Networks* **8**, 1379-1408.
- [2] Boyles, R.A.(1983). On the convergence of the EM algorithm, *Journal of The Royal Statistical Society* **B 45**, 47-50.
- [3] Dempster, A.P., Laird, N.M. and Rubin, D.B.(1977). Maximum likelihood from incomplete data via the EM algorithm(with Discussion), *Journal of The Royal Statistical Society* **B 39**, 1-38.
- [4] Flury, B. and Zoppe, A.(2000). Excercises in EM, *The American Statisticians*, **54**, 207-209.
- [5] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.
- [6] Jamshidian, M. and Jennrich (1997). Acceleration of the EM algorithm by using Quasi-Newton methods, *Journal of the Royal Statistical Society* **B 59**, 569-587.
- [7] Laird, N.W. (1978). Emprical Bayes methods for two-way tables, *Biometrika* **69**, 581-590.

BIBLIOGRAPHY

- [8] Lange, K. (1995). A gradient algorithm locally equivalent to the EM algorithm, *Journal of The Royal Statistical Society B* **57**, 425-437.
- [9] Liu, C. and Rubin, D.B. (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence, *Biometrika* **81**, 633-648.
- [10] Liu, C. and Rubin, D.B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME, *Statistica Sinica* **5**, 19-39.
- [11] Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society B* **44**, 226-233.
- [12] McLachlan, G.J. and Krishnan T. (1997). *The EM algorithm and extensions* Wiley-Interscience.
- [13] Meng, X.L. (1997). The EM algorithm and medical studies: a historical link. *Statistical Methods in Medical Research* **6**, 3-23.
- [14] Meng, X.L. and van Dyke, D.B. (1997). The EM algorithm-an old folk-song sung to a fast new tune, *Journal of the Royal Statistical Society B* **59**, 511-567.
- [15] Meng, X.L. and Rubin, D.B. (1991). Using EM to obtain asymptotic variance-covariance matrices - the SEM algorithm, *Journal of the American Statistical Association* **86**,899-909.
- [16] Meng, X.L. and Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework, *Biometrika* **80**, 267-278.
- [17] Oakes, D.(1999). Direct calculation of the information matrix via the EM algorithm, *J. R. Statist. Soc. B*, **61**, Part2, 479-482.

BIBLIOGRAPHY

- [18] Orchard, T. and Woodbury, M.A. (1972). A missing information principle: theory and applications. *Proc.of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, **Vol.1**, 697-715.
- [19] Rao.C.R. (1965). *Linear Statistical Inference and its Applications*,New York: Wiley.
- [20] Rubin, D.B. (1976). Inference with missing data. *Biometrika*,**63** 581-592.
- [21] Rubin, D.B. (1987a). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley
- [22] Rubin, D.B. (1987b). The SIR algorithm. *Journal of the American Statistical Association*, **82** 543-546.
- [23] Rubin, D.B. (1991). EM and beyond. *Psychometrika* **56**, 241-254.
- [24] Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data* Chapman and Hall/CRC.
- [25] Sorensen, D. and Gianola, D. (2002). *Likelihood, Bayesian, and MCMC Methods in Quantative Genetics* Springer.
- [26] Steel, B.M. (1996). A modified EM algorithm for estimation in generalized mixed models. *Biometrika* **52**, 1295-1310.
- [27] Tanner, M.A. (1996). *Tools for Statistical Inference*, Springer.
- [28] Tanner, M.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation *Journal of the American Statistical Association*, **82**, 528-550.
- [29] Yamaguchi, K.(1990). Generalized EM algorithm for models with contaminated normal error terms, *Statistical Methods and Data Analysis*,(Niki,N. ed),Tokyo: Scientist Inc.

BIBLIOGRAPHY

- [30] Watanabe, M. and Yamaguchi, K.(2004). *The EM Algorithm and Related Statistical Models*,, Marcel Dekker, New York.
- [31] Wei, G.C.and Tanner, M.A. (1990). A Monte carlo implementation of the EM algorithm and the poorman's data augmentation algorithms. *Journal of the American Statistical Association* **85**, 699-704.
- [32] Wu,C.F.J.(1983). On the convergence properties of the EM algorithm. *Annals of Statistics*, **11**, 95-103.

第10章 マルコフ連鎖モンテカルロ法入門

古澄英男¹

マルコフ連鎖モンテカルロ法は、ベイズ統計学を中心に幅広い分野で利用されている計算手法である。この方法は、様々な期待値を計算する場合によく用いられ、マルコフ連鎖と呼ばれる確率過程の収束に関する性質を利用することによって所与の確率分布から確率変数を発生させていることに特徴がある。最近では、マルコフ連鎖モンテカルロ法に関する理論的研究や拡張が活発に行われている。本章では、モンテカルロ積分とマルコフ連鎖について概観した後、マルコフ連鎖モンテカルロ法について解説する。

¹神戸大学大学院経営学研究科

目次

1	はじめに	1
1.1	モンテカルロ積分	1
1.2	モンテカルロ法からマルコフ連鎖モンテカルロ法へ	3
2	マルコフ連鎖	4
2.1	マルコフ連鎖と推移行列	4
2.2	マルコフ連鎖の性質	6
2.3	詳細釣り合い条件	8
3	メトロポリス-ヘイスティングス法	9
3.1	メトロポリス-ヘイスティングスアルゴリズム	9
3.2	MHアルゴリズムの収束	12
3.3	MHアルゴリズムの組み合わせ	12
4	ギブス・サンプリング	14
4.1	ギブス・サンプリングアルゴリズム	14
4.2	多重ブロックMHアルゴリズムとギブス・サンプリング	15
4.3	データ拡大法	17
5	実際の利用について	18
5.1	収束の判定	18
5.2	効率性	19
5.3	混合の改善	20
6	応用例	21
6.1	ロジットモデルのベイズ推定	22
6.2	隠れマルコフモデルのベイズ推定	24

マルコフ連鎖モンテカルロ法入門

1 はじめに

近年の計算機のめざましい発達に伴い、統計科学における計算手法の役割はますます重要となってきた。本稿の目的は、もともと統計物理の分野において提案され、その後、ベイズ統計学、計量経済学、生物統計、ゲノム解析、空間統計など統計科学の幅広い分野で応用されているマルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo method: 以下 MCMC 法) について説明を行うことである。MCMC 法がよく用いられるのは、平均値や分散など様々な期待値 (あるいは積分) の計算が必要とされる場合である。そこでまず、期待値の基本的な計算手法の 1 つであるモンテカルロ法について振り返ることとする。

1.1 モンテカルロ積分

確率変数 \mathbf{x} の確率分布を $\pi(\mathbf{x})$ と表し¹、 \mathbf{x} の関数 $h(\mathbf{x})$ の期待値

$$(1) \quad I = E_{\pi} [h(\mathbf{x})] = \int_{\mathcal{X}} h(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}$$

を考えることにする。ここで、 $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$ であり、 $E_{\pi}[\cdot]$ は $\pi(\mathbf{x})$ に関する期待値を表す。もし、 $\pi(\mathbf{x})$ が正規分布やガンマ分布などのよく知られた確率分布であれば、計算機上で直接 $\pi(\mathbf{x})$ からサンプリングを行うことが可能である²。いま、 $\pi(\mathbf{x})$ からの独立なサンプルを $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})$ と表せば、(1) 式で与えられる期待値は、

$$(2) \quad \hat{I} = \frac{1}{n} \sum_{i=1}^n h(\mathbf{x}^{(i)})$$

¹以下では、記号 π を確率密度関数及び確率分布の両方に用いる。

²本稿では、与えられた確率分布から乱数を生成させることをサンプリング (sampling) と呼ぶことにする。また、得られた乱数をサンプルと呼ぶことにする。確率分布から乱数を発生させる一般的な方法としては、逆関数法や棄却法などがある。これらについては、Devroye (1986), Ripley (1987), Gentle (2003) などに詳しい。

によって推定することができるであろう。このとき、大数の法則によって $n \rightarrow \infty$ のとき \hat{I} は I に収束する。つまり、 n が十分大きいときには、 I を \hat{I} によって近似できることになる。このようにして期待値を求める方法を **モンテカルロ積分** (Monte Carlo integration) と呼ぶ。

統計科学が扱う多くの問題では、対象とする確率分布 $\pi(\mathbf{x})$ が複雑であるため、 $\pi(\mathbf{x})$ から直接サンプリングするのが困難である場合がある。そこで、別の確率分布 $q(\mathbf{x})$ を考え、そこからサンプリングすることを考える³。このとき、(1) 式を書き直すと、

$$I = \int_{\mathcal{X}} h(\mathbf{x}) \frac{\pi(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) d\mathbf{x} = E_q \left[h(\mathbf{x}) \frac{\pi(\mathbf{x})}{q(\mathbf{x})} \right]$$

となる。よって、 $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})$ を $q(\mathbf{x})$ からの独立なサンプルとすれば、 I の推定量として

$$(3) \quad \hat{I}_{IS} = \frac{1}{n} \sum_{i=1}^n h(\mathbf{x}^{(i)}) w(\mathbf{x}^{(i)})$$

が得られる。ここで、 $w(\mathbf{x}^{(i)}) = \pi(\mathbf{x}^{(i)})/q(\mathbf{x}^{(i)})$ であり、 $\mathbf{x}^{(i)}$ に対する重みと見なすことができる。(3) 式のように、重みが付いたサンプルを用いて期待値を求める方法のことを、**重点サンプリング** (importance sampling) という。

重点サンプリングにおいても、 $q(\mathbf{x})$ が一定の条件を満たしていれば、 $n \rightarrow \infty$ のときに \hat{I}_{IS} が I に収束することが分かっている。また、重点サンプリングの精度は $q(\mathbf{x})$ の選択に依存しており、 $q(\mathbf{x}) \propto |h(\mathbf{x})| \pi(\mathbf{x})$ であるときに精度が最も良くなることが知られている (Rubinstein (1981), Geweke (1989), Evans and Swartz (2000) を参照)⁴。

ベイズ分析などでは、 $\pi(\mathbf{x})$ の正規化定数が未知であることが多い。そこで、 $\pi(\mathbf{x}) = p(\mathbf{x}) / \int_{\mathcal{X}} p(\mathbf{x}) d\mathbf{x} \propto p(\mathbf{x})$ と表し⁵、(1) 式を

$$I = \int_{\mathcal{X}} h(\mathbf{x}) \frac{p(\mathbf{x})}{\int_{\mathcal{X}} p(\mathbf{x}) d\mathbf{x}} d\mathbf{x} = \frac{\int_{\mathcal{X}} h(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}}{\int_{\mathcal{X}} p(\mathbf{x}) d\mathbf{x}}$$

³ここで、 $q(\mathbf{x})$ の台 (support) は $\pi(\mathbf{x})$ の台を含んでいる必要がある。

⁴記号 \propto は比例を表す。

⁵確率分布は積分すると必ず 1 にならないといけないので、任意の非負の関数 $p(\mathbf{x})$ が与えられたとき、対応する確率分布は $p(\mathbf{x}) / \int_{\mathcal{X}} p(\mathbf{x}) d\mathbf{x}$ と表すことができる。ここで、 $\int_{\mathcal{X}} p(\mathbf{x}) d\mathbf{x}$ は \mathbf{x} に依存しない定数であり正規化定数と呼ぶ。

と書き直す。さらに、 $q(\mathbf{x})$ からサンプリングすることにすれば、

$$(4) \quad I = \frac{\int_{\mathcal{X}} h(\mathbf{x}) \frac{p(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) d\mathbf{x}}{\int_{\mathcal{X}} \frac{p(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) d\mathbf{x}} = \frac{E_q \left[h(\mathbf{x}) \frac{p(\mathbf{x})}{q(\mathbf{x})} \right]}{E_q \left[\frac{p(\mathbf{x})}{q(\mathbf{x})} \right]}$$

と表すことができる。 (4) 式の分母と分子に対して重点サンプリングを適用すれば、

$$\tilde{I}_{IS} = \frac{\frac{1}{n} \sum_{i=1}^n h(\mathbf{x}^{(i)}) w(\mathbf{x}^{(i)})}{\frac{1}{n} \sum_{i=1}^n w(\mathbf{x}^{(i)})} = \sum_{i=1}^n h(\mathbf{x}^{(i)}) w^*(\mathbf{x}^{(i)})$$

が得られる。ここで、 $w(\mathbf{x}^{(i)}) = p(\mathbf{x}^{(i)})/q(\mathbf{x}^{(i)})$ 、 $w^*(\mathbf{x}^{(i)}) = w(\mathbf{x}^{(i)})/\sum_{j=1}^n w(\mathbf{x}^{(j)})$ である。また、 $\sum_{i=1}^n w^*(\mathbf{x}^{(i)}) = 1$ であることから、 $w^*(\mathbf{x}^{(i)})$ は基準化された重みとなっている。

\hat{I} や \hat{I}_{IS} は I の不偏推定量であるが、 \tilde{I}_{IS} は不偏ではない。しかし、 $n \rightarrow \infty$ のときには \tilde{I}_{IS} も I に収束することが分かっている (Geweke (1989))。また、問題によっては \hat{I}_{IS} よりも \tilde{I}_{IS} の方が精度がよくなることがある (Robert and Casella (2004))。

1.2 モンテカルロ法からマルコフ連鎖モンテカルロ法へ

重点サンプリングは、統計科学における重要な計算手法の1つである。しかし、 \mathbf{x} が高次元であるときや $\pi(\mathbf{x})$ が複雑な確率分布であるときには、サンプリングが容易でしかも $\pi(\mathbf{x})$ をうまく近似する $q(\mathbf{x})$ を選ぶことは非常に困難である。このような理由により、重点サンプリングの利用は \mathbf{x} の次元が低い場合などに限られ、複雑な統計モデルへの適用には限界があった。そのため、重点サンプリングに代わる計算手法が望まれていた。

モンテカルロ積分や重点サンプリングが、独立なサンプルを利用していたのに対し、1990年頃からサンプル間の相関を許したモンテカルロ法が統計科学の分野において利用されるようになってきた。この方法が、本稿の主題の MCMC 法である。MCMC 法には数多くの方法があり、その中でも MH アルゴリズムとギブス・サンプリングと呼ばれる方法が実際の統計解析におい

てよく用いられている。そこで本稿では、この2つの方法を中心にMCMC法について説明を行うことにする⁶。

本稿の構成は以下の通りである。次節においてマルコフ連鎖について簡単に振り返り、第3節でMHアルゴリズムについて説明する。第4節ではギブス・サンプリングについて説明し、第5節では実際に利用するときの注意点を述べる。最後に第6節でMCMC法の応用例を示す。

2 マルコフ連鎖

MCMC法は、その名が示す通り「マルコフ連鎖」と「モンテカルロ法」を組み合わせた方法である。具体的には、マルコフ連鎖 (Markov chain) と呼ばれる確率過程の性質を利用して確率分布からのサンプリングを行い、様々な計算を行う方法である。この節では、マルコフ連鎖とその性質について簡単に説明する⁷。

2.1 マルコフ連鎖と推移行列

時点 t における確率変数を $\mathbf{x}^{(t)}$ と表し、確率過程 $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots)$ を考えることにする。また、 $\mathbf{x}^{(t)}$ ($t = 0, 1, \dots$) の取りうる値の集合を記号 \mathcal{X} によって表し、これを状態空間 (state space) と呼ぶ。以下では説明を簡単にするために、 $\mathbf{x}^{(t)}$ は1次元の確率変数とし、その状態空間は $\mathcal{X} = \{1, \dots, k\}$ であるとする。

確率過程 $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots)$ が、すべての $i, j \in \mathcal{X}$ と $t \geq 0$ に対して、

$$\begin{aligned} & \Pr(\mathbf{x}^{(t+1)} = j | \mathbf{x}^{(0)} = i_0, \mathbf{x}^{(1)} = i_1, \dots, \mathbf{x}^{(t-1)} = i_{t-1}, \mathbf{x}^{(t)} = i) \\ (5) \quad & = \Pr(\mathbf{x}^{(t+1)} = j | \mathbf{x}^{(t)} = i) \end{aligned}$$

を満たすとき、 $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots)$ はマルコフ連鎖であるという。ここで、 $i_k \in \mathcal{X}$ は時点 k の状態を表す。一般の確率過程では、 $\mathbf{x}^{(t+1)}$ の従う確率分布は、時

⁶MCMC法を解説したテキストとして、Liu (2001), Robert and Casella (2004), Gamerman and Lopes (2006) を挙げておく。日本語による解説としては、大森 (2001), 中妻 (2003), 伊庭 (2005), 和合 (2005) などがある。

⁷マルコフ連鎖については、Karlin and Taylo (1975) や Ross (1995) などに詳しい。

点 t までの履歴 $\{\mathbf{x}^{(0)} = i_0, \mathbf{x}^{(1)} = i_1, \dots, \mathbf{x}^{(t-1)} = i_{t-1}, \mathbf{x}^{(t)} = i\}$ に依存して決まると考えられる。しかし、マルコフ連鎖は、 $t+1$ 期における条件付確率分布が時点 t よりも前の履歴には依存しない確率過程となっており、この性質はマルコフ性 (Markov property) と呼ばれる。

(5) 式の条件付確率を $p(i, j) = \Pr(\mathbf{x}^{(t+1)} = j | \mathbf{x}^{(t)} = i)$ と表し、これを**推移確率** (transition probability) と呼ぶ。また、 $p(i, j)$ を第 (i, j) 要素とする $k \times k$ 行列を

$$\mathbf{T} = \begin{pmatrix} p(1,1) & p(1,2) & \cdots & p(1,k) \\ p(2,1) & p(2,2) & \cdots & p(2,k) \\ \vdots & \vdots & \ddots & \vdots \\ p(k,1) & p(k,2) & \cdots & p(k,k) \end{pmatrix}$$

と表す。この行列は**推移行列** (transition matrix) と呼ばれ、マルコフ連鎖がどのように推移していくかを決定する。推移確率 $p(i, j)$ は、現在の状態が i であるとき、次の期に状態 j へ移る確率を表しているので、推移行列の各行については、 $p(i, j) \geq 0$ と $\sum_{j=1}^k p(i, j) = 1$ が成立している。

例 2.1. 状態空間を $\mathcal{X} = \{1, 2, 3\}$ とし、推移行列

$$T = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 3/4 & 0 & 1/4 \\ 0 & 1 & 0 \end{pmatrix}$$

を持つマルコフ連鎖を考える。もし、現在の状態が 1 であるときには、次の期には確率 $1/2$ で現在の状態にとどまり、確率 $1/3$ で状態 2 に移り、確率 $1/6$ で状態 3 に推移する。現在の状態が 2 であれば、状態 1 もしくは状態 3 に移動する。また、状態 3 からは必ず状態 2 に推移する。□

初期状態 $\mathbf{x}^{(0)}$ の確率分布を行ベクトル $\boldsymbol{\pi}^{(0)}$ を用いて、

$$\boldsymbol{\pi}^{(0)} = (\pi_1^{(0)}, \pi_2^{(0)}, \dots, \pi_k^{(0)}) = (\Pr(\mathbf{x}^{(0)} = 1), \Pr(\mathbf{x}^{(0)} = 2), \dots, \Pr(\mathbf{x}^{(0)} = k))$$

と表すことにする。同様に、行ベクトル $\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}, \dots$ によって、 $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$ の確率分布を表す。つまり、

$$\boldsymbol{\pi}^{(t)} = (\pi_1^{(t)}, \pi_2^{(t)}, \dots, \pi_k^{(t)}) = (\Pr(\mathbf{x}^{(t)} = 1), \Pr(\mathbf{x}^{(t)} = 2), \dots, \Pr(\mathbf{x}^{(t)} = k))$$

である。ここで、 $\mathbf{x}^{(1)}$ の確率分布 $\boldsymbol{\pi}^{(1)}$ について考えると、

$$\begin{aligned}\pi_j^{(1)} &= \Pr(\mathbf{x}^{(1)} = j) = \sum_{i=1}^k \Pr(\mathbf{x}^{(0)} = i, \mathbf{x}^{(1)} = j) \\ &= \sum_{i=1}^k \Pr(\mathbf{x}^{(0)} = i) \Pr(\mathbf{x}^{(1)} = j | \mathbf{x}^{(0)} = i) = \sum_{i=1}^k \pi_i^{(0)} p(i, j)\end{aligned}$$

であることから、 $\boldsymbol{\pi}^{(1)} = \boldsymbol{\pi}^{(0)} \mathbf{T}$ を得る。また、 $\mathbf{x}^{(2)}$ についても、 $\boldsymbol{\pi}^{(2)} = \boldsymbol{\pi}^{(1)} \mathbf{T} = \boldsymbol{\pi}^{(0)} \mathbf{T}^2$ となる。これを逐次繰り返していけば、 $\boldsymbol{\pi}^{(t)} = \boldsymbol{\pi}^{(0)} \mathbf{T}^t$ が導かれる。

2.2 マルコフ連鎖の性質

先に得た関係より、マルコフ連鎖の確率的振る舞いは、初期分布 $\boldsymbol{\pi}^{(0)}$ と推移行列 \mathbf{T} によって完全に決定されることが分かる。このことから、初期分布と推移行列が与えられれば、マルコフ連鎖に従う確率変数は次のようにして発生させることができる：

- (1) 初期状態 $\mathbf{x}^{(0)}$ を $\boldsymbol{\pi}^{(0)}$ からサンプリングする。
- (2) $t = 0, 1, \dots$ に対して、 $\mathbf{x}^{(t+1)}$ を $(\mathbf{T})_{\mathbf{x}^{(t)}}$ からサンプリングする。

