科研費シンポジウム
「不偏性をはずした時系列推定と因果性の観光統計学への応用」報告書

開催責任者：高木 祥司（奈良教育大学）
天野 友之（和歌山大学）
日時：2014年9月1日（月）～9月3日（水）
場所：奈良教育大学　講義１号棟　102講義室

本シンポジウムは
科学研究費・基盤研究(A)「非対称・非線形統計理論と経済・生体科学への応用」
研究代表者：谷口正信（早稲田大学基幹理工学部）
の助成により開催された。

プログラム

＜９月１日（月）＞

＜セッション１　座長：天野友之（和歌山大学）＞

[14:20-15:00]
村上享（松江工業高等専門学校）
「中海・宍道湖・大山圏域における観光動向調査に関する分析」
[15:00-15:40]
新村秀一（成蹊大学）
「DEAによる47都道府県の観光業に関する分析」
[15:40-16:20]
柳本武美（総合研究大学院大学）
「初期ヤマト政権の地政要因ー時空間統計学の視点から」
[16:20-17:00]
持元江津子（聖泉大学、大阪産業大学、天理大学）
「現役大学生の基本３グラフ作成スキルの現状と問題点：３大学150名を事例として」

＜９月２日（火）＞

＜セッション２　座長：西山陽一（統計数理研究所）＞

[10:00-10:40]

佃康司（総合研究大学院大学）

「The Ewens sampling formula および random mappings に対する新たな汎関数中心極限定理」

[10:40-11:20]

今田恒久（東海大学熊本校舎）

「順序制約下の正規母平均列間の差に関する多重比較法」

[11:20-12:00]

筑瀬靖子（香川大学）

「Distribution Theory Considered on the Three Statistical Manifolds」

＜セッション３　座長：筑瀬靖子（香川大学）＞

[13:30-14:10]

H. C. Jimbo（早稲田大学）、

M.J.Craven（University of Plymouth）

「A Novel Approach to Optimal Portfolio Selection with Loss Risk Constraints」

[14:10-14:50]

西山陽一（統計数理研究所）

「A stochastic maximal inequality, strict countability, and related topics」

[14:50-15:30]

柿沢佳秀（北海道大学）

「Bootstrap-based Bartlett-type adjustment」

＜特別セッション　座長：谷口正信（早稲田大学）＞

[16:00-17:00（特別講演）]

Yoon-Jae Whang（Seoul National University）

「A Nonparametric Test of a Strong Leverage Hypothesis」

＜９月３日（水）＞

＜セッション４　座長：柿沢佳秀（北海道大学）＞

[10:00-10:40]

清水優祐（九州大学大学院）

「正則化推定におけるモーメント収束」

[10:40-11:20]

明石郁哉（早稲田大学大学院）

「Higher-order asymptotic properties of frequency domain GMM estimators」

[11:20-12:00]

劉　言（早稲田大学大学院）

「Robust estimation of frequencies」

＜セッション５　座長：高木祥司（奈良教育大学）＞

[13:30-14:10]

須藤慶大（早稲田大学大学院）、

劉言（早稲田大学大学院）、

谷口正信（早稲田大学）

「補間誤差をコントラスト関数とする母数推定」

[14:10-14:50]

田中秀和（大阪府立大学）、

Nabendu Pal (University of Louisiana at Lafayette)、

Wooi K. Lim (William Paterson University)

「On Improved Estimation of Gamma Parameters」

[14:50-15:30]

阿部俊弘（東京理科大学）、

藤澤洋徳（統計数理研究所）

「最頻値の位置を変えない非対称分布族の生成」

中海・宍道湖・大山圏域における観光動向調査に関する分析

―観光統計を活用した地方の観光構造に関する空間分析を中心として―

松江工業高等専門学校　村上享

1. はじめに

観光に関する客観的情報は，有効な観光施策の実施，意思決定のためにも非常に重要な要素である．しかし，多くの観光統計は日本全体の動向調査もしくは県単位の個別・独自調査に限られるため，空間的に統一された観光統計は整備されていないのが現状である．特に，地方部では，自動車を交通手段とする周遊観光が中心となることから，地域間の空間的移動特性に関する情報は観光圏の有機的連携のためにも重要な要素となってくる．また，このような空間的移動は道路ネットワークなどの社会資本の整備状況と密接に関わってくることから，単に一時点の社会環境下の観光行動に着目した情報だけでなく，社会資本などの外部環境が変化した場合の影響についても把握しておくことが重要であるといえる．

そこで，今回，既存の観光統計では分析が難しい狭域的な地域として鳥取県と島根県を跨ぐ越境観光圏を対象に，新たな観光周遊調査を行い観光特性に関する空間的分析を行った．これは，空間を連続的に移動する観光周遊行動を客観的に把握するためには，行政区域単位などの空間が断絶された観光情報ではなく，地域全体の空間的連続性を考慮した観光情報の整備が重要となるためである．この調査結果により，山陰地方の宍道湖・中海周辺地域（越境観光圏域）を対象に、観光周遊性や観光消費の関係性等に着目した分析を行い，空間的観光情報の整備重要性と得られた知見を示す．

2. 観光行動特性

1)県境越えの観光流動の実態

地域間の流動を流動率（ある地域に訪問した観光客数のうち他地域に訪問した観光客数の割合と定義する）を用いて分析する．日帰り客（図１上）では，松江-出雲，米子-境港間の相互流動が多いが，全体として地域間相互の動きは少ない．一方，宿泊客（図１下）では，地域間相互の流動が活発化している．宿泊客の流動から県境の動きに着目すると，鳥取側の境港や米子

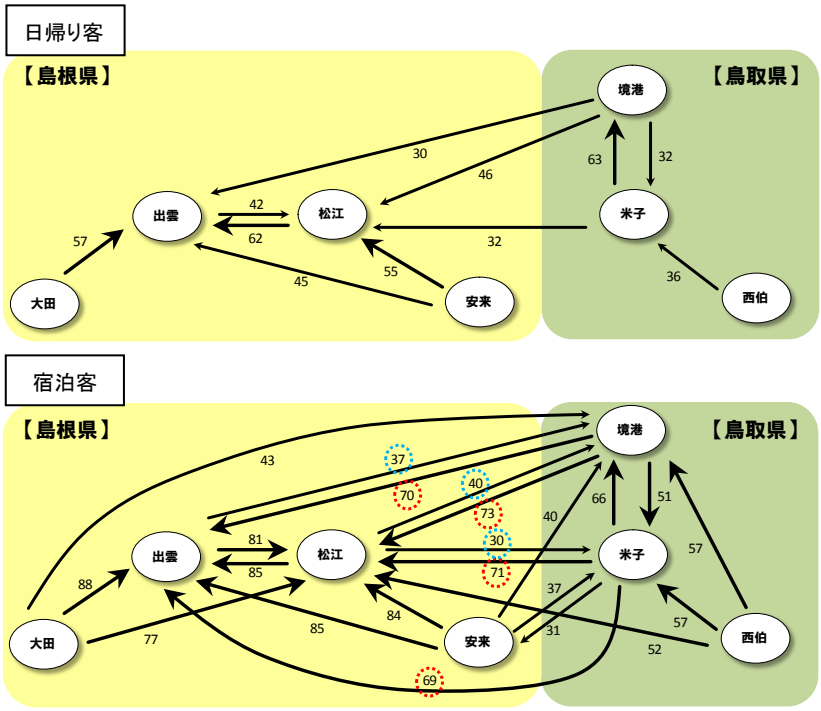

図１　地域間流動（数値は流動率 30%以上を表示）　(単位:%)

に来訪した人の約 7 割は島根側の松江, 出雲に訪問している. 逆に松江や出雲に来訪した人の内, 鳥取側の境港や米子に訪問した割合は 3～4 割と相対的に少ない実態が浮き彫りとなった. 従って, 滞在型観光の促進にあわせて、島根県側の観光地から鳥取県側への誘客促進も当圏域の活性化には重要なポイントと考えられる.

２)観光地別の観光客特性

島根側から鳥取側への流動が少ない構造について，その要因を探るため，島根側と鳥取側の観光客特性の違いについて分析を行った．具体的には，コレスポンデンス分析を用いて，圏域内の主要観光地を対象に, 年齢別, 居住地別に観光客特性の分析を行った. コレスポンデンス分析は, 複数の観光地間の来訪パターンの類似度や関係の深さを調べる手法で，その結果は２次元マップでグラフィカルに表現できる特徴がある. また, 各属性もマップに同時プロットすることができ, 属性別の特徴も把握することができる．結果は紙面の都合上省略し，考察結果を示す.

鳥取・島根の各県の来訪者属性の違いをみると，島根県内の施設は，松江市内や出雲大社などが原点付近にプロットされている一方で, 原点から離れている施設は, 全体的に第 1～3 象限にプロットされる傾向が多いことから，相対的に若年来訪者もしくは遠方地域の高齢来訪者が多い傾向がうかがえた．一方で鳥取県内の施設をみると，水木しげるロードと鳥取砂丘以外の観光地の多くは，第４象限に多くプロットされており，周辺地域の高齢来訪者が多い傾向がうかがえる．このように鳥取県と島根県で観光施設の来訪者特性が異なることは，各県がそれぞれニーズの異なる観光施設を提供していることを示している一方で，上述の１）の結果をふまえると，日帰り宿泊客別に以下の点について考察することが出来る.

日帰り観光は, 観光周遊時間が限定的であり立ち寄り観光施設がニーズによって絞られる一方, 島根・鳥取両県でニーズの異なる観光地を提供していることを考えると，現況の観光サービスの提供状況では有効な連携を得ることが難しいといえる．そのため，越境観光圏としての一体的な連携施策を検討する際の一つの視点としては，例えば島根県側には周辺地域の高齢来訪者にターゲットを絞った観光地づくりを支援する一方で，鳥取県側には若年来訪者もしくは遠方地域の高齢来訪者でターゲットを絞った観光地づくりを行うことが，観光圏域内の周遊性・一体性を高めるキッカケになるものと考えられる.

宿泊観光は，流動ベースでは鳥取県側から島根県側への流動が多い一方で，逆の流動は相対的に少ない傾向にある．これは，島根県が鳥取県に比べてある程度多様な観光ニーズに対応したサービスを提供する施設が立地する一方で，鳥取県側はニーズが限定的であり観光サービスの多様性が少ないことが要因として考えられる．図 3-6 でみると，近年来訪客が増大している境港市の水木しげるロード（鳥取県）は，地理的に近い皆生温泉（鳥取県）よりも玉造温泉（島根県）の方が近くにプロットされている．また，境港の魚市場も同一市にあるが離れてプロットされている．この結果から，集客力のある水木しげるロードは，鳥取県側の観光地よりも島根県側の観光地との結びつきが強く，観光客が鳥取県側の観光地に思っているほど流動していないことが分かる．従って，水木しげるロードと鳥取側観光地との連携強化も重要な課題であることも示唆される.

DEAによる47都道府県の観光業に関する分析

新村秀一
成蹊大学経済学部

DEA法はｐ個の投入される経営資源の入力に対して、ｑ個の出力される経営成果を用いて、組織の経営効率性を評価する手法として開発された。数理計画法で定式化されているために、統計ユーザーにとってなじみが少ない。しかし筆者はすでに誤分類数最小（Minimum Number of Misclassifications, MNM）化基準による最適線形判別関数を整数計画法でアプローチし成果を得た[10-12]。DEA法はｐ個の入力変数とｑ個の出力変数の総合化された比でもって回帰型のデータを分析するこれまでの統計手法にない新しい手法と考えている。しかしDEA研究は数理計画法によるモデル開発に重点があるのは仕方がないが、応用分析までがそれに引きずられている傾向が強い。

そこでDEAが実際に応用研究で役立つための実践的な応用法を考えた。そして平井[16]が47都道府県の滞在型の観光の分析に用いたデータに、筆者の作成した汎用DEAプログラムと標準化された分析手法の適用を行った結果を報告する。

1. DEA法に関する問題点の改善と標準的な解析方法

DEAはｐ個の入力 $\mathbf{x}$ を企業(Decision Making Unit, DMUただし評価対象ということにする)などの組織運営に投入される経営資源と考え、ｑ個の出力 $\mathbf{y}$ をその結果達成される成果と考える。そして個々の評価対象に最適な重みを，式(1)の数理計画法モデルで求める手法である。DEA効率値を ${}^t\mathbf{b_i}*\mathbf{y_i}/{}^t\mathbf{a_i}*\mathbf{x_i}$ と定義し、それを各DMUが最大の値になる重み($\mathbf{b_i}$、$\mathbf{a_i}$)を求める。この重みを分析対象の47都道府県に適用しクロス効率値 ${}^t\mathbf{b_i}*\mathbf{y_j}/{}^t\mathbf{a_i}*\mathbf{x_j}$ を制約式で1以下に制約しているのがCCRモデルである。各評価対象に最適な重みを求めて評価しているのに、47都道府県で1になるDEA効率的な県と、1未満になる非効率な県に分かれる。非効率な県は、自分に最適な重みを求めているが、その重みを他県にも適用して1になるためである。この場合にDEA効率的な県を手本にして非効率な県は改善すればよいという考え方である。

$$\mathrm{MAX}= {}^t\mathbf{b_i}*\mathbf{y_i}/{}^t\mathbf{a_i}*\mathbf{x_i}\ ;\quad {}^t\mathbf{b_i}*\mathbf{y_j}/{}^t\mathbf{a_i}*\mathbf{x_j}\leqq 1\ ;\quad j=1,\cdots,n \qquad (1)$$

