さを評価したものであり，モデル $f(x|\hat{\omega})$ の AIC にほかならないと見ることができる．そこで，パラメトリック・モデルもベイズ型モデルもどちらも適用可能な状況では，両方の AIC（あるいは ABIC）を比較してモデル選択を行なってよいとの見方も可能である．しかし，回帰分析の場合のシミュレーションの結果によると，状況によっては，この見方が論理の不整合をもたらす．この難点は EIC によって一応解決可能ではあるが，ABIC について理論的な研究がもっと行なわれるべきである．

原点未知の3母数対数正規分布について

金 藤 浩 司

データ解析に用いられるデータはその値が正値しか取らないものは少なくなく，これらのデータは分布の立ち上り点（原点）を零と見なすよりも正の未知な値（母数）α と考えることが適切な場合が多いと思われ，それにより原点推定の問題が生じる．このような場合，確率変数 X に対して $x-\alpha$ を確率密度関数に代入し，$\alpha(\alpha<x<\infty)$ を分布の他の母数と同時に推定することになる．しかし，この様な母数の入れ方では分布によって α をデータの最小値に近付けることでいくらでも likelihood が大きくなることが知られている．

本研究では，従来用いられてきた2母数対数正規分布ではなく，岩瀬・平野（1990）が定義した2母数対数正規分布に対して，既存の母数の入れ方とは異なる分布の立ち上がり点の推定も行えるリパラメトリゼイションにより3母数対数正規分布を新たに提案した．母数の推定量として，moment estimators, modified moment estimators 及び maximum likelihood estimators を示し，modified moment estimates 及び maximum likelihood estimates に対するアルゴリズムを示した．

参 考 文 献

岩瀬晃盛，平野勝臣（1990）．べき逆ガウス型分布とその応用，応用統計学，**19**，163-176．

地理情報を用いたデータリンケージと統計解析

馬 場 康 維

1. 地理情報の利用

コンピュータの発達とともに画像処理技術の進展も著しく，種々の画像情報のデジタル化が進んだ．この様な環境のもとで，紙の上の情報であった地図が新たな意味を持ち始めている．画像情報のデジタル化により，地図の数値化が進み種々の地理情報を組合わせて利用することが可能になり，行政から日常生活まで様々なところで地図の利用が盛んになりつつある．

ところで，近年様々な統計数値情報が蓄積され，一般の利用に供されるようになってきた．官庁統計データもその中の一つである．例えば，国勢調査の集計結果や，学校基本調査の集計結果など各種の報告書の内容が，数値データとして磁気媒体等で販売されている．官庁による統計情報の多くは，都道府県あるいは市区町村単位の集計からなっている．この様な情報を地図上で表現するという利用法が地理情報の利用の中で最もポピュラーなものであろう．進学，就職，結婚等，人の社会移動を引起こす行動をあげるまでもなく，多くの社会科学的問題には地理的環境を考慮する必要があり，地域差のある事項の特徴を探索的・発見的にとらえるのにこの様な情報の視覚化が有効である．

2. データリンケージ

社会現象の解明には，数値データを地図上で表現するだけでなく，人口，環境等，当該地域の特性を

もからめた分析が必要である．しかし，残念ながら，各種の統計情報を結合して利用するのは，実際には難しい．これは，データの整合に時間を割かれるためである．そこで，「統計情報の統合化」が地理情報利用の重要なポイントになる．

統計情報を地図上で表現するには，都道府県境界等の図形を描きこれに統計情報を割り付ける．そのために，境界を表わす図形に識別コード（図形ID）を付ける．したがって，図形IDを共通コードとして用いることができれば，出処の異なる統計調査の結果を結合して利用できることになる．いわば，境界で囲まれた一つの領域が個体に対応し，人口，面積等の情報を，その個体の属性と見なすことに対応する．

エリアとして用いることのできるものには 1）都道府県，2）市区町村，3）街区，4）郵便番号，5）電話局番，6）メッシュ，等がある．

統計情報の中には，エリアの属性ではなく位置（点）の属性と考えた方が良いものもある．教育に関連した例でいえば，学校，スポーツ施設等，施設に関連した情報がこれにあたる．学校の生徒数，教員数等は，いわば，点（これはエリアの大きさに比較して点とみなせるという意味である）に与えられた属性と考えれば良い．

エリアにせよ点にせよ，図形情報には包含関係のあるものがある．施設の情報は，それが大きな施設でない限り，街区に含まれ，したがって，市区町村，都道府県に含まれる．この様な包含関係を扱えるシステムを構築することによって，地理情報を利用し，出処の異なる種々のデータを結合利用することが可能になる．

データからのモードの推定（2）

川 合 伸 幸

昨年に引き続き，連続データからのモードの推定について考えた．

さて昨年の年度研究報告会で報告したモード推定量は，variable partition histogramによるdensity estimateで$k=1$としたもののモードになっていることがわかった（Variable partition histogramについてはIzenman（1991）参照）．Devroye and Gyorfiの証明によりこのモード推定量は，真のdensity fに関する条件なしに真のモードに対するL_1の意味でのstrong consistent estimatorになっていることがわかる（Izenman (1991)）．

さて，昨年の推定量はこのように理論的にはすっきりしているが実際的ではない．第1点は，現実のデータというのは有効数字3桁か4桁に丸められているもので，そういうデータに対してはこのモード推定量は一意に決まらないことが多い．そこで一意に決まる範囲でなるべくwindowを狭くとるいくつかの試みについて考えた．しかしbiasは確かに小さくなるがvarianceが大きくなるということがある．これはいわばkernel的な方法の宿命であって，windowを小さくとっているためいわばその外のデータを捨てていることになりvarianceが大きくなるのは当然のことなのである．そこで一意に決まる範囲でなるべくwindowを狭くとるのではなく，左右対称とみなせる範囲でなるべくwindowを広くとらねばならないと考えるようになった．分布は一山と仮定し，左右対称性を検出するために正規分布をあてはめ，一つの正規分布と孤立点として分布の特徴を抽出することを考えた．正規分布をいくつにとればよいのか，どれを孤立点とするかの判断には，自由パラメータの数をいくつにしたらよいかの問題が起こり，そのためAICを利用することにした．そして一番たくさんのサンプルを含む山の平均をモードの推定量とすることを考えた．これをn-modeとよぶ．また分布は一山と仮定するなら最小間隔でつながっているサンプルの平均値をとるというシンプルな推定量が考えられる．これをs-modeとよぶ．Weibull分布$\alpha=1, M=2, \gamma=0$でシミュレーションをやってみて，MSE（Mean Squared Error）を評価したところ，中央値が一番良くs-modeは中央値にわずかに劣り次いでn-modeという結果になった．しかし，中央値は明らかに一致性はなく，一致性にこだわるならs-modeを使うべきであろう．