が統計学やコンピュータの知識があるものと仮定しなかった．すなわち，平均値は分かるが，分散は知っている程度，コンピュータの電源は入れられる程度で，使用解説書に書かれていることが理解できればよい，と考えた．データ解析の基礎的知識なしで，実際の解析が行えるように意図したものであり，いくつかの問題も生じたが，教育的な効果もあったものと考えられる．

HALBAUは基本的には，質問紙調査データなどを想定し，その解析に有効なように構成されている．HALBAUの特徴は，(1) メニュー画面の採用，(2) データ入力のエディタがある，(3) 極めて多変数を扱える，(4) 実用的なケース数を扱える，という点にあるだろう．データエディタでは最大約5,000変数，約32,000ケースまで入力可能である（ただし，この制限はパソコンの周辺機器の構成で若干変わる）．実際の解析では，それぞれの手法で制限が異なるが，多変量解析では一度に89変数，約32,000ケースという制限が一般的である．

1986年の開発当初では，機能も手法もかなり限られていたが，継続的にバグの取り除きや方法の追加などを行い，1989年にバージョン3とした（高木 他 (1989)）．以後は細かな修正を継続的に行ってきたが，大きな改訂は行わなかった．現在も自分自身で調査研究を行っているので，実際に自分で集めたデータの解析はHALBAUで行っているが，HALBAUだけで実際の研究を行うのは，現在のところ必ずしも十分でない．ユーザーの苦情や要望，実際の使用で不便な点などを考慮しながら，バージョン4の改訂作業を現在行っている．

参考文献

高木廣文，佐伯圭一郎，中井里史 (1989).『HALBAUによるデータ解析入門』，現代数学社，京都．
柳井晴夫，高木廣文 編著 (1986).『多変量解析ハンドブック』，現代数学社，京都．

平滑化について

柏木宣久

等間隔にとられた系列データに対し折れ線回帰を当てはめる問題について考察した．折れ線回帰の目的は得られたデータから最適な接合した切り替わる線形トレンドを推定することである．それ故，折れ線回帰は線形トレンドにおける構造的変化を内在すると予測されるデータのトレンド解析にとって有用である．線形トレンドの構造的変化は経済データの解析でしばしば問題とされている．また，画像のエッジを線形トレンドにおける構造的変化とみなして検知することも考えられる．本報告では，多くの構造的変化が存在する場合に折れ線回帰の目的を近似的に達成するための非ガウス状態空間モデルを提案した．あわせて線形トレンドにおける構造的変化を検知するための手続きを提案した．そして提案の方法を経済時系列データに適用した．

ベクトル系列の尤度解析

種村正美

一昨年の年度研究報告会に発表した方位相互作用の尤度推定の方法（統計数理，Vol. 38，No. 1，pp. 123-124参照）をベクトル系列の解析に適用し，特に，アミノ酸配列の解析に応用したことを報告した．

有界領域 V の空間に散布された N 個の点の位置座標 $\boldsymbol{X}\equiv(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)$ が与えられているとし，各点には方位ベクトル $\boldsymbol{S}\equiv(\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N)$ が付随しているとする（$\boldsymbol{s}_i\in\Omega, |\boldsymbol{s}_i|=1, i=1, \ldots, N$；$\Omega$ は角度空間）．いま，点 i と j の位置座標と方位ベクトルに対して相互作用ポテンシャル $\Phi_\theta(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{s}_i, \boldsymbol{s}_j)$ が働き，与えられた $\boldsymbol{X}$ および $\boldsymbol{S}$ がこのポテンシャルの下でのGibbsカノニカル分布に従うと仮定する．簡単のために，相互作用は「隣接する」点間 i, j にのみ働き，しかもそれらの距離には依存せず，二点