一方，CCRモデルは各評価対象に一番有利な評価を行うため，入出力の変数が増えてくると手本が増え、手本が全て1に抑えられて優先順位がつけられない欠点がある．そこで式(2)のInverted DEAモデル[14]の利用を考えた．DEA効率値に代わってDEA非効率値(${}^t\mathbf{b_i}*\mathbf{y_i}/{}^t\mathbf{a_i}*\mathbf{x_i}$)を考え，他の評価対象が1以上になるという制約で最小化する重みを求める．このモデルの有用性は，DEA非効率値が1になる非効率な手本に注目することではなく，CCRモデルで手本になった評価対象の中でDEA非効率値が最大の評価対象を最初の改善目標にできる点である．

$$\mathrm{MIN}= {}^t\mathbf{b_i}*\mathbf{y_i}/{}^t\mathbf{a_i}*\mathbf{x_i}\ ;\quad {}^t\mathbf{b_i}*\mathbf{y_j}/{}^t\mathbf{a_i}*\mathbf{x_j}\geqq 1\ ;\quad j=1,\cdots,n \qquad (2)$$

すなわち、DEA法を統計手法と同じく次のような改善と汎用のプログラム[5]を開発した。

1) 改善目標を2入力1出力あるいは1入力2出力の2個の比（1出力/1入力）を計算し、その散布図でDEA効率フロンティアを説明し、非効率な評価対象と原点を結ぶ直線とDEA効率フロンティアの交点を理想的な改善点と説明している。理想点を改善目標とすることは計算が煩雑であり、かつ現

実応用で理想点を改善目標にできない。またこの説明は、p+q≧4 以上には適用できない。

2) そこで新村[6]はクロス効率値から DEA クラスターと呼ぶ方法を開発した。この方法で分析するとすぐに分かるが、変数の数が増えると DEA クラスターが増えて改善活動が細分化される。そこで Inverted DEA モデルで DEA 効率的な評価対象の中で優先順位をつけて改善の手本を一つに絞ることを考えた。

3) DEA は経営効率分析といわれながら、どのように具体的に改善するかの方法がなかった。そこで改善目標の構成比を実現することが非効率な評価対象の目標と考え「1 入力固定改善法」という手法を考えた。

2. 47 都道府県の滞在型観光の DEA による分析

平井[16]は，47 都道府県の一人当たり県民所得と他県民所得，ホテル・旅館 1 施設あたり客室数，面積(km^2)あたり温泉地数と観光施設数を 5 個の入力変数と考え，県内観光客，県外観光客と外国人観光客(各 100 万人泊)の 3 個の出力変数を用いて，DEA 法で 47 都道府県の効率性を CCR モデルで分析した先駆的研究と評価できる．Google で検索しても観光に関する DEA によるアプローチは見当たらない．これは平井も指摘しているが，観光に関するデータの作成に手間隙がかかるためと考えられる．その点で，彼の作成したデータを利用して観光に関する分析を新村の提案する汎用 DEA プログラム[5]と解析方法[8]で分析することは，この分野の研究の入門として意義があると考える．その結果、滞在型観光では北海道が突出で異質であり、世界遺産の富士山を抱える静岡を手本として分析することで、温泉地や観光施設数の多さが効果を持っていないこと、同じ世界遺産を共有する山梨県が非効率であるのは別荘や保養施設が統計に含まれていない可能性、そして京都がまったく評価されないのは日帰り型観光にアプローチしていないか宗教施設の宿坊が統計からもれている可能性が伺えた。DEA 分析が効果を発揮する、全国の美術館や博物館の分析、あるいは大型テーマパークの効率性や世界遺産の分析がよりよい結果が得られることが考えられるので、この分野の研究を提案した。

参考文献

[3]新村秀一(2004)．JMP 活用　統計学とっておき勉強法．講談社.

[4]新村秀一(2007)．Excel と LINGO で学ぶ数理計画法．丸善.

[5]新村秀一(2011).数理計画法による問題解決法．日科技連出版社.

[6] 新村秀一(2011).DEA による回帰型データのクラスター分析，成蹊大学一般研究報告，45/3,1-37.

[8] 新村秀一(2013)．DEA 利用のための実践的な解説書.成蹊大学経済学部論集,44/1，15-41.

[10]S. Shinmura(2014)．End of Discriminant Functions Based on Variance Covariance Matrices. 2014 ICORE，5-16.

[11] S. Shinmura(2014). Comparison of Linear Discriminant Functions by K-fold Cross Validation. Data Analytic 2014，1-6.(2014 年 8 月).

[12] S. Shinmura(2014).Three Serious Problems and New Facts of the Discriminant Analysis.1-16.　Spring のレクチャーシリーズの叢書に収録.

[14]杉山学(2010).『経営効率分析のための DEA と inverted DEA』．静岡学術出版.

[16]平井貴幸(2010).観光客を「効率的」に誘致している都道府県を探る－DEA による効率性分析―.地域と経済，7．111－116.

初期ヤマト政権の地政要因 - 時空間統計学の視点から

柳 本 武 美: 総合研究大学院大学

1. 古代史研究における統計的推論

統計学の進展の近年に見られる特徴は、複雑なモデルを用いることにより適用対象を柔軟に拡大する点にある。推論における誤りを受け入れて、入手可能なあらゆるデータを利用することにより、推論の誤りを減少させることに主眼が置かれる。古代の歴史研究では文献情報が乏しい。その結果多様な証拠を統合して真実を探ることになるので統計的な側面が大きい。しかし従来の研究では統計的な視点に乏しく素朴な科学的推論形式に止まっているように見える。そこで水田稲作の導入とクニあるいは国と呼ばれる地域政権が成立する時期である弥生時代 (BC1000-AD250 を想定) について興味ある話題について統計的視点の適用を試みている。

このときに強調すべき点は推定問題では推定量のリスクを小さくすることであって、推定値の損失を小さくすることではない。特に次元の高いベクトル母数を推定するときにはこの違いが大きくなる。

1.1 時空間モデル 統計学の進展における今日の特徴の一つは、時間と場所を特に選ばない知識の発見からある特定の時間と場所を特定したときに必要な知識の追究がある。従来は時間と空間に依存する命題を普遍性がなく刹那的でレベルの低い知見と捉えられてきたが、実際に我々が求めている知見の多くは確かに時間と場所に依存している。

歴史の研究はまさに場所と時間に依存した事実に関する研究である。具体的なデータの解析方法は提案されていないが、近年培われてきたデータの扱いに関する発想は十分に適用が可能である。

1.2 データの多様性 通常では歴史の研究は文献・記録など文章による記録が情報の中心になる。しかし、弥生時代では文章による記録は量として限られる上に、その資料の信頼性の吟味が必要になる。一方で遺跡から発掘物、遺伝情報など生理学的データ、C_{14}・鉛同位対比など物理化学的測定データ、日本語の特徴、伝承・風俗など多様なデータが多い。篠田 (2007) にも見られるように、ミトコンドリアと Y-遺伝子のハプロタイプの分布に著しい違いがあったことは資料の理解に柔軟な思考が求められることを示している。多様なデータを統合する方法はメタ・アナリシスと教育学・医療の分野で発展してきた。古代史の研究では遙かに複雑なメタ・アナリシスが必要になるが、決して特殊な問題ではないことを示唆している。

2. 重要な仮説とその検証

不十分な証拠から過去を復元するためには、適切な仮説 (あるいは前提) を設けることである。その仮説の妥当性は十分に検証すべきである。仮説の検証に当たっては、その仮説を支持する証拠を見いだすことも大切であるが、当該仮説を否定した仮説を設けてその仮説が妥当でないことを示すことが必要である。

2.1 二重構造モデル ここで採用した仮説の防御を検討するために二重構造モデル (例えば埴原編, 1994) を改めて見直す。このモデルは当時議論が分かれていた、「優秀な渡来民が基層集団を圧倒した (置換モデル)」「基層集団が渡来民の技術を吸収した (転換モデル)」に対して提示された。特に、転換モデルを意識していたようである。二重構造モデルにおける基層集団の想定など詳細に関しては議論があるが、既存の二つのモデルが否定されるのは明瞭である。実際置換モデルでは Y-遺伝子ハプロタイプの分布が説明できない。また、転換モデルでは稲作の普及とか北方系の言語構造と南方系の単語と風俗を説明できない。つまり、二重構造モデルは確固たる地位を占めている。次節で

提案する仮説は、二重構造モデルの一つの詳細化である。

2.2 集団差仮説 古代史における命題の構築において見られる仮説として、ある特定集団が他の集団に対して何らかの意味で優越しているとされている。ある集団が優秀であると考えられる理由として結果を無理なく説明することにある。勝った人は強いとの言説には一理があるが、一度だけの結果ではその言説は必ずしも信頼できない。トーナメント戦の勝者は大抵は強いが、実力に差がなくても何れかのチームが勝つ。歴史な事実は過去を引き継ぐから、他の集団と能力は同等であっても見かけ上の結果は大きく異なり得る。しかし、見かけ上のばらつきは Maringle 系列の変動とか、極値統計学が示す関係者の人数でも説明できることが多い。

3. 稲作・漁猟民仮説

縄文時代の末期から稲作・漁労民が渡来して西日本に進出した。その後半島南岸から後続の人々は既にある程度開発された後に、多くがしてきたと考える。従来は特に区別はしない。安本美典 (2000) とか鳥越憲三郎 (2007) を含めて漁労民の役割の大きさを指摘する論があるが、その吟味と含意についての考察は十分ではない。

従来説を否定する観察事実として以下が挙げられる。１）筑後川・吉備・出雲・河内湾流域など稲作と漁労の好適地で発展した。２）食糧としての家畜飼育の定着が著しく遅れた。３）Y-遺伝子ハプロタイプ O47z と O-SRY$_{465}$ の比率が日本と韓国で大きく異なる (Hammer ら, 2005)。４）列島での進出経路で海岸沿いが目立つ。５）中京地区・南九州への進出が遅れた。

また、仮説を支持する事実として a) 漁労民は悪い条件の下での渡海に耐える技術と経験がある b) 着岸して稲の収穫を得るまでの間海岸の近くでの生存能力が高い。c) 半島で東海岸を北上するより、南に渡海した方が稲作と漁労に有利である。d) 我が国では後世でも漁民が多い

一方で提案した仮説と既存の観察とに齟齬を来す事実にも注意する必要がある。a) 北九州での水田稲作は突然に現れている。b) 渡来民は骨格・顔つきが基層人とは大きく異なる。比率も渡来系が $7-80\%$ を占める。 こうした仮説と相反する観察があるのもある意味ではやむを得ない面がある。しかし、ある程度の説明は可能である。

4. 初期ヤマト政権

改めて飛鳥地域の特徴を確認する。九州からは離れて瀬戸内海を東進すると河内湾に到達する。河内湾には淀川と大和川が流入していた。大和川を遡上して生駒山系の南を超えて少し南進した位置にある。九州に到達した稲作・漁労民がある程度の年月があれば、ごく普通に到達できる位置にある。両河川の流域が広大であり、流域は生駒山系を廻って奈良盆地で近接する。また淀川は、宇治川を通じて京都に更に瀬田川を通じて琵琶湖もその流域に含まれる。琵琶湖は天然の貯水池でもあって稲作にとっての適地である。こうした恵まれた環境なのかで大和の初期政権は飛鳥地域に生まれた。紆余曲折を経ながら勢力を拡大していった。飛鳥地域の特徴は、つまり、飛鳥地域である程度大きな政権ができると、遠くに遠征しなくても近くでその勢力が拡大できる可能性を秘めている。

古代史で余り研究が進んでいない地域に伊勢・尾張・美濃の中京地域がある。海岸が入り組んで河川も多く纏まった稲作と漁労に好適である宏大な地域である。従って、初期ヤマト政権の強力なライバルになり得たはずである。しかし、稲作の普及が地勢要因から遅れたと推認されている。

文献

1) Hammer, M. F. *et al.* J. Hum. Genet., 2005. 篠田謙一 日本放送協会, 2007. 鳥越憲三郎 雄山閣, 2009. 埴原和郎 朝日選書, 1994. 宝来聰 岩波書店, 1997. 安本美典 宝島社新書, 2000.