ここで、 $(\mathbf{T})_{\mathbf{x}^{(t)}}$ は推移行列 \mathbf{T} の第 $\mathbf{x}^{(t)}$ 行からなる確率分布を表す。また、このようにして得られる $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots)$ では、明らかに $\mathbf{x}^{(t)} \sim \boldsymbol{\pi}^{(t)}$ となっている。

MCMC 法との関連を考えたとき、 $\boldsymbol{\pi}^{(t)}$ は収束するのか、収束するとすればその条件は何であるか、収束先はどのような確率分布であるのか、ということが問題となる。この問いに対する鍵となるのが、マルコフ連鎖の (1) 既約性 (irreducibility), (2) 非周期性 (aperiodicity), (3) 不変分布 (invariant distribution) である。

マルコフ連鎖の既約性とは、連鎖がどのような状態から出発したとしても、有限回のステップで別の状態にたどり着くことができる性質のことである。より厳密には、推移行列が \mathbf{T} であるマルコフ連鎖を考えたとき、すべての $i, j \in \mathcal{X}$ に対して、 $(\mathbf{T}^n)_{ij} > 0$ を満たす有限の n が存在すれば、マルコフ連鎖は既約であるという。ここで、 $(\mathbf{T}^n)_{ij}$ は \mathbf{T}^n の第 (i, j) 要素を表す。

例 2.2. 例 2.1 のマルコフ連鎖は既約であるので，ここでは既約でないマルコフ連鎖を示そう．例えば，推移行列が

$$\mathbf{T} = \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.3 & 0.7 & 0 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}$$

であるマルコフ連鎖を考えてみる．状態 3 からは，すべての状態に推移することができるが，状態 1 あるいは状態 2 からは状態 3 にたどり着くことができない．よって，この推移行列を持つマルコフ連鎖は既約ではない． □

次に，状態 $i \in \mathcal{X}$ に対して， $\{n \geq 1 : (\mathbf{T}^n)_{ii} > 0\}$ で定義される集合を考える．これは，元の状態に戻るのに必要なステップ数の集合を表している．この集合の最大公約数を状態 i の周期 (period) といい，すべての状態の周期が 1 であるとき，マルコフ連鎖は非周期的であるという．文字通り，非周期性は一定の時間間隔で訪れる状態がないことを意味している．

例 2.3. マルコフ連鎖の推移行列が

$$\mathbf{T} = \begin{pmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

であるとする．このマルコフ連鎖では， $\{n \geq 1 : (\mathbf{T}^n)_{ii} > 0\} = \{2, 4, 6, \dots\}$ ($i = 1, \dots, 4$) である．よって，連鎖の周期は 2 なので非周期的ではない． □

最後にマルコフ連鎖の不変分布についてである．推移行列が \mathbf{T} であるマルコフ連鎖に対して，行ベクトル $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$ が

$$(1) \pi_i \geq 0 \ (i \in \mathcal{X}), \sum_{i=1}^k \pi_i = 1, \quad (2) \boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{T}$$

を満たすとき， $\boldsymbol{\pi}$ は \mathbf{T} の不変分布であるという．

例 2.4. 例 2.1 において， $\boldsymbol{\pi} = (1/2, 1/3, 1/6)$ を考えると， $\sum_{i=1}^3 \pi_i = 1$ であり， $\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{T}$ を満たす．よって，この $\boldsymbol{\pi}$ は \mathbf{T} の不変分布である． □

マルコフ連鎖が既約性と非周期性を満たしているときには，不変分布が一意に存在する．またこのとき， $\boldsymbol{\pi}^{(t)}$ が不変分布 $\boldsymbol{\pi}$ に収束することも証明できる (Häggström (2002) を参照)．この結果を定理としてまとめておく．

定理 (マルコフ連鎖の収束). マルコフ連鎖 $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots)$ は既約で非周期であると、その推移行列を \mathbf{T} とする。また、 π を \mathbf{T} の不変分布とする。このとき、任意の初期分布 $\pi^{(0)}$ に対して、 $t \rightarrow \infty$ であるとき $\frac{1}{2} \sum_{i=1}^k |\pi_i^{(t)} - \pi_i| \rightarrow 0$ となる⁸。

2.3 詳細釣り合い条件

これまでの議論から、マルコフ連鎖の収束を利用して確率分布 π からサンプリングを行うためには、既約性と非周期性を満たし、 π を不変分布とするような推移行列を作ればよいことになる。そして、先の方法によって $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots)$ をサンプリングし十分大きい m をとれば、 $(\mathbf{x}^{(m+1)}, \mathbf{x}^{(m+2)}, \dots)$ は π からのサンプルとなる。

実際には、既約性や非周期性を満たす推移行列を作成することはそれほど難しいことではない。むしろ問題となるのは、所与の π を不変分布とする推移行列をどのように設計すればよいかということである。このとき重要な役割を果たすのが、**詳細釣り合い条件** (detailed balance condition) と呼ばれるもので、

$$(6) \quad \pi_i p(i, j) = \pi_j p(j, i) \quad (i, j \in \mathcal{X})$$

によって与えられる。この条件が満たされているとき、マルコフ連鎖は**可逆** (reversible) であるとも言う。詳細釣り合い条件は、 π が不変分布となるための十分条件となっている⁹。MCMC 法の多くのアルゴリズムでは、この詳細釣り合い条件を満たすように推移行列が設計されている。

これまで状態空間が離散であるマルコフ連鎖について説明してきた。状態空間が連続の場合でも、数学的に面倒な点はあるが、これまでの結果は基本

⁸任意の初期分布に対して収束するということは、初期状態を適当な確率分布から選んでも、あるいは固定した値を選んでもよいことを意味する。

⁹詳細釣り合い条件が十分条件であることは、(6) 式の両辺を i に関して和をとれば、 $\sum_{i=1}^k \pi_i p(i, j) = \sum_{i=1}^k \pi_j p(j, i) = \pi_j$ となることから確認できる。

的に成立する¹⁰. 連続な状態空間では, 推移行列の代わりに

$$\Pr(\mathbf{x}^{(t+1)} \in A | \mathbf{x}^{(t)} = \mathbf{x}) = \int_A T(\mathbf{x}, \mathbf{y}) d\mathbf{y} \quad (\mathbf{x} \in \mathcal{X}, A \subset \mathcal{X})$$

を満たす条件付確率分布 $T(\mathbf{x}, \mathbf{y})$ を考える. この $T(\mathbf{x}, \mathbf{y})$ のことを**推移核** (transition kernel) と呼ぶ. また, $T(\mathbf{x}, \mathbf{y})$ の不変分布は,

$$\pi(\mathbf{y}) = \int_{\mathcal{X}} \pi(\mathbf{x}) T(\mathbf{x}, \mathbf{y}) d\mathbf{x}$$

を満たす確率分布 $\pi(\mathbf{x})$ として定義される. さらに, 詳細釣り合い条件は,

$$\pi(\mathbf{x}) T(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y}) T(\mathbf{y}, \mathbf{x})$$

によって与えられることになる.

3 メトロポリス–ヘイスティングス法

MCMC 法の中で最も代表的な方法が, メトロポリス–ヘイスティングス (Metropolis–Hastings : MH) アルゴリズムである. これは, 半世紀以上も前に Metropolis *et al.* (1953) により提案され, その後 Hastings (1970) によって一般化されたアルゴリズムである¹¹. 本節では, この MH アルゴリズムについて説明を行う.

3.1 メトロポリス–ヘイスティングスアルゴリズム

いま, 確率分布 $\pi(\mathbf{x})$ からサンプリングを行いたいとしよう. この $\pi(\mathbf{x})$ のことを**目標分布** (target distribution) と呼ぶ. MH アルゴリズムでは, **提案分布** (proposal distribution) と呼ばれる確率分布を利用して目標分布からのサンプリングを行う¹². 提案分布は, 現在の状態を所与としたときの条件付確率分布であり, 以下では $q(\mathbf{y}|\mathbf{x})$ と表すことにする.

¹⁰一般的な状態空間におけるマルコフ連鎖については, Nummelin (1984) や Meyn and Tweedie (1993) に詳しい.

¹¹Dongarra and Sullivan (2000) では 20 世紀におけるトップ 10 アルゴリズムが示されており, その 1 つに MH アルゴリズムが挙げられている.

¹²提案分布を候補発生分布 (candidate generating distribution) と呼ぶこともある.

MH アルゴリズムの特徴は、提案分布から次の状態の候補となる値を発生させ、それを採択確率 (acceptance probability) と呼ばれる値に従って採択するか棄却するかを決める点にある。具体的には、以下のようにして $\pi(\mathbf{x})$ からのサンプリングを行う。

MH アルゴリズム

- (1) 初期値 $\mathbf{x}^{(0)}$ を決める。
- (2) $t = 0, 1, \dots$ に対して次を繰り返す。
 - (i) \mathbf{y} を $q(\mathbf{y}|\mathbf{x}^{(t)})$ から発生させる。
 - (ii) u を一様分布 $\mathcal{U}(0, 1)$ から発生させ¹³,

$$\mathbf{x}^{(t+1)} = \begin{cases} \mathbf{y} & u \leq \alpha(\mathbf{x}^{(t)}, \mathbf{y}) \text{ の場合} \\ \mathbf{x}^{(t)} & \text{その他の場合} \end{cases}$$

とする。ただし、 $\alpha(\mathbf{x}, \mathbf{y}) = \min \left\{ 1, \frac{\pi(\mathbf{y})q(\mathbf{x}|\mathbf{y})}{\pi(\mathbf{x})q(\mathbf{y}|\mathbf{x})} \right\}$ である。

ここで、 $\alpha(\mathbf{x}, \mathbf{y})$ が採択確率である。採択確率は目標分布や提案分布の比にしか依存していないので、正規化定数が分からない場合でも MH アルゴリズムが適用できる。また、候補の値が棄却されれば $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)}$ とすることから、MH アルゴリズムでは同じ値が連続する場合がある。

MH アルゴリズムを実行するには、候補の値を発生させる提案分布 $q(\mathbf{y}|\mathbf{x})$ を決定しなければならない。提案分布の選択は、MH アルゴリズムの効率性や収束の速度に大きく影響を与えるので、十分注意する必要がある。

例 3.1. 目標分布 $\pi(\mathbf{x})$ が連続分布であるとする。そこで、現在の状態が $\mathbf{x}^{(t)} = \mathbf{x}$ であるとき、候補の値を

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

にしたがって発生させるとする¹⁴。この方法を酔歩連鎖 (random walk chain) と呼ぶ。このとき $q(\mathbf{y}|\mathbf{x}) = q(\mathbf{x}|\mathbf{y})$ が成立するので、採択確率は $\alpha(\mathbf{x}, \mathbf{y}) =$

¹³ $\mathcal{U}(a, b)$ は、区間 (a, b) における一様分布を表す。

¹⁴ $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ は、平均が $\boldsymbol{\mu}$ 、分散行列が $\boldsymbol{\Sigma}$ である正規分布を表す。

$\min \{1, \pi(\mathbf{y})/\pi(\mathbf{x})\}$ と簡略化される. Metropolis *et al.* (1953) によって提案された方法は, $q(\mathbf{y}|\mathbf{x}) = q(\mathbf{x}|\mathbf{y})$ であるときの MH アルゴリズムである.

酔歩連鎖では, ϵ の分布として正規分布以外に, 一様分布 $U(-\sigma, \sigma)$ や多変量 t 分布 $\mathcal{T}(\mathbf{0}, \sigma^2 \mathbf{I})$ などもよく用いられる¹⁵. また, σ は現在の状態からの変化分を決定するパラメータであり, **ステップサイズ** (step size) と呼ばれる. ステップサイズが小さいと採択確率は 1 に近くなるが, 現在の状態が少ししか変化せず, 状態空間全体を推移するのに時間がかかる. 逆にステップサイズが大ききときには, 現在の状態は大きく変化するが採択確率が低下してしまう¹⁶. □

例 3.2. 酔歩連鎖では, 目標分布の情報を全く使わずに候補となる値を発生させている. そこで, 候補となる値を

$$\mathbf{y} = \mathbf{x} + \frac{\sigma^2}{2} \frac{\partial \log \pi(\mathbf{x})}{\partial \mathbf{x}} + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

にしたがって発生させることを考える. この方法を, **ランジェヴァン連鎖** (Langevin chain) と呼ぶ (Roberts and Rosenthal (1998), Christensen *et al.* (2001), Christensen and Waagepetersen (2002) を参照). 現在の状態 \mathbf{x} が目標分布のモード付近にあるときには $\partial \log \pi(\mathbf{x})/\partial \mathbf{x} \approx 0$ であり, ランジェヴァン連鎖は酔歩連鎖とほぼ同じになる. しかし, \mathbf{x} が目標分布のモードから離れているときには $\partial \log \pi(\mathbf{x})/\partial \mathbf{x} \neq 0$ であるため, ランジェヴァン連鎖では酔歩連鎖よりもモードに近い領域に候補の値を発生させることができる. □

例 3.3. 候補となる \mathbf{y} を, 現在の状態 \mathbf{x} に依存することなく発生させるとする. つまり, $q(\mathbf{y}|\mathbf{x}) = q(\mathbf{y})$ とするのが**独立連鎖** (independent chain) である. この場合, 採択確率は $\alpha(\mathbf{x}, \mathbf{y}) = \min \left\{ 1, \frac{\pi(\mathbf{y})/q(\mathbf{y})}{\pi(\mathbf{x})/q(\mathbf{x})} \right\}$ となる. ここで, $\pi(\mathbf{x})/q(\mathbf{x})$ は, 重点サンプリングの重みに対応しており, \mathbf{y} の重みが \mathbf{x} のそれよりも大きければ必ず \mathbf{y} が選択され, そうでなければ重みの比に応じて \mathbf{y} が選択される. 提案分布 $q(\mathbf{y})$ としては, 採択確率が高くなるように $\pi(\mathbf{x})$ を

¹⁵ $\mathcal{T}_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ は, 自由度が ν , 平均が $\boldsymbol{\mu}$, 尺度行列が $\boldsymbol{\Sigma}$ である多変量 t 分布を表す. また, 1次元の t 分布のときには, 記号 $t_\nu(\mu, \sigma^2)$ を用いる.

¹⁶最適なステップサイズの選択については, Roberts *et al.* (1997) や Roberts and Rosenthal (2001) らが議論している.

近似する確率分布を用いる必要がある。実際の問題では、 $\pi(\mathbf{x})$ のモードを平均に、 $\{-\partial \log \pi(\mathbf{x}) / \partial \mathbf{x} \partial \mathbf{x}'\}^{-1}$ を分散行列とする多変量 t 分布などがよく用いられる（例えば、Chib and Greenberg (1995) を参照）。□

3.2 MH アルゴリズムの収束

MH アルゴリズムによってサンプリングされる $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots)$ が、マルコフ連鎖を形成していることは明らかであろう。また、MH アルゴリズムでは、 \mathbf{x} から \mathbf{y} へ推移するときの推移核は、

$$(7) \quad T(\mathbf{x}, \mathbf{y}) = q(\mathbf{y}|\mathbf{x})\alpha(\mathbf{x}, \mathbf{y}) + r(\mathbf{x})\delta_{\mathbf{x}}(\mathbf{y})$$

によって与えられる。ここで、右辺の第 2 項は候補の値が棄却される場合に対応しており、 $r(\mathbf{x}) = \int_{\mathcal{X}} q(\mathbf{y}|\mathbf{x}) \{1 - \alpha(\mathbf{x}, \mathbf{y})\} d\mathbf{y}$ 、 $\delta_{\mathbf{x}}(\mathbf{y}) = I(\mathbf{x} = \mathbf{y})$ である¹⁷。また、(7) 式の各項については、 $q(\mathbf{y}|\mathbf{x})\alpha(\mathbf{x}, \mathbf{y})\pi(\mathbf{x}) = q(\mathbf{x}|\mathbf{y})\alpha(\mathbf{y}, \mathbf{x})\pi(\mathbf{y})$ 、 $r(\mathbf{x})\delta_{\mathbf{x}}(\mathbf{y})\pi(\mathbf{x}) = r(\mathbf{y})\delta_{\mathbf{y}}(\mathbf{x})\pi(\mathbf{y})$ が成立する。したがって、

$$\begin{aligned} \pi(\mathbf{x})T(\mathbf{x}, \mathbf{y}) &= q(\mathbf{y}|\mathbf{x})\alpha(\mathbf{x}, \mathbf{y})\pi(\mathbf{x}) + r(\mathbf{x})\delta_{\mathbf{x}}(\mathbf{y})\pi(\mathbf{x}) \\ &= q(\mathbf{x}|\mathbf{y})\alpha(\mathbf{y}, \mathbf{x})\pi(\mathbf{y}) + r(\mathbf{y})\delta_{\mathbf{y}}(\mathbf{x})\pi(\mathbf{y}) = \pi(\mathbf{y})T(\mathbf{y}, \mathbf{x}) \end{aligned}$$

を得る。つまり、MH アルゴリズムによって生成されるマルコフ連鎖は詳細釣り合い条件を満たしており、不変分布として $\pi(\mathbf{x})$ を持つことになる。また、Roberts and Smith (1994) や Tierney (1994) が示しているように、MH アルゴリズムではほとんどの場合で既約性と非周期性が満たされている。したがって、MH アルゴリズムによって $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots)$ を発生させれば、十分大きい m 以降のサンプル $(\mathbf{x}^{(m+1)}, \mathbf{x}^{(m+2)}, \dots)$ は $\pi(\mathbf{x})$ からのサンプルと見なすことができる。このマルコフ連鎖が不変分布に収束するまでの時間 m のことを、**稼働検査期間** (burn-in period) と呼ぶ。

3.3 MH アルゴリズムの組み合わせ

MH アルゴリズムは簡潔なアルゴリズムであり、その実装は比較的容易である。また、MH アルゴリズムは、拡張性の高いアルゴリズムとなっている。

¹⁷ $I(\cdot)$ は指標関数を表す。

例えば、 $\pi(\mathbf{x})$ からのサンプリングを行うのに、2つのMHアルゴリズムがあるとしよう。ここでは、各アルゴリズムの推移核を $T_1(\mathbf{x}, \mathbf{y})$ 、 $T_2(\mathbf{x}, \mathbf{y})$ と表すことにする。そして、確率 w で推移核が $T_1(\mathbf{x}, \mathbf{y})$ であるMHアルゴリズムを使い、確率 $1-w$ で推移核が $T_2(\mathbf{x}, \mathbf{y})$ であるMHアルゴリズムを使ってサンプリングを行うとする。これは、例えば確率 w で酔歩連鎖を利用し、確率 $1-w$ で独立連鎖を利用するような場合に相当する。

この方法によって生成されるマルコフ連鎖の推移核は、

$$(8) \quad T(\mathbf{x}, \mathbf{y}) = wT_1(\mathbf{x}, \mathbf{y}) + (1-w)T_2(\mathbf{x}, \mathbf{y})$$

と表すことができる。この推移核のことを、**混合型推移核** (mixture of transition kernels) と呼ぶ。容易に確認されるように、 $T_i(\mathbf{x}, \mathbf{y})$ ($i = 1, 2$) の不変分布が $\pi(\mathbf{x})$ であれば、(8)式の推移核 $T(\mathbf{x}, \mathbf{y})$ の不変分布も $\pi(\mathbf{x})$ となる。したがって、このようにMHアルゴリズムを組み合わせても、そこから生成されるマルコフ連鎖は目標分布に収束することになる。混合型推移核を用いる利点は、性質の異なるMHアルゴリズムを組み合わせることによって、より効率的なサンプリング方法を簡単に構築できる所にある。

MHアルゴリズムの組み合わせ方には、混合型の他に循環型と呼ばれる方法もある。この方法では、最初に $T_1(\mathbf{x}, \mathbf{x}')$ を推移核とするMHアルゴリズムによって状態を \mathbf{x} から \mathbf{x}' に推移させる。そして次に、推移核が $T_2(\mathbf{x}', \mathbf{y})$ であるMHアルゴリズムを用いて \mathbf{x}' から \mathbf{y} に推移させる。このとき、 \mathbf{x} から \mathbf{y} へ移動するときの推移核は、

$$(9) \quad T(\mathbf{x}, \mathbf{y}) = \int_{\mathcal{X}} T_1(\mathbf{x}, \mathbf{x}')T_2(\mathbf{x}', \mathbf{y})d\mathbf{x}'$$

となる。(9)式の推移核のことを、**循環型推移核** (cycle of transition kernels) という。循環型推移核についても、各推移核の不変分布が $\pi(\mathbf{x})$ であれば、

$$\begin{aligned} \int_{\mathcal{X}} \pi(\mathbf{x})T(\mathbf{x}, \mathbf{y})d\mathbf{x} &= \int_{\mathcal{X}} \int_{\mathcal{X}} \pi(\mathbf{x})T_1(\mathbf{x}, \mathbf{x}')T_2(\mathbf{x}', \mathbf{y})d\mathbf{x}d\mathbf{x}' \\ &= \int_{\mathcal{X}} \pi(\mathbf{x}')T_2(\mathbf{x}', \mathbf{y})d\mathbf{x}' = \pi(\mathbf{y}) \end{aligned}$$

となり、 $\pi(\mathbf{x})$ が $T(\mathbf{x}, \mathbf{y})$ の不変分布となる¹⁸。循環型推移核を利用した重要

¹⁸混合型と循環型の組み合わせは、2つ以上のMHアルゴリズムに対しても適用できる。このときの収束条件の詳細については、Tierney (1994) を参照されたい。

