

モンテカルロ法による客観ベイズ法

尾形良彦

ベイズ模型の適合度を測る ABIC 等には事後分布のパラメタに関する積分（規格化定数）が計算される必要がある。また時には事前確率分布をペナルティ関数から自在に定義するための規格化定数を求める積分を実行しなければならない。これらの積分は高次元（高多重）なので、計算実行可能なベイズ模型は、ガウス型とかその近似が良いものなどのように、強く制限されている。このためにモンテカルロ積分の高速化の必要性がさげばれている（統計数理研究所 5 年計画；情報資源と統計ソフトウェアの組織化）。しかし通常の（crude）モンテカルロ積分法では相当の乱数を使用しても十分な精度を得られないばかりかバイアスが無視できないことが大きな困難であった。

これらの規格化定数の積分の対数が Fisher Score 型統計量の Scale 母数に関する 1 次元積分で表されることを利用すると、不偏で十分な精度のものが安定して得られ、その様な困難が解消されることが分かった。

超パラメタを決めたとして、その後のもう 1 つの困難は事後分布から大量のパラメタを推定することである。これもガウス型ならば最小二乗法によって効率よく推定できるが、一般に非線形最適化で求めるのでは、安定した収束が期待できないことである。

そこで事後分布のモード（最大）を実現するパラメタを求めるのではなく、事後分布の平均であるところのベイズ推定量（ベクトル）を求めることにした。確率分布を定義するための規格化定数が分かっていない分布の標本を実現するシミュレーションには、マルコフ連鎖の遷移確率の性質を利用した Metropolis らのモンテカルロ法が適している。これによって、事後分布や事前分布のパラメタの平均はもとより、誤差、Fisher Score 型統計量など、必要なあらゆる情報をシミュレーション出来ることになる。

以上の精度のチェックや非線形非ガウスのベイズ模型の選択と推定法の実用性を示す幾つかの例について計算してみた。

調査実験解析研究系

集団構造の統計 —— 特異分布の解析システム ——

田口時夫

集中曲面（[2], [3] 参照）の解析結果から、次のような記述統計量が得られる。

A. 線形構造について（exact な線形回帰をもつ 2 変量分布）

$\Delta_{10}, \Delta_{11}, \Delta_{12}(=\Delta_x), \Delta_{20}, \Delta_{21}(=\Delta_y), \Delta_{22}, \Delta_{xy}$ （[1], [2], [3] 参照）などを集中基本量とする。

従来の原点の周りのモーメントを内積率と仮称すると、これらは外積率といえる演算形式をもっている。外積率に幾何学的表現を与えるものに、豊田 敬（昭和 63 年 1 月 統計数理研究所 講演会報告）のフラクタイル・グラフがある。但し、これは 1 変量に関するものであるが、容易に拡張する事ができる。

Δ_{ij} 間の和や商等を用いると、内積率間の諸演算で構成される統計量の系に対応して、自立した一連の統計量の系が得られる。正規分布や Mardia 第 1 種パレート分布が良い例証を与える。

B. 非線形構造について（定弾力性値をもつ回帰曲線を与える分布）

観測値ベクトル (x, y) の代わりに、これを標準化して $(x/\mu_x, y/\mu_y)$ を用いてその外積率を求め、その 1/2 をとることにより標準化外積率 G_{ij} の系を得ることが出来る。 G_{ij} に幾何学的表現を与えるものに筆者の集中曲面がある（[2], [3] 参照）。

(x, y) の代わりに $(\log x, \log y)$ を用いた内積率は、その間の演算によって非線形記述統計量の系を構

成するが、標準化外積率間の単純な演算はそれに対応した別種の記述統計量の系を与えることが出来る。これらの統計量について、Gibrat 分布や Mardia 第 2 種パレート分布が良い例証を与える。

参 考 文 献

- [1] Taguchi, T. (1981). On a multiple Gini's coefficient and some concentrative regressions, *Metron*, Vol. XXXIX-N, 1-2.
- [2] 田口時夫 (1984). 経済分析と多次元解析 — 新しい計量空間の形成と展望 —.
- [3] Taguchi, T. (1987). On the structure of multivariate concentration — Some relationships among the concentration surface and two variate mean difference and regressions — *Comput. Statist. Data Anal.*, 6, North-Holland, Amsterdam-New York.

ライントランセクトサンプリング

岸 野 洋 久

生物、生態学における野外調査では、しばしばライントランセクトサンプリングが行われる。トラックラインを決めてその上を走り、そこから発見されたものをサンプリングして行く。対象が不均質に分布しているとき各場所で密度に比例してサンプリングされるので、偏りのない推定が保証される。予め設定された幅の中をすべて調べるストリップトランセクトに比し実施が容易であり、船や飛行機、車による調査に適している。

これをトラックラインの決定を第一段とし、そこからの発見を第二段とする二段抽出とみると、その特性、注意すべき問題点が整理される。第二段の抽出率は未知であるが、多くの船を用いた大規模調査から対象の地理的分布を推定する場合や、資源の経年変化を見る場合には、この抽出率の船間比較が必要となる。トラックラインからの横距離別の発見確率を $g(y)$ とすると、最終的な抽出率は、 $2L \cdot \int_0^L g(y) dy / A$ である。ここで、 L はトラックラインの長さ、 A は対象の分布する面積である。 $g(0)=1$ が保証される場合は発見の横距離分布からこれを知ることができる。そうでないときは二人の観察者が独立に観察し、二重発見率を調べるなど別に付加情報が必要となる (Kishino (1988))。

推定量の精度は第一次抽出単位 (部分トランセクト) の間の分散により評価される。資源密度 D の推定の場合、 $\text{var}(\hat{D}) = \sum_{i=1}^k L_i (\hat{D}_i - \hat{D})^2 / L(k-1)$ として評価される。 L_i, \hat{D}_i はそれぞれ i 番目の部分トランセクトの長さ及びその推定密度である。発見数が少ないときはジャックナイフ法を用いる。部分トランセクトをどの程度の大きさにするかは解析の際に事後的に決められる。系統抽出である為、これをパッチスケールより小さく取ると、部分トランセクト間で正の相関を持ち、分散を過小評価してしまう (Kishino (1987))。

参 考 文 献

- [1] Kishino, H. (1987). Notes on sampling design and evaluation of the survey of wildlife population in the contagious distribution, Paper SC/39/Mi 6 presented to the IWC Scientific Committee, June 1987.
- [2] Kishino, H., Kasamatu, F. and Toda, T. (1988). On double line transect method, *Rep. Int. Whal. Commn.*, Vol. 38 (in press).

組み合わせ的階層分類法における距離の性質

大 隅 昇

階層的分類法とは、分類対象間の距離 (一般には類似度・非類似度) にもとづく分類操作のことである。この種の分類法のいくつかの手法 (少なくとも 9 つの手法) は、クラスター生成過程における距離