i, j の方位ベクトルの関数として表されるというモデルを考え，$\Phi_\theta(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{s}_i, \boldsymbol{s}_j)=f_\theta(\boldsymbol{s}_i, \boldsymbol{s}_j), (i, j)$：隣接対とする．このとき，対数尤度は $\log L=-\sum_{i<j;\,\mathrm{n.n.}} f_\theta(\boldsymbol{s}_i, \boldsymbol{s}_j)-\log Z(f_\theta; N)$ と表され，$Z(f_\theta; N)$ は規格化因子である．ここで $\sum_{i<j;\,\mathrm{n.n.}}$ は隣接対についての和を表す．また，上記の仮定により，$Z(f_\theta; N)$ には位置座標に関する積分は現われない．各点の隣接点数が一定の場合，この $Z(f_\theta; N)$ が厳密に求められることがある．いま，$\boldsymbol{X}, \boldsymbol{S}$ が直線線分 V 上に配置するとして，$|\Omega|=2\pi, f_\theta(\boldsymbol{s}_i, \boldsymbol{s}_{i+1})\equiv f_\theta(\boldsymbol{s}_i\cdot\boldsymbol{s}_{i+1})$ の場合を考えると $\boldsymbol{s}_i\cdot\boldsymbol{s}_{i+1}=\cos(\phi_{i+1}-\phi_i)$ に注意して，

$$Z(f_\theta; N)=\left[\frac{1}{2\pi}\int_0^{2\pi}\exp\{-f_\theta(\cos\phi)\}d\phi\right]^{N-1}$$

となり，規格化因子 Z が一重積分のみで表される．これから数値積分によって対数尤度は容易に計算できる．

さて，DNA や蛋白質などの分子配列は，20 種のアミノ酸の配列として翻訳されるが，アミノ酸間の類似度を物理化学的性質から特徴づけることができ，この類似度がアミノ酸の置換頻度のモデルとよく符合することが知られている．われわれは，個々のアミノ酸をベクトルと見立てて，二つのアミノ酸間の類似度を隣接するベクトルの間の角度に変換することによって，アミノ酸配列をベクトル系列に置き換えた．この考え方に基づいて，上記の尤度法をいくつかのアミノ酸配列データに当てはめた．その結果，従来行なわれている経験的方法に基づくデータ解析の結果と定性的によく一致する結果が得られることが分かった．従って，われわれの方法はアミノ酸配列の新しい尤度解析法として有用であると考えられる．

Wilcoxon 統計量の分散の推定

（客員）大阪大学 教養部　白 旗 慎 吾

分布関数 $F(x)$ と $G(x)$ を持つ 2 つの母集団を考える．Wilcoxon 統計量は代表的なノンパラメトリック統計量であり，2 つの母集団の間に差があるかどうかの検定，すなわち帰無仮説 H：$F=G$ の検定に多用されている．しかしながら，Wilcoxon 統計量は，それと同値な Mann-Whitney 統計量 U の形を見れば，F と G からの確率変数をそれぞれ X, Y とするとき，$\theta=\Pr(X>Y)=1/2$ の検定と解釈する方が自然であり，H はその特殊な場合にあたる．ただし，H'：$\theta=1/2$ の下では U の分布は F, G に関係し，そのままではノンパラメトリックな検定はできない．一方，U の漸近正規性は成立しているので，漸近的には H' の検定のためには U の分散が精度良く推定できればよい．

ここでは，U の分散，標準偏差の推定を考え，さらに，それらの推定量を用いたときの θ の信頼区間の正確さについて調べる．分散の推定量としては，最小分散不偏推定量（U 推定量，U 統計量になる），Bootstrap 推定量（B 推定量），Fligner-Policello の推定量（F 推定量），Jackknife 推定量（J 推定量）の 4 つを考える．これらの推定量の平均 2 乗誤差は分布を指定すれば厳密に計算可能であるが，それには大変な労力が必要となるので，比較はコンピュータ・シミュレーションを用いる．さらに，これらの推定量の平方根を U の標準偏差の推定量とし，標準偏差の推定量を用いて，信頼区間を構成する．

以下のことが示される．分散の推定では，B 推定量は平均 2 乗誤差の意味で最良であり，U 推定量がそれに次ぐ．ただし，その差は小さく，不偏性を重視するなら U 推定量が優れている．F 推定量は 4 つの推定量の中では最も悪い．標準偏差の推定でも，U 推定量の Bias は小さく，平均 2 乗誤差は同じ結論となる．一方，θ の信頼区間のシミュレーションによる信頼度の測定では，良さはその逆であり，F 推定量が最良であり，B 推定量はあまり良くない．B 推定量は一種の縮小推定量となっており，分散を小さく推定しているため信頼区間が狭くなりすぎているように思える．なお，信頼区間の推定では正規近似，t 分布による近似，有限修正の有り無しも考えたが，t 近似が効果的であり，有限修正はあまり効果がない．