

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 Györfi 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れ、そのため AIC を利用することにした。そして一番たくさん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。

参 考 文 献

Izenman, A.J. (1991). Recent developments in nonparametric density estimator, *J. Amer. Statist. Assoc.*, 86, 205-224.

統計メタ情報ナビゲーション・システムについて

大 隅 昇

研究所内で収集・蓄積されてきた統計情報資源（共同研究情報，図書登録情報，各種出版刊行物，研究所の歴史・沿革等を含む諸活動など）は，それぞれが個別的に利用できても，保存形態やメディアの種類が多様である。しかも各情報の質や密度が異なるために，相互に関連づけて利用するという環境が十分とはいえない。これら統計情報資源の共有化を図ることは重要な課題である。この要請に応える支援システムとして，昨年からの仕様の検討や試作を進めてきた“統計メタ情報ナビゲーション・システム”の開発を行なった。開発に際して，所内統計情報資源を，そのメディア形態，情報相互の関連，管理状況等にしながら整理区分し，これに基づきシステムの開発指針を検討した。このとき，既存の統計情報資源取得方法や利用環境を継承保持しながら，集約し共有化することが重要であるとの視点から，ハイパーテキスト化やハイパーメディア環境の利用を前提とした設計指針を立案した。統計情報資源をハイパーメディア環境上に実現するために，① ハイパーテキスト化の技法を用いて多様な形態の統計情報資源を共有化する，② 統計情報資源の重要度に応じて段階的にデータベース化を進める，③ ネットワーク機能を用いてデータベースの一元化を行なう，などを本年度の方針としてシステムの基本部分を開発した。開発システムの特徴として，① データベース機能を，主として文字情報を扱うテキスト・データベースと，画像（動画，静止画）・音声を扱うイメージ・データベースとに分離した管理方式（デュアル・システム方式），② ワークステーションをサーバとしパーソナルコンピュータをクライアントとするクライアント・サーバ・ネットワーク環境の導入，③ リレーショナル・データベースによる統計情報資源の一元管理，④ システム利用モードの設定機能（利用者・参照者モード，管理者モードの切り換えによるデータ保護管理，データベースの一貫性保持），⑤ ヒューマン・インターフェースへの配慮，とくに統計情報資源の全容を概観する機能（ハイパーブラウザー），情報相互の関連を照会するナビゲーション機能やヘルプ機能，簡易エディター機能，等を挙げることができる。この支援環境を利用して，希望する統計情報資源（たとえば共同研究情報，図書情報，刊行物等のデータベース）を同一プラットフォームからアクセスできるだけでなく，研究所来訪者への情報サービスなどが可能となる。今後は，本システムを強化し，多数クライアントへの情報サービス，パーソナルコンピュータへの移植用の標準モジュールの開発，広域ネットワーク・サービス技法の研究等を進める。

HALBAU の 開 発 に つ い て

高 木 廣 文

HALBAU（はるぼう）は当初，柳井・高木（1986）に掲載されている多変量解析用のプログラム集として，その作成が計画されたものである。しかしながら，サブプログラムだけでは何の役にも立たないので，これにデータ入力，解析のための変数指定メニュー，分析結果の出力などのプログラムを加え，さらに多変量解析以外のデータ解析のための基本的な手法を多数加えて，統合的な統計学パッケージとして開発された。

開発当初の主な目的は，大型計算機などの設備が使用できない研究者，学生などに，同等の統計学の解析をパソコン上で行えるような環境を提供することであった。HALBAUの開発にあたって，使用者