現役大学生の基本３グラフ作成スキルの現状と問題点：３大学１５０名を事例として

聖泉大学・大阪産業大学・天理大学（非常勤講師）　持元　江津子

要旨：小学校で学ぶ基本３グラフの作成スキル調査を行った結果，少なからずの現役大学生が原点「0」のない棒グラフを正しいと判断したり，階層別データの円グラフ表現が不得手であったり，そのスキルはまだまだ未熟であることが明らかになった．報道や宣伝をつうじてジャンクグラフやバッドチャートをたびたび目にせざるを得ない環境下で，高等教育において基本３グラフを軸にグラフ作成スキルを改めて指導する余地があり，その教育効果もいくらか見込めることを本研究は示唆する．

キーワード：大学生　グラフリテラシー　基本３グラフ　グラフ作成スキル　教育効果

1.　はじめに

筆者のこれまでの調査研究[1]において大学生にほぼ定着していると結論した小学校で学ぶ初歩的なグラフ作成スキルについて，棒グラフ，折線グラフ，円グラフの基本３グラフに的を絞り，本研究ではさらに踏み込んで調べることにした．

また，OECD の学習到達度調査 PISA2009 および PISA2012 の結果より，日本の若者のグラフリテラシーは作成面でも読み取り面でも芳しくないことが示唆されている[2]．よって，本研究にも相応の意義があると考える．

2.　調査方法

本研究では，高等教育機関において筆者が講義担当している広い意味での統計学関連科目の受講生を対象に，講義中に実施した小テストの結果に注目した．表１は，被験者の構成や調査実施科目と教室，設問の表示方法，表２に示す学習ポイントの事前指導の有無を示している．小テストは 20 問で構成され，うち８問が本研究の調査対象である．設問はすべて二者択一式で２個のグラフを並べてみせ，設問の意図に合う選択を期待し，解答のために１問あたり 30 秒を配分した．

表１　被験者について

大学		A大学専攻a	B大学	A大学専攻b	C大学	計
学年	B1			5	79	**84**
	B2	20	13	15	15	**63**
	B3	3		1	1	**5**
	B4			3		**3**
	B5-			1		**1**
	計	**23**	**13**	**25**	**95**	***156***
科目		統計関連科目	統計学	統計関連科目	コンピュータ入門	
実施教室		PC教室	講義室	講義室	PC教室	
表示方法		モニター	プロジェクター	印刷した紙	モニター	
事前指導		**十分ある**	**かなりある**	**まったくなし**	**一部ある**	

表２　調査で問われた学習ポイント

番号	学習ポイント
(1)	データの分類項目に序列がない場合，棒を数量の降順に並べるとわかりやすい棒グラフになる．
(3)	3D もどきの棒グラフは棒の長さを不明瞭にし，座標軸が直交していないため，棒の長さが項目間の数量比を正確に反映せず，グラフを読む者にとって棒の長さで数量を直感的に比較しづらく，正しくない．
(5)	分類項目に序列のある階層別データを棒グラフ化する場合，棒をデータの降順ではなく階層に合わせて並べることが概ね優先される．
(7)	円グラフにおいて分類項目に序列がない場合，割合すなわちパイの角度の大きさで降順にパイをとる．
(9)	円グラフは正円で描かれ，分類項目間の構成比をパイの角度の比または面積の比で表すものであるが，3D 円グラフに描くと楕円になり，角度や面積の比が構成比を正しく反映しないため，3D 円グラフは正しくない．
(15)	分類項目に序列のある階層別データを円グラフで表現する場合，パイの取り方は構成比の降順よりも階層の順序を概ね優先する．
(18)	比例尺度のデータを棒グラフで正確に表現するためには，グラフを原点「0」から始めなければならない．
(20)	間隔尺度のデータには「0」に絶対的な意味がなく，比率で数量を比較することができないため，棒グラフはふさわしくない．折線グラフで表現するのが適切である．

3.　調査結果

各設問の有効回答数は 153～155 と高く，正答率（正答またはより望ましい解答の選択者数／有効回答数）は表３とおりである．

表３　正答率

	(1)	(3)	(5)	(7)	(9)	(15)	(18)	(20)
A大学専攻a	1.000	0.955	0.909	0.913	0.957	0.478	0.609	0.957
B大学	0.923	1.000	0.692	1.000	0.769	0.692	0.846	0.846
A大学専攻b	1.000	0.960	0.840	0.920	***0.480***	0.542	***0.042***	***0.667***
C大学	0.958	0.937	0.916	0.968	0.830	0.581	0.495	0.774
all	**0.968**	**0.948**	**0.884**	**0.955**	**0.787**	**0.569**	**0.471**	**0.791**
B1	0.964	0.929	0.905	0.964	0.810	0.585	0.476	0.795
B2	0.968	0.968	0.839	0.935	0.758	0.548	0.500	0.770
B3-	1.000	1.000	1.000	1.000	0.778	0.556	0.222	0.889

4.　考察

4.1.　正答率の主な傾向

表３より，(1)，(3)，(7)の正答率は「1」に近く極めて高い．次に，(5)，(9)，(20)の正答率が比較的高いが，(9)のみA大学専攻 b グループの正答率が５割を切っている．(15)は全般的に正答率が低い．残る(18)は，正答率が総じて低いのみならず，グループ間の差が極めて大きいという特徴がある．

また，約半数の被験者が７問以上正答し，約９割が５問以上正答していた（表４）．そして，正答率の信頼区間は表５のとおりで，(15)と(18)の正答率は他の設問よりも信頼度 99%で有意に低いことが分った．

表４　被験者別の正答数

正答数	人数	累積相対人数
8	25	0.160
7	52	0.494
6	48	0.801
5	18	0.917
4～0	13	1.000

表５　全被験者の正答率の信頼区間

	信頼度95%信頼区間		信頼度99%信頼区間	
	下限	上限	下限	上限
(1)	0.9267	0.9861	0.9078	0.9892
(3)	0.9015	0.9736	0.8814	0.9785
(5)	0.8239	0.9253	0.8012	0.9350
(7)	0.9092	0.9778	0.8894	0.9821
(9)	0.7161	0.8442	0.6914	0.8592
(15)	0.4894	***0.6445***	0.4646	***0.6669***
(18)	0.3932	***0.5494***	0.3699	***0.5737***
(20)	0.7197	0.8478	0.6948	0.8627

まず棒グラフについて，(1)の結果より，一般に棒が降順に並んでいればわかりやすく感じる大学生が多いが，(5)の結果より，短時間では分類項目ラベルに注意が向かない大学生や，

階層的な分類の知識に乏しい大学生が若干名いることが伺えた．(3)の結果より，３D もどきの棒グラフが正確さに欠けることを知っていたり気付いたりする大学生の多いことがわかった．また，(18)の正答率が 5 割を下回ったため，棒グラフが数量の比率を視覚的にわかりやすくするのに優れ，棒の長さが数量を反映して決まるという原則の身に付いている大学生は多くないことが分かった．

つぎに円グラフに関して，(7)の結果より，割合の降順でパイを並べる一般的な表現方法の定着はよいものの，(15)の結果より，序列のある階層的な分類項目の場合については，知識が定着していないか知識がない，あるいは無関心であるような大学生が約 4 割を占めていることが分かった．そして，(9)の結果より，３D グラフの方が正円の２D グラフよりも正しいとする大学生が少なからずいることが分かった．

折線グラフの場合，(20)の結果より，気温の変化のような間隔尺度のデータについて，折線グラフを正しいとする選択が期待されていたが，約２割の大学生が棒グラフを選んだ．

4.2. 比例尺度と間隔尺度

棒グラフに関する設問(18)と(20)の結果から，被験者が比例尺度のデータと間隔尺度のデータを十分には区別していない可能性が考えられる．折線グラフは数値的な変化変動を伝えるのに適し，間隔尺度と比例尺度の両方の時系列またはそれに類するデータのグラフ化に用いられる．データの尺度によらず原点「0」は重要でない．これに対して棒グラフは，比例尺度のデータのグラフ表現にのみ用い，原点「0」は表２で解説しているように重要な意味をもっている．

この区別を知らない，あるいは理解していなければ，グラフの下部を省略した原点「0」のない棒グラフを正しく感じたり，間隔尺度のデータの棒グラフ化を正しいと判断したり，マイナスから始まっている棒グラフに疑問を感じないことはおおいにありうる．本研究の被験者である大学生の少なからずが，これに当てはまったものと推察される．

4.3. 階層別データ

階層別データの棒グラフ化に関する(5)の正答率はやや低く，円グラフ化に関する(15)の正答率は明らかに低かった．後者はどの被験者グループにおいても正答率が低く，被験者が階層別に分類されたデータの扱いをいくらか苦手とし，事前指導の状況にもあまり左右されなかったと推察される．

4.4. 教育効果の可能性

A大学の事例に注目すると，グラフ作成スキルに関する小テスト実施までの事前指導が，専攻ａの学生に対しては十分に行われたが，専攻ｂの学生にはまったく行われなかった（表１）．両グループの正答率の差異は，(9) ３D 円グラフ不正確さと，(18) 原点「0」から始まる棒グラフ，(20) 間隔尺度データのグラフ化に関して顕著であり，事前指導のなかった専攻ｂグループの正答率が，事前指導の充実した専攻ａグループを下回った（表３）．

これらの正答率の検定結果より，３D 円グラフの問題と棒グラフの原点「0」を巡る問題については信頼度 99%で，また，間隔尺度データのグラフ化の問題については信頼度 95%で，専攻ａグループの正答率が専攻ｂグループを有意に上回っていることが分かる（表６）．これは，大学生がより正しくきめ細かなグラフ作成スキルを習得するのに，教育の役立つ可能性を示唆する．

表６　正答率（母割合）の差の検定結果

		(9)	(18)	(20)
正答率	A大学専攻a	0.957	0.609	0.957
	A大学専攻b	0.480	0.042	0.667
	A大学合計	0.708	0.319	0.809
標準得点Z_0		***3.629***	***4.169***	***2.525***
有意差の有無		**有：信頼度99%**	**有：信頼度99%**	**有：信頼度95%**

5. 結論

以上より，現役大学生の基本３グラフ作成スキルの現状について，少なからぬ学生が原点「0」のない棒グラフを正しいと判断したり，気温の変化を棒グラフで表現したりで，驚くほど初歩的なところでつまずく可能性のあることが分った．３D 円グラフを正円グラフよりも正しく感じる傾向も根強い．階層別データのグラフ化に不慣れな様子もうかがえる．小学校から中学校，高等学校までの授業の中で基本３グラフに触れる機会はいくらでもあったと推測されるものの，現役大学生の作成スキルはまだまだ未熟である実態がみえてきた．

また，３D 円グラフのようなジャンクグラフやバッドチャートと呼ばれるグラフと，正しいグラフの区別が曖昧な大学生も少なくないかもしれない．例えば３D 円グラフは，最近の政府統計にも使用例があり，大学の講義で実際に使われているような MS Office 教則本などにおいて見栄えのよいプレゼンテーション技法の一つとして推奨されている事例もある．日々の報道や宣伝活動における視覚アピール用のツールとして，グラフ風のイラストが活用されている事例も少なくない．このように現代社会は，正しいグラフの知識や作成スキルを身につけるのに必ずしもよい環境ではない．誤ったグラフを正しく感じる習慣は，誤ったグラフを用いた印象操作に左右されやすい傾向を伴う危険性をもつであろう．

とはいえ前章４節で考察したように，よい意味での教育効果の可能性を求める余地はまだまだあり，高等教育の場で改めてグラフの初歩について抑えておくことは有効と思われる．

なお，本研究はごくごく狭い範囲で収集したデータに基づいている．日本の大学生全体について，上述と同等のことを言えるかどうかは，今後新たに幅広く調べる必要があるだろう．本研究がそのための一歩一助となれば幸いである．

注・参照文献

[1] 持元江津子「現役大学生のグラフ作成スキルの現状と課題について：関西二つの大学を事例として」聖泉論叢(19)，137-144(2011)．持元江津子，武藤，長島，榎，小黒，小山田「データ可視化技術としてのグラフ作成スキルをめぐる現役大学生の現状と課題」第６回日本統計学会春季集会，ポスター発表，2012/3/4．持元江津子，武藤，長島，榎，小黒，小山田「大学生のグラフ・スキルの現状分析：グラフを作る力と読む力」2013 年度　統計関連学会連合大会，口頭発表，2013/9/11．持元江津子「大学生のグラフを読む力にみるグラフリテラシーと情報リテラシーの現状についての一考察」総合教育センター紀要(12)，91-96，(2014)，天理大学人間学部．
[2] 「OECD 生徒の学習到達度調査 PISA 2009 年デジタル読解力調査～国際結果の概要～」文部科学省国立教育政策研究所，53-54．および「OECD 生徒の学習到達度調査～2012 年調査分析資料集～」同上，69-73．国立教育政策研究所; http://www.nier.go.jp/kokusai/pisa/ よりダウンロード可．最終情報確認日 2014/8/15.

The Ewens sampling formula 及び random mappings に対する新たな汎関数中心極限定理

総合研究大学院大　佃　康司 ⋆

非負整数値確率変数 $(C_1^n, C_2^n \ldots, C_n^n)$ が $C_1^1 + 2C_n^n + \ldots + nC_n^n = n$ を満たすものとする．すなわち，$(C_1^n, C_2^n, \ldots, C_n^n)$ は自然数 n の確率分割において，下の添字で表されるサイズをもつ要素の個数であり，size index と呼ばれることもある．明らかに $(C_1^n, C_2^n, \ldots, C_n^n)$ は独立ではないが，多くの場合に $(C_1^n, C_2^n, \ldots)$ の有限次元マージナルは独立なポアソン確率変数 $(Z_1, Z_2, \ldots)$ の対応するものに分布収束することが知られている．具体的には，Ewens sampling formula と random mappings の場合に，それぞれ

$$\mathbb{E}[Z_j] = \frac{\theta}{j}, \quad \mathbb{E}[Z_j] = \frac{e^{-j}}{j} \sum_{i=0}^{j-1} \frac{j^i}{i!}$$

となる．この弱収束はある固定された b に関して

$$(C_1^n, \ldots, C_b^n) \to^d (Z_1, \ldots, Z_b)$$

という結果を意味するが，応用上は，$b = b(n)$ が n と同時に大きくなるような状況にも近似が正しいこと，すなわち高次元の極限理論が重要な場合も多いである．特に，$n \to \infty$ における汎関数中心極限定理

$$u \rightsquigarrow \frac{\sum_{j=1}^{[n^u]} C_j^n - \theta u \log n}{\sqrt{\theta \log(n)}} \to^d B \quad \text{in } D[0,1]$$

の証明において本質的である (Arratia and Tavaré[5])．ここで，B は標準ブラウン運動であり，D は Skorokhod 空間（cádlág 関数の空間）である．

本報告では The Ewens sampling formula 及び random mappings に対する新しい汎関数 CLT

$$u \rightsquigarrow \frac{\sum_{j=1}^{[n^u]} C_j^n - \sum_{j=1}^{[n^u]} \mathbb{E}[Z_j]}{\sqrt{\sum_{j=1}^{[n^u]} \mathbb{E}[Z_j]}} \to^d G \quad \text{in } L^2(du, [0,1]),$$

を証明の概略を示した．ここで，$u \rightsquigarrow G(u) = B(u)/\sqrt{u}$ であり，$L^2(du, [0,1])$ は $[0,1]$ 上で定義された du に関して二乗可積分な関数の同値類の空間である．

ランダム組み合わせ論では，DeLaurentis and Pittel[7] が random permutation に対して（D における；以下同様）汎関数 CLT を 1985 年に証明した．Random permutation の一般化であ

⋆ 日本学術振興会特別研究員 DC

る the Ewens sampling formula 及び random mappings に対するものはそれぞれ Hansen[9, 8] が，Billingley[6] にあるタイトネス基準をチェックし，有限次元マージナルの分布収束を示すことで行った．その後， Arratia et al.[1] が示した Poisson process approximation を経由した全変動距離を用いた証明が Arratia and Tavaré[5] によってなされた．この方法は汎関数 CLT 以外の漸近論にも有用である．この証明法を用いて logarithmic assemblies に対する漸近論が Arratia et al.[4] によって確立され，より一般的な logarithmic structures に対する漸近論が Arratia et al.[2] により確立された．ここで， logarithmic structures とは $(C_1^n, C_2^n, \ldots)$ と独立な確率変数列 $(Z_1, Z_2, \ldots)$ の間に the conditioning relation: $\mathbb{P}[C_1^n = c_1, \ldots, C_n^n = c_n] = \mathbb{P}\left[Z_1 = c_1, \ldots, Z_n = c_n \,\middle|\, \sum_{j=1}^n jZ_j = n\right]$ と the logarithmic condition: $\lim_{j\to\infty} j\mathbb{P}[Z_j = 1] = \lim_{j\to\infty} j\mathbb{E}[Z_j] = \theta$ が成り立つような確率法則のことをいう．この分野におけるまとめられた書籍としては Arratia et al.[3] がある．本発表で議論した 2 つの問題は上の 2 条件を満たす重要な例であり，これらの L^2 空間における漸近挙動を考察することによって一般的な構造に対する議論へ向けた原型となることが期待される．

References

[1] Arratia, R., Barbour, A. D., Tavaré, S. 1992. Poisson process approximations for the Ewens sampling formula, *Ann. Appl. Probab.* **2**, 519–535.

