データからのモードの推定

川 合 伸 幸

初等統計学に残された問題の一つに，データからどのようにしてモードを推定したらよいかということがある．情報を集約する仕方としてモードをとり，モードを比較するのが適切な場合というのはかなりあるので，データから直接モードを推定する方式が確定すれば有用な道具になると考えられる．

モードは分布の中心を示す．そこで次のような素朴な推定量が求まる．各点 x_0 を中心とする長さ $a>0$ の区間 $(x_0-a,\ x_0+a)$ を考え，この中に含まれるデータ点の数がいちばん多くなるような x_0 を中心あるいはモードと考える．これは長さ a に依存するので $Mo(a)$ と書くことにする．まず前提として，母集団の分布 $F(x)$ は連続な密度関数 $f(x)$ をもち，モードは一意に与えられるものとする．さらに数学的議論には関係ないが実際的な観点から密度関数は一山だと仮定する．そして $f(x)$ はモードのまわりで2階微分可能とする．$X_1, X_2,\dots, X_n$ を $F(x)$ からのランダム・サンプル，$X_1, X_2,\dots, X_n$ から得られる経験分布関数を $F_n(x)$ とする．漸近的には $f(x)$ が mode の a 近傍で左右対称でないと，$Mo(a)$ はかたよりをもつことがわかる．そして $a\to 0$ とするとき $Mo(a)$ は母集団分布 $F(x)$ のモードに収束する．そこで，

1° a はできるかぎり0に近い

2° データから $Mo(a)$ は確率1で一意に定まる

という二つの条件を満たすように a を求める．それに対応するモード推定量 $Mo(a_0)$ を求めると，

$$\min_{1\le k\le n-1}\{X(k+1)-X(k)\}$$

(ただし $X(k)$ は $X_1, X_2,\dots, X_n$ の順序統計量) を与えるような k を k_0 として，

$$Mo(a_0)=\{X(k_0+1)+X(k_0)\}/2$$

となる．$X(k_0+1)-X(k_0)$ は確率1で0に収束するから，$Mo(a_0)$ は Chernoff (1964) より一致性があることがわかる．一方，漸近分散は複雑な量になり，推定量の変動を評価するにはブートストラップ法を使うことが考えられる．

参 考 文 献

Chernoff, H. (1964). Estimation of the mode, *Ann. Inst. Statist. Math.*, **16**, 31-41.

統計情報資源の統合化の試み
—— 統計メタ情報ナビゲーション・システム ——

大 隅 昇

研究機関で収集・蓄積された統計情報資源（共同研究情報，図書登録情報，各種の出版刊行物，統計ソフトウェア，各種調査データなど）は，それぞれは自由に利用できても，保存形態やメディアの種類が多様で相互に関連づけて利用する支援環境が必ずしも十分とはいえない．これら統計情報資源の統合化を進め共有化を図ることは重要な課題である．その共通した特徴として，① 統計情報の形態が多様なメディアの上にある，② 統計情報が統合化されていない，③ 統計情報が一元化されず散在している，④ 各情報が断片的で相互の関連が必ずしも明らかでない，などがある．このことから，統計数理研究所を例として，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統計情報資源活用のためのナビゲーション・システムの基本構想を検討した．この際，既存の統計情報取得の方法や利用環境を