な MCMC 法が、次節で説明するギブス・サンプリングである。

4 ギブス・サンプリング

MH アルゴリズムと並んでよく用いられるのが、ギブス・サンプリング (Gibbs sampling) と呼ばれる方法である。この方法は、Geman and Geman (1984) による画像復元のためのアルゴリズムとして知られていたが、Gelfand and Smith (1990) がベイズ推定における事後分布からのサンプリング法として用いたことによって、統計科学の分野でも急速に広まっていった。

4.1 ギブス・サンプリングアルゴリズム

ギブス・サンプリングは、 \mathbf{x} を k 個のブロック $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ に分割し、各 \mathbf{x}_i をその条件付確率分布 $\pi(\mathbf{x}_i | \mathbf{x}_{-i})$ からサンプリングする方法である¹⁹。ここで、 $\mathbf{x}_{-i} = (\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_k)$ であり、 $\pi(\mathbf{x}_i | \mathbf{x}_{-i})$ のことを完全条件付分布 (full conditional distribution) と呼ぶ。

ギブス・サンプリングアルゴリズム

- (1) 初期値 $\mathbf{x}^{(0)} = (\mathbf{x}_1^{(0)}, \dots, \mathbf{x}_k^{(0)})$ を決める。
- (2) $t = 0, 1, \dots$ に対して次を繰り返す。
 - (i) $\mathbf{x}_1^{(t+1)}$ を $\pi(\mathbf{x}_1 | \mathbf{x}_2^{(t)}, \dots, \mathbf{x}_k^{(t)})$ からサンプリングする。
 - (ii) $\mathbf{x}_2^{(t+1)}$ を $\pi(\mathbf{x}_2 | \mathbf{x}_1^{(t+1)}, \mathbf{x}_3^{(t)}, \dots, \mathbf{x}_k^{(t)})$ からサンプリングする。
 - ⋮
 - (k) $\mathbf{x}_k^{(t+1)}$ を $\pi(\mathbf{x}_k | \mathbf{x}_1^{(t+1)}, \dots, \mathbf{x}_{k-1}^{(t+1)})$ からサンプリングする。

ギブス・サンプリングでは、MH アルゴリズムのように提案分布を選択する必要がない。そのため、すべての $\pi(\mathbf{x}_i | \mathbf{x}_{-i})$ からサンプリングできるのであれば、実装するのは非常に容易である。また、ここで示したアルゴリズムでは、決められた順番で \mathbf{x}_i をサンプリングしているが、ランダムな順番で \mathbf{x}_i のサンプリングを行っても構わない (Liu *et al.* (1995))。

¹⁹ブロックの選び方については、第 5 節で議論している。実際の問題においては、分析する統計モデルによってブロックが自然と決まることが多い。

例 4.1. 目標分布が $\pi(x_1, x_2) \propto {}_n C_{x_1} x_2^{x_1+1} (1-x_2)^{n-x_1+1}$ であるとする。ただし, $x_1 \in \{0, 1, \dots, n\}$, $x_2 \in [0, 1]$ である。このとき, x_1 と x_2 の完全条件付分布はそれぞれ,

$$\pi(x_1|x_2) \propto {}_n C_{x_1} x_2^{x_1+1} (1-x_2)^{n-x_1+1}, \quad \pi(x_2|x_1) \propto x_2^{x_1+1} (1-x_2)^{n-x_1+1}$$

である。したがって, ギブス・サンプリングでは, $x_1|x_2 \sim \text{Bi}(n, x_2)$, $x_2|x_1 \sim \text{Be}(x_1 + \alpha, n - x_1 + \beta)$ のサンプリングを繰り返せばよい²⁰. \square

4.2 多重ブロック MH アルゴリズムとギブス・サンプリング

ギブス・サンプリングによって生成されるマルコフ連鎖は, k 回のステップを経て次の状態へと推移する。ここでは, 説明を簡単にするために $k = 2$ として, ギブス・サンプリングと MH アルゴリズムとの関係を見ることにする。そのために, 多重ブロック MH アルゴリズム (multiple-block MH algorithm) と呼ばれる方法について説明する。

いま, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ から $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$ への推移を, $(\mathbf{x}_1, \mathbf{x}_2) \rightarrow (\mathbf{y}_1, \mathbf{x}_2) \rightarrow (\mathbf{y}_1, \mathbf{y}_2)$ の順番で行うとする。多重ブロック MH アルゴリズムでは, $(\mathbf{x}_1, \mathbf{x}_2)$ から $(\mathbf{y}_1, \mathbf{x}_2)$ の推移を, 不変分布が $\pi(\mathbf{x}_1|\mathbf{x}_2)$ である MH アルゴリズムによって行う。これは, 候補の値を提案分布 $q_1(\mathbf{y}_1|\mathbf{x}_1, \mathbf{x}_2)$ から発生させ, 採択確率を

$$\alpha_1(\mathbf{x}_1, \mathbf{y}_1) = \min \left\{ 1, \frac{\pi(\mathbf{y}_1|\mathbf{x}_2)q_1(\mathbf{x}_1|\mathbf{y}_1, \mathbf{x}_2)}{\pi(\mathbf{x}_1|\mathbf{x}_2)q_1(\mathbf{y}_1|\mathbf{x}_1, \mathbf{x}_2)} \right\}$$

とする MH アルゴリズムを考えればよい。このときの推移核を $T_1(\mathbf{x}_1, \mathbf{y}_1|\mathbf{x}_2)$ と表すことにする。同様に, $(\mathbf{y}_1, \mathbf{x}_2)$ から $(\mathbf{y}_1, \mathbf{y}_2)$ への推移も, 不変分布が $\pi(\mathbf{x}_2|\mathbf{x}_1)$ となるように, 提案分布が $q_2(\mathbf{y}_2|\mathbf{y}_1, \mathbf{x}_2)$, 採択確率が

$$\alpha_2(\mathbf{x}_2, \mathbf{y}_2) = \min \left\{ 1, \frac{\pi(\mathbf{y}_2|\mathbf{y}_1)q_2(\mathbf{x}_2|\mathbf{y}_1, \mathbf{y}_2)}{\pi(\mathbf{x}_2|\mathbf{y}_1)q_2(\mathbf{y}_2|\mathbf{y}_1, \mathbf{x}_2)} \right\}$$

である MH アルゴリズムによって行い, このときの推移核を $T_2(\mathbf{x}_2, \mathbf{y}_2|\mathbf{y}_1)$ と表す。

²⁰ $\text{Bi}(n, p)$ は二項分布を表し, その確率関数は $\pi(x) \propto p^x (1-p)^{n-x}$ である。また $\text{Be}(a, b)$ はベータ分布を表し, 確率密度関数は $\pi(x) \propto x^{a-1} (1-x)^{b-1}$ である。

多重ブロック MH アルゴリズムでは、 \mathbf{x} から \mathbf{y} への推移核が、

$$T(\mathbf{x}, \mathbf{y}) = T_1(\mathbf{x}_1, \mathbf{y}_1 | \mathbf{x}_2) T_2(\mathbf{x}_2, \mathbf{y}_2 | \mathbf{y}_1)$$

で与えられ、循環型推移核の特殊な場合と見なすことができる。さらに、 $T_1(\mathbf{x}_1, \mathbf{y}_1 | \mathbf{x}_2)$ の不変分布が $\pi(\mathbf{x}_1 | \mathbf{x}_2)$ であること用いれば、

$$\begin{aligned} \int_{\mathcal{X}} \pi(\mathbf{x}) T(\mathbf{x}, \mathbf{y}) d\mathbf{x} &= \iint \pi(\mathbf{x}_1, \mathbf{x}_2) T_1(\mathbf{x}_1, \mathbf{y}_1 | \mathbf{x}_2) T_2(\mathbf{x}_2, \mathbf{y}_2 | \mathbf{y}_1) d\mathbf{x}_1 d\mathbf{x}_2 \\ &= \int \left[\int \pi(\mathbf{x}_1 | \mathbf{x}_2) T_1(\mathbf{x}_1, \mathbf{y}_1 | \mathbf{x}_2) d\mathbf{x}_1 \right] \pi(\mathbf{x}_2) T_2(\mathbf{x}_2, \mathbf{y}_2 | \mathbf{y}_1) d\mathbf{x}_2 \\ &= \int \pi(\mathbf{y}_1 | \mathbf{x}_2) \pi(\mathbf{x}_2) T_2(\mathbf{x}_2, \mathbf{y}_2 | \mathbf{y}_1) d\mathbf{x}_2 \end{aligned}$$

を得る。ここで、積分内の最初の確率分布は、ベイズの定理より $\pi(\mathbf{y}_1 | \mathbf{x}_2) = \pi(\mathbf{y}_1) \pi(\mathbf{x}_2 | \mathbf{y}_1) / \pi(\mathbf{x}_2)$ と書き直すことができるので、

$$\begin{aligned} \int_{\mathcal{X}} \pi(\mathbf{x}) T(\mathbf{x}, \mathbf{y}) d\mathbf{x} &= \int \pi(\mathbf{y}_1) \pi(\mathbf{x}_2 | \mathbf{y}_1) T_2(\mathbf{x}_2, \mathbf{y}_2 | \mathbf{y}_1) d\mathbf{x}_2 \\ &= \pi(\mathbf{y}_1) \pi(\mathbf{y}_2 | \mathbf{y}_1) = \pi(\mathbf{y}) \end{aligned}$$

が成立する。したがって、多重ブロック MH アルゴリズムによって生成されるマルコフ連鎖は不変分布として $\pi(\mathbf{x})$ を持ち、既約性や非周期性などの条件が満たされれば目標分布に収束することになる。多重ブロック MH アルゴリズムを用いると、高次元のサンプリングを低次元のサンプリングに分解することができ非常に便利である。

ギブス・サンプリングとの関係を調べるために、多重ブロック MH アルゴリズムにおいて、提案分布が完全条件付分布に一致している場合を考える。例えば、 $q_1(\mathbf{y}_1 | \mathbf{x}_1, \mathbf{x}_2) = \pi(\mathbf{y}_1 | \mathbf{x}_2)$ であるとすると、対応する採択確率は

$$\alpha_1(\mathbf{x}_1, \mathbf{y}_1) = \min \left\{ 1, \frac{\pi(\mathbf{y}_1 | \mathbf{x}_2) \pi(\mathbf{x}_1 | \mathbf{x}_2)}{\pi(\mathbf{x}_1 | \mathbf{x}_2) \pi(\mathbf{y}_1 | \mathbf{x}_2)} \right\} = 1$$

となる。同様に、 $q_2(\mathbf{y}_2 | \mathbf{y}_1, \mathbf{x}_2) = \pi(\mathbf{y}_2 | \mathbf{y}_1)$ とすれば $\alpha_2(\mathbf{x}_2, \mathbf{y}_2) = 1$ となる。つまり、ギブス・サンプリングは、すべての採択確率が 1 である多重ブロック MH アルゴリズムの特別な場合と見なすことができる²¹。これより、ギブス・

²¹ギブス・サンプリングでは $\pi(\mathbf{x})$ を不変分布として持つが、詳細釣り合い条件は必ずしも満たされないことに注意する必要がある。

サンプリングもある条件の下では目標分布 $\pi(\mathbf{x})$ に収束することになる²².

多重ブロック MH アルゴリズムに関する結果から、ギブス・サンプリングにおいて一部の完全条件付分布からのサンプリングが行えないときには、それを適切な MH アルゴリズムで置き換えてもよいことも分かる。ギブス・サンプリングと MH アルゴリズムとを組み合わせさせた方法は、Metropolis within Gibbs アルゴリズムとして知られている (Müller (1991)) .

4.3 データ拡大法

目標分布 $\pi(\mathbf{x})$ から直接サンプリングするよりも、別の確率変数 $\mathbf{z} \in \mathcal{Z}$ を導入して、 (\mathbf{x}, \mathbf{z}) の同時確率分布 $\pi(\mathbf{x}, \mathbf{z})$ からサンプリングを行った方が容易になることがよくある。このとき、同時確率分布が $\int_{\mathcal{Z}} \pi(\mathbf{x}, \mathbf{z}) d\mathbf{z} = \pi(\mathbf{x})$ を満たしていれば、MCMC 法によってサンプリングされた \mathbf{x} は $\pi(\mathbf{x})$ からのサンプルとなる。このように、新たな変数を導入してサンプリングを行う方法のことを、**データ拡大法** (data augmentation method) と呼ぶ (Tanner and Wong (1987)) . データ拡大法は、プロビットモデルや打ち切りのある回帰モデルを推定するときによく用いられる手法である (Chib (1992), Albert and Chib (1993)) . また、不完全データを分析する場合にも有効な方法である。

データ拡大法において、 $\pi(\mathbf{x}|\mathbf{z})$ と $\pi(\mathbf{z}|\mathbf{x})$ の完全条件付分布から構成されるギブス・サンプリングを考えることにする。このとき、 \mathbf{x} は統計モデルのパラメータ、 \mathbf{z} は欠損データ (missing data) を表しているとするれば、 $\pi(\mathbf{z}|\mathbf{x})$ からのサンプリングは、Rubin (1987) の代入法 (imputation) に対応している。さらに、このときのギブス・サンプリングは、第 xx 章で説明されている EM アルゴリズム (Dempster *et al.* (1970)) と関連していることが知られており、 $\pi(\mathbf{z}|\mathbf{x})$ からのサンプリングが E ステップに、 $\pi(\mathbf{x}|\mathbf{z})$ からのサンプリングが M ステップに対応する (Liu (2002) を参照) .

いま、 x と z はスカラーであるとし、目標分布 $\pi(x)$ が確率分布 $p(x)$ と非負関数 $l(x)$ によって、 $\pi(x) \propto p(x)l(x)$ と表すことができるとしよう。また、

²²ギブス・サンプリングの収束に関しては、Chan (1993) や Roberts and Smith (1994) を参照。

z は非負の確率変数であるとし、 (x, z) の同時確率分布

$$(10) \quad \pi(x, z) \propto I[z < l(x)]p(x)$$

を考えることにする。このとき、 x の周辺確率分布が目標分布に一致していることは容易に確認できる。そこで、この同時確率分布に対してギブス・サンプリングを行うのがスライス・サンプリング (slice sampling) と呼ばれる方法である。(10) 式より、 $\pi(z|x)$ は一様分布 $\mathcal{U}(0, l(x))$ なので、 z のサンプリングは簡単である。一方、 $\pi(x|z)$ は $\{x : l(x) > z\}$ 上に制約された確率分布 $p(x)$ であるため、サンプリングが困難であるように思われる。しかし、Damien *et al.* (1999) では、スライス・サンプリングが適用できる確率分布の例が多く示されている。また、スライス・サンプリングの応用や拡張については、Besag and Green (1993), Higdon (1998), Damien and Walker (2001), Neal (2003) などを参照して欲しい。

5 実際の利用について

5.1 収束の判定

MCMC 法では、マルコフ連鎖が目標分布に収束するまである程度時間を要する。そのため、サンプリングした $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots)$ の最初の部分は捨て、残りのサンプルを利用して期待値計算などを行う。収束するまでの時間は扱う問題によって様々であり、連鎖が収束しているかどうかをその都度判断しなければならない。

収束を判定する最も基本的な方法は、MCMC 法によって得られたサンプルの時系列プロットを調べることである。つまり、時系列プロットを作成して、サンプルの変動が安定的になる時点を分析者の目で判断する。また、得られたサンプルから統計量を計算し、客観的に収束を判定する方法も数多く提案されている (Cowles and Carlin (1996) や Mengersen *et al.* (1999) を参照)。その中でよく用いられているのは、Heidelberger and Welch (1983), Gelman and Rubin (1992), Geweke (1992), Raftery and Lewis (1992) ら

による方法であろう²³。しかし、これらの収束判定法も問題によってはうまく機能しないこともあるので、実際に収束を判定するときには、複数の方法を組み合わせて総合的に判定するのがよい。

ところで、MCMC 法が使われ始めた頃は、異なる初期値から複数のサンプル列を発生させる多重連鎖 (multiple chain) と、1つの初期値から1つのサンプル列を発生させる単一連鎖 (single chain) のどちらを使うべきかという議論があった。多重連鎖では独立なサンプルを利用できるが、計算時間などの問題から比較的短いサンプル列を発生させることが多く、目標分布に収束していない恐れがある。それに対して、単一連鎖では独立なサンプルを得ることはできないが、十分長いサンプル列を発生させることができ、多くの場合で目標分布からのサンプルを得ることができる。現在では、連鎖が収束していることを重視し、多重連鎖よりも単一連鎖の方がよく用いられている。

5.2 効率性

連鎖が収束していると判定されれば、MCMC 法によって得られたサンプル $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m+n)})$ を用いて、目標分布に関する期待値 $E_\pi[h(\mathbf{x})]$ を

$$(11) \quad \hat{I} = \frac{1}{n} \sum_{i=1}^n h(\mathbf{x}^{(m+i)})$$

によって推定することができる。ここで、 m は収束するまでの期間を表しており、 $n \rightarrow \infty$ のとき \hat{I} は $E_\pi[h(\mathbf{x})]$ に収束することが知られている (Tierney (1994))。また、 \hat{I} の分散は、

$$\text{Var}(\hat{I}) = \frac{\sigma^2}{n} \left\{ 1 + 2 \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \rho_j \right\} \approx \frac{\sigma^2}{n} \left(1 + 2 \sum_{j=1}^{\infty} \rho_j \right)$$

であることも分かっている。ここで、 $\sigma^2 = \text{Var}[h(\mathbf{x})]$ であり、 ρ_j は $h(\mathbf{x}^{(t)})$ と $h(\mathbf{x}^{(t+j)})$ の相関係数を表す。

実際の問題では $\rho_j > 0$ である場合がほとんどなので、MCMC 法による期待値計算は、独立なサンプルを使うときよりも分散が大きくなることが分か

²³これらの収束判定方法は、R で作成された CODA というソフトウェアで実行できる。

る。そこで、MCMC法の効率性を測る1つの尺度として、標本自己相関係数 $\hat{\rho}_j$ がよく用いられる。標本自己相関係数の値が高く、なかなかゼロに減少しないときには、(11)式の精度が悪いことを意味するので次節で説明する方法などによってMCMC法を改善する必要がある。また、標本自己相関係数の代わりに、 $1 + 2 \sum_{j=1}^{\infty} \hat{\rho}_j$ を用いて効率性を評価することもある。これを、非効率性因子 (inefficiency factor) と呼んでいる²⁴。サンプルが独立であるとき \hat{I} の分散は σ^2/n であるので、非効率性因子は得られたサンプルを用いたときと独立なサンプルを用いたときの分散の比を表している。

5.3 混合の改善

サンプル間の自己相関が高く効率性が悪くなる原因としては、マルコフ連鎖が状態空間をゆっくりとしか推移していないことが考えられる。このようなとき、連鎖の混合 (mixing) が悪いという。また、MHアルゴリズムで採択確率が低いような場合にも、同じ値が続くことになり自己相関が高くなる。マルコフ連鎖の混合が悪いときには、目標分布への収束も遅くなるので、より多くのサンプルを発生させるか、あるいは提案分布を変えるなどしてサンプリングの方法を改善する必要がある。

連鎖の混合を改善するとき、ギブス・サンプリングや多重MHアルゴリズムを使っているのであれば、別々にサンプリングしている変数を1つのブロックにまとめることができないか検討するとよい。このとき、なるべく相関の高い変数を同じブロックにまとめると改善の効果が高くなる (Liu (1994))。また、例えば $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ のブロックを考え、 $\pi(\mathbf{x}_1, \mathbf{x}_2)$ を解析的に求めることができるとする。このような場合、 $(\mathbf{x}_1, \mathbf{x}_2)$ はギブス・サンプリングによって $\pi(\mathbf{x}_1|\mathbf{x}_2)$ と $\pi(\mathbf{x}_2|\mathbf{x}_1)$ からサンプリングし、 \mathbf{x}_3 は $\pi(\mathbf{x}_3|\mathbf{x}_1, \mathbf{x}_2)$ からサンプリングを行うと、ブロック間の相関が減少し連鎖の混合を改善することができる。

ブロック数を少なくするとアルゴリズムが複雑になったり、ブロック化を行ってもブロック間の相関が高いこともある。このようなとき、適当な変数変

²⁴実際には無限級数を計算することはできないので、十分大きい L を選択して $1 + 2 \sum_{j=1}^L \hat{\rho}_j$ を計算したり、適当な平滑化ウィンドウを用いて非効率性因子を計算をする。

換を行うことによって、アルゴリズムを複雑にすることなくブロック間の相関を減少させ、連鎖の混合を改善できる場合がある。この方法は、変量効果モデルや階層モデルで有効であることが知られている (Gelfand *et al.* (1995), Roberts and Sahu (1997)) .