[2] Arratia, R., Barbour, A. D., Tavaré, S. 2000. Limits of logarithmic combinatorial structures, *Ann. Probab.* **28**, 1620–1644.

[3] Arratia, R., Barbour, A. D., Tavaré, S. 2003. *Logarithmic Combinatorial Structures: a Probabilistic Approach*, EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich.

[4] Arratia, R., Stark, D., Tavaré, S. 1995. Total variation asymptotics for Poisson process approximations of logarithmic combinatorial assemblies, *Ann. Probab.* **23**, 1347–1388.

[5] Arratia, R., Tavaré, S. 1992. Limit theorems for combinatorial structures via discrete process approximations, *Random Structures Algorithms* **3**, 321–345.

[6] Billingsley, P. 1999. *Convergence of Probability Measures. Second Edition*, Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York.

[7] DeLaurentis, J. M., Pittel, B. 1985. Random permutations and Brownian motion. *Pac. J. Math.* **119**, 287–301.

[8] Hansen, J. C. 1989. A functional central limit theorem for random mappings. *Ann. Probab.* **17**, 317–332. Correction: Hansen, J.C. 1991. *Ann. Probab.* **19**, 1393–1396.

[9] Hansen, J. C. 1990. A functional central limit theorem for the Ewens sampling formula. *J. Appl. Probab.* **27**, 28–43.

順序制約下の正規母平均列間の差に関する多重比較法

東海大学 熊本教養教育センター 今田 恒久

1. はじめに

正規分布に従う K 個の独立な確率変数 $X_1, X_2, \ldots, X_K$ があり，$X_k \sim N(\mu_k, \sigma^2)$ $(k = 1, 2, \ldots, K)$ と仮定する．母平均列間に順序 $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_K$ を仮定し，$\mu_k < \mu_{k+1}$ を満たす k を見つけたい．そのため，仮説

$$H_k : \mu_k = \mu_{k+1} \quad \text{v.s.} \quad H_k^A : \mu_k < \mu_{k+1}.$$

を設定し，$H_1, H_2, \ldots, H_{K-1}$ の同時検定を考える．各仮説の検定には Lee and Spurrier (1995) が選定した統計量を用いる．すなわち，母集団 $N(\mu_k, \sigma^2)$ $(k = 1, 2, \ldots, K)$ からの標本 $x_{k1}, x_{k2}, \ldots, x_{kn_k}$ に対して，

$$\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki}, \quad s = \sqrt{\frac{1}{\phi} \sum_{k=1}^{K} \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)^2}, \quad \phi = \sum_{k=1}^{K} n_k - K.$$

とし，H_k を検定するための統計量は

$$t_{k,k+1} = \frac{\bar{x}_{k+1} - \bar{x}_k}{\sqrt{\frac{1}{n_{k+1}} + \frac{1}{n_k}} s}$$

を用いる．ここでは，シングルステップ法，および総称してステップワイズ法と呼ばれる閉検定手順に基づくステップダウン法，逐次棄却型ステップダウン法，ステップアップ法について考え，手法を比較する．

2. Lee and Spurrier (1995) のシングルステップ式多重比較法

仮説 $H_1, H_2, \ldots, H_{K-1}$ に対する Lee and Spurrier (1995) のシングルステップ式多重比較法では，棄却限界値 c を指定し，$t_{k,k+1} > c$ ならば H_k を棄却し，そうでなければ保留とする．c は指定した有意水準 α を満たすように決定する．$H_1, H_2, \ldots, H_{K-1}$ がすべて保留となる確率は

$$P(t_{1,2} \leq c, t_{2,3} \leq c, \ldots, t_{K-1,K} \leq c). \tag{1}$$

c は 有意水準 α を指定して，繰り返し計算により (1) が $1 - \alpha$ に等しくなるように決定する．Lee and Spurrier (1995) で説明されているように，$H_1, H_2, \ldots, H_{K-1}$ がすべて真であると仮定したとき，(1) は $K - 1$ 変量 t 分布の確率密度関数を被積分関数とした多重積分として表される．

3. 閉検定手順に基づくステップダウン式多重比較法

白石 (2014) は，より検出力が高い手法を得るため，閉検定手順に基づくステップダウン式多重比較法を構築した．この方法の概要を述べる．$H_1, H_2, \ldots, H_{K-1}$ のうち，あらゆる複数個の仮説の積集合，および $H_1, H_2, \ldots, H_{K-1}$ から構成される仮説の族 F は閉じており，F に含まれる各仮説に対して指定した有意水準に対応した検定方式を構築する．以上の取り決めに基づき，族 F に対する閉検定手順に基づくステップダウン式多重比較法では F に含まれる各仮説に対して，その仮説を誘導するすべての仮説が棄却され，さらに，その仮説自身が棄却されるとき，その仮説を棄却することにする．この多重比較法の最大タイプ I FWE は α 以下となる．

4. 逐次棄却型ステップダウン多重比較法

Dunnett and Tamhane (1991) は複数個の正規母平均に対する対照比較のための多重比較において逐次棄却型ステップダウン式多重比較法を提案している．道家，今田，中村 (2006) は，このステップダウン法を $H_1, H_2, \ldots, H_{K-1}$ の同時検定に適用した．この方法の概要を述べる．標本より統計量 $t_{1,2}, t_{2,3}, \ldots, t_{K-1,K}$ を計算し，大きさ順に並べたものを

$$t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(K-1)} \tag{2}$$

と表し，これに対応する仮説を $H_{(1)}$, $H_{(2)}$, ..., $H_{(K-1)}$ と表す．逐次棄却型ステップダウン法は最大 $K-1$ 回検定を行う．各段階の棄却限界値 $c_1, c_2, \ldots, c_{K-1}$ （ただし，$c_1 < c_2 < \cdots < c_{K-1}$）を指定して以下のように検定を実行する．

Step 1.
Case 1. $t_{(K-1)} \leq c_{K-1}$ ならば，$H_1, H_2, \ldots, H_{K-1}$ をすべて保留として検定を終了する．
Case 2. $t_{(K-1)} > c_{K-1}$ ならば，$H_{(K-1)}$ を棄却して Step 2 へ進む．
Step 2.
Case 1. $t_{(K-2)} \leq c_{K-2}$ ならば，$H_{(1)}, H_{(2)}, \ldots, H_{(K-2)}$ をすべて保留として検定を終了する．
Case 2. $t_{(K-2)} > c_{K-2}$ ならば，$H_{(K-2)}$ を棄却して Step 3 へ進む．

以下同様にして最大 Step $K-1$ まで検定を続ける．各段階の棄却限界値 $c_1, c_2, \ldots, c_{K-1}$ は指定した有意水準に対して，Dunnett and Tamhane (1991) と同様な方法で決定される．

5. ステップアップ式多重比較法

Dunnett and Tamhane (1992) は複数個の正規母平均に対する対照比較のための多重比較においてステップアップ式多重比較法を提案している．ここでは，このステップアップ法を $H_1, H_2, \ldots, H_{K-1}$ の同時検定に適用する．この方法の概要を述べる．標本より統計量 $t_{1,2}, t_{2,3}, \ldots, t_{K-1,K}$ を計算し，大きさ順に並べたものを (2) のように表し，対応する仮説を $H_{(1)}, H_{(2)}, \ldots, H_{(K-1)}$ と表す．

ステップアップ法は最大 $K-1$ 回検定を行う．各段階の棄却限界値 $c_1, c_2, \ldots, c_{K-1}$ （ただし，$c_1 < c_2 < \cdots < c_{K-1}$）を指定して以下のように検定を実行する．

Step 1.
Case 1. $t_{(1)} > c_1$ ならば，$H_1, H_2, \ldots, H_{K-1}$ をすべて棄却して検定を終了する．
Case 2. $t_{(1)} \leq c_1$ ならば，$H_{(1)}$ を保留して Step 2 へ進む．
Step 2.
Case 1. $t_{(2)} > c_2$ ならば，$H_{(2)}, H_{(3)}, \ldots, H_{(K-1)}$ をすべて棄却して検定を終了する．
Case 2. $t_{(2)} \leq c_2$ ならば，$H_{(2)}$ を保留して Step 3 へ進む．

以下同様にして最大 Step $K-1$ まで検定を続ける．各段階の棄却限界値 $c_1, c_2, \ldots, c_{K-1}$ は指定した有意水準に対して，Dunnett and Tamhane (1992) と同様な方法で決定される．

6. 発表内容

発表では上述のシングルステップ法，3 種のステップワイズ法に対して指定した有意水準を満たす棄却限界値の決定方法と検出力の計算方法について議論する．さらに，検出力の数値例を与え，手法を比較する．

参考文献

[1] 道家暎幸，今田恒久，中村智洋 (2006)．用量反応試験のための逐次型ステップワイズ法の開発．日本統計学会誌, 第 36 巻 (第 1 号), 45-64.

[2] Dunnett, C.W. Horn, M. and Vollandt, R. (2001). Sample size determination in step-down and step-up multiple tests for comparing treatments with a control. *Journal of Statistical Planning and Inference*, **97**, 367-384.

[3] Dunnett, C.W. and Tamhane, A.C. (1991). Step down multiple tests for comparing treatments with a control in unbalanced one-way layouts. *Statistics in Medicine*, **10**, 939-947.

[4] Dunnett, C.W. and Tamhane, A.C. (1992). A step-up multiple test procedure. *Journal of the American Statistical Association*, **87**, 162-170.

[5] Hayter, A.J. and Tamhane, A.C. (1991). Sample size determination for step-down multiple test procedures; orthogonal contrasts and comparisons with a control. *Journal of Statistical Planning and Inference*, **27**, 271-290.

[6] Lee, R. E. and Spurrier, J. D. (1995): Successive comparisons between ordered treatments. *J. Statist. Plann. Inference*, **43**, 323-330.

[7] 永田靖, 吉田道弘 (1997). 統計的多重比較法の基礎. サイエンティスト社.

[8] 白石高章 (2014)．多群連続モデルにおける位置母数に順序制約のある場合の閉検定手順．日本統計学会誌, 第 43 巻 (第 2 号), 215-245.

Distribution Theory Considered on the Three Statistical Manifolds

Yasuko Chikuse, Professor Emeritus, Kagawa University

In this paper, we give some discussions on the statistical distributions which are considered on each of the three manifolds, that is, Stiefel, Grassmann and shape manifolds. These manifolds are useful for analyzing practical data, in orientation statistical analyses and shape statistical analyses.

The Stiefel manifold V(k,m) is the space whose points are k-frames in the m-dimensional real space R(m), which is represented by the set of mxk matrices X such that X'X is equal to the kxk identity matrix. The case m=k gives the orthogonal group O(m) of mxm orthonormal matrices. A point of V(k,m) may be called an orientation, extending the notion of a direction for k=1 (the unit hypersphere of general dimension; circle (m=2), sphere (m=3)).

The Grassmann manifold G(k,m-k) is the space whose points are k-planes (k-dimensional hyperplanes in R(m)).The case k=1 is the real projection space consisting of all lines through the origin. To each k-plane in G(k,k-m) corresponds a unique mxm orthogonal projection matrix P idempotent of rank k. If the k-columns of an mxk matrix Y in V(k,m) span the k-plane we have YY'=P. Thus we consider the matrix space P(k,m-k), the set of all mxm orthogonal projection matrices idempotent of rank k, which is equivalent to the Grassmann manifold G(k,m-k).

The Stiefel and Grassmann manifolds for the special case k=1 are treated as the directional statistics and may be put into two categories, directed (spherical) and undirected (axial, lines), respectively. There exists some large literature of theoretical treatment on directional statistics for k=1.

The directional statistics are treated in earth (or geological) sciences (the origin of the earth magnetic field, crystallography, palaeoccurrents), astrophysics, biology, meteorology, animal behavior.... The analyses for the case k¡m¡4 occur in practice in medical sciences, astronomy (vectorcardiogram, orbits of comets). Examples of observations on G(k,m-k) arise in the signal processing of radar with m elements observing k targets....

