これらのすぐれた統計学者はベイズ的方法の良さを十分認識していた．しかしその実用化には成功しなかったのである．

この問題に対して，エントロピー最大化原理は容易に解答を与える．客観性あるいは情報有効性の獲得は，多くのモデルの比較結果を通じて実現される．ここではそれぞれのベイズモデルの対数尤度あるいはエントロピーの推定値が，客観性あるモデル比較上の情報を提供することになる．

ベイズモデルに不確定なパラメータが存在する場合にこれを最尤法によって決定することにすれば，

$$\begin{aligned}\mathrm{ABIC}=&(-2)\log \max \text{likelihood}\\&+2(\text{number of parameters})\end{aligned}$$

によって情報量規準を定義することができる．ただし，ここに likelihood はベイズ模型の尤度で

$$\int g(x \mid \theta,\ S_1)p(\theta \mid S_2)d\theta$$

によって定義される．$g(x \mid \theta, S_1)$ はデータ分布，$p(\theta \mid S_2)$ は事前分布，S_1, S_2 は未知パラメータである．

ABIC の実用上の有効性は，このシンポジウムで報告される多くの例がこれを利用して満足すべき結果を与えていることから明らかである．ベイズ的方法の実用化という従来の統計学の枠組では解決し得なかった問題に対して組織的な接近を可能にするというこの事実は，エントロピー最大化原理が従来の統計学の枠組みを超える新しい視点を与えていることの証明である．

統計的モデルの幾何学

東京大学工学部　甘　利　俊　一

統計学は，確率的な機構にもとづいて発生したとみなせる観測データをもとに，もとの確率的な構造についての推論を行なう科学である．確率変数を x とし，その分布密度関数を $p(x)$ としよう．また，適当な正則条件を持つ密度関数全体の集合を S としよう．確率変数 x の k 個の実現値 $x_1, \cdots, x_k$ をもとにして，そのもとにある S の一つの要素を求めるのが推定の問題である．このとき，集合 S はどんな構造をしているのか，たとえば S の二つの要素 $p(x)$ と $q(x)$ の間には意味のある分離度が定義できるのかが問題となる．通常，集合 S の中から真の分布 p を探すには，S は広すぎてうまくいかない．このため，統計学者は統計的モデル M を考える．これは有限個のパラメータ $\theta=(\theta_1, \cdots, \theta_n)$ で指定される確率分布 $p(x, \theta)$ の集まりで，S の中に n 次元の部分空間として入っている．

M として何を採用するか，これはモデルを構成する問題であって，確率変数 x を生成する物理的機構など種々の先験的な知識に基づいて M を設定することになる．古典的な統計学はモデル M は正しい，すなわち真の分布 $p(x)$ は M の中に入っている，という仮定のもとで理論を展開する．しかし，通常 M は近似にすぎず，真の分布は M に入っていないことも多い．赤池の情報量基準は多数の合理的なモデルを考えて，どのモデルを選ぶべきかをデータから推論しようという，従来の枠を破る試みであった．モデル M を与えられたものとして考えるのではなく，このように一度対象化して考えるならば，統計的モデル M 自体の持つ性質を議論する必要

性が明らかになる．M は S の中でどのような部分を占めるのか，M の形状はどのようになっているのか，M の幾何学的構造は統計的推論にどのようにかかわっているのか，こうした問題に答えるのが統計的モデルの幾何学である．この場合，M の内の立場から M の構造を調べる内的幾何学と，S の幾何学との関係で M の性質を調べる外的幾何学があって，両者が密接に関連している．

微分幾何は与えられた空間を局所的に線形化してその線形幾何としての構造を調べるのであるが，この局所線形化を空間のあらゆる点で行うと共に，こうした局所線形化空間すなわち接空間を互に接続することにより，全体としての曲がった空間の構造を把握する．観測するデータの数の多い場合を扱う漸近理論では，推論は比較的正確であるから，モデル M を線形近似した接空間の線形幾何学によって理論が構成できる．通常の統計的漸近理論が美しい統一的な形をしているのはこのためである．高次の漸近理論を展開しようとなるとこのような線形化だけではだめで，線形化空間を互に結びつけていくアフィン接続という手法が必要になる．統計的モデルの空間には α-接続，とくに $\alpha=1$ の指数接続と $\alpha=-1$ の混合接続が導入できて，これらが双対的な構造を有し，どちらも統計的推論の特性と深く結びついていることが示せる．こうして，高次の漸近理論は統計的モデルの微分幾何の上に統一的に建設できる．

一方，空間 S 自体の性質はどうなっているであろうか．無限次元の空間 S 自体の幾何学を作るには未だに多少の数学的困難が避けられない．しかし，S は $\alpha=\pm 1$ のとき平担，他の α に対しても S の拡大が平担で，このことから S の二点すなわち二つの確率分布 $p(x)$ と $q(x)$ の間に α 擬距離 $D_\alpha(p, q)$ というものが導入できて，具体的には

$$D_\alpha(p, q)=\begin{cases} \dfrac{4}{1-\alpha^2}\{1-\int p^{\frac{1-\alpha}{2}} q^{\frac{1+\alpha}{2}} dx\}, & \alpha\neq\pm 1 \\ -\int p \log(q/p)dx, & \alpha=-1 \\ -\int q \log(p/q)dx, & \alpha=1 \end{cases}$$