上述した以外にも、連鎖の混合を改善するための方法が数多く提案されている。例えば、EM アルゴリズムとギブス・サンプリングの類似性に注目して、Liu and Wu (1999), Meng and van Dyk (1999), van Dyk and Meng (2001) らは、EM アルゴリズムの加速方法を MCMC 法に適用した方法を提案している。Neal (1996) では、ランジェヴァン連鎖を特殊な場合として含むハミルトニアン・モンテカルロ法について議論している。また、変数変換を通してサンプリング方法の改善を図るものとして、Liu and Sabatti (2000) の一般化ギブス・サンプリングや Liu (2003) によって提案されたアルゴリズムなどがある。目標分布が多峰分布であるときには、目標分布を変えて多重連鎖を利用する parallel tempering (Geyer (1991)) や simulated tempering (Geyer and Thompson (1995)) などの方法が有用であろう。さらに、候補の値を多数発生させる multiple try 法 (Liu *et al.* (2000)) や multiple point 法 (Qin and Liu (2001)) も、連鎖の混合の改善をねらった方法である。最近では粒子アルゴリズム (第 xxx 章) の考えを MCMC 法に取り込んだ方法も Moral *et al.* (2006) によって提案されている。

階層モデルや複雑な統計モデルを扱うときには、連鎖の混合が悪くなることが多い。しかし、残念ながらこれまでに紹介した方法の中で最良の方法というものはない。マルコフ連鎖の混合の改善は扱う問題の構造に大きく依存しており、どの方法を用いるかはその都度検討しなければならない。

6 応用例

この節では MCMC 法の応用例として、ロジットモデルと隠れマルコフモデルのベイズ推定を考える。ベイズ推定では、尤度関数とパラメータの事前分布から事後分布と呼ばれる確率分布を導出し、そこからパラメータに関する

る推論を行う²⁵。多くの場合、事後分布を解析的に分析することが不可能であるため、最近ではMCMC法によって事後分布からパラメータをサンプリングして様々な計算を行っている。

6.1 ロジットモデルのベイズ推定

まず、Breslow and Clayton (1993) によって分析された種のデータを用いて変量効果を伴うロジットモデルをベイズ推定する。このデータには、21のプレートに配置された種の発芽割合が記録されている²⁶。そこで、第*i*プレートの種の数 n_i 、発芽した種の数 y_i と表し、Breslow and Clayton (1993) にならって次のモデルを考えることにする：

$$(12) \quad y_i \sim \text{Bi}(n_i, p_i), \quad \log \frac{p_i}{1-p_i} = \mathbf{x}'_i \boldsymbol{\beta} + b_i, \quad b_i \sim \mathcal{N}(0, \sigma^2)$$

ここで、 \mathbf{x}_i は説明変数ベクトルを、 b_i は変量効果を表す。さらに、 $\boldsymbol{\beta}$ と σ^2 の事前分布として $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$ 、 $\sigma^2 \sim \text{IG}(n_0/2, s_0/2)$ を仮定する²⁷。

モデルが非線形であるため、すべてのパラメータを同時に事後分布からサンプリングするのは困難である。そこで、 $\boldsymbol{\beta}$ 、 σ^2 、 b_i ($i = 1, \dots, 21$) とに分けてサンプリングすることにする。このとき、 σ^2 の完全条件付分布は容易に導出することができ、

$$\sigma^2 | \boldsymbol{\beta}, \{b_i\} \sim \text{IG} \left(\frac{21 + n_0}{2}, \frac{\sum_{i=1}^{21} b_i^2 + s_0}{2} \right)$$

となる。

次に、 $\boldsymbol{\beta}$ の完全条件付分布は

$$\pi(\boldsymbol{\beta} | \sigma^2, \{b_i\}) \propto \prod_{i=1}^{21} p_i^{y_i} (1-p_i)^{n_i-y_i} \times \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \mathbf{B}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\}$$

と表され、MHアルゴリズムによってサンプリングすることにする。そこで、提案分布を構築するために、修正線形モデル

$$(13) \quad \tilde{y}_i = \mathbf{x}'_i \boldsymbol{\beta} + b_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \tau_i^2)$$

²⁵ベイズ推定に関するテキストとしては、例えば Berger (1980) がある。

²⁶データの詳細については、Crowder (1978) を参照されたい。

²⁷ $\text{IG}(a, b)$ は逆ガンマ分布を表し、確率密度関数は $\pi(x) \propto (1/x)^{a+1} \exp(-b/x)$ である。

を考える (McCullagh and Nelder (1989) を参照). ここで, $\tilde{y}_i = \mathbf{x}'_i\boldsymbol{\beta} + b_i + (y_i - n_i p_i)/\{n_i p_i(1 - p_i)\}$, $\tau_i^2 = 1/\{n_i p_i(1 - p_i)\}$ であり, MH アルゴリズムを適用するときには現在の $\boldsymbol{\beta}$ の値で評価する. (13) 式と $\boldsymbol{\beta}$ の事前分布を組み合わせると, 完全条件付分布は正規分布 $\mathcal{N}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})$ によって近似することができる. ただし,

$$\hat{\mathbf{V}}^{-1} = \sum_{i=1}^{21} \frac{\mathbf{x}_i \mathbf{x}'_i}{\tau_i^2} + \mathbf{B}_0^{-1}, \quad \hat{\boldsymbol{\beta}} = \hat{\mathbf{V}} \left\{ \sum_{i=1}^{21} \frac{\mathbf{x}_i (\tilde{y}_i - b_i)}{\tau_i^2} + \mathbf{B}_0^{-1} \boldsymbol{\beta}_0 \right\}$$

である. ここでは正規分布を多変量 t 分布に置き換え, 提案分布として多変量 t 分布 $\mathcal{T}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})$ を用いることにする.

最後に, b_i の完全条件付分布は

$$\pi(b_i | \boldsymbol{\beta}, \sigma^2) \propto p_i^{y_i} (1 - p_i)^{n_i - y_i} \exp\left(-\frac{b_i^2}{2\sigma^2}\right)$$

となる. 変量効果についても, $\boldsymbol{\beta}$ と同様にして提案分布が $t(\hat{b}_i, \hat{v}_i^2)$ である MH アルゴリズムによってサンプリングを行う. ただし, $\hat{v}_i^2 = \sigma^2 \tau_i^2 / (\sigma^2 + \tau_i^2)$, $\hat{b}_i = \hat{v}_i^2 (\tilde{y}_i - \mathbf{x}'_i \boldsymbol{\beta}) / \tau_i^2$ である.

Breslow and Clayton (1993) と同じ説明変数 (定数項, 種の種類, 希釈液の種類, 交叉項) を用いて, (12) 式のモデルを推定した結果が図 1 に示されている. 推定に際しては, 事前分布を $\boldsymbol{\beta}_0 = 0$, $\mathbf{B}_0 = 100\mathbf{I}$, $n_0 = 1$, $s_0 = 0.01$ とし, 提案分布の自由度は $\nu = 20$ とした²⁸. 図 1 より, 本節のアルゴリズムによって得られるサンプルには自己相関はほとんどなく, 効率的であることが分かる²⁹. また, 採択確率もすべてのパラメータで 90% 以上であり, 修正線形モデルにもとづく近似がうまくいっていると言える. 表 1 には, パラメータの事後平均と事後標準偏差が, Breslow and Clayton (1993) の罰則付疑似最尤推定の結果と合わせて示されている³⁰. ベイズ推定値と疑似最尤推定値がほぼ同じであることが見て取れる.

(12) 式は, 次のように書き直すことができる:

$$y_i \sim \mathcal{B}i(n_i, p_i), \quad \log \frac{p_i}{1 - p_i} = \mu_i, \quad \mu_i \sim \mathcal{N}(\mathbf{x}'_i \boldsymbol{\beta}, \sigma^2)$$

²⁸ 自由度 ν の値は, 採択確率が高くしかも提案分布の裾が厚くなるように調整した.

²⁹ コレログラムでは, 横軸にラグをとり, 縦軸には標本自己相関係数の値が示されている.

³⁰ サンプルの時系列プロットから収束を判断し, 最初の 5000 個のサンプルを捨て残りの 10000 個のサンプルから事後平均と事後標準偏差を計算している

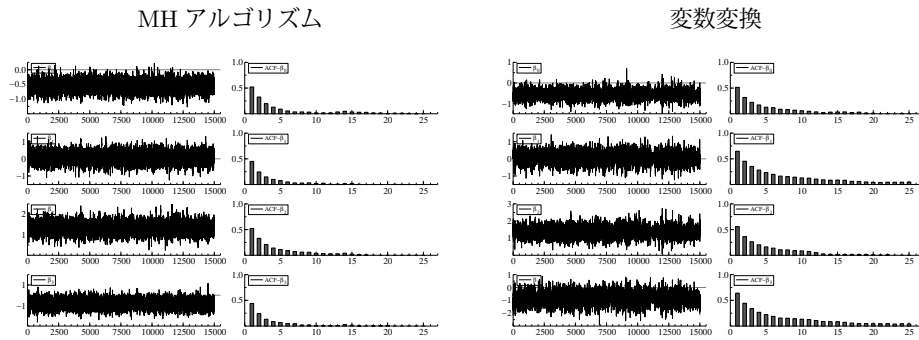


図 1: 種のデータ：サンプルの時系列プロットとコレログラム

表 1: 種のデータ：パラメータの推定値

	ベイズ推定		罰則付疑似最尤推定	
	事後平均	事後標準偏差	推定値	標準誤差
定数項	-0.546	0.177	-0.542	0.190
種の種類	0.091	0.293	0.146	0.308
希釈液の種類	1.333	0.250	1.339	0.270
交叉項	-0.803	0.399	-0.825	0.430
σ	0.242	0.125	0.313	0.121

このとき、 β のサンプリングは MH アルゴリズムを用いることなく、 $\beta | \sigma^2, \{\mu_i\} \sim \mathcal{N}(\tilde{\beta}, \tilde{\mathbf{V}})$ によって行うことができる³¹。ただし、 $\tilde{\mathbf{V}}^{-1} = \sum_{i=1}^{21} \mathbf{x}_i \mathbf{x}_i' / \sigma^2 + \mathbf{B}_0^{-1}$ 、 $\tilde{\beta} = \tilde{\mathbf{V}} (\sum_{i=1}^{21} \mathbf{x}_i \mu_i / \sigma^2 + \mathbf{B}_0^{-1} \beta_0)$ である。第 5 節で述べたように、問題によってはこのような定式化を行った方が混合がよくなることもある。しかし、図 1 から、変数変換を行った方が自己相関の値が大きくなっており、ここでの分析では混合の改善は見られなかった。

6.2 隠れマルコフモデルのベイズ推定

次の応用例として、Leroux and Puterman (1992) の胎動データを考えることにする。このデータには、ある羊の胎児が母体内で動いた回数が 5 秒間隔で記録されている。そこで、時点 t における胎動回数を y_t と表記し、ポア

³¹これに伴い、 μ_i と σ^2 のサンプリングも若干の修正が必要である。

ソン分布に従うと仮定する：

$$y_t \sim \mathcal{P}o(\mu_{s_t}) \quad (t = 1, \dots, T)$$

ここで、 $\mathcal{P}o(\mu_{s_t})$ は平均が μ_{s_t} であるポアソン分布を表す。また、 $s_t \in \{1, 2\}$ は潜在変数を表し、Leroux and Puterman (1992) と同様に、推移行列が

$$\mathbf{T} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

であるマルコフ連鎖に従うとする³²。また、 \mathbf{T} の不変分布を $\boldsymbol{\pi}_0 = (\pi_{01}, \pi_{02})$ と表し、 $s_1 \sim \boldsymbol{\pi}_0$ であると仮定する。 s_t は観測されない変数であり、マルコフ連鎖に従うことから、上のようなモデルを隠れマルコフモデルと呼ぶ。

ベイズ推定を行うため、パラメータ ($\{\mu_i\}, \{p_{ii}\}, \{s_t\}$) をギブス・サンプリングによって事後分布からサンプリングする。いま、 μ_i の事前分布をガンマ分布 $\mathcal{G}a(a_i, b_i)$ とすれば、完全条件付分布は

$$\mu_i | \{p_{ii}\}, \{s_t\} \sim \mathcal{G}a \left(a_i + \sum_{t \in S_i} y_t, b_i + n_i \right) \quad (i = 1, 2)$$

となる³³。ここで、 $S_i = \{t : s_t = i\}$ であり、 n_i は S_i の要素数を表す。

次に、 p_{ii} の事前分布を $\mathcal{B}e(\alpha_i, \beta_i)$ とすれば、 p_{ii} の完全条件付分布は、

$$p_{ii} | \{\mu_i\}, \{s_t\} \sim \mathcal{B}e(\alpha_i + n_{ii}, \beta_i + n_{ij}) \quad (i = 1, 2)$$

で与えられる。ここで、 n_{ij} は状態 i から状態 j に推移した s_t の数を表す。

最後に s_t の完全条件付分布は、ベルヌーイ分布

$$\pi(s_t = i | \{\mu_i\}, \{p_{ii}\}, \{s_{t'}\}_{t' \neq t}) \propto \begin{cases} f(y_t | \mu_i) \pi_{0i} p_{i, s_{t+1}} & t = 1 \text{ の場合} \\ f(y_t | \mu_i) p_{s_{t-1}, i} p_{i, s_{t+1}} & 1 < t < T \text{ の場合} \\ f(y_t | \mu_i) p_{s_{t-1}, i} & t = T \text{ の場合} \end{cases}$$

によって与えられる。ここで、 $f(y|\mu)$ はポアソン分布の確率関数を表す。この方法では、 s_t を一つずつサンプリングするため連鎖の混合が悪くなる場合

³²Leroux and Puterman (1992) では、 $\{s_t\}$ が独立な場合も分析している。

³³ガンマ分布 $\mathcal{G}a(a, b)$ の確率密度関数は、 $\pi(x) \propto x^{a-1} \exp(-bx)$ で与えられる。

がある。Chib (1996) では $\{s_t\}$ を同時にサンプリングする方法が提案されており、比較のためこれら 2つの方法を用いてモデルを推定することにする。

図 2 には、 $\{s_t\}$ を同時にサンプリングしたとき（左側）と 1 つずつサンプリングしたとき（右側）の μ_i の時系列プロットと標本自己相関係数のプロットが示されている。標本自己相関係数のプロットを比べると、 $\{s_t\}$ を同時に

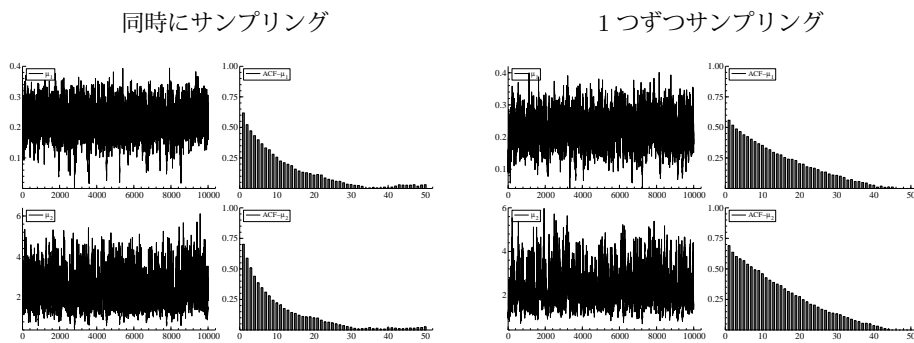


図 2: 胎動データ：サンプルの時系列プロットとコレログラム

サンプリングした方が自己相関が速くゼロに減少し、効率性がよくなることが確認される。また、表 2 にはパラメータの事後平均、事後標準偏差、非効率性因子 (IF) が示されている³⁴。この表からも、同時にサンプリングした方が効率性が改善されることが分かる。

表 2: 胎動データ：推定値と非効率性因子

	同時にサンプリング			1 つずつサンプリング		
	事後平均	事後標準偏差	IF	事後平均	事後標準偏差	IF
μ_1	0.220	0.050	14.346	0.220	0.048	19.289
μ_2	2.282	0.769	13.445	2.273	0.744	24.257
p_{11}	0.968	0.024	16.195	0.968	0.023	20.295
p_{22}	0.672	0.154	2.6689	0.671	0.148	4.7609

³⁴ここでは、Chib (1996) と同じ事前分布を用いて分析を行った。また各統計量の値は、最初の 5000 個のサンプルを捨て残りの 10000 個のサンプルから計算し、非効率性因子については 50 までのラグを用いた。

参考文献

- [1] 伊庭幸人 (2005). “マルコフ連鎖モンテカルロ法の基礎,” 計算統計 II (マルコフ連鎖モンテカルロ法とその周辺), 3–106, 岩波書店.
- [2] 大森裕浩 (2001). “マルコフ連鎖モンテカルロ法の最近の展開,” 日本統計学会誌 **31**, 305–344.
- [3] 中妻照雄 (2003). 「ファイナンスのための MCMC 法によるベイズ分析」三菱経済研究所
- [4] 和合肇 (編) (2005) 「ベイズ計量分析 – マルコフ連鎖モンテカルロ法とその応用」東洋経済新報社
- [5] Albert, J. and Chib, S. (1993). “Bayesian analysis of binary and polychotomous response data,” *Journal of the American Statistical Association* **88**, 669–679.
- [6] Berger, J.O. (1980). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). Springer, New York.
- [7] Besag, J. and Green, P. (1993). “Spatial statistics and Bayesian computation,” *Journal of the Royal Statistics Society* **B55**, 25–37.
- [8] Breslow, N.E. and Clayton, D.G. (1993). “Approximate inference in generalized linear mixed models,” *Journal of the American Statistical Association* **88**, 9–25.
- [9] Chan, K.S. (1993). “Asymptotic behavior of the Gibbs sampler,” *Journal of the American Statistical Association* **88**, 320–326.
- [10] Chib, S. (1992). “Bayes regression for the tobit censored regression model,” *Journal of Econometrics* **51**, 79–99.
- [11] Chib, S. (1996). “Calculating posterior distributions and modal estimates in Markov mixture models,” *Journal of Econometrics* **75**, 79–97.
- [12] Chib, S. and Greenberg, E. (1995). “Understanding the Metropolis-Hastings algorithm,” *American Statistician* **49**, 327–335.
- [13] Christensen, O.F., Møller, J., and Waagepetersen, R.P. (2001). “Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models”, *Methodology and Computing in Applied Probability* **3**, 309–327.
- [14] Christensen, O.F. and Waagepetersen, R. (2002). “Bayesian prediction of spatial count data using generalized linear mixed models,” *Biometrics* **58**, 280–286.
- [15] Cowles, M.K. and Carlin, B.P. (1996). “Markov chain Monte Carlo convergence diagnostics: A comparative review”, *Journal of the American Statistical Association* **91**, 883–904.

- [16] Crowder, M.J. (1978). “Beta-Binomial ANOVA for proportions,” *Applied Statistics* **27**, 34–37.
- [17] Damien, P., Wakefield, J., and Walker, S.G. (1999). “Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables,” *Journal of the Royal Statistics Society* **B61**, 331–344.
- [18] Damien, P. and Walker, S.G. (2001). “Sampling truncated normal, beta, and gamma densities,” *Journal of Computational and Graphical Statistics* **10**, 206–215.
- [19] Dempster, A. P., Laird, N.M., and Rubin, D.B. (1977). “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistics Society* **B39**, 1–38.
- [20] Devroye, L. (1896). *Non-Uniform Random Variate Generation*. Springer–Verlag, New York.
- [21] Dongarra, J. and Sullivan, F. (2000). “Guest editors’ introduction: The top 10 algorithms,” *Computing in Science and Engineering* **2**, 22–23.
- [22] Evans, M. and Swartz, T. (2000). *Approximating Integrals Via Monte Carlo and Deterministic Methods*. Oxford University Press, Oxford.
- [23] Gamerman, D. and Lopes, H. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2nd ed.). Chapman & Hall/CRC, London.
- [24] Gelfand, A.E., Sahu, S.K., and Carlin, B.P. (1995). “Efficient parametrisations for normal linear mixed models,” *Biometrika* **82**, 479–488.
- [25] Gelfand, A.E. and Smith, A.F.M. (1990). “Sampling-based approaches to calculating marginal densities,” *Journal of the American Statistical Association* **85**, 398–409.
- [26] Gelman, A. and Rubin, D.B. (1992). “Inference from iterative simulation using multiple sequences (with discussion),” *statistical Science* **7**, 457–511.
- [27] Geman, S. and Geman, D. (1984). “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- [28] Gentle, J.E. (2003). *Random Number Generation and Monte Carlo Methods*. Springer, New York.
- [29] Geweke, J. (1989). “Bayesian inference in econometric models using Monte Carlo integration,” *Econometrica* **57**, 1317–1340.
- [30] Geweke, J. (1992). “Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments,” in J. M. Bernardo *et al.*(eds.), *Bayesian Statistics 4*, 169–193, Oxford University Press, Oxford.
- [31] Geyer, C.J. (1991). “Markov chain Monte Carlo maximum likelihood,” in

- E.Keramigas (eds.), Computing Science and Statistics: The 23rd symposium on the inference, Interface Foundation, Fairfax, 156–163.
- [32] Geyer, C.J. and Thompson, E. (1995). “Annealing Markov chain Monte Carlo with applications to ancestral inference”, *Journal of the American Statistical Association* **90**, 909–920.
- [33] Hastings, W.K. (1970). “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika* **57**, 97–109.
- [34] Häggström, O. (2002). *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press, Cambridge.
- [35] Heidelberger, P. and Welch, P.D. (1983). “Simulation run length control in the presence of an initial transient,” *Operations Research* **31**, 1109–1144.
- [36] Higdon, D. (1998). “Auxiliary variable methods for Markov chain Monte Carlo with applications,” *Journal of the American Statistical Association* **93**, 398–409.
- [37] Karlin, S. and Taylor, J. (1975). *A First Course in Stochastic Processes* (2nd ed.). Academic Press, New York.
- [38] Leroux, B.G. and Puterman, M.L. (1992). “Maximum-penalized likelihood estimation for independent and Markov-dependent mixture models,” *Biometrics* **48**, 545–558.
- [39] Liu, C. (2002). “An example of algorithm mining: Covariance adjustment to accelerate EM and Gibbs,” in J. Huang and H. Zhang (eds.), *Development of Modern Statistics and Related Topics*, 74–88, World Scientific, New Jersey.
- [40] Liu, C. (2003). “Alternating subspace-spanning resampling to accelerate Markov chain Monte Carlo simulation,” *Journal of the American Statistical Association* **98**, 110–117.
- [41] Liu, J.S. (1994). “The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem,” *Journal of the American Statistical Association* **89**, 958–966.
- [42] Liu, J.S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer, New York.
- [43] Liu, J.S., Liang, F., and Wong, W.H. (2000). “The use of multiple-try method and local optimization in Metropolis sampling,” *Journal of the American Statistical Association* **95**, 121–134.
- [44] Liu, J.S. and Sabatti, C. (2000). “Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation,” *Biometrika* **87**, 353–369.
- [45] Liu, J.S., Wong, W.H., and Kong, A. (1995). “Covariance structure and convergence rate of the Gibbs sampler with various scans”, *Journal of the Royal Statistical Society* **B57**, 157–169.