The shape of an object can be considered as the geometrical information on the object that remains after allowance is made for changes of location, scale and orientation. Statistical problems in shape analysis arise in biology, medicine, image

analysis, animal sciences.... Every set of k (not totally coincident) labelled points in R(m) can be centred and scaled to give the mx(k-1) pre-shape matrix Z (with trZ'Z=1) in the pre-shape space S(m,k-1).

We discuss distributions considered for each of the mxk X in the Stiefel manifold V(k,m), of the mxm P=XX' in the projection space P(k,m-k) and of the mx(k-1) Z in the pre-shape space S(m,k-1). There exists the uniform distribution on each space, and the test of the hypothesis of uniformity may be a natural one. We shall consider distributions other than the uniform distribution. They are the Fisher (Langevin) type, the Bingham type, ACG (angular central Gaussian) type, and the combinations of these distributions.

REFERENCES

1. Y. Chikuse and P. E. Jupp, A test of uniformity on shape spaces, J. Multivariate Anal. 88 (2004) 163-176.

2. Y. Chikuse and P. E. Jupp, Some families of distributions on higher shape spaces, To be submitted (2014).

3. Y. Chikuse, Statistics on Special Manifolds, Lecture Notes in Statistics 174, Springer (2003).

4. Y. Chikuse and G. S. Watson, Large sample asymptotic theory of tests for uniformity on the Grassmann manifold, J. Multivariate Anal. 54 (1995) 18-31.

5. A. W. Davis, Invariant polynomials with two matrix arguments, extending the zonal polynomials: Applications to multivariate distribution theory, Ann. Inst. Statist. Math. 31, A (1979) 465-485.

6. A. T. James, Distributions of matrix variates and latent roots dereived from normal samples, Ann. Math. Statist. 35 (1964) 475-501.

7. T. D. Downs, Orientation statistics, Biometrika 59 (1972) 665-676.

8. C. Bingham, An antipodally symmetric distributions on the sphere, Ann. Statist. 2 (1974) 1201-1225.

9. I. L. Dryden and K. V. Mardia, Statistical Shape Analysis, Wiley (1998).

10. K. V. Mardia and P. E. Jupp, Directional Statistics, Wiley (2000).

Bootstrap-based Bartlett-type adjustment

柿沢 佳秀 (北大経済)

1. はじめに バートレット調整と知られている LR 統計量の帰無分布の近似を改良する話題は, Bartlett (1937; Proc.R.Soc.Lond.) による分散均一性の具体例が始まりで, その一般論は, Lawley (1956; Biometrika) が $(\boldsymbol{\theta}'_{(1)}, \boldsymbol{\theta}'_{(2)})' \in \mathbf{R}^p$ に関する帰無仮説 $\boldsymbol{\theta}_{(1)} = \boldsymbol{\theta}_{(1)0}$ の LR 統計量の平均 $E^{(N)}_{\theta_{(1)0},\theta^\dagger_{(2)}}[\mathrm{LR}^{(N)}] = p_1 + (\mathcal{E}_p - \mathcal{E}_{p-p_1})/N + \mathrm{o}(N^{-1})$ を計算, 及び, $p_1\mathrm{LR}^{(N)}/E^{(N)}_{\theta_{(1)0},\theta^\dagger_{(2)}}[\mathrm{LR}^{(N)}]$ の任意次のキュムラントは $\mathrm{o}(N^{-1})$ を無視したときカイ 2 乗確率変数 $\chi^2_{p_1}$ のキュムラントに一致することを示したことによる. 実際, これらの結果は LR 統計量の漸近展開 (Hayakawa (1977,1987 訂正; AISM)) に整合する:

$$\begin{aligned} P^{(N)}_{\theta_{(1)0},\theta^\dagger_{(2)}}[\mathrm{LR}^{(N)} \le x] &= G_{p_1}(x) + \frac{A_1}{24N}\{G_{p_1+2}(x) - G_{p_1}(x)\} + \mathrm{o}(N^{-1}) \\ &= G_{p_1}(x) - \frac{A_1}{12N}\, g_{p_1+2}(x) + \mathrm{o}(N^{-1})\,. \end{aligned}$$

$\mathrm{LR}^{(N)}$ がバートレット調整可能である, すなわち, $\mathrm{BLR}^{(N)} = \mathrm{LR}^{(N)}/\{1+A_1/(12p_1N)\}$, $\{1 - A_1/(12p_1N)\}\mathrm{LR}^{(N)}$ の漸近展開の N^{-1} 項が消失する; $P^{(N)}_{\theta_{(1)0},\theta^\dagger_{(2)}}[\mathrm{BLR}^{(N)} \le x] = G_{p_1}(x) + \mathrm{o}(N^{-1})$ という性質は, Rao/Wald 統計量などで一般に成立せず, 異なる調整法が 1991 年の 3 論文 [Chandra and Mukerjee; JMVA, Cordeiro and Ferrari; Biometrika, Taniguchi; JMVA] で論じられた [今日にバートレット型調整と知られるようになった原型である]. CM/T 法の多母数拡張 [Kakizawa(2010,2012a; JMVA)] としての一般化バートレット型調整 (GB 法), 及び, CF 法の一般化 [Kakizawa(2012b; SPL)] としての GCF 法も提案された. さらに, バートレット型調整後の $N^{-1/2}$-(average) 局所検出力の意味において GB 法では LR 検定に同等になることが示されており, 調整前の統計量の特性が失われてしまうことが分かっている.

本報告では, 漸近カイ 2 乗型の統計量 $T^{(N)}$;

$$P^{(N)}[T^{(N)} \le x] = G_q(x) - \frac{2}{N}\sum_{\ell=1}^{3}\pi_\ell g_{q+2\ell}(x) + \mathrm{o}(N^{-1}) \tag{1}$$

について「CF 法[1]」と「ブートストラップ法」を組み合わせることで帰無分布の改良を行った[2].

[1]CF 法では, 漸近展開 (1) をもつ場合に

$$T^{\mathrm{CF}(N)} = \Big[1 - \frac{2}{N}\Big\{\frac{\pi_3}{q(q+2)(q+4)}\,(T^{(N)})^2 + \frac{\pi_2}{q(q+2)}\,T^{(N)} + \frac{\pi_1}{q}\Big\}\Big]T^{(N)}$$

と調整するので, 係数 π_ℓ's の計算が不可欠であった. Cordeiro and Ferrari(1991) 以降のおおよそ２０年で Rao 統計量に関する CF 法は非正規/非線形回帰モデルで膨大な論文があるものの, 2 次形式の行列の取り方が複数あり, Wald 統計量も含めて検討の余地があった.

[2]ブートストラップ・バートレット調整, すなわち, $\mathrm{BLR}^{(N)}$ の係数 A_1 をブートストラップ法で推定するという方法は文献に見られるが, LR 統計量以外のバートレット型調整に対してブートストラップ法の適用は新しい; Kojima and Kubokawa(2013; JMVA) は正規回帰モデルを仮定しているが, 本報告は一般漸近理論として考察している.

2. 主な結果　ある統計量 $T^{(N)}$ について

(i)[3] $T^{(N)} = (\mathbf{U}^{(N)})'\mathbf{V}^{-1}\mathbf{U}^{(N)} + \mathrm{o}_p(N^{-1})$; ここに, $\mathbf{V} = (v_{i_1,i_2})$ は $q \times q$ の正定値行列, かつ, $P^{(N)}[T^{(N)} \le x] = P^{(N)}[(\mathbf{U}^{(N)})'\mathbf{V}^{-1}\mathbf{U}^{(N)} \le x] + \mathrm{o}(N^{-1})$,

(ii) B-G の意味での N^{-1}-エッジワース展開をもつ $\mathbf{W}^{(N)} = (W_1^{(N)}, \ldots, W_m^{(N)})'$ $(m \ge q)$ が存在し, $U_i^{(N)} = \sum_{k=0}^{2} N^{-k/2}\pi_i^{[k]}(\mathbf{W}^{(N)})$ のキュムラントは次で与えられる (ここに, $\pi_i^{[k]}(\cdot)$ は $k+1$ 次斉次多項式):

$$
\begin{aligned}
Cum^{(N)}(U_i^{(N)}) &= \frac{\kappa_i}{N^{1/2}} + \mathrm{o}(N^{-1})\,, \\
Cum^{(N)}(U_{i_1}^{(N)}, U_{i_2}^{(N)}) &= v_{i_1,i_2} + \frac{\kappa_{i_1,i_2}}{N} + \mathrm{o}(N^{-1})\,, \\
Cum^{(N)}(U_{i_1}^{(N)}, U_{i_2}^{(N)}, U_{i_3}^{(N)}) &= \frac{\kappa_{i_1,i_2,i_3}}{N^{1/2}} + \mathrm{o}(N^{-1})\,, \\
Cum^{(N)}(U_{i_1}^{(N)}, U_{i_2}^{(N)}, U_{i_3}^{(N)}, U_{i_4}^{(N)}) &= \frac{\kappa_{i_1,i_2,i_3,i_4}}{N} + \mathrm{o}(N^{-1})
\end{aligned}
$$

ならば, $P^{(N)}[\mathbf{U}^{(N)} \in A]$ の漸近展開 (特に, $A = \{\mathbf{u} : \mathbf{u}'\mathbf{V}^{-1}\mathbf{u} \le x\}$, $x > 0$) を経由して, 漸近展開 (1) が正当化される (ここに, 係数 π_ℓ's はキュムラントの係数 $\kappa_{i_1,\ldots,i_R}$, $R = 1, 2, 3, 4$ で表現される; 詳細は省略) のような状況を考えた.

命題

$$
\begin{pmatrix} E^{(N)}[T^{(N)}] - q \\ E^{(N)}[(T^{(N)})^2] - q(q+2) \\ E^{(N)}[(T^{(N)})^3] - q(q+2)(q+4) \end{pmatrix}
= \begin{pmatrix} 1 & 1 & 1 \\ 2(q+2) & 2(q+4) & 2(q+6) \\ 3(q+2)(q+4) & 3(q+4)(q+6) & 3(q+6)(q+8) \end{pmatrix}
\begin{pmatrix} \frac{2\pi_1}{N} \\ \frac{2\pi_2}{N} \\ \frac{2\pi_3}{N} \end{pmatrix} + \mathrm{o}(N^{-1})\,.
$$

命題の左辺は, B 回のブートストラップ繰り返しにより推定できるので, 係数 π_ℓ's のブートストラップ推定値 π_ℓ^*'s が求まり, ブートストラップ・バートレット型調整

$$
T^{\mathrm{BCF}(N)} = \Big[1 - \frac{2}{N}\Big\{\frac{\pi_3^*}{q(q+2)(q+4)}\,(T^{(N)})^2 + \frac{\pi_2^*}{q(q+2)}\,T^{(N)} + \frac{\pi_1^*}{q}\Big\}\Big]T^{(N)}
$$

を提案した [$T^{(N)}$ に関して単調にするのは容易である; Kakizawa(1996; Biometrika)]. これにより, 漸近展開の係数を手計算することなく, サイズ $\alpha + \mathrm{o}(N^{-1})$ が保証された手続きを計算機上で構成でき, シミュレーション実験結果も報告した.

古典的な母数モデルに対する尤度原理からの Rao/Wald 統計量 (LR はバートレット調整可能な例外である) は勿論, 現代的な EL/GEL 法によるノンパラメトリック推論の諸統計量への応用も議論した.

[3]統計量の square-root 分解とそのキュムラント計算は Lawley(1956) にも形式的な扱いでみられた. しかし漸近展開の正当化 (Bhattacharya and Ghosh(1978; AS)) が確立された現在, (M 推定量のような正規型だけでなく) カイ 2 乗型漸近展開の正当化に対しても「基礎量 $\mathbf{W}^{(N)}$ のエッジワース展開＋ B-G 変換」のプログラムが利用できる. なお, (i) の後者を保証する十分条件として $\mathrm{o}_p(N^{-1})$ 項を詳細に記述してもよい.

A Nonparametric Test of a Strong Leverage Hypothesis

Yoon-Jae Whang

Seoul National University

The so-called leverage hypothesis is that negative shocks to prices/returns affect volatility more than equal positive shocks. Whether this is attributable to changing financial leverage is still subject to dispute but the terminology is in wide use. There are many tests of the leverage hypothesis using discrete time data. These typically involve fitting of a general parametric or semiparametric model to conditional volatility and then testing the implied restrictions on parameters or curves. We propose an alternative way of testing this hypothesis using realized volatility as an alternative direct nonparametric measure. Our null hypothesis is of conditional distributional dominance and so is much stronger than the usual hypotheses considered previously. We implement our test on individual stocks and a stock index using intraday data over a long span. We find only very weak evidence against our hypothesis.

正則化推定におけるモーメント収束

九州大学 清水 優祐
九州大学 増田 弘毅

統計的推測手法で得られる推定量のモーメントの収束条件の定式化を行うことで, AIC 型のモデル評価基準の方法論等を理論的に保証でき, 推定量依存型統計量のモーメントの収束を必要とする様々な議論の正当性を与えることが可能となる. 近年, データの大規模化に伴って正則化推定法が以前に増して注目されているが, そのようなモデルにおける推定量のモーメント収束の理論的根拠は未だ明らかにされていない. そこで本研究は, Knight and Fu [1] の Bridge-LSE と, Radchenko [2] の Sparse-Bridge-LSE のモーメントの収束条件を与える.

