

学統計学部機関誌 *Statistica* に受理された。

此の拡張された所得分布の特性量は、パレート係数や局所エンゲル弾力性係数の拡張形式を与えるものであり、ジブラ係数がジーニ係数に標準形式の正規分布 $\Phi(x)$ の逆関数を適用して表現されるように、相対的外積統計量にガンマ分布 $\Gamma(x)$ やベータ分布 $B(x)$ の逆関数によって表現される。この点における共通性から、此等のパラメータ表現形式は今後「グレート統計量」として所得分布の各種特性量の推測に活用されることが期待され、且つ展望出来る。

D & D における画像データの扱い

丸山直昌

D & D (Data and Description) は統計データをリスト形式のテキストファイルによって記述するデータ記述文法である。これは慶応大学の渋谷政昭・柴田里程両氏によって提案され、共同研究のテーマとなっている EJDA (統計データの電子ジャーナル) でもデータ記述の標準方式として位置付けている。D & D では統計データを特定のコンピュータソフトウェアには依存しない形で表現し、人間も読むことができることを目標としているため、内容としてはテキストデータとして表現できるデータのみを扱ってきた。ところがここ一・二年、画像データを扱うコンピュータのハードウェア及びソフトウェアが急速に発展してきており、D & D 文法の面でも画像データの取扱いは避けて通ることができない問題となってきた。

画像データは一般にデータ量が多く、またテキストデータとして表現することは適切ではないので、D & D の文法にある外部データ宣言 (external 宣言) を用いてバイナリーデータとして扱う他はない。しかし画像データをバイナリーデータとして表現する方法は現在多くの方式が提案されていて、どれも決定版とは言い難い。又 D & D において画像データを扱う場合、(1) データ本体の説明などの必要上データ記述の一部として画像データを取り入れる場合、(2) 画像データそのものをデータ解析の対象として扱う場合、の2通りがあり得る。これらの問題点をふまえて D & D における画像データの取扱い方法について現在研究中であり、関連ソフトウェアの設計と試作を行っている。

調査と解析の一話題

坂元慶行

たとえば回帰分析で多項式回帰モデルを用いたとすると、低次のモデルでデータの構造を表現しきれない場合、不適切な次数の上昇は推定値の不安定を招いたりパラメトリック・モデルに特有のくせをデータに押しつけることになり、望ましい結果は得られない。このような場合、推定値が滑らかに変わるといふ条件の下で尤度を最大化することによってパラメータの推定値を決める方法が有効である。問題は滑らかさのウェイトをどう決めるかであるが、赤池は、この問題をベイズ型モデルの事前分布のパラメータ、いわゆるハイパー・パラメータの選択の問題と見なして、周辺尤度最大化による決定法を提案した。これを基礎にした“ABIC 最小化事後モード法”とでも称すべき推定法は多くの現実の問題に適用され、極めて有用な解析結果を生み出してきた。

ところで、“ABIC 最小化事後モード法”は、通常、冒頭で述べたような状況、すなわち、パラメータ θ でデータ x を規定するモデル $f(x|\theta)$ の推定が問題であるが、推定法設計上の技法としてハイパー・パラメータ ω をもつ事前分布 $\pi(\theta|\omega)$ を想定し、この事前分布を利用することによって $f(x|\theta)$ のよい推定値を得ることをめざすという状況で用いられる。しかし、ABIC の定義式の第1項の尤度 $\int f(x|\theta)\pi(\theta|\omega)d\theta$ は $f(x|\omega)$ と書けるから、ABIC は、パラメータ ω を最尤法で決めたときの $f(x|\hat{\omega})$ のよ

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值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がり点 (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、このよう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かでも 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ーションにより 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クージュと統計解析

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。このような環境のもとで、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。このよう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. データリンクージュ

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