- [46] Liu, J.S. and Wu, Y.N. (1999). “Parameter expansion for data augmentation,” *Journal of the American Statistical Association* **94**, 1264–1274.
- [47] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models* (2nd ed.). Chapman & Hall, London.
- [48] Meng, X.-L. and van Dyk, D.A. (1999). “Seeking efficient data augmentation schemes via conditional and marginal augmentation,” *Biometrika* **86**, 301–320.
- [49] Mengersen, K.L., Robert, C.P., and Guihenneuc-Jouyaux, C. (1999). “MCMC convergence diagnostics: A review,” in J. M. Bernardo *et al.* (eds.), *Bayesian Statistics 6*, 415–440, Clarendon Press, Oxford.
- [50] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). “Equations of state calculations by fast computing machines,” *Journal of Chemical Physics* **21**, 1087–1091.
- [51] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. Springer–Verlag, London.
- [52] Moral, P.D., Doucet, A., and Jasra, A. (2006). “Sequential Monte carlo samplers”, *Journal of the Royal Statistical Society* **B68**, 411–436.
- [53] Müller, P. (1991). “A generic approach to posterior integration and Gibbs sampling,” Technical report 91-09, Institute of Statistics and Decision Sciences, Duke University.
- [54] Neal, R.M. (1996). *Bayesian Learning for Neural Networks*. Lecture Notes 118, Springer–Verlag, New York.
- [55] Neal, R.M. (2003). “Slice sampling,” *Annals of Statistics* **31**, 705–767.
- [56] Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, Cambridge.
- [57] Qin, Z. and Liu, J.S. (2001). “Multi-point Metropolis method with application to hybrid Monte Carlo,” *Journal of Computational Physics* **172**, 827–840.
- [58] Raftery, A.E. and Lewis, S. (1992). “How many iterations in the Gibbs sampler?” in J. M. Bernardo *et al.* (eds.), *Bayesian Statistics 4*, 763–773, Oxford University Press, Oxford.
- [59] Ripley, W. (1987). *Stochastic Simulation*. Wiley, New York.
- [60] Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd ed.). Springer–Verlag, New York.
- [61] Roberts, G.O., Gelman, A., and Gilks, W.R. (1997). “Weak convergence and optimal scaling of random walk Metropolis algorithms,” *Annals of Applied Probability* **7**, 110–120.
- [62] Roberts, G.O. and Rosenthal, J.S. (1998). “Optimal scaling of discrete ap-

- proximations to Langevin diffusions,” *Journal of the Royal Statistical Society* **B60**, 255–268.
- [63] Roberts, G.O. and Rosenthal, J.S. (2001). “Optimal scaling for various Metropolis-Hastings algorithms,” *Statistical Science* **16**, 351–367.
- [64] Roberts, G.O. and Sahu, S.K. (1997). “Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler,” *Journal of the Royal Statistical Society* **B56**, 377–384.
- [65] Roberts, G.O. and Smith, A.F.M. (1994). “Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms,” *Stochastic Processes and Their Applications* **49**, 207–216.
- [66] Ross, S.M. (1995). *Stochastic Processes* (2nd ed.). Wiley, New York.
- [67] Rubin, D.B. (1987) *Multiple Imputation or Non-response in Surveys*. Wiley, New York.
- [68] Rubinstein, R.Y. (1981). *Simulation and the Monte Carlo Method*. Wiley, New York.
- [69] Tanner, M. A. and Wong, W. H. (1987). “The calculation of posterior distributions by data augmentation,” *Journal of the American Statistical Association* **82**, 528–549.
- [70] Tierney, L. (1994). “Markov chains for exploring posterior distributions (with discussion),” *Annals of Statistics* **22**, 1701–1762.
- [71] van Dyk, D.A. and Meng, X.-L. (2001). “The art of data augmentation (with discussions),” *Journal of Computational and Graphical Statistics* **10**, 1–111.

第11章 逐次モンテカルロ法とパーティクル フィルタ

生駒哲一¹

「パーティクルフィルタ」とは逐次モンテカルロ法の最適フィルタ問題への適用と位置づけられ、動的システムの状態推定問題を状態空間中の多数の粒子の数値計算により近似的に解く方法である。コンピューターの高い計算能力を駆使したパーティクルフィルタはカルマンフィルタに比べると格段に広いクラスの動的システムを扱うことができる魅力的なフィルタで、今後も様々な分野で積極的に活用されるであろう。本章では状態推定問題を定式化した後、逐次モンテカルロ法の枠組みでパーティクルフィルタを説明する。その後、もっとも簡単な特殊系であるモンテカルロフィルタと高度な特殊形のラオ＝ブラックウエル化を解説する。後半では、幾つかの応用事例により、パーティクルフィルタの特長を活かした利用方法を紹介する。

¹九州工業大学工学部

第11章 逐次モンテカルロ法とパーティクルフィルタ

¹ パーティクルフィルタ (Particle filters, 粒子フィルタとも呼ばれる) とは, コンピュータを駆使した状態推定法の一つである. 状態空間中の多数の粒子により状態の分布を近似し, 粒子の値の数値計算により状態の分布を時間更新する. 近年のコンピュータ計算性能の向上に伴い, かなり実用的になってきた手法である. そのアイデア自体は古くから (カルマンフィルタが発明された 1958 年より後, 1960 年代終盤から) 提案されてはいたが [2], [3], [14], [15], [28], [29], [32], 当時のコンピュータは十分な性能を持たなかった為, 一旦は忘れ去られた. その後, コンピュータの性能が向上してきた 1990 年代初頭に, 日本と英国で同時期に再発明された [12], [20].

ここでの興味の対象は, 時間的に変化する信号を扱うシステム, すなわち動的システムである. 動的システムが出力する信号の挙動を定めるのは, システムの内部状態とその時間変化規則, および各種の入力情報である. このうち内部状態が主たる役割を果たす. 今後この内部状態を, 単に「状態」と呼ぶことにする. ある時刻での状態が分かれば, 以降の対象システムの挙動が分かる. 状態の時間変化規則は, 対象についての経験的な知識に基づき, 数学モデルで記述する. これが「システムモデル」と呼ばれるものである. ただし, 我々は対象システムの時間変化を完全には把握できない為, 決定論的要素と確率論的要素の二つでシステムモデルを記述する. 時間変化に関して既知の事柄は決定論的に記述され, 未知の事柄は確率論的に記述される.

状態は, 直接には観測できないことがよくある. 仮に直接観測ができたと

¹Sep.21,2007

しても、観測にまつわるノイズを伴うことが多い。これを、ノイズを含む観測過程として、数学モデルで記述する。このモデルは「観測モデル」と呼ばれる。例えば、状態に、観測にまつわるノイズが状態に足しあわされたものが観測値として得られる、といった観測モデルを立てる。また、状態の全てが直接的に観測できるわけではなく、その一部分や、あるいは変換された値しか観測できない状況も多い。観測モデルでは、このような状況も必要に応じて記述する。

システムモデルと観測モデルの対で「状態空間モデル」が構成される。状態空間モデルが既知で、初期値（初期分布）が既知の場合に、観測値が時々刻々と与えられる状況を取り扱う。この状況で、観測値の時系列が与えられた下での状態の事後分布を求めることが「状態推定」である。パーティクルフィルタとは、この状態推定を、状態空間中の多数の粒子を用いた数値計算により、近似的に行う手法である。

本文では、実時間で動作するオンラインアプリケーションを念頭において、主にフィルタ（ろ波）分布の推定について重点的に論じる。

1 状態空間モデルと状態推定

パーティクルフィルタを説明する前に、まず、状態空間モデルと状態推定についての一般的事柄を定式化しておくことにする。以下では、離散時間システムを考え、離散時刻を整数 $k = 0, 1, 2, \dots$ で表す。時刻 k の状態ベクトルを \mathbf{x}_k 、観測ベクトルを \mathbf{y}_k と表記する。

状態の時間変化、すなわち状態遷移を、システムモデル

$$\mathbf{x}_k \sim f(\cdot | \mathbf{x}_{k-1}) \quad (1.1)$$

で表す。状態から観測値が得られる過程を、観測モデル

$$\mathbf{y}_k \sim h(\cdot | \mathbf{x}_k) \quad (1.2)$$

で表す。二つのモデル中の条件付き確率分布 $f(\mathbf{x}_k | \mathbf{x}_{k-1})$ および $h(\mathbf{y}_k | \mathbf{x}_k)$ は、どちらも既知である。これらシステムモデルと観測モデルの対で、状態空間

モデルが構成される。状態の初期分布を

$$p_0(\mathbf{x}_0) \quad (1.3)$$

とし、これも既知であるものとする。以降、やや特殊な表記法として、観測や状態の時系列を、

$$\mathbf{y}_{1:k} \equiv \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k\} \quad (1.4)$$

$$\mathbf{x}_{1:k} \equiv \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\} \quad (1.5)$$

と表すことにする。

式 (1.1) のシステムモデルは、状態遷移のマルコフ性

$$p(\mathbf{x}_k | \mathbf{x}_{1:k-1}, \mathbf{y}_{1:k-1}) = f(\mathbf{x}_k | \mathbf{x}_{k-1}) \quad (1.6)$$

を仮定したことになる。すなわち、現時刻 k の状態 \mathbf{x}_k の確率分布は、1 時刻前の状態 \mathbf{x}_{k-1} が与えられれば定まり、それよりも過去の状態や過去の観測の有無に関わらず状態 \mathbf{x}_k の確率分布は同じである。これは、数学モデルを扱う際に、式 (1.6) 左辺の確率分布が出てきた場合には、それをシステムモデルで置き換えてよいことを意味する。

一方、式 (1.2) の観測モデルは、

$$p(\mathbf{y}_k | \mathbf{x}_{1:k}, \mathbf{y}_{1:k-1}) = h(\mathbf{y}_k | \mathbf{x}_k) \quad (1.7)$$

を仮定したことになる。すなわち、現在の観測 \mathbf{y}_k の確率分布は、現在の状態 \mathbf{x}_k が与えられれば定まり、それより過去の状態や観測の有無には影響されない。システムモデルの場合と同様に、式 (1.7) の左辺が数学モデルの式中出现してきた場合には、それを観測モデルで置き換えてよい。

状態推定とは、観測時系列 $\mathbf{y}_{1:k}$ が与えられた下での状態の事後分布を求めることである。この事後分布は、観測時系列の最終時刻 k と状態の時刻 k' との関係で、予測 ($k' > k$)、フィルタ (あるいは「ろ波」) ($k' = k$)、平滑化 ($k' < k$) の3つに分類することができる。本文では主にフィルタを扱うことにする。これは、動的システムを実時間で扱う場合を想定していることに相当する。

動的システムの実時間的扱に向いたものとして、観測値 \mathbf{y}_k が与えられる度に状態系列の事後分布を時間更新する逐次ベイズの式

$$p(\mathbf{x}_{1:k} | \mathbf{y}_{1:k}) = p(\mathbf{x}_{1:k-1} | \mathbf{y}_{1:k-1}) \frac{f(\mathbf{x}_k | \mathbf{x}_{k-1}) h(\mathbf{y}_k | \mathbf{x}_k)}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1})} \quad (1.8)$$

がある。この式は、次のように求めることができる。

$$\begin{aligned} p(\mathbf{x}_{1:k} | \mathbf{y}_{1:k}) &= \frac{p(\mathbf{x}_{1:k}, \mathbf{y}_k | \mathbf{y}_{1:k-1})}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1})} \\ &= p(\mathbf{x}_{1:k-1} | \mathbf{y}_{1:k-1}) \frac{p(\mathbf{x}_k, \mathbf{y}_k | \mathbf{x}_{1:k-1}, \mathbf{y}_{1:k-1})}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1})} \\ &= p(\mathbf{x}_{1:k-1} | \mathbf{y}_{1:k-1}) \frac{f(\mathbf{x}_k | \mathbf{x}_{k-1}) h(\mathbf{y}_k | \mathbf{x}_k)}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1})} \end{aligned}$$

2 逐次モンテカルロ法

逐次モンテカルロ法とは、分布の系列に従う粒子系列を効率的に生成する方法の総称である。分布の系列として状態系列の事後分布、すなわち状態推定を扱う特殊な場合がパーティクルフィルタである。パーティクルフィルタにもいくつかのバリエーションがある。すなわち、パーティクルフィルタとは、逐次モンテカルロ法により動的システムの状態推定を行う方法の総称である。逐次モンテカルロ法やその関連手法についての解説は、例えば [30] を参照されたい。本節では、パーティクルフィルタについて述べた後、そのバリエーションの一つであるモンテカルロフィルタを説明する。

2.1 パーティクルフィルタ

現在の時刻を k とし、1時刻前 ($k-1$) までの状態系列の事後分布に対し重み付き粒子群による近似

$$\left\{ \left(\mathbf{x}_{1:k-1}^{(i)}, w_{k-1}^{(i)} \right) \right\}_{i=1}^M \sim p(\mathbf{x}_{1:k-1} | \mathbf{y}_{1:k-1}) \quad (2.1)$$

が与えられているものとする。パーティクルフィルタは、観測 \mathbf{y}_k が得られたとき、1時刻前の重みつき粒子群 (式 (2.1)) から、現在の時刻 k の重み付き

粒子群

$$\left\{ \left(\mathbf{x}_{1:k}^{(i)}, w_k^{(i)} \right) \right\}_{i=1}^M \sim p(\mathbf{x}_{1:k} | \mathbf{y}_{1:k}) \quad (2.2)$$

を数値計算により得る手続きから成る。すなわち、事後分布に従う重み付き粒子を時間更新するアルゴリズムから成る。

更新アルゴリズムでは、時刻 k の粒子をプロポーザル分布から生成し、重み値を適切な値に更新する。その後、必要に応じてリサンプリングと呼ばれる復元抽出手続きを行う。

状態系列に対するプロポーザル分布は、

$$q(\mathbf{x}_{1:k} | \mathbf{y}_{1:k}) := q(\mathbf{x}_{1:k-1} | \mathbf{y}_{1:k-1}) q(\mathbf{x}_k | \mathbf{x}_{1:k-1}, \mathbf{y}_{1:k}) \quad (2.3)$$

と分解される。右辺第一項 $q(\mathbf{x}_{1:k-1} | \mathbf{y}_{1:k-1})$ は、過去の粒子系列を生成したプロポーザル分布であり、その条件部分には現在の観測 \mathbf{y}_k が含まれないことに注意したい。過去の粒子を生成した時点では、現在の観測 \mathbf{y}_k はまだ得られていない為、このような形になるのである。なお、右辺第一項に対しても、式 (2.3) の関係を再帰的に適用し、分布を分解していくことになる。

現時刻 k の粒子を、プロポーザル分布で

$$\tilde{\mathbf{x}}_k^{(i)} \sim q(\cdot | \mathbf{x}_{1:k-1}^{(i)}, \mathbf{y}_{1:k}), \quad i = 1, 2, \dots, M \quad (2.4)$$

と生成する。生成した粒子 $\tilde{\mathbf{x}}_k^{(i)}$ を、過去の粒子系列と合併して

$$\tilde{\mathbf{x}}_{1:k}^{(i)} := \mathbf{x}_{1:k-1}^{(i)} \cup \left\{ \tilde{\mathbf{x}}_k^{(i)} \right\}, \quad i = 1, 2, \dots, M \quad (2.5)$$

とし、これを現時刻 k までの暫定的な粒子系列とする。ただし、このように合併するのはあくまで数式上の扱いであって、実際にコンピュータで数値計算を行う際には、特に必要がなければ粒子系列全体をメモリには保存せず、現時刻の粒子のみを保存するようにする。

こうして得られた粒子群が、式 (2.2) のように、現時刻までの状態系列の事後分布を近似するようにしたい。そのためには、インポートランスサンプリングの考え方に従い、粒子の重み $w_{k-1}^{(i)}$ を適切に更新する。インポートランスサンプリングとは、近似したい分布 (目的分布 $p(\mathbf{x})$) とは異なる分布である

プロポージル分布 $q(\mathbf{x})$ から粒子を生成するが、各粒子に適切な重み値を与えることで重み付き粒子群が目的分布の近似となるようにする方法である。インポートランスサンプリングの利点の一つとして、目的分布が複雑でその分布に従う乱数生成が難しい場合、これを回避できる点がある。すなわち、目的分布よりも単純な分布（例えば正規分布）をプロポージル分布とし、それに従う乱数を生成する。これに適切な重み値を付与して、目的分布を近似するようにする。また、インポートランスサンプリングでは、サンプル（乱数）の配置が同じであっても、重みを調整することで、異なる分布を近似表現することができる。

プロポージル分布はある程度自由に定めることができるが、目的分布に近いことが望ましい。目的分布の確率非零領域に対してプロポージル分布から粒子が生成されるようにするため、条件

$$\forall \mathbf{x}, p(\mathbf{x}) > 0 \Rightarrow q(\mathbf{x}) > 0 \quad (2.6)$$

を満たす必要がある。また、数値計算上の留意事項として、後述する重み（式 (2.7)）の値がオーバーフローしないように注意が必要で、そのためには $p(\mathbf{x}) \gg q(\mathbf{x})$ とならないようにする。

インポートランスサンプリングにおける重みは、目的分布とプロポージル分布との比で与えられる。パーティクルフィルタにおいては、目的分布は事後分布 $p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})$ であり、プロポージル分布が $q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})$ であるから、重みは

$$w_k \equiv w_k(\mathbf{x}_{1:k}) \propto \frac{p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})}{q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})} \quad (2.7)$$

となる。重みは粒子の相対的関係を表すものなので、重みの絶対的値は重要ではない。そのため、重みは分布の比に比例する値をとっている。

重みの更新式は、事後分布の更新式 (1.8) から次のように得ることができる。式 (1.8) の両辺をプロポージル分布で割り、式 (2.3) を使って右辺のプロポージル分布を分解すると

$$\frac{p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})}{q(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})} = \frac{p(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k-1})}{q(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k-1})} \frac{f(\mathbf{x}_k|\mathbf{x}_{k-1})h(\mathbf{y}_k|\mathbf{x}_k)}{q(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{y}_{1:k})p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} \quad (2.8)$$

を得る。重みの定義式 (2.7) より、左辺および右辺の第 1 項はそれぞれ時刻 k および $k-1$ の重みであるので、

$$w_k(\mathbf{x}_{1:k}) \propto w_{k-1}(\mathbf{x}_{1:k-1}) \frac{f(\mathbf{x}_k | \mathbf{x}_{k-1}) h(\mathbf{y}_k | \mathbf{x}_k)}{q(\mathbf{x}_k | \mathbf{x}_{1:k-1}, \mathbf{y}_{1:k})} \quad (2.9)$$

となり、重みの更新式が導出された。パーティクルフィルタの手続きでは、この式を用いて、各粒子の重み値 $\tilde{w}_k^{(i)} \equiv w_k(\tilde{\mathbf{x}}_{1:k}^{(i)})$ を

$$\tilde{w}_k^{(i)} \propto w_{k-1}^{(i)} \frac{f(\tilde{\mathbf{x}}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}) h(\mathbf{y}_k | \tilde{\mathbf{x}}_k^{(i)})}{q(\tilde{\mathbf{x}}_k^{(i)} | \mathbf{x}_{1:k-1}^{(i)}, \mathbf{y}_{1:k})}, \quad i = 1, 2, \dots, M \quad (2.10)$$

と更新する。

以上で述べた粒子の生成、重みの更新の後、リサンプリングを行うことで、パーティクルフィルタの時間更新アルゴリズムは完了する。ただしリサンプリングは、重み値が粒子間で大きくばらついている時のみ行えば十分である。例えば極端だが、全粒子の重み値が等しい場合には、リサンプリングは全く必要ない。このような状況でリサンプリングを行うと、モンテカルロ誤差が無用に混入することになり、むしろ推定の精度が悪くなる。ここでのモンテカルロ誤差は、リサンプリングの手続き中の復元抽出で用いる乱数系列の、たまたまの値（実現値）について生じる。また、リサンプリングの手続きは乱数生成などの複雑な処理を必要とすることから、なるべくならリサンプリングを行わない方が計算量が減り好都合である。よって、重み値が均等かあるいは偏っているかをまず判断して、リサンプリングを行うかどうか決定するのが好ましい。

重み値が均等であるかどうかを測るひとつの尺度として、有効サンプルサイズ (Effective Sample Size)

$$ESS = \frac{1}{\sum_{i=1}^M (w_k^{(i)})^2} \quad (2.11)$$

がある [23]。これは、有効に活用されている粒子の数を与える指標である。重み値が全粒子について均等の場合、すなわち、全粒子が等しく活用されている場合、 $ESS = M$ となる。一方、ある一つの粒子のみが非零重み値を持ち、