線形回帰モデル

$$Y_i = \theta_0^\top X_i + \epsilon_i, \quad i = 1, \dots, n \tag{1}$$

を考える. θ_0 は p 次元パラメータの真の値で, コンパクト集合 $\Theta \subset \mathbb{R}^p$ の中にあるものとする. $\epsilon_1, \epsilon_2, \cdots$ はノイズを表す. このときコントラスト関数

$$Z_n(\theta) = \sum_{i=1}^{n} (Y_i - \theta^\top X_i)^2 + \lambda_n \sum_{j=1}^{p} |\theta_j|^\gamma \tag{2}$$

の最小点 $\hat{\theta}_n$ で θ_0 を推定することを Bridge-LSE と呼ぶ. $\lambda_n > 0$ はチューニングパラメータであり, $\gamma > 0$ である. 特に θ_0 が,

$$\theta_0 = (z_0, \rho_0) = (z_{0,1}, \dots, z_{0,p_0}, \rho_{0,1}, \dots, \rho_{0,p_1})$$

と表される状況を考える. ここで $p_0 + p_1 = p$, 任意の $k \in \{1, \dots, p_0\}$, $l \in \{1, \dots, p_1\}$ に対して, $z_{0,k} = 0$, $\rho_{0,l} \neq 0$ である. 線形回帰モデル (1) とコントラスト関数 (2) は次のように書き直せる.

$$Y_i = \rho_0^\top X_i^{(\rho)} + \epsilon_i, \quad i = 1, \dots, n,$$
$$Z_n(\theta) = Z_n(z, \rho) = \sum_{i=1}^{n} \left(Y_i - z^\top X_i^{(z)} - \rho^\top X_i^{(\rho)}\right)^2 + \lambda_n \sum_{k=1}^{p_0} |z_k|^\gamma + \lambda_n \sum_{l=1}^{p_1} |\rho_l|^\gamma.$$

ここで, $X_i^{(z)} := (X_{i,1}, \dots, X_{i,p_0})$, $X_i^{(\rho)} := (X_{i,p_0+1}, \dots, X_{i,p_0+p_1})$ である. またコンパクト集合 $\Theta_0 \times \Theta_1 \subset \mathbb{R}^{p_0} \times \mathbb{R}^{p_1}$ 上の Z_n の最小点を $\hat{\theta}_n = (\hat{z}_n, \hat{\rho}_n)$ とする. このとき,

$$P(\hat{z}_n = 0) \to 1$$

が成り立つ推定を, Sparse-Bridge-LSE と呼ぶ.

推定量のモーメントの収束について簡単に説明する. $\hat{u}_n := \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \hat{u}_0$ なる確率変数 $\hat{u}_0$ の存在を仮定する. このとき任意の高々多項式増大度の連続関数 f に対して,

$$E[f(\hat{u}_n)] \to E[f(\hat{u}_0)] \tag{3}$$

を保証する条件を考える. このためには, 統計的確率場の一様裾確率評価を与える Yoshida [4] の多項式型大偏差不等式 (PLDI), $\forall L > 0, \exists c = c_L > 0, \forall r > 0$,

$$\sup_{n>0} P(|\hat{u}_n| > r) \leq \frac{c}{r^L} \tag{4}$$

が重要な役割を演じる. すなわちモーメントの収束 (3) は, PLDI (4) の成立条件により保証されるのである. そこで本研究は, 先述の Bridge-LSE, Sparse-Bridge-LSE に対して PLDI の成立条件の考察を行う. 特に Sparse-Bridge-LSE では, 零および非零パラメーターの推定量の漸近的な性質が異なり, 局所漸近二次構造の仮定が崩れることに注意する.

さらに Bridge の枠組みを広げるために, モデルの一般化を行う. 具体的には, $\mathfrak{p}_n(0) = 0$ を満たす非負の関数 $\mathfrak{p}_n$ に対して, コントラスト関数を,

$$Z_n(\theta) = Z_n(z, \rho) = \sum_{i=1}^{n} (Y_i - \rho^\top X_i^{(\rho)})^2 + \sum_{k=1}^{p_0} \mathfrak{p}_n(z_k) + \sum_{l=1}^{p_1} \mathfrak{p}_n(\rho_l)$$

とする. この一般化は, Fan and Li の SCAD (2001) や, Dicker et al. の seamless-L_0 (2012),

$$\mathfrak{p}_n(\theta_j) := \frac{2n\lambda_n}{\log 2} \log\Big(\frac{|\theta_j|}{|\theta_j| + \tau_n} + 1\Big)$$

を含む. 本研究の結果により, Radchenko [3] のような混合収束率を有する統計的確率場における PLDI に対する 1 つのアプローチが得られる.

参考文献

[1] Knight, K., Fu, W. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics*, 1356-1378.

[2] Radchenko, P. (2005). Reweighting the lasso. *In 2005 Proceedings of the American Statistical Association [CD-ROM]*, http://www-rcf.usc.edu/~radchenk/Lasso.pdf

[3] Radchenko, P. (2008). Mixed-rates asymptotics. *The Annals of Statistics*, 36(1), 287-309.

[4] Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Annals of the Institute of Statistical Mathematics*, 63(3), 431-479.

Higher-order asymptotic properties of frequency domain GMM estimators

Fumiya Akashi*

1 Fundamental settings

In this talk, we consider the frequency domain generalized method of moment (GMM) estimator, and discuss the higher-order asymptotic properties of the estimator as Akahira and Takeuchi [1] or Taniguchi [3] did for the maximum likelihood estimator of parametric models. Suppose that $\{X(t) : t \in \mathbb{Z}\}$ is a Gaussian stationary process with the usual spectral density function $f(\omega)$ and we consider the estimation problem of $\theta_0 \in \mathbb{R}^p$, a nonparametric important quantity of the process. We also assume that the information about θ_0 is summarized in the form

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} G(\omega;\theta_0)f(\omega)d\omega = 0_m,$$

where $G(\omega;\theta)$ is an $m \times 1$ vector-valued function, called a score function, and 0_m is the m-dimensional zero-vector. It is well known that many important quantities of time series models (such as autocorrelations of processes, coefficients of the best linear predictor/interpolator) can be expressed as θ_0 by choosing an appropriate score function $G(\omega;\theta)$. We call this θ_0 an over-identified pivotal quantity of the process since the dimension m of the spectral restriction can exceed $\dim(\theta) = p$. Kakizawa [2] proposed the frequency domain GMM estimator of θ_0 as

$$\widehat{\theta}_{\mathrm{GMM}} := \arg\max_{\theta\in\Theta} q_n(\theta), \quad q_n(\theta) := -\frac{n}{2}\widehat{R}_n(\theta)'\widehat{W}_n\widehat{R}_n(\theta),$$

where

$$\widehat{R}_n(\theta) := \frac{1}{n}\sum_{t=1}^{n} G(\lambda_t;\theta)I_{n,X}(\lambda_t)d\omega, \quad I_{n,X}(\omega) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X(t)\exp(\mathbf{i}t\omega)\right|^2$$

and $\widehat{W}_n$ is an $m\times m$ symmetric matrix such that $\widehat{W}_n \xrightarrow{\mathcal{P}} W$, with W being an $m\times m$ positive definite matrix. We call $\widehat{W}_n$ a sequence of wighting matrices. Kakizawa [2] derived consistency and asymptotic normality of the GMM estimator, and showed that the optimal weighting matrix is given as

$$V^{-1} = \left[\frac{1}{\pi}\int_{-\pi}^{\pi} G(\omega;\theta_0)G(\omega;\theta_0)'f(\omega)^2 d\omega\right]^{-1}. \tag{1.1}$$

In this paper, we derive the Edgeworth expansion of the distribution function of the GMM estimator with optimal weight (1.1), and discuss the bias adjustment of the GMM estimator in the sense of Akashira and Takeuchi [1].

*Graduate School of Fundamental Science and Engineering, Waseda University.

2 Main results

Now we derive the asymptotic expansion of the GMM estimator with the optimal weight;

$$\widehat{\theta}_n := \arg\max_{\theta\in\Theta} q_n^{(\mathrm{opt})}(\theta), \quad q_n^{(\mathrm{opt})}(\theta) := -\frac{n}{2}\widehat{R}_n(\theta)'V^{-1}\widehat{R}_n(\theta),$$

where V is defined as (1.1). Hereafter we restrict ourselves to the situation where θ is scalar. Under some regularity conditions, we have the following theorem.

Theorem 1.

$$P_{\theta_0}^n\left\{\sqrt{I(\theta_0)n}\left(\widehat{\theta}_n-\theta_0\right)\le y\right\} = \Phi(y) - \frac{\phi(y)}{\sqrt{n}}\left\{c_1^{(1)} + \frac{c_{11}^{(2)}}{2}y + \frac{c_{111}^{(1)}}{6}(y^2-1)\right\} + O\left(\frac{1}{n}\right), \tag{2.1}$$

where the coefficients $I(\theta)$, $c_1^{(1)}$, $c_{11}^{(2)}$ and $c_{111}^{(1)}$ are given as functionals of the spectral density function $f(\omega)$ and score function $G(\omega;\theta_0)$.

Based on Theorem 1, we discuss the asymptotic median unbiasedness of the estimator, which is a concept introduced by Akahira and Takeuchi [1]. Substituting $c_1^{(1)}$, $c_{11}^{(2)}$, $c_{111}^{(1)}$ and $y=0$ into (2.1), we have

$$\begin{aligned} &P_{\theta_0}^n\left\{\sqrt{I(\theta_0)n}\left(\widehat{\theta}_n-\theta_0\right)\le 0\right\} \\ &= \frac{1}{2} + \frac{\phi(0)}{\sqrt{n}}\left\{\frac{B(\theta_0)}{I(\theta_0)^{1/2}} - \frac{4K(\theta_0)+3Q^{(1)}(\theta_0)'V^{-1}W(\theta_0)V^{-1}Q^{(1)}(\theta_0)}{3I(\theta_0)^{3/2}}\right\} + O\left(\frac{1}{n}\right), \end{aligned} \tag{2.2}$$

where $B(\theta)$, $I(\theta)$, $K(\theta)$, $Q^{(1)}(\theta)$ and $W(\theta)$ are expressed as functionals of $f(\omega)$ and $G(\omega;\theta)$. (2.2) implies that the GMM estimator $\widehat{\theta}_n$ is not second-order asymptotic median unbiased (AMU) in the sense of Akashira and Takeuchi [1]. If we put

$$\begin{aligned} \widehat{\theta}_n^* &= \widehat{\theta}_n + \frac{1}{n}\left\{\frac{B(\widehat{\theta}_n)}{I(\widehat{\theta}_n)} - \frac{4K(\widehat{\theta}_n)+3Q^{(1)}(\widehat{\theta}_n)'V^{-1}W(\widehat{\theta}_n)V^{-1}Q^{(1)}(\widehat{\theta}_n)}{3I(\widehat{\theta}_n)^2}\right\} \\ &= \widehat{\theta}_n + \frac{1}{n}\left\{\frac{B(\theta_0)}{I(\theta_0)} - \frac{4K(\theta_0)+3Q^{(1)}(\theta_0)'V^{-1}W(\theta_0)V^{-1}Q^{(1)}(\theta_0)}{3I(\theta_0)^2}\right\} + o\left(\frac{1}{n}\right), \end{aligned} \tag{2.3}$$

then, the Edgeworth expansion of the modified GMM estimator $\widehat{\theta}_n^*$ is given as

$$P_{\theta_0}^n\left\{\sqrt{n}\left(\widehat{\theta}_n^*-\theta_0\right)\le x\right\} = \Phi\left(x\sqrt{I(\theta_0)}\right) + \frac{x^2}{\sqrt{nI(\theta_0)}}c_2^*(\theta_0)\phi\left(x\sqrt{I(\theta_0)}\right) + O\left(\frac{1}{n}\right),$$

where the coefficient $c_2^*(\theta_0)$ is described by $B(\theta)$, $I(\theta)$, $K(\theta)$, etc. That is, the modified GMM estimator $\widehat{\theta}_n^*$ is second-order AMU. Note that we can construct consistent estimators of the coefficients appearing (2.3) without assuming that the true $f(\omega)$ and θ_0 are known. Thus, we can construct the second-order AMU GMM estimator even if we know neither the true spectral density function of the process nor the true value of the pivotal quantity.

References

[1] Akahira, M. and Takeuchi, K. (1981). *Asymptotic efficiency of statistical estimators: Concepts and higher order asymptotic efficiency.* Springer.

[2] Kakizawa, Y. (2013). Frequency domain generalized empirical likelihood method. *Journal of Time Series Analysis.*

[3] Taniguchi, M. (1991). Higher order asymptotic theory for time series analysis. *Lecture Notes in Statistics*, 68.

Robust estimation of frequencies

Yan LIU[1]
(Waseda University)

Abstract

Nowadays, the quantile estimation becomes a notable method in statistics for its robustness against the moments of random variables. In this talk, we extend the idea of quantile in time domain to that in frequency domain. The objective function for the quantile estimator in time domain can be naturally extended into frequency domain. The quantile estimator in frequency domain has the consistency for the true value. However, asymptotic normality of the quantile estimator based on the bare periodogram does not hold, which is obviously different from the quantile theory for time domain. We give the asymptotic properties of the estimator. The modified estimator for asymptotic normality will be also provided.

keywords: quantile estimator, frequency domain, asymptotic properties of estimators

1. Quantiles in frequency domain

Suppose $\{X_t\,;\,t\in\mathbb{Z}\}$ is a second order stationary process. The observation stretch for the process is defined by $\{X_t\,;\,1\le t\le n\}$. From Herglotz's theorem, there exists a right continuous distribution function $F(\mu)$ for the covariance function $R(h)$ of the process such that

$$R(h)=\int_{-\pi}^{\pi}e^{ih\mu}F(d\mu),\quad (h\in\mathbb{Z}).$$

For simplicity from now on, write $R(0)=\Sigma_X$. Suppose the ψth quantile λ for the distribution function $F(\mu)$ over $0\le\psi\le\Sigma_X$ defined by

$$\lambda=F^{-1}(\psi)=\inf\{\mu\,;F(\mu)\ge\psi\}. \tag{1.1}$$

Equivalently, for $0\le p\le 1$,

$$\lambda=\inf\{\mu\,;F(\mu)\Sigma_X^{-1}\ge p\}.$$

2. Estimation for the quantiles in the frequency domain

In the estimation for quantiles in time domain, the check function $\rho_\tau(u)$ defined in the following way is usually used (e.g. Koenker (2005)):

$$\rho_\tau(u)=u(\tau-\mathbb{1}(u<0)).$$

[1]This work was supported by Grant-in-Aid for JSPS Fellows: 26·7404 (Liu, Y.)