他の全ての粒子は重み値 0 を持つ極端な場合には、 $ESS = 1$ となる。これは一つの粒子のみが活用され他の粒子は全く活用されていないことに相当する。 ESS の値について、実際には適当な閾値を設けて、リサンプリングを行うかどうかを判断することになる。

パーティクルフィルタは実時間のオンラインアプリケーションを志向している。その計算量は、実時間実行に見合う程度に少ないことが必要である。また、観測データ y_k が与えられる度に行う処理は、時刻によらず一定のオーダーの計算量でなければ現実的ではない。すなわち、時間と共に計算量が増えていくような処理は、実時間のオンラインアプリケーションには不向きである。これらの点について、パーティクルフィルタは手続きが工夫されている。その工夫の一つとして、過去の粒子を再利用するようになっている。すなわち、過去の時刻の粒子については、現時刻では特に値の変更はせず、単に再利用する。過去に遡って計算しない為、計算量を粒子数のオーダーに抑えることができる。よって、計算量のオーダーは、粒子数 M の比例であり、また時間に関して一定である。計算量が粒子数の累乗になったりせず、時間と共に増大しないので、実時間のオンラインアプリケーションに向いている。ただし、粒子数の多さがパーティクルフィルタの難点の一つであるが、粒子ごと個別に行う計算が多いため、並列計算による高速化が期待される。

パーティクルフィルタの理論的裏づけと処理手順は以上の通りであるが、実時間オンラインアプリケーションを想定した実際的情況では、次のような、より簡単な手続きを用いることが多い。プロポザル分布の条件部分には、1 時刻前の状態と現時刻の観測のみを用い、それよりも過去の状態や観測は用いないようにする。すなわち、 $q(\mathbf{x}_k | \mathbf{x}_{1:k-1}, \mathbf{y}_{1:k}) = q(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k)$ とする。これに伴い、保持する粒子も最新の 1 時刻分のみでよくなり、粒子群が近似する分布は状態系列の事後分布ではなく、ある 1 時刻の状態の事後分布、すなわちフィルタの分布となる。このことを考慮して、パーティクルフィルタの実用的なアルゴリズムをまとめると、図 2.1 のようになる。ただし、より実際的なアルゴリズムとして、式 (2.14) の計算は数値計算上の都合から対数をとって行う。その外にも、重み値を対数重みから非対数重みへ戻す際に、値がオーバーフローしないような工夫も必要である。

0. 時刻 $k-1$ のフィルタ分布について、重み付き粒子群による近似が与えられているものとする。

$$\left\{ \left(\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)} \right) \right\}_{i=1}^M \sim p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) \quad (2.12)$$

1. プロポザル分布より粒子を生成する。

$$\tilde{\mathbf{x}}_k^{(i)} \sim q(\cdot | \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_k), \quad i = 1, 2, \dots, M \quad (2.13)$$

2. 重みを更新する。得られた重みは正規化されるものとする。

$$\tilde{w}_k^{(i)} \propto w_{k-1}^{(i)} \frac{f(\tilde{\mathbf{x}}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}) h(\mathbf{y}_k | \tilde{\mathbf{x}}_k^{(i)})}{q(\tilde{\mathbf{x}}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_k)}, \quad i = 1, 2, \dots, M \quad (2.14)$$

3. リサンプリングを行う場合は

$$\mathbf{x}_k^{(i)} \sim \begin{cases} \tilde{\mathbf{x}}_k^{(1)} & \text{with prob. } \tilde{w}_k^{(1)} \\ \tilde{\mathbf{x}}_k^{(2)} & \text{with prob. } \tilde{w}_k^{(2)} \\ \vdots & \vdots \\ \tilde{\mathbf{x}}_k^{(M)} & \text{with prob. } \tilde{w}_k^{(M)} \end{cases} \quad i = 1, 2, \dots, M \quad (2.15)$$

と復元抽出を行い、重みを次の通り均等化する。

$$w_k^{(i)} := 1/M, \quad i = 1, 2, \dots, M \quad (2.16)$$

リサンプリングを行わない場合は、粒子と重みはそのまま、

$$\mathbf{x}_k^{(i)} := \tilde{\mathbf{x}}_k^{(i)}, \quad w_k^{(i)} := \tilde{w}_k^{(i)}, \quad i = 1, 2, \dots, M \quad (2.17)$$

とする。

4. 以上の手続きにより、時刻 k のフィルタ分布を近似する重み付き粒子群が得られる。

$$\left\{ \left(\mathbf{x}_k^{(i)}, w_k^{(i)} \right) \right\}_{i=1}^M \sim p(\mathbf{x}_k | \mathbf{y}_{1:k}) \quad (2.18)$$

図2.1: パーティクルフィルタのアルゴリズム

統計解析においては、尤度や情報量規準を用いてモデルパラメータの推定やモデル選択を行いたい場合がある。パーティクルフィルタでは、時刻 k での尤度を

$$\begin{aligned} p(\mathbf{y}_k | \mathbf{y}_{1:k-1}) &= \int h(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) d\mathbf{x}_k \\ &\simeq \sum_{i=1}^M h(\mathbf{y}_k | \tilde{\mathbf{x}}_{k|k-1}^{(i)}) w_{k-1}^{(i)} \end{aligned} \quad (2.19)$$

と計算する。ここで、粒子 $\tilde{\mathbf{x}}_{k|k-1}^{(i)}$ はシステムモデルにより $\tilde{\mathbf{x}}_{k|k-1}^{(i)} \sim f(\mathbf{x}_k | \mathbf{x}_{k-1}^{(i)})$ と生成したもので、正規化重み $w_{k-1}^{(i)}$ との組にて1期先予測の分布 $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ を近似している。実際のアロリズムでは、式 (2.19) は対数を取って計算する。これにより対数尤度が得られるので、そこから情報量規準も計算することができる。

本文はフィルタ分布の推定に重点を置いたものであるが、アプリケーションによっては平滑化 $p(\mathbf{x}_{k-L} | \mathbf{y}_{1:k})$ (ただし L は正の整数) が必要になる場合もあるので、これについて少し触れておく。パーティクルフィルタによる平滑化は、一見すると、単に過去の時刻の粒子を保持しておき、それに現時刻 k の重みが伴う形 $\left\{ \left(\mathbf{x}_{k-L}^{(i)}, w_k^{(i)} \right) \right\}_{i=1}^M$ で良いように思えるが、これはうまく機能しない。その理由は、リサンプリングの手続きにより粒子のバリエーションが減るからである。時刻を k から遡るにつれて粒子のバリエーションはますます減ってゆき、遡る時刻 L が大きいとき、極端な場合には全て同じ値の粒子ばかりになってしまう。そのため、平滑化分布を表すのに十分な粒子のバリエーションが得られなくなるのである。

平滑化分布をうまく得るための方法がいくつか工夫されている。 $N > k$ について、例えば、固定区間平滑化の式

$$p(\mathbf{x}_k | \mathbf{y}_{1:N}) = p(\mathbf{x}_k | \mathbf{y}_{1:k}) \int \frac{p(\mathbf{x}_{k+1} | \mathbf{y}_{1:N}) f(\mathbf{x}_{k+1} | \mathbf{x}_k)}{p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})} \quad (2.20)$$

に基づき、粒子の配置はそのまま、時刻を遡りながら重みを調整する方法 [6] や、重みの調整だけでなく過去の粒子を新たに生成する方法として、Two-filter-formula

$$p(\mathbf{x}_k | \mathbf{y}_{1:N}) \propto p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) p(\mathbf{y}_{k:N} | \mathbf{x}_k) \quad (2.21)$$

を後退アルゴリズムで時刻を遡りながら粒子と重みで近似する方法 [20] [31] がある。また、状態系列の事後分布を

$$p(\mathbf{x}_{1:N}|\mathbf{y}_{1:N}) = p(\mathbf{x}_N|\mathbf{y}_{1:N}) \prod_{k=1}^{N-1} p(\mathbf{x}_k|\mathbf{x}_{k+1:N}, \mathbf{y}_{1:N}) \quad (2.22)$$

と分解し、右辺積の各項が

$$p(\mathbf{x}_k|\mathbf{x}_{k+1:N}, \mathbf{y}_{1:N}) = p(\mathbf{x}_k|\mathbf{x}_{k+1}, \mathbf{y}_{1:k}) \propto p(\mathbf{x}_k|\mathbf{y}_{1:k}) f(\mathbf{x}_{k+1}|\mathbf{x}_k) \quad (2.23)$$

となることを用いて、比較的少ない計算量で平滑化分布に従う重み付き粒子を求める方法 [11] もある。

アプリケーションによっては、分布から特性値を求める必要がある。パーティクルフィルタにおいて得られた重み付き粒子から、分布の特性値を求める方法について若干述べておく。特性値の種類によってその方法は異なるが、求めやすいのは分布の期待値で、粒子の重み付き平均を計算すればよい。分散についても、重みを考慮した分散値を計算すればよい。これらの計算は、リサンプリング後であれば重みが均一になるので、もっと簡単な計算で済むが、モンテカルロ誤差がリサンプリング前よりも増す点に注意しよう。メジアンについては、状態ベクトルのうちメジアンを取りたい一変量について、粒子の持つ値を大きさの順に並べて、重みが均一の場合には中央に位置する粒子の値をとればよい。リサンプリング前であれば、正規化重みの累積和を計算してゆき、それが0.5になるときの変量の値を求めればよい。区間推定や確率計算についても、メジアンの計算と同様の原理で可能である。ただし、メジアンの計算には、整列や累積和といったやや複雑な計算を含むことに注意しよう。また、分布の裾の領域について確率を計算することは、その領域に粒子があまり存在しないので不向きである。MAP（事後確率最大）推定値を求めるには、例えばカーネル密度関数推定を用い、推定した密度関数の最大値を求めることで原理的には可能だが、計算量が非常に大きくなる点が問題である。これを粒子近似のレベルで効率よく求める方法も研究されている [10].

2.2 モンテカルロフィルタ

モンテカルロフィルタ [20] とは、パーティクルフィルタの一種で、事後分布に従う多数の粒子を用いて状態推定を行う方法の一つである。そのアルゴリズムは時間更新の形式で、カルマンフィルタと同様に 1 期先予測とフィルタリングの手続きを交互に、時刻を進めながら実行するものとなっている。モンテカルロフィルタはアルゴリズムが簡単であるため、パーティクルフィルタの中でもよく使われている。モンテカルロフィルタの分かり易い解説としては、[33] を参照されたい。

モンテカルロフィルタはパーティクルフィルタの特殊な場合になっており、プロポーザル分布としてシステムモデルを用いる。

$$q(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k) = f(\mathbf{x}_k | \mathbf{x}_{k-1}) \quad (2.24)$$

これより、重みの更新式 (2.14) では、分子のシステムモデルと分母のプロポーザルが同一となり消去される。また、リサンプリングを毎回行うので、前時刻の重みは粒子間で均等である。これらより、モンテカルロフィルタの重みの更新式は、前時刻の重みを用いずに、観測モデル h のみの簡潔な形となる。

モンテカルロフィルタのアルゴリズムは次の通りである。現在の時刻を k とし、1 時刻前 (時刻 $k-1$) のフィルタ分布の粒子群が次で与えられている。

$$\left\{ \mathbf{x}_{k-1}^{(i)} \right\}_{i=1}^M \sim p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) \quad (2.25)$$

まず、1 期先予測の粒子を、システムモデルに従い

$$\tilde{\mathbf{x}}_k^{(i)} \sim f(\mathbf{x}_k | \mathbf{x}_{k-1}^{(i)}), \quad i = 1, 2, \dots, M \quad (2.26)$$

と生成する。そして、現時刻の観測データ \mathbf{y}_k を観測モデルに代入して各粒子の尤度を計算し、これに比例する重み値を得る。

$$w_k^{(i)} \propto h(\mathbf{y}_k | \tilde{\mathbf{x}}_k^{(i)}), \quad i = 1, 2, \dots, M \quad (2.27)$$

ここで計算した重み値は、正規化されているものとする。すなわち、重みの総和は 1 である。最後に、重みに比例する確率で粒子を復元抽出するリサン

プリングを行う。適当な乱数系列を用いて

$$\mathbf{x}_k^{(i)} \sim \begin{cases} \tilde{\mathbf{x}}_k^{(1)} & \text{with prob. } w_k^{(1)} \\ \tilde{\mathbf{x}}_k^{(2)} & \text{with prob. } w_k^{(2)} \\ \vdots & \vdots \\ \tilde{\mathbf{x}}_k^{(M)} & \text{with prob. } w_k^{(M)} \end{cases} \quad i = 1, 2, \dots, M \quad (2.28)$$

と復元抽出を行い、これにより現時刻 k のフィルタ分布の粒子群

$$\{\mathbf{x}_k^{(i)}\}_{i=1}^M \sim p(\mathbf{x}_k | \mathbf{y}_{1:k}) \quad (2.29)$$

を得る。以上がモンテカルロフィルタの時間更新アルゴリズムである。これを図 2.1 に概念的に示す。

図 2.1 では、一時刻前 ($k-1$) のフィルタ分布に従う粒子群からスタートし、システムモデルに基づく一期先予測をまず行う (式 (2.26))。ここでは、典型的には、システムモデルが差分方程式

$$\mathbf{x}_k = \mathbf{F}(\mathbf{x}_{k-1}) + \mathbf{v}_k, \quad \mathbf{v}_k \sim N(\mathbf{0}, \mathbf{Q}) \quad (2.30)$$

で表され、決定論的な状態遷移 $F(\cdot)$ と、システムノイズ \mathbf{v}_k の確率論的な変動とが、各粒子に与えられる。次に、一期先予測の各粒子について尤度を計算し、これを粒子の重みとする (式 (2.27))。尤度関数は、観測したデータと観測モデルから与えられる。最後に、リサンプリングとして、重みに比例する確率で粒子の復元抽出を行い、現時刻 k のフィルタ分布に従う粒子群を得る。

モンテカルロフィルタの利点として、プロポーザル分布選択の必要がなく、アルゴリズムが簡単なことなどが挙げられる。重みの計算式が式 (2.27) のように観測モデル h だけから成り、システムモデル f の値を計算する必要がない。また、尤度の計算についても

$$p(\mathbf{y}_k | \mathbf{y}_{1:k-1}) \simeq \frac{1}{M} \sum_{i=1}^M h(\mathbf{y}_k | \tilde{\mathbf{x}}_k^{(i)}) \quad (2.31)$$

となり、重みの計算式 (2.27) の結果を単に用いるだけで済む。

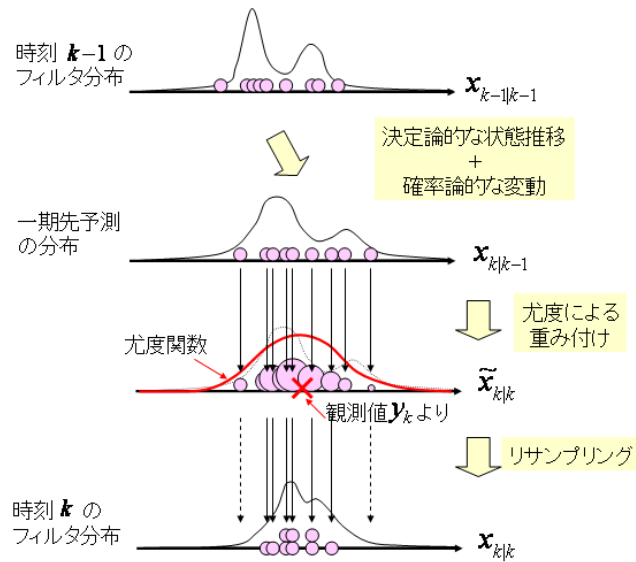


図 2.1: モンテカルロフィルタの概念図

システムモデルが対象を十分よく記述している場合には、式 (2.26) の 1 期先予測で生成される粒子がフィルタ分布の高確率領域に多数配置し、その結果モンテカルロフィルタはうまく動作する。しかし、システムモデルが対象とかけ離れている場合には問題が生じる。このような対象とモデルとの乖離は、実世界の課題を扱う際にはよく生じることである。システムモデルが対象と異なるため、1 期先予測で生成される粒子がフィルタ分布の高確率領域とは異なる領域に集中すると、リサンプリングの際には高確率領域に存在するごく少数の粒子が多数回復元抽出されてフィルタ分布を近似表現することになる。粒子のバリエーションが少ないことから、近似精度が悪く、場合によっては 1 点に縮退した分布となってしまう。このような状況では、粒子数を非常に多くしなければ、十分な近似性能を得られないことになる。このことを図 2.1 に概念的に示す。モンテカルロフィルタを用いる場合には、この点に注意する必要がある。

粒子生成 = 1期先予測

粒子の存在しない領域は、分布を表現できない

予測が正しい場合はOK. 予測が外れた場合に問題発生.

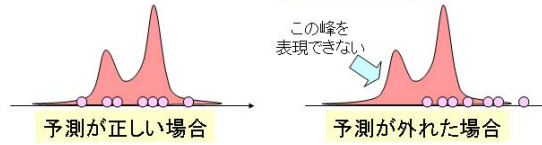


図 2.2: モンテカルロフィルタの問題点

3 ラオ - ブラックウェル化

パーティクルフィルタとは、逐次モンテカルロ法の最適フィルタ問題への適用で、いくつかの特殊形や改良を含むフィルタリング手法の総称であった。前節で述べたモンテカルロフィルタは、パーティクルフィルタの最も簡単な形の特特殊形であった。ここでは別の特殊形として、ラオブラックウェル化 [4] と呼ばれるテクニックについて説明する。

状態ベクトル \mathbf{x} の要素を二つの部分に分け

$$\mathbf{x} = \{\mathbf{z}, \theta\} \quad (3.1)$$

と書くことにしよう。これに対応して、目的分布 $p(\mathbf{x})$ も

$$p(\mathbf{x}) = p(\theta)p(\mathbf{z}|\theta) \quad (3.2)$$

と分解する。このとき、右辺において、第二項の分布が解析的に求められるならば、第一項の分布についてのみ粒子を用いて推定すれば十分である。右辺第二項の分布については、粒子を条件部分に代入し、解析的手続きにより求める。このようにすることで、粒子近似で扱う状態空間の次元を減らすことができ、近似の精度も良くなる。これがラオブラックウェル化の基本的な考え方である。

ラオブラックウェル化されたパーティクルフィルタ [7] では、解析的に求める分布（すなわち、式 (3.2) 右辺第二項の分布）は、典型的には、カルマ

ンフィルタで求めることになる。この状況は、 θ_k が与えられた下での線形ガウス状態空間モデル

$$\begin{cases} \mathbf{z}_k &= \mathbf{F}(\theta_k)\mathbf{z}_{k-1} + \mathbf{G}(\theta_k)\mathbf{v}_k, & \mathbf{v}_k &\sim N(\mathbf{0}, \mathbf{Q}_k) \\ \mathbf{y}_k &= \mathbf{H}(\theta_k)\mathbf{z}_k + \mathbf{w}_k, & \mathbf{w}_k &\sim N(\mathbf{0}, \mathbf{R}_k) \end{cases} \quad (3.3)$$

で表すことができる。ここで、 \mathbf{Q}_k や \mathbf{R}_k も θ_k に依存して構わない。一方、状態 $\mathbf{x}_k = \{\mathbf{z}_k, \theta_k\}$ のうち粒子近似で求める部分 θ_k については、システムモデル

$$\theta_k \sim f_\theta(\cdot | \theta_{k-1}) \quad (3.4)$$

に従うものとする。式 (3.3) の状態空間モデルは方程式のスタイルで記述されたものであるが、これを統計モデル（確率分布）のスタイルで記述すれば

$$\begin{cases} \mathbf{z}_k &\sim f_z(\cdot | \mathbf{z}_{k-1}, \theta_k) \\ \mathbf{y}_k &\sim h(\cdot | \mathbf{z}_k, \theta_k) \end{cases} \quad (3.5)$$

となる。これらのモデル式は、式 (1.6) や式 (1.7) と同様、

$$p(\theta_k | \mathbf{z}_{1:k-1}, \theta_{1:k-1}, \mathbf{y}_{1:k-1}) = f_\theta(\theta_k | \theta_{k-1}) \quad (3.6)$$

$$p(\mathbf{z}_k | \mathbf{z}_{1:k-1}, \theta_{1:k}, \mathbf{y}_{1:k-1}) = f_z(\mathbf{z}_k | \mathbf{z}_{k-1}, \theta_k) \quad (3.7)$$

$$p(\mathbf{y}_k | \mathbf{z}_{1:k}, \theta_{1:k}, \mathbf{y}_{1:k-1}) = h(\mathbf{y}_k | \mathbf{z}_k, \theta_k) \quad (3.8)$$

を仮定したことになる。

さて、状態のうち粒子近似で計算を行う部分 θ_k について、状態系列の事後分布を時間更新する逐次ベイズの式は、式 (1.8) と同様に

$$p(\theta_{1:k} | \mathbf{y}_{1:k}) = p(\theta_{1:k-1} | \mathbf{y}_{1:k-1}) \frac{f_\theta(\theta_k | \theta_{k-1}) p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \theta_{1:k})}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1})} \quad (3.9)$$

となる。ただし、右辺分子第二項の分布 $p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \theta_{1:k})$ が、式 (1.8) とは異なる点に注意する。これは尤度であり、状態 $\mathbf{x}_k = \{\mathbf{z}_k, \theta_k\}$ のうち \mathbf{z}_k をカルマンフィルタで求める際に得ることができる。