We minimize this check function to estimate the τth quantile. The idea can be naturally extended into frequency domain, i.e., define $\hat{\lambda}_n$ for λ by

$$\hat{\lambda}_n = \arg\min_{\theta} \int_{-\pi}^{\pi} \rho_p(\omega - \theta) I_{n,X}(\omega) d\omega. \tag{2.1}$$

Here, the periodogram $I_{n,X}(\omega)$ based on $\{X_t\,;\, 1 \le t \le n\}$ is defined by

$$I_{n,X}(\omega) = \frac{1}{2\pi n} \Big| \sum_{j=1}^{n} X(j) e^{ij\omega} \Big|^2. \tag{2.2}$$

3. Asymptotic theory for the quantile estimator in frequency domain

Under the settings in the last section, we have the following results. The asymptotic results mainly depend on Hosoya and Taniguchi (1982) and the proofs of Theorems in this section, based on Pollard (1991) and Knight (1998), are omitted.

Theorem 3.1. *Suppose the ψth quantile λ of the spectral density of $\{X_t\,;\, t \in \mathbb{Z}\}$ is defined by* (1.1) *and $\hat{\lambda}_n$ is defined by* (2.1)*. Under regular conditions on the process $\{X_t\,;\, t \in \mathbb{Z}\}$, then we obtain*

$$\hat{\lambda}_n \to_p \lambda.$$

Note that the spectral density for the real-valued process is symmetry around the origin. For the fourth order uncorrelated process $\{X_t\,;\, t \in \mathbb{Z}\}$, the following theorem holds. We remark that the asymptotic distribution is normal given a exponential distributed random variable independent of the normal distribution. This is not expected only from the quantile estimation in time domain.

Theorem 3.2. *Suppose the ψth quantile λ of the spectral density of $\{X_t\,;\, t \in \mathbb{Z}\}$ is defined by* (1.1) *and $\hat{\lambda}_n$ is defined by* (2.1)*. Under regular conditions on the process, then for $-\pi < \lambda < 0$,*

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \to_d \mathcal{N}(0, \mathscr{E}^{-2}\sigma^2),$$

where $\mathscr{E}$ is an exponential random variable with mean $f(\lambda)$ and

$$\sigma^2 = 4\pi p^2 \int_{-\pi}^{\pi} f(\omega)^2 d\omega + 2\pi(1-4p) \int_{-\pi}^{\lambda} f(\omega)^2 d\omega + \frac{\kappa_4}{(2\pi)^2}(4\pi^2 p^2 - 4\pi p(\pi+\lambda) + (\pi+\lambda)^2),$$

κ_4 is the fourth order cumulant of the process.

4. Discussion

Adding a harmonic component m_t to the stationary time series model, we have

$$Y_t = m_t + X_t, \quad m_t = A\cos(\omega_0 t + \phi).$$

The model has been considered for a long time. As already known, the maximum $\hat{\omega}_0$ of the periodogram is the maximum likelihood estimator of the certain frequency ω_0 if $\{X_t\}$ is Gaussian. Also, the quantity behaves well under the assumption that $\{X_t\}$ is stationary from Rice and Rosenblatt (1988). In our numerical result, however, the estimator is not so good for the certain frequency ω_0 although the estimated quintiles are pulled around to the certain frequency ω_0.

補間誤差をコントラスト関数とする母数推定

須藤　慶大　（早稲田大学　基幹理工学研究科）
劉　言　（早稲田大学　基幹理工学研究科）
谷口　正信　（早稲田大学　理工学術院）

1　はじめに

統計の補間問題は欠損値解析などの様々な分野に対して重要な問題である．本研究では，正規定常過程のスペクトル密度が事前にわからない場合に母数付きのスペクトル密度 $f_{\boldsymbol{\theta}}(\lambda)$ であてはめたとき，$f_{\boldsymbol{\theta}}(\lambda)$ と真のスペクトル密度 $g(\lambda)$ との乖離度の基準（コントラスト関数）として誤特定補間誤差を提案し，それを最小にする母数の推定量の漸近的性質を明らかにした．以下では $\boldsymbol{Z}$ を整数全体，$\boldsymbol{R}$ を実数全体とする．

2　1時点の補間

$\{X_t, t \in \boldsymbol{Z}\}$ を定常過程として，その真のスペクトル密度を $g(\lambda)$ とする．Taniguchi (1980) は $g(\lambda)$ が事前にわからない場合に $\{X_t\}$ のスペクトル密度として $f_{\boldsymbol{\theta}}(\lambda)\,, \boldsymbol{\theta} \in \Theta \subset \boldsymbol{R}^m$ をあてはめたとき $\{X_t\}$ の p 時点 $\{X_1, \cdots, X_p\}$ に対する誤特定補間誤差行列

$$\Sigma_1 = \left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_{\boldsymbol{\theta}}(\lambda)}F(\lambda)d\lambda\right\}^{-1}\left\{\int_{-\pi}^{\pi}\frac{g(\lambda)}{f_{\boldsymbol{\theta}}(\lambda)^2}F(\lambda)d\lambda\right\}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_{\boldsymbol{\theta}}(\lambda)}F(\lambda)d\lambda\right\}^{-1} \tag{1}$$

を求めた．ただし，$F(\lambda) = \boldsymbol{e}(\lambda)\boldsymbol{e}(\lambda)^*$, $\boldsymbol{e}(\lambda) = (e^{-\boldsymbol{i}\lambda}, \ldots, e^{-\boldsymbol{i}p\lambda})'$ である．これより $\{X_t\}$ の 1 点 X_0 に対する補間誤差は，$F(\lambda) = 1$ として，$\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_{\boldsymbol{\theta}}(\lambda)}d\lambda\right\}^{-2}\left\{\int_{-\pi}^{\pi}\frac{g(\lambda)}{f_{\boldsymbol{\theta}}(\lambda)^2}d\lambda\right\}$ で与えられる．そこで，

$$D(f_{\boldsymbol{\theta}}, g) = \left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_{\boldsymbol{\theta}}(\lambda)}d\lambda\right\}^{-2}\left\{\int_{-\pi}^{\pi}\frac{g(\lambda)}{f_{\boldsymbol{\theta}}(\lambda)^2}d\lambda\right\} \tag{2}$$

として $D(f_{\boldsymbol{\theta}}, g)$ を最小にするような $\boldsymbol{\theta}$ の推測を行うことを考える．以下，$\{X_t\}$ は平均 0 の正規定常過程と仮定して，$g(\lambda)$ の代わりに $g(\lambda)$ の漸近不偏推定量である観測系列 $\{X_1, \ldots, X_n\}$ のピリオドグラム $I_n(\lambda) = \dfrac{1}{2\pi n}\left|\sum_{t=1}^{n} X_t e^{\boldsymbol{i}t\lambda}\right|^2$ を用いて，

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} D(f_{\boldsymbol{\theta}}, I_n) \tag{3}$$

とする．$\hat{\boldsymbol{\theta}}_n$ の漸近的性質を議論するために，適切な正則条件を仮定すると以下が成り立つ．

定理　推定する母数 $\boldsymbol{\theta}$ の真の値を $\boldsymbol{\theta}_0$ とする．このとき

$$(1)\quad \hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0, \quad \text{(確率収束)} \tag{4}$$

$$(2)\quad \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, 4\pi\, U(\boldsymbol{\theta}_0)^{-1}V(\boldsymbol{\theta}_0)U(\boldsymbol{\theta}_0)^{-1})\,. \quad \text{(分布収束)} \tag{5}$$

ただし, $U(\boldsymbol{\theta}_0)$ は正則な行列で, $U(\boldsymbol{\theta}_0)$, $V(\boldsymbol{\theta}_0)$ は $f_{\boldsymbol{\theta}}(\lambda)$ とその $\boldsymbol{\theta}$ に対する微分を用いた積分で表現される行列である. 次に補間誤差を最小にする $\boldsymbol{\theta}$ の推定量が漸近有効になるか考える. 時系列解析における正規 Fisher 情報量行列を

$$\mathcal{F}(\boldsymbol{\theta}) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\lambda) \frac{\partial}{\partial \boldsymbol{\theta}'} \log f_{\boldsymbol{\theta}}(\lambda) d\lambda \tag{6}$$

と定義する. $\hat{\boldsymbol{\theta}}_n$ の漸近分散が Fisher 情報量行列の逆行列 $\mathcal{F}(\boldsymbol{\theta}_0)^{-1}$ に到達するときに $\hat{\boldsymbol{\theta}}_n$ は漸近有効と呼ぶ. 以下の定理は $\hat{\boldsymbol{\theta}}_n$ が漸近有効になるのは極めて限定的であることを示している.

定理

$$4\pi\, U(\boldsymbol{\theta}_0)^{-1} V(\boldsymbol{\theta}_0) U(\boldsymbol{\theta}_0)^{-1} \geq \mathcal{F}(\boldsymbol{\theta}_0)^{-1}. \tag{7}$$

ここで, 不等式 $\{*\} \geq \{\cdot\}$ は行列 $\{*\}$, $\{\cdot\}$ に対して $\{*\} - \{\cdot\}$ が半正定値になることを意味する. なお等号が成立するのは, λ に依存しないある $m \times m$(m は母数の次元) 定数行列 C に対して,

$$\left\{\int_{-\pi}^{\pi} \frac{1}{f_{\boldsymbol{\theta}_0}(\mu)} d\mu\right\} \frac{1}{f_{\boldsymbol{\theta}_0}(\lambda)} \frac{\partial}{\partial \boldsymbol{\theta}} f_{\boldsymbol{\theta}_0}(\lambda) - \int_{-\pi}^{\pi} \frac{1}{f_{\boldsymbol{\theta}_0}(\mu)^2} \frac{\partial}{\partial \boldsymbol{\theta}} f_{\boldsymbol{\theta}_0}(\mu) d\mu + C \frac{\partial}{\partial \boldsymbol{\theta}} f_{\boldsymbol{\theta}_0}(\lambda) = \mathbf{0} \tag{8}$$

a.e. $\lambda \in [-\pi, \pi]$ となるときである. 本研究では AR(1) と MA(1) の具体例を用いて上式 (7) を確かめ, その場合 $\hat{\boldsymbol{\theta}}_n$ が漸近有効になるときは $\{X_t\}$ が単純な i.i.d. 正規イノベーション過程のときであることを示した. さらに AR(1) の場合において $\hat{\boldsymbol{\theta}}_n$ が, 漸近有効である Whittle 推定量に比べて有効でないことを数値実験して確かめた.

3 p 時点の補間

1 時点補間は応用を考えると限定的なので, 本研究ではさらに p 時点補間を議論する. 補間誤差 (1) を trace の意味で最小にする $\boldsymbol{\theta}$ の推定を行うことを考え, (3) で定義される推定量 $\hat{\boldsymbol{\theta}}_n$ に対して, 定理 1 と同様, 推定量の一致性, 漸近正規性を示すことができた.

4 まとめ

予測の観点から導かれる推定量は Whittle 推定量 (例えば Taniguchi and Kakizawa (2000)) と呼ばれ, その推定量は漸近有効になることは先行研究で扱われているが, 今回の研究では補間誤差の観点から導かれる推定量は一般に漸近有効にならないことが示された. 予測は過去の情報から現時点 (予測したい 1 時点) の情報を求めるものであるのに対して, 補間では現時点に対して過去と未来の情報から構成されるものにも関わらず, 今回のような意外な結果を得た.

参考文献

[1] Taniguchi, M. (1980). Regression and interpolation for time series. *Recent Developments in Statistical Inference and Data Analysis* : K. Matusita, Ed. 311-321. North-Holland Publishing Company.

[2] Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic theory of statistical inference for time series*. New York : Springer-Verlag.

On Improved Estimation of Gamma Parameters

Hidekazu Tanaka (Osaka Prefecture Univ.)
Nabendu Pal (Univ. Louisiana at Lafayette)
Wooi K. Lim (William Paterson Univ.)