\mathbf{z}_k に対するカルマンフィルタの計算は、1期先予測とフィルタリングから成る。これらを確認率分布の式で表すと、1期先予測は

$$\begin{aligned} p(\mathbf{z}_k | \mathbf{y}_{1:k-1}, \theta_{1:k}) &= \int p(\mathbf{z}_{k-1} | \mathbf{y}_{1:k-1}, \theta_{1:k}) p(\mathbf{z}_k | \mathbf{z}_{k-1}, \theta_{1:k}, \mathbf{y}_{1:k-1}) d\mathbf{z}_{k-1} \\ &= \int p(\mathbf{z}_{k-1} | \mathbf{y}_{1:k-1}, \theta_{1:k-1}) f_z(\mathbf{z}_k | \mathbf{z}_{k-1}, \theta_k) d\mathbf{z}_{k-1} \end{aligned} \quad (3.10)$$

と書ける。ここで、二つめの等号では、式(3.7)と、関係

$$p(\mathbf{z}_{k-1} | \mathbf{y}_{1:k-1}, \theta_{1:k}) = p(\mathbf{z}_{k-1} | \mathbf{y}_{1:k-1}, \theta_{1:k-1}) \quad (3.11)$$

を用いた。この式は、式(3.6)を用いて導出することができる。一方、フィルタリングは

$$p(\mathbf{z}_k | \mathbf{y}_{1:k}, \theta_{1:k}) = \frac{p(\mathbf{z}_k | \mathbf{y}_{1:k-1}, \theta_{1:k}) h(\mathbf{y}_k | \mathbf{z}_k, \theta_k)}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \theta_{1:k})} \quad (3.12)$$

と書ける。この式の分母より、式(4.48)で用いる尤度が求まる。

重みの更新式は、時刻 k の粒子をプロポーザル分布 $q_\theta(\theta_k | \theta_{k-1}, \mathbf{y}_k)$ から生成するとき、式(2.9)の導出と同様にして、式(3.9)から

$$w_k(\theta_{1:k}) \propto w_{k-1}(\theta_{1:k-1}) \frac{f_\theta(\theta_k | \theta_{k-1}) p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \theta_{1:k})}{q_\theta(\theta_k | \theta_{k-1}, \mathbf{y}_k)} \quad (3.13)$$

と得られる。

以上の原理に従い、ラオブラックウェル化されたパーティクルフィルタのアルゴリズムをまとめると図3.1-3.2の通りになる。ここで、記号を明確にしておく。状態空間中の粒子は、パーティクルフィルタで計算を行う部分 $\theta_k^{(i)}$ と、カルマンフィルタで解析的に計算を行う部分 $\mathbf{z}_{k|k}^{(i)}$ とに分けられる。解析的に計算を行う部分 $\mathbf{z}_{k|k}^{(i)}$ は、 $\theta_k^{(i)}$ が与えられた下では正規分布に従うので、その平均ベクトル $\bar{\mathbf{z}}_{k|k}^{(i)}$ と分散共分散行列 $\mathbf{V}_{k|k}^{(i)}$ で表す。ここで添え字の $k|k$ は、縦棒 ' | ' の左が状態の時刻、右が観測時系列の最終時刻を表す。すなわち、 $k|k$ はフィルタ分布であることを表し、 $k|k-1$ は1期先予測であることを表す。これらの記号を使って粒子を表すと、時刻 k のフィルタ分布については次式となる。

$$\left(\theta_k^{(i)}, \bar{\mathbf{z}}_{k|k}^{(i)}, \mathbf{V}_{k|k}^{(i)} \right) \quad (3.14)$$

0. 時刻 $k-1$ のフィルタ分布の重み付き粒子群が与えられている.

$$\left\{ \left(\theta_{k-1}^{(i)}, \bar{\mathbf{z}}_{k-1|k-1}^{(i)}, \mathbf{V}_{k-1|k-1}^{(i)}, w_{k-1}^{(i)} \right) \right\}_{i=1}^M \quad (3.15)$$

1. プロポーザル分布から粒子を生成する.

$$\tilde{\theta}_k^{(i)} \sim q_{\theta}(\cdot | \theta_{k-1}^{(i)}, \mathbf{y}_k), \quad i = 1, 2, \dots, M \quad (3.16)$$

2. 粒子系列 $\tilde{\theta}_{1:k}^{(i)}$ が与えられた下でのカルマンフィルタの手続きを, 各粒子 ($i = 1, 2, \dots, M$) について以下の通り行う.

2-1 1期先予測

1期先予測 $p(\mathbf{z}_k | \mathbf{y}_{1:k-1}, \tilde{\theta}_{1:k}^{(i)})$ の平均と分散を求める. (\mathbf{A}' は行列 \mathbf{A} の転置を表す)

$$\begin{cases} \bar{\mathbf{z}}_{k|k-1}^{(i)} &= \mathbf{F}(\tilde{\theta}_k^{(i)}) \bar{\mathbf{z}}_{k-1|k-1}^{(i)} \\ \mathbf{V}_{k|k-1}^{(i)} &= \mathbf{F}(\tilde{\theta}_k^{(i)}) \mathbf{V}_{k-1|k-1}^{(i)} \mathbf{F}'(\tilde{\theta}_k^{(i)}) + \mathbf{G}(\tilde{\theta}_k^{(i)}) \mathbf{Q}_k \mathbf{G}'(\tilde{\theta}_k^{(i)}) \end{cases} \quad (3.17)$$

2-2 尤度

尤度 $p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \tilde{\theta}_{1:k}^{(i)})$ の平均と分散を求める.

$$\begin{cases} \bar{\mathbf{y}}_{k|k-1}^{(i)} &= \mathbf{H}(\tilde{\theta}_k^{(i)}) \bar{\mathbf{z}}_{k|k-1}^{(i)} \\ \boldsymbol{\Sigma}_{k|k-1}^{(i)} &= \mathbf{H}(\tilde{\theta}_k^{(i)}) \mathbf{V}_{k|k-1}^{(i)} \mathbf{H}'(\tilde{\theta}_k^{(i)}) + \mathbf{R}_k \end{cases} \quad (3.18)$$

2-3 フィルタ分布

フィルタ分布 $p(\mathbf{z}_k | \mathbf{y}_{1:k}, \tilde{\theta}_{1:k}^{(i)})$ の平均と分散を求める.

$$\begin{cases} \bar{\mathbf{z}}_{k|k}^{(i)} &= \bar{\mathbf{z}}_{k|k-1}^{(i)} + \mathbf{K}_k^{(i)} \left[\mathbf{y}_k - \bar{\mathbf{y}}_{k|k-1}^{(i)} \right] \\ \tilde{\mathbf{V}}_{k|k}^{(i)} &= \left[\mathbf{I} - \mathbf{K}_k^{(i)} \mathbf{H}(\tilde{\theta}_k^{(i)}) \right] \mathbf{V}_{k|k-1}^{(i)} \end{cases} \quad (3.19)$$

図 3.1: ラオーブラックウェル化パーティクルフィルタのアルゴリズム (前半)

ここで $\mathbf{K}_k^{(i)}$ は第 i 粒子のカルマンゲインである.

$$\mathbf{K}_k^{(i)} = \mathbf{V}_{k|k-1}^{(i)} \mathbf{H}'(\tilde{\theta}_k^{(i)}) \left[\boldsymbol{\Sigma}_{k|k-1}^{(i)} \right]^{-1} \quad (3.20)$$

3. 重みの更新を行う. $i = 1, 2, \dots, M$ について

$$\tilde{w}_k^{(i)} \propto w_{k-1}^{(i)} \frac{f_{\theta}(\tilde{\theta}_k^{(i)} | \theta_{k-1}^{(i)}) N(\mathbf{y}_k; \bar{\mathbf{y}}_{k|k-1}^{(i)}, \boldsymbol{\Sigma}_{k|k-1}^{(i)})}{q_{\theta}(\tilde{\theta}_k^{(i)} | \theta_{k-1}^{(i)}, \mathbf{y}_k)} \quad (3.21)$$

ここで $N(\mathbf{y}; \bar{\mathbf{y}}^{(i)}, \boldsymbol{\Sigma}^{(i)})$ は, 平均ベクトル $\bar{\mathbf{y}}^{(i)}$ 分散共分散行列 $\boldsymbol{\Sigma}^{(i)}$ の正規分布の確率密度関数に, \mathbf{y} を代入したものである.

4. リサンプリングを行う場合は, 粒子 $\tilde{\theta}_k^{(i)}$, 平均ベクトル $\tilde{\mathbf{z}}_{k|k}^{(i)}$, 分散共分散行列 $\tilde{\mathbf{V}}_{k|k}^{(i)}$ をまとめて扱い, これを重み $\tilde{w}_k^{(i)}$ に比例する確率で M 回復元抽出を行う. リサンプリングで抽出された粒子を $\theta_k^{(i)}$, 平均ベクトルを $\bar{\mathbf{z}}_{k|k}^{(i)}$, 分散共分散行列を $\mathbf{V}_{k|k}^{(i)}$ と表す. 重みは均等化し $w_k^{(i)} = 1/M$ とする.

リサンプリングを行わない場合は, 粒子と重みはそのままとする.

$$\begin{aligned} \theta_k^{(i)} &:= \tilde{\theta}_k^{(i)}, \bar{\mathbf{z}}_{k|k}^{(i)} := \tilde{\mathbf{z}}_{k|k}^{(i)}, \mathbf{V}_{k|k}^{(i)} := \tilde{\mathbf{V}}_{k|k}^{(i)}, \\ w_k^{(i)} &:= \tilde{w}_k^{(i)}, \quad i = 1, 2, \dots, M \end{aligned} \quad (3.22)$$

5. 以上の手続きにより, 時刻 k のフィルタ分布の重み付き粒子群が得られる.

$$\left\{ \left(\theta_k^{(i)}, \bar{\mathbf{z}}_{k|k}^{(i)}, \mathbf{V}_{k|k}^{(i)}, w_k^{(i)} \right) \right\}_{i=1}^M \quad (3.23)$$

図 3.2: ラオーブラックウェル化パーティクルフィルタのアルゴリズム (後半)

4 応用事例

パーティクルフィルタの特長を活かした応用事例をいくつか紹介する。各々の事例にて、用いている状態空間モデルと、その特徴や性質を簡単に解説する。なお、状態推定の方法については、原理的には、前節までで述べた各種方法から一つを適宜選んで用いればよい。粒子による事後分布の近似がうまく機能していれば、パーティクルフィルタの計算手続きに対して特に気を使う必要はない。最も簡単な特殊形であるモンテカルロフィルタを用いれば十分であることも多い。特別なプロポーザル分布を用いる場合には、そこに分析者の工夫を入れることができるが、その良し悪しによって推定性能も変わるので注意が必要である。紙面の都合上、各モデルを詳細に解説することはできないので、より詳しい内容については、個々の参考文献や、逐次モンテカルロ法について包括的書かれた書物 [8] などを参照されたい。

4.1 非線形モデル

パーティクルフィルタがその威力を発揮するのは、典型的には、システムの非線形性により状態の分岐が生じる場合や、観測過程に強い非線形性があり、それにより縮退等が生じる場合である。これらの場合のデモンストレーションとしてよく引用される非線形モデルが

$$\begin{cases} x_k = \frac{1}{2}x_{k-1} + \frac{25x_{k-1}}{1+(x_{k-1})^2} + 8 \cos(1.2k) + v_k, & v_k \sim N(0, \tau^2) \\ y_k = \frac{(x_k)^2}{20} + w_k, & w_k \sim N(0, \sigma^2) \end{cases} \quad (4.1)$$

である [1]。上の式がシステム方程式、下の式が観測方程式である。システム方程式は、状態 x_k の時間変化について非線形となっている。この式は、状態値が正または負の領域にてしばらくの間時間変化を繰り返し、時々、正の領域から負の領域へ（あるいはその逆へ）急変するという性質を持つ。一方、観測方程式は、状態 x_k の 2 乗から観測値を得ており、状態値の符号情報が欠落して観測される点が特徴的である。このため、1 時刻の観測値だけからは、

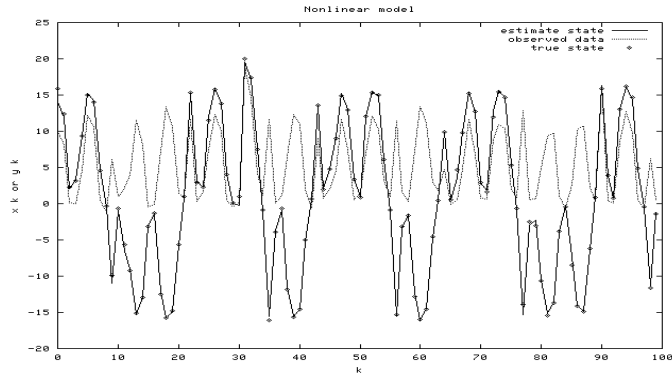


図 4.1: 非線形モデルの推定例

状態の正負は判らないようになっている。

現時刻の観測値が状態の符号情報を直接的に与えなくても、状態推定では観測時系列が与えられた下での状態の事後分布を求めるので、観測時系列を総合して考えることになり、状態の符号が判る。推定の際、状態には正および負の二つの可能性がある為、状態推定の分布には峰が二つあるものを想定することが必要である。カルマンフィルタは状態推定を正規分布で表すので、このような2峰の分布は扱えない。カルマンフィルタにおいて、状態の正負の分岐が生じるような推定方法を考えることもできるが、現時刻の正負2状態から次時刻の正負2状態への遷移は4通りの経路があり、遷移を繰り返すと経路の組み合わせ数は大きなものになってしまう。一方、パーティクルフィルタでは、このような組み合わせ的状况も、上述したアルゴリズムを単に実行するだけで粒子群が自動的に適応して配置され、適切に扱うことができる。

非線形モデルについて、コンピュータシミュレーションにより推定した結果を図 4.1 に示す。図にて、点線が観測値（データ）、実線が推定結果、○印が真の状態値である。推定結果をみると、観測値は常に正であるにも関わらず、真の状態値に近い推定値が得られているのが分かる。

4.2 非ガウスモデル

非ガウス分布を用いると、状態推定に多峰の分布が現れて、興味深い結果を得る場合がある [19]. 1 階のトレンドモデル

$$\begin{cases} x_k = x_{k-1} + v_k \\ y_k = x_k + w_k \end{cases} \quad (4.2)$$

において、もしシステムノイズ v_k と観測ノイズ w_k が共にガウス分布であれば、線形ガウス状態空間モデルとなりカルマンフィルタで状態推定が可能である。しかしどちらか一方（あるいは両方とも）を非ガウス分布にすると、状態推定は大きく異なる挙動を示すようになる。ここでは特に、コーシー分布などの裾の重い分布を用いる場合を説明する。

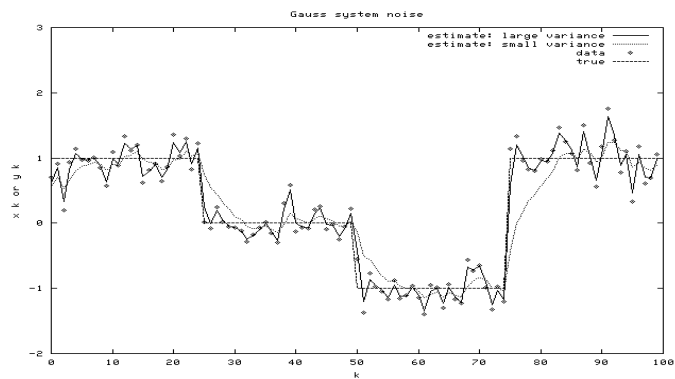
コーシー分布は単峰の確率密度関数を持ち、主として峰付近の値を取るが、稀に裾の値を取る。この特性は、状態推定にも影響を与える。具体的には、コーシー分布の峰と裾の二つの可能性を反映した、2 峰の分布が状態推定の結果として現れる。パーティクルフィルタによる状態推定では、現時刻の分布を 2 峰のまま推定し、どちらか一方に決めてしまわないようにする。ここでは、粒子集合が二つのクラスタを形成する。時間が経つにつれてデータが増えていくと、2 峰のうちどちらであったかが判るようになるので、その時点で 2 峰のうち一方が正解であったことが分かる。これに対しカルマンフィルタでは単峰の正規分布で推定するため、2 峰のうちどちらか一方に正規分布を配置する。選んだ峰がたまたま正解であればよいが、そうでない場合には時間を遡って推定をやり直さなければならない。あるいは、2 峰の中間のような無意味な領域に正規分布を配置する場合もあり、その結果、推定精度や追従性能が悪くなる。

以下、システムノイズおよび観測ノイズに、それぞれコーシー分布を用いた場合の推定例について述べる。まず、システムノイズにコーシー分布を用いると、状態の時間変化について前時刻とほぼ同じであるが稀にジャンプが生じる状況を表すことになる。この状況について、コンピュータシミュレーションにより推定を行った結果を図 4.2 に示す。図にて、真の状態は太線の点線で示され、時刻 $k = 25, 50, 75, 100$ において値の突発的な変化を持ち、そ

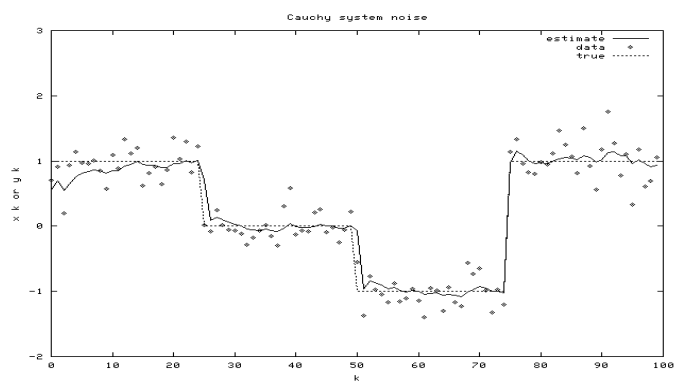
れ以外の時刻では一定値を取る。この真の状態値に、独立なガウス分布に従う観測ノイズを加えたものを観測値（データ）とし、図ではこれを○印で表した。システムノイズがガウス分布の場合と、コーシー分布の場合とについてそれぞれ推定を行い、結果を図中の (a) および (b) に示した。ガウス分布の場合には、分散が大きい時と小さい時について推定を行った。それぞれの結果は、図 (a) 中の実線および点線である。

図 (a) を見ると、分散の大きなガウス分布をシステムノイズに用いると、状態値の突発的な変化に追従しやすくなる反面、値の変化しないところでは観測値に追従するような推定値となり、安定した推定が行えていない。一方、分散の小さなガウス分布を用いた場合には、状態の値が変化しないところでは安定した推定が行えるものの、突発的な変化への追従が遅くなる。これに対し、コーシー分布をシステムノイズに用いた場合には、図 (b) の実線で示した推定値のように、安定した推定と突発的な変化への速い追従の、両方の性質が得られているのが分かる。

次に、観測ノイズにコーシー分布を用いると、主として真値近傍の値が観測ノイズを伴い観測されるが、稀に外れ値が生じる状況を表すことになる。この状況について、コンピュータシミュレーションにより推定を行った結果を図 4.3 に示す。ここでは正弦波のトレンドを持つ状態値が、ガウス分布の観測ノイズを伴って観測され、稀に外れ値が混入する状況を扱った。図では、外れ値が、時刻 $k = 13, 38, 63, 88$ において生じている。観測ノイズにガウス分布を用いる場合と、コーシー分布を用いる場合とについて推定を行った。推定結果を図中の (a) および (b) にそれぞれ示す。ガウス分布を用いる場合には、その分散が大きい時と小さい時について推定した。分散が大きい時には観測値をあまり信用しないような推定結果が得られ、その為外れ値に対しては頑健になるものの、外れ値でないデータに対しての追従性は悪くなる。一方、分散が小さい時には、データに対する追従性は確保されるが、外れ値に対しても追従してしまい、推定値が外れ値に引きずられるような結果となる。これに対し、コーシー分布を観測ノイズに用いた場合が図 4.3(b) である。図を見ると、追従性と、外れ値に対する頑健性の両方が同時に実現されているのが分かる。

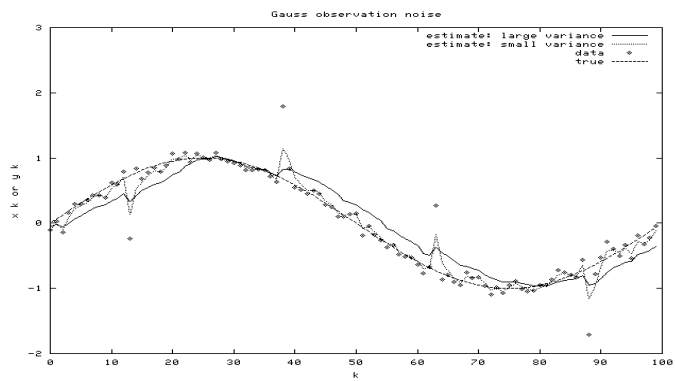


(a) システムノイズにガウス分布を用いた場合

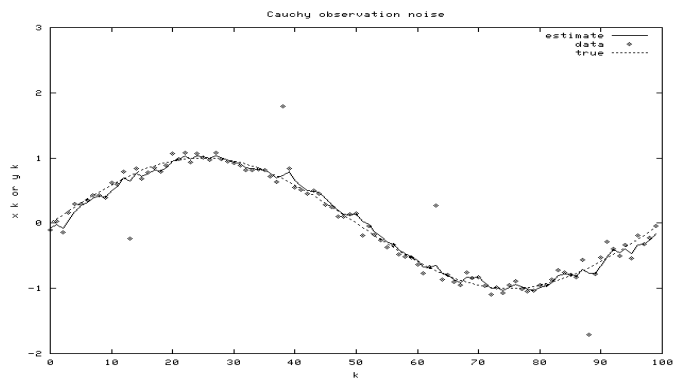


(b) システムノイズにコーシー分布を用いた場合

図 4.2: 稀にジャンプが生じるデータの推定例



(a) 観測ノイズにガウス分布を用いた場合



(b) 観測ノイズにコーシー分布を用いた場合

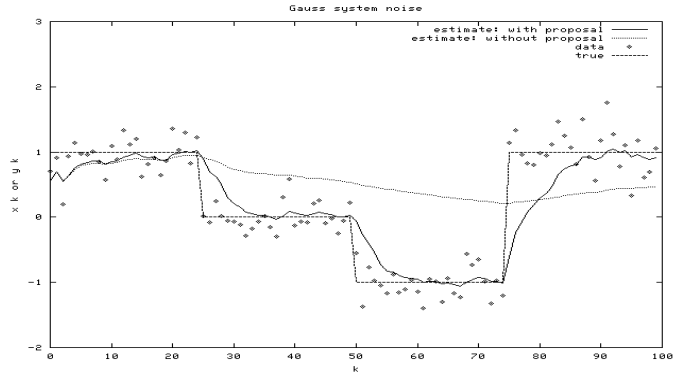
図 4.3: 外れ値を含むデータの推定例

さて、プロポーザル分布の設計例とその効果を示すために、システムノイズにコーシー分布を用いる場合を例に挙げて説明しよう。図 4.2 に示した推定よりも、さらにシステムノイズの分散を小さくしてみる。コーシー分布については、分布の広がりパラメータを小さくする。そうすると、システムモデルで生成される粒子は前の時刻とほぼ同じ位置のものが大半となり、時刻 $k = 25, 50, 75, 100$ における値の突発的な変化に追従できるような粒子は生成されないことになる。

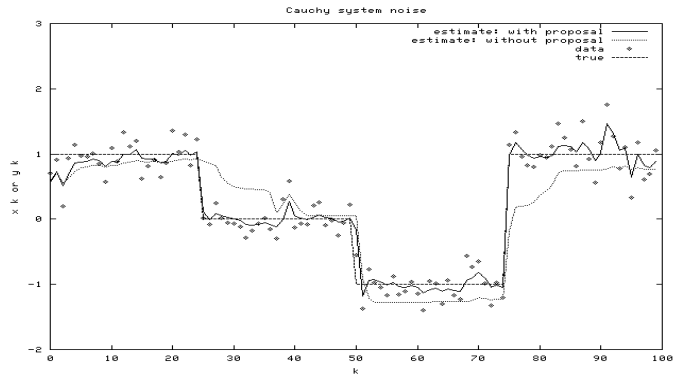
これに対し、プロポーザル分布に、観測データ y_k を中心とするガウス分布を補助的に用いると、推定結果が改善される。プロポーザル分布としては、2つのコンポーネントを持つ混合分布を使い、0.2の確率で観測データを中心とするガウス分布を、残りの確率でシステムモデルを用いた。観測データを中心とするガウス分布は、ある程度大きな分散値を持つものとする。図4.4に推定の結果を示す。実線がプロポーザル分布を用いた場合、破線が用いない場合である。図 4.2 と同様に、データは○印で示し、真値も破線で示してある。

システムノイズにガウス分布を用いた場合はカルマンフィルタでも推定可能であるが、ここではパーティクルフィルタにより推定を行った。システムノイズの分散が小さいため、破線の推定結果のように、時刻 $k = 25$ での最初の突発的な変化に追従できずに、粒子群が真値を見失ってさまよっているのがわかる。これに対し、プロポーザル分布を使った場合が実線の推定結果である。プロポーザル分布により粒子が真値の付近にも生成されるため、分散の小さなシステムノイズを用いた推定に特有の遅い追従を粒子により正しく表している。

システムノイズにコーシー分布を用いた場合についても、同様のプロポーザル分布を使って推定を行った。プロポーザル分布を使わない推定結果が破線であるが、時刻 $k = 25, 50, 75, 100$ において値が突発的に変化しているのに追従できない場合がある。また、真値が一定の箇所については、一旦生じたバイアスがなかなか回復しない現象も見られる。一方、プロポーザル分布を使えば、実線で示したような真値付近の推定値が正しく得られる。



(a) システムノイズにガウス分布を用いた場合



(b) システムノイズにコーシー分布を用いた場合

図 4.4: 小さなシステムノイズでの推定：プロポーザル分布に観測データを中心とするガウス分布とシステムモデルの混合分布を用いた場合

4.3 ターゲット・トラッキング

レーダー等の観測に基づく飛行機や宇宙船等のターゲット追跡は、動的システムにおける最適フィルタリングの古典的かつ代表的な問題である [27]. カルマンフィルタが提案された時期にはちょうどアポロ計画が進行中であり、月へ航行する宇宙船の位置速度を可能な限り精確に求めるためにカルマンフィルタが用いられた [13]. これには、遠方に位置する宇宙船の位置が地上からでは正確には観測できないことと、月軌道投入や大気圏再突入の際に失敗しない為に正確な情報が必要なことが大きな要因としてある. 宇宙船の軌道に関する方程式は一般に非線形であるが、これを線形化して用いる拡張カルマンフィルタが考案され使われた. どちらのフィルタも状態推定にはガウス分布が仮定されており、ここがパーティクルフィルタとは大きく異なる点である.

ターゲット・トラッキングの概要を示すため、簡単な2次元での追跡モデルを例に挙げよう. 追跡対象の2次元位置を \mathbf{p}_k , 速度を \mathbf{s}_k , 加速度を \mathbf{a}_k と表す. 追跡対象のダイナミクスに関して特に先験的知識がない場合を考えることにする. この場合は、典型的なシステムモデルとして、確率差分方程式

$$\begin{bmatrix} \mathbf{p}_k \\ \mathbf{s}_k \\ \mathbf{a}_k \end{bmatrix} = \begin{bmatrix} \mathbf{I} & T\mathbf{I} & T^2\mathbf{I} \\ \mathbf{0} & \mathbf{I} & T\mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{p}_{k-1} \\ \mathbf{s}_{k-1} \\ \mathbf{a}_{k-1} \end{bmatrix} + \begin{bmatrix} T^3\mathbf{I} \\ T^2\mathbf{I} \\ T\mathbf{I} \end{bmatrix} \mathbf{v}_k, \quad \mathbf{v}_k \sim N(\mathbf{0}, \mathbf{Q}) \quad (4.3)$$

が用いられる. ここで \mathbf{I} および $\mathbf{0}$ は、 2×2 の単位行列および零行列である. また T は、連続時間システムを離散時間化した際のサンプリングタイムである. 離散時間化する前の連続時間システムは、確率微分方程式

$$\frac{d}{dt} \begin{bmatrix} \mathbf{p}(t) \\ \mathbf{s}(t) \\ \mathbf{a}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{p}(t) \\ \mathbf{s}(t) \\ \mathbf{a}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{I} \end{bmatrix} \mathbf{v}(t), \quad \mathbf{v}(t) \sim N(\mathbf{0}, \mathbf{Q}) \quad (4.4)$$

である.

システムモデルの意味するところは、式 (4.4) の連続時間モデルから解釈すると分かりやすい. 式中の状態ベクトルの要素に着目すると、加速度 $\mathbf{a}(t)$ の微分 (変化量) が白色ガウス過程 $\mathbf{v}(t)$ に従っており、これがモデルの本質

部分である。その他の要素、速度 $s(t)$ および位置 $p(t)$ は、それぞれの微分が、微分階数の一つ高い加速度および速度に等しくなることを表しているに過ぎない。よってこのシステムモデルは、加速度の変化に対して時間的滑らかさを想定した式となっている。なお、加速度は位置の2階微分であるが、滑らかさを想定する要素の微分階数 d を変更することで、ターゲット位置の時間変化の滑らかさを調節することができる。より高い微分階数の要素を用いれば、より滑らかな時間変化を仮定することになる [18], [21]。

観測モデルでは、まず観測者の位置を原点としよう。ターゲットの位置 $\mathbf{p}_k = [p_k^x, p_k^y]'$ のみが観測され、それに観測ノイズが加わるモデルでもよいが、より実際的には、レーダー等による観測を想定して、極座標の観測方程式

$$\begin{bmatrix} r_k \\ \theta_k \end{bmatrix} = \begin{bmatrix} \sqrt{(p_k^x)^2 + (p_k^y)^2} \\ \text{atan}(p_k^y, p_k^x) \end{bmatrix} + \begin{bmatrix} w_k^r \\ w_k^\theta \end{bmatrix}, \quad \begin{bmatrix} w_k^r \\ w_k^\theta \end{bmatrix} \sim N(\mathbf{0}, \mathbf{R}) \quad (4.5)$$

が用いられる。ここで $\text{atan}(y, x)$ は $\tan^{-1}(y/x)$ を適宜計算する関数である。

拡張カルマンフィルタでは、式 (4.5) の非線形な観測過程を線形化した上でガウス観測ノイズを加えたものを観測モデルとし、推定にはカルマンフィルタを用いる。しかし、このようにすると、観測ノイズの分布が元の式とは異なった形状になることに注意しよう。元の式 (4.5) では真値を通る円周上の三日月形のような分布になるのに対し、拡張カルマンフィルタでは真値の周りの楕円になる。よって、精度良く追跡を行おうとするならば、式 (4.5) を線形化せずそのまま用いる必要があり、ここでパーティクルフィルタが役に立つことになる。

更には、式 (4.5) の観測モデルにて、角度 θ_k のみが観測され距離 r_k は観測されない場合も、パーティクルフィルタにより適切に扱うことができる [24]。また、システムモデルにおいて滑らかさを想定する要素の微分階数 d についても、実際の状況に応じて適応的に定めることができる。微分階数を時間変化するものとして d_k と表し、これを状態ベクトルに加えて状態推定を行う。もし観測モデルが線形であれば、この場合にはラオーブラックウェル化されたパーティクルフィルタを使うと効果的である [16]。これらの内容の詳細については、それぞれの文献を参照されたい。

4.4 動画像におけるビジュアル・トラッキング

コンピュータやデジタル機器の性能向上と低価格化に伴い、デジタル画像を扱う高性能な装置が容易に使えるようになってきた。静止画像の処理のみに留まらず、動画像をリアルタイムで処理し、カメラで撮影した外界対象物の動的情報を得ることができるようになってきている。特に、画像中の対象物を把握し追跡すること、すなわちビジュアルトラッキングは、動画像処理における主要な研究課題の一つである。

パーティクルフィルタを用いたビジュアルトラッキングとしては、CONDENSATION (Conditional density propagation) [17] が代表的である。CONDENSATION では、画像から粒子の尤度を計算するプロセスが工夫されている。時刻 k の画像を \mathbf{I}_k と表すとき、これを単純に観測データとみなすと、尤度は $h(\mathbf{I}_k|\mathbf{x}_k)$ となる。画像 \mathbf{I}_k は一般に多数の要素から成る 2次元配列であるが、これを 1次元に並べ直して確率ベクトルとみなすと、尤度を求めるには非常に高い次元の空間での確率分布を扱うことになる。しかし動画像における追跡では、画像全体の情報が必要になることは稀であり、追跡対象の周辺部分の画像情報があれば十分であることが多い。このことを考慮して、CONDENSATION では尤度計算が工夫されている。その概念図を図 4.10 に示す。以下、この図に従い説明をする。

追跡対象の典型的形状は既知とし、これを「テンプレート」として予め与えておく。テンプレートはいくつかの制御点 S_j で表され、それらの座標が与えられる。テンプレートの形状は、これら制御点を定められた順に通るスプライン曲線と与えられる。追跡の際には、テンプレートをアフィン変換して画像フレーム（動画像中の 1 枚の画像）にあてはめる。アフィン変換とは平行移動 (x, y) 、回転 θ 、拡大縮小 s を与えるもので、これらを担うパラメータ (x, y, θ, s) で表される。アフィン変換のパラメータを粒子の値 $\mathbf{x}_k^{(i)} = (x_k^{(i)}, y_k^{(i)}, \theta_k^{(i)}, s_k^{(i)})$ とし、パーティクルフィルタで状態推定してフィルタ分布 $p(\mathbf{x}_k|\mathbf{I}_{1:k})$ を得ることで追跡が行われる。追跡に先立って、システムモデル $f(\mathbf{x}_k|\mathbf{x}_{k-1})$ と観測モデル $h(\mathbf{I}_k|\mathbf{x}_k)$ を定めておく必要がある。

観測モデル $h(\mathbf{I}_k|\mathbf{x}_k)$ は、各粒子の尤度を与える。各粒子の尤度 $h(\mathbf{I}_k|\mathbf{x}_k^{(i)})$ としては、粒子の持つアフィン変換パラメータ $\mathbf{x}_k^{(i)}$ によりテンプレートを画

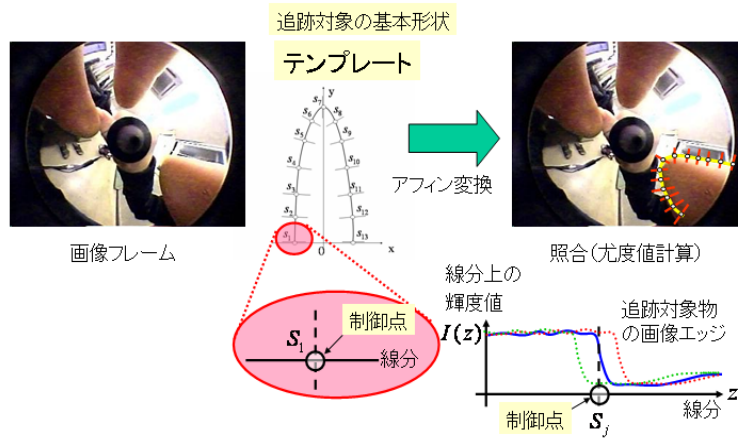


図 4.10: CONDENSATION における尤度計算の概念図

像にあてはめたときの一致度合いが用いられる。この一致度合いは、テンプレートの各制御点を通りテンプレート形状に対し垂直な線分を考え、その線分上の画像エッジから計算する。線分上の画素値から画像エッジの位置を計算し、これがテンプレートで想定している対象物のエッジ位置（すなわち制御点の位置）に近ければ、尤度は高い値となるようにする。より具体的には、制御点からの距離 z について、ガウス分布 $N(0, \sigma^2)$ を用いる。テンプレートは複数の線分を持つので、全ての線分についての同時分布を考え、これを全体の尤度として用いる。ただし、実際のビジュアルトラッキングでは、画像エッジ位置が欠落したり、外れ値を取ったりすることも多い。そのような状況でも追跡を円滑に進めるために、線分上での尤度計算では、画像エッジ位置が線分長さを超えた場合には尤度はそれ以上悪い値を取らないようにすることが行われる。

システムモデル $f(\mathbf{x}_k | \mathbf{x}_{k-1})$ により、追跡対象の動きに関する先験的知識を記述する。もし対象物の動きに関して何の先験的知識もなければ、対象物は滑らかに動くという想定を行う。これには、ターゲット追跡の場合と同様

な、ランダムウォークモデル，すなわち1階の差分方程式

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{v}_k, \quad \mathbf{v}_k \sim N(\mathbf{0}, \mathbf{Q}) \quad (4.6)$$

がよく用いられる．あるいは2階の差分方程式を用いることで，より滑らかな動きを想定することもできる．追跡対象のダイナミクスが既知であったり，あるいは他の入力情報が得られそれを効果的に使える場合には，それらを反映したシステムモデルを構築してより正確な追跡を行うことができる．

関連図書

- [1] Andrede Netto, M.L., Gimeno, L., and Mendes, M.J. (1978). "On the Optimal and Suboptimal Nonlinear Filtering Problem for Discrete-Time Systems", *IEEE Trans. on Automatic Control*, Vol.23, 1062–1067.
- [2] Akashi, H., Kumamoto, H., and Nose, K. (1975). "Application of Monte Carlo method to optimal control for linear systems under measurement noise with Markov dependent statistical property", *International Journal of Control*, Vol.22, No.6, 821–836.
- [3] Akashi, H. and Kumamoto, H. (1977). "Random Sampling Approach to State Estimation in Switching Environments", *Automatica*, Vol.13, 429-434.
- [4] Casella, G. and Robert, C.P. (1996). "Rao-Blackwellisation of sampling schemes", *Biometrika*, Vol.83, No.1, 81–94.
- [5] Clapp, T.C., and Godsill, S.J., (1999). "Fixed-Lag Smoothing Using Sequential Importance Sampling", *Bayesian Statistics*, Vol.6, 743–752.
- [6] Doucet, A., Godsill, S.J., and Andrieu, C. (2000). "On sequential Monte Carlo sampling methods for Bayesian filtering", *Statistics and Computing*, Vol.10, Nov.3, 197–208.

- [7] Doucet, A., de Freitas, N., Murphy, K., and Russell, S. (2000). "Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks", *Uncertainty in AI*.
- [8] Doucet, A., de Freitas, N., Gordon, N.J. (eds) (2001). *Sequential Monte Carlo Methods in Practice*, New York, Springer.
- [9] Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*, London, Chapman & Hall/CRC.
- [10] Godsill, S.J., West, M., and Doucet, A. (2001). "Maximum a Posteriori Sequence Estimation using Monte Carlo Particle Filters", *Annals of the Institute of Statistical Mathematics*, Vol. 53, No. 1, 82–96.
- [11] Godsill, S.J., West, M., and Doucet, A. (2004). "Monte Carlo Smoothing for Nonlinear Time Series", *Journal of the American Statistical Association*, Vol. 99, No. 465, 156–168.
- [12] Gordon, N.J., Salmond, D.J., and Smith, A.F.M. (1993). "Novel approach to nonlinear/non-Gaussian Bayesian state estimation", *IEE Proceedings-F*, Vol.140, No.2, 107-113.
- [13] Grewal, M.S. and Andrews, A.P. (2001). *Kalman Filtering, Theory and Practice using MATLAB*, New York, Wiley-Interscience Publication, John Wiley & Sons, Inc.
- [14] Handschin, J.E. and Mayne, D.Q. (1969). "Monte Carlo Techniques to Estimate the Conditional Expectation in Multi-stage Non-linear Filtering", *International Journal of Control*, Vol.9, 547–559.
- [15] Handschin, J.E. (1970). "Monte Carlo Techniques for Prediction and Filtering of Nonlinear Stochastic Process", *Automatica*, Vol.6, 555–563.
- [16] Ikoma, N., Higuchi, T., and Maeda, H. (2002). "Tracking of maneuvering target by using switching structure and heavy-tailed distribution

- with particle filter method”, *Proc.of the 2002 IEEE Conference on Control Applications*, 1283–1287.
- [17] Isard, M., and Blake, A. (1998). ”CONDENSATION – Conditional Density Propagation for Visual Tracking”, *International Journal of Computer Vision*, Vol.29, No.1, 5–28.
- [18] Kitagawa, G. and Gersch, W. (1984). ”A smoothness prior state space approach to the modeling of time series with trend and seasonality,” *Journal of the American Statistical Association*, vol.79, no.386, 378–389.
- [19] Kitagawa,G., (1987). ”Non-Gaussian State-Space Modeling of Nonstationary Time Series”, *Journal of the American Statistical Association*, Vol.82, No.400, 1032-1041.
- [20] Kitagawa,G., (1996). ”Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models”, *Journal of Computational and Graphical Statistics*, Vol.5, No.1, 1–25.
- [21] Kitagawa,G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, New York, Springer.
- [22] Liu, J.S. and Chen, R. (1998). ”Sequential Monte Carlo methods for dynamic systems”, *Journal of the American Statistical Association*, Vol.93, No.443, 1032–1044.
- [23] Liu, J.S. (2001). *Monte Carlo Strategies in Scientific Computing*, New York, Springer.
- [24] McGinnity, S. and Irwin, G.W. (2000). ”Multiple model bootstrap filter for manoeuvring target tracking”, *IEEE Trans. on Aerospace and Electronic Systems*, Vol.36, No.3, 1006–1012.

- [25] Moral,P.D. (2004). *Feynman-Kac Formulae, Genealogical and Interacting Particle Systems with Applications*, New York, Springer.
- [26] Robert, C.P. and Casella, G. (1999). *Monte Carlo Statistical Methods* New York, Springer.
- [27] Singer, R.A. (1970). "Estimating Optimal Tracking Filter Performance for Manned Maneuvering Targets", *IEEE Trans. on Aerospace and Electronic Systems*, Vol.AES-6, No.4, 473-483.
- [28] Yoshimura, T. and Soeda, T. (1972). "The Application of Monte Carlo Methods to the Nonlinear Filtering Problem", *IEEE Trans. on Automatic Control*, 681-684.
- [29] 明石, 熊本, (1975). "分散減少法を導入したモンテカルロ法による離散時間非線形フィルタの構成", *システムと制御*, Vol.19, No.4, 211-222.
- [30] 伊庭, (2005). 逐次モンテカルロ法入門, 計算統計 マルコフ連鎖モンテカルロ法とその周辺, 岩波書店.
- [31] 北川, (1996). "遺伝的アルゴリズムとモンテカルロフィルタ", *統計数理*, 第44巻, 第1号, 19-30.
- [32] 辻, 中村, (1973). "統計処理の手法による非線形フィルタ", *電気学会論文誌*, Vol.93-C, No.5, 109-116.
- [33] 樋口, (2005). 粒子フィルタ, *電子情報通信学会誌*, Vol.88, No.12, 989-994.