One of the most widely used positively skewed probability distribution is the gamma distribution $Ga(\delta, \sigma)$ with pdf $f(x|\delta, \sigma)$ given as

$$f(x|\delta, \sigma) = (\Gamma(\delta)\sigma^{\delta})^{-1} \exp(-x/\sigma)x^{\delta-1} \quad (x > 0),$$

where $\delta, \sigma > 0$ are the shape and scale parameters, respectively. In this talk, we consider the estimation problem of δ and σ from the point of view of second order admissibility. Based on iid observations $X_1, X_2, \ldots, X_n$ from $Ga(\delta, \sigma)$, the maximum likelihood estimators (MLEs) of δ and σ, denoted by $\hat{\delta}_0$ and $\hat{\sigma}_0$, are found by solving the equations

$$g(\delta) := \ln \delta - \psi(\delta) = \ln R, \quad \hat{\sigma}_0 = \bar{X}/\hat{\delta}_0,$$

where

$$\psi(\delta) := (\partial/\partial\delta) \ln \Gamma(\delta), \quad R := \bar{X}/\widetilde{X}, \quad \bar{X} := \textstyle\sum_{i=1}^{n} X_i/n, \quad \widetilde{X} := \left(\prod_{i=1}^{n} X_i\right)^{1/n}.$$

The bias corrected versions of δ and σ suggested by Anderson and Ray (1975) are given as

$$\hat{\delta}_1 = \hat{\delta}_0 + c_1(\hat{\delta}_0)/n \quad \text{and} \quad \hat{\sigma}_1 = \hat{\sigma}_0/h_n(\hat{\delta}_1),$$

where $c_1(\delta) := -3\delta + (2/3)$ and

$$h_n(\delta) := [1 + 3n\delta \{(n-3)(n\delta-1)\}^{-1}\{1 + (9\delta)^{-1} - (n\delta)^{-1} + (n-1)(27n\delta^2)^{-1}\}]^{-1}.$$

The conditional MLE of δ suggested by Yanagimoto (1988) has the asymptotic expression $\hat{\delta}_2 = \hat{\delta}_0 + c_2(\hat{\delta}_0)/n + o_p(1/n)$ as $n \to \infty$, where $c_2(\delta) := 1/\{2\delta g'(\delta)\}$.

First, we consider the estimation problem of δ. Here, we reparameterize (δ, σ) to $\theta := (\delta, \eta)$, where $\eta = \delta\sigma$. The second order properties of the concerned estimators of δ will be studied in the class $\mathcal{C} := \{\hat{\delta}_0 + c(\hat{\theta}_0)/n + o_p(1/n) \,|\, c \in C^1(\mathbb{R}^2_+)\}$, where $\hat{\theta}_0$ is the MLE of θ. Let

$$\zeta_i(\delta) := \exp\{\textstyle\int_{\delta_0}^{\delta} I_{11}^{\delta}(u) b_{\delta,i}(u) du\}$$

for some constant $\delta_0 > 0$, where $I_{11}^{\delta}(\delta) = \psi'(\delta) - 1/\delta$ is the $(1,1)$ element of $I^{\delta}(\theta)$, the Fisher information matrix of θ per observation, and $b_{\delta,i}(\delta)$, $i = 0, 1, 2$, are the first order bias coefficient of $\hat{\delta}_i \in \mathcal{C}$.

Theorem 1 All the three estimators $\hat{\delta}_0$, $\hat{\delta}_1$ and $\hat{\delta}_2$ are found to be second order inadmissible (SOI). Also, the proposed new estimators $\hat{\delta}_i^* := \hat{\delta}_0 + c_i^*(\hat{\delta}_0)/n \in \mathcal{C}$, $i = 0, 1, 2$, are second order better (SOB) than $\hat{\delta}_i$ and second order admissible (SOA) in $\mathcal{C}$, where $c_0 := 0$ and

$$c_i^*(\delta) := c_i(\delta) - \left\{\zeta_i(\delta)\overline{K}_i(\delta)\right\}^{-1}, \quad \overline{K}_i(\delta) := \textstyle\int_{\delta}^{\infty} \{I_{11}^{\delta}(u)/\zeta_i(u)\} du.$$

Remark 1 In Takagi (2012), the second order admissibilities and/or inadmissibilities of a general class of estimators have been discussed, and using that result the second order inadmissibilities of the three estimators, $\hat{\delta}_i$, $i = 0, 1, 2$, can be obtained. A comprehensive simulation shows that $\hat{\delta}_1^*$ has the best improvement over $\hat{\delta}_0$.

Next, we consider the estimation problem of σ. Using the maximum likelihood restriction $\hat{\delta}\,\hat{\sigma} = \bar{X}$, which can be justified by the moment condition $E(\bar{X}) = \delta\sigma$, it is tempting to derive new estimators of σ from those of δ. Thus one may propose

$$\hat{\sigma}_2 = \bar{X}/\hat{\delta}_2 \quad \text{and} \quad \hat{\sigma}_i^* = \bar{X}/\hat{\delta}_i^*, \quad i = 0, 1, 2,$$

apart from $\hat{\sigma}_0$ and $\hat{\sigma}_1$ as stated earlier. Similarly, one may propose six new estimators of σ as

$$\tilde{\sigma}_i = \tilde{X}/\phi_n(\hat{\delta}_i) \quad \text{and} \quad \tilde{\sigma}_i^* = \tilde{X}/\phi_n(\hat{\delta}_i^*), \quad i = 0, 1, 2,$$

since $E(\tilde{X}) = \sigma\phi_n(\delta)$, where $\phi_n(\delta) := \{\Gamma(\delta + 1/n)/\Gamma(\delta)\}^n$. Here, we use the reparameterization $(\sigma, \delta) \to \vartheta = (\sigma, \lambda)$ such that $\psi(\delta) = -\ln(\sigma\lambda)$. In estimating σ, we restrict estimators to the class $\mathcal{D} := \{\hat{\sigma}_0 + d(\hat{\vartheta}_0)/n + o_p(1/n) \mid d \in C^1(\mathbb{R}_+^2)\}$, where $\hat{\vartheta}_0$ is the MLE of $\vartheta = (\sigma, \lambda)$.

Lemma 1 As $n \to \infty$, the twelve estimators we deal with have the asymptotic expressions

$$\hat{\sigma}_i = \hat{\sigma}_0 + d_i(\hat{\vartheta}_0)/n + o_p\,(1/n)\,, \quad \hat{\sigma}_i^* = \hat{\sigma}_0 + d_i^*(\hat{\vartheta}_0)/n + o_p\,(1/n)\,,$$
$$\tilde{\sigma}_i = \hat{\sigma}_0 + \tilde{d}_i(\hat{\vartheta}_0)/n + o_p\,(1/n)\,, \quad \tilde{\sigma}_i^* = \hat{\sigma}_0 + \tilde{d}_i^*(\hat{\vartheta}_0)/n + o_p\,(1/n)$$

for $i = 0, 1, 2$, where

$$d_0(\vartheta) := 0, \quad d_1(\vartheta) := 3\sigma\{1 + 1/(9\delta) + 1/(27\delta^2)\}, \quad d_2(\vartheta) := -\sigma c_2(\delta)/\delta,$$
$$d_i^*(\vartheta) := -\sigma c_i^*(\delta)/\delta, \quad \tilde{d}_i(\vartheta) := -\sigma\psi'(\delta)\{1 + 2c_i(\delta)\}/2,$$
$$\tilde{d}_i^*(\vartheta) := -\sigma\psi'(\delta)\{1 + 2c_i^*(\delta)\}/2 \quad \text{and} \quad \delta = \delta(\sigma, \lambda).$$

Let

$$\pi_d(\vartheta) := \exp\{\textstyle\int_{\sigma_0}^{\sigma} I_{11}^{\sigma}(u, \lambda) b_{\sigma,d}(u, \lambda) du\}$$

for some constant $\sigma_0 > 0$, where $I_{11}^{\sigma}(\vartheta) = (\delta\psi'(\delta) - 1)/\{\sigma^2\psi'(\delta)\}$ is the $(1, 1)$ element of $I^{\sigma}(\vartheta)$, the Fisher information matrix of ϑ per observation, and $b_{\sigma,d}(\vartheta)$ is the first order bias coefficient of $\hat{\sigma}_d \in \mathcal{D}$.

Theorem 2 The estimators $\hat{\sigma}_i$ and $\hat{\sigma}_i^*$ are SOI, and $\tilde{\sigma}_i$ and $\tilde{\sigma}_i^*$ are SOA in $\mathcal{D}$ for $i = 0, 1, 2$. Also,

$$\hat{\sigma}_i^+ := \hat{\sigma}_0 + d_i^+(\hat{\vartheta}_0)/n \quad \text{and} \quad \hat{\sigma}_i^{*+} := \hat{\sigma}_0 + d_i^{*+}(\hat{\vartheta}_0)/n$$

are SOB than $\hat{\sigma}_i$ and $\hat{\sigma}_i^*$, respectively, and are SOA in $\mathcal{D}$, where

$$d_i^+(\vartheta) := d_i(\vartheta) - \{\pi_i(\vartheta)\overline{H}_i(\vartheta)\}^{-1}, \qquad \overline{H}_i(\vartheta) := \textstyle\int_{\sigma}^{\infty}\{I_{11}^{\sigma}(s, \lambda)/\pi_i(s, \lambda)\}ds,$$
$$d_i^{*+}(\vartheta) := d_i^*(\vartheta) - \{\pi_i^*(\vartheta)\overline{H}_i^*(\vartheta)\}^{-1}, \quad \overline{H}_i^*(\vartheta) := \textstyle\int_{\sigma}^{\infty}\{I_{11}^{\sigma}(s, \lambda)/\pi_i^*(s, \lambda)\}ds.$$

References

H. Tanaka, N. Pal and W. K. Lim. On Improved Estimation of a Gamma Scale Parameter, submitted for publication.

最頻値の位置を変えない非対称分布族の生成

阿部 俊弘 (東京理科大学)
藤澤 洋徳 (統計数理研究所)

1　導入

対称な連続分布は正規分布をはじめとして様々な分布が提案されている (Johnson, Kotz and Balakrishnan, 1995; Kotz, Balakrishnan and Johnson, 2000). 2000 年以降になると, Azzalini (1985) による非対称分布が改めて注目を浴びるようになり, その手法を用いた新しい非対称分布が提案されるようになってきた. この Azzalini 型の分布族は生成の仕方が容易であり, 扱いやすいが単峰性が自然に保証されていないなどの欠点を持っている. 本研究では, この問題点を克服した様々な特徴をもつ分布族を提案する.

Azzalini (1985) による非対称分布は次のように与えられる: $f(x)$ を 0 に関して対称な密度関数とし, $G(x)$ を G' が偶関数であるような分布の分布関数とする. このとき,

$$f_{AZ}(x) = 2G(\lambda x)f(x) \tag{1}$$

は密度関数となる. ここで, $f(x)$ は原点対称の密度関数であり, $G(x)$ は原点対称な密度関数をもつ分布関数である.

Azzalini and Capitanio (2003) はこのような分布のクラスを次のように一般化した. $G_0(x)$ をその導関数 G_0' が偶関数である分布関数とし, $w(x)$ を奇関数とする. このようにして生成される分布

$$f_{AC}(x) = 2G_0(w(x))f(x)$$

は確率密度関数となる. Genton (2004) や Azzalini (2005) でもこの一般化された非対称分布族の解説がされている. また, このクラスの分布族の円周版は Abe and Pewsey (2011) で研究されている. 特に $f(x)$ と $G(x)$ が標準正規分布の密度関数と分布関数である場合の密度関数は, 正規分布の特殊性から単峰性は保証されているが, 一般的には単峰性は保証されない (Ma and Genton, 2004). モードが同じ場所に留まらずに右に左に動くことも気になる点である.

一方, Jones (2014) は

$$f^*(x) = f(s^{-1}(x)) \tag{2}$$

を密度関数に持つ非対称分布族を提案している. ここで, $f(x)$ は原点対称の密度関数, $s(x)$ は $s'(x) > 0$ であり, 偶関数 $H(x)$ を用いて

$$s(x) = x + H(x) \tag{3}$$

により与えられる.

この分布族はパラメータの選び方によらず, 常に単峰性が保証されている, 正規化定数が基の対称分布と等しい, 乱数の発生が簡単である等の理由から, 注目を浴び始めている. ここまでの分布族を含む近年の非対称分布族の発展に関しては, Jones (2015) にまとめられている.

我々は密度関数 (2) で (3) の $H(x)$ に

$$H(0) = 0$$

の条件を追加した非対称分布族を考える. これにより, 変換後の分布の最頻値の位置は変わらないことがわかる. 位置パラメータ μ は (2) において $x \mapsto x - \mu$ として導入される.

2　最頻値を変えない非対称変換の例

$H(x)$ として，

$$H(x) = \begin{cases} a_\lambda \dfrac{\sqrt{1+\lambda^2 x^2}-1}{\lambda} & \text{for } \lambda \neq 0, \\ 0 & \text{for } \lambda = 0, \end{cases}$$

を考える．ここで, $a_\lambda = 1 - e^{-\lambda^2}$ である．このとき, $H(0) = 0$ を満たす．$s(x) = x + H(x)$ とすると，その逆関数は

$$s^{-1}(x) = \begin{cases} \dfrac{\lambda x + a_\lambda - a_\lambda \sqrt{(\lambda x + a_\lambda)^2 + 1 - a_\lambda^2}}{\lambda(1-a_\lambda^2)} & \text{for } \lambda \neq 0, \\ x & \text{for } \lambda = 0. \end{cases} \tag{4}$$

と陽的な形で与えられる．

上の例では非対称パラメータ λ が大きくなるにつれて，歪みが大きくなる．このような性質は様々な歪度について証明が可能である．

参考文献

[1] Abe, T. and Pewsey, A. (2011). Sine-skewed circular distributions. *Stat. Pap.*, **52**, 683–707.

[2] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–178.

[3] Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scand. J. Statist.* **32**, 159–200.

[4] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. Roy. Stat. Soc., Series B*, **65**, 367–389.

[5] Genton, M. G., editor (2004). *Skew-elliptical distributions and their applications.* Chapman & Hall/CRC, Boca Raton, FL.

[6] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distributions. Volume 2, Wiley, New York.

[7] Jones, M.C. (2014). Generating distributions by transformation of scale. *Statist. Sinica*, **24**, 749–771.

[8] Jones, M.C. (2015). On families of distributions with shape parameters. *Int. Stat. Rev.*, in press.

[9] Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). Continuous Multivariate Distributions. Volume 1, Wiley, New York.

[10] Ma, Y. and Genton, M. G. (2004). Flexible class of skew-symmetric distributions. *Scand. J. Statist.* **31**, 459–468.