である．明らかに $\alpha=-1$ の場合の $D_{-1}(p, q)$ は Kullback-Leibler の情報量，$\alpha=1$ の場合がその双対になっていて，これは赤池の理論の基礎になるエントロピーである．

α 平担な空間には二種のポテンシャル関数が存在し，ルジャンドル変換の核をなしている．また，α と $-\alpha$ の双対性が存在する．α 平担な空間はユークリッド空間の拡張であって，α 擬距離，$\pm\alpha$ 測地線を用いると拡張ピタゴラスの定理が成立する．ここから α 近似と α 測地線の関係が明らかにされるが，これがエントロピー最大化原理および最尤推定法の理論的な根拠を与え，理論の枠組みを明らかにする．

これらを詳細に説明するには，微分幾何学的な諸量を統計的モデルの空間の上で定義し，さらに従来なかった新しい双対接続の概念を導入しなければならない．こうすると，統計的推論の高次漸近理論や，多数の局外母数を含む場合の推定など，多くの議論が新しい立場からできる．ここでは，統計的モデルの幾何学の立場から，定常ガウス時系列の全体のなす空間 S およびその部分空間である p 次 AR 過程の時系列全体のなす空間 M_p^{AR}，p 次 MA 過程の時系列全体のなす空間 M_p^{MA} などの幾何学的構造についてふれてみよう．全空間 S は，正則なパワースペクトラム $P(\omega)$ の全体のなす関数空間に等しい．S の幾何学的構造を調べると，これはすべての α に対し α 平担であることがわかる．二つのパワースペクトラム $P_1(\omega)$ と $P_2(\omega)$ の間の α 擬距離を求めると，

$$D_\alpha(P_1, P_2)=\begin{cases}(1/\alpha^2)\int[\{P_2(\omega)/P_1(\omega)\}^\alpha-1-\alpha\log(P_2/P_1)]d\omega, & \alpha\neq 0\\ \dfrac{1}{2}\int[\log\{P_2(\omega)/P_1(\omega)\}]^2 d\omega, & \alpha=0\end{cases}$$

である．さらに，$P(\omega)$ のフーリエ変換である自己共分散系列 c_t, $(t=0,1,\cdots)$ を考えると，$c=(c_0, c_1, \cdots)$ は空間 S の無限次元座標系をなすが，これは $\alpha=-1$ の基準の直線座標系をなす．また，$P^{-1}(\omega)$ のフーリエ展開係数である逆自己共分散の系列 $\tilde{c}=(\tilde{c}_0, \tilde{c}_1, \cdots)$ を考えると，これは $\alpha=1$ の基準の直線座標系を構成する．

AR モデルの空間 M_p^{AR} を考えよう．明らかに，次数 $p(p=0,1,2,\cdots)$ のモデルの集りは入れ子構造

$$M_0^{AR}\subset M_1^{AR}\subset M_2^{AR}\subset\cdots$$

になっている．モデル M_p^{AR} の幾何学的構造を調べると，これは $\alpha=1$ の平坦空間である．さて，与えられた時系列 $P(\omega)$ を p 次の AR モデルで近似する問題を考えてみよう．近似の基準は $\alpha=-1$ 擬距離(Kullback-Leibler の情報量)最小，すなわちエントロピー最大原理を用いる．いま，p 次 AR モデルを用いた $P(\omega)$ の最良近似を $\hat{P}_p(\omega)$ としよう．次数 p を上げれば近似の度合は良くなっていく．このとき，近似度に関する次の誤差の分解定理が成立する．

$$D_{-1}(P. \hat{P}_0)=\sum_{q=0}^{\infty} D_{-1}(\hat{P}_{q+1}, \hat{P}_q)$$

ここで，

$$\lim_{q\to\infty}\hat{P}_q=P$$

さらに形をかえると

$$D_{-1}(P, \hat{P}_q)=D_{-1}(P, \hat{P}_{q+1})+D_{-1}(\hat{P}_{q+1}, \hat{P}_q)$$

が成立する．これらがピタゴラスの定理の形をしていることが直ちにわかるであろう．こうした定理が成立するのは，AR モデルの空間が $\alpha=1$ 平担で，情報量が $\alpha=-1$ 擬距離であったからに他ならない．

MA モデルについても同様のことが成立する．ただし，MA モデルは $\alpha=-1$ 平担であるから，今度は双対情報量（情報量の式で p と q とを入れかえたもの）を近似の基準として用いると，誤差の分解定理が成立する．また，Bloomfield の指数型モデルの空間は $\alpha=0$ で平担で，これは自己相対であるから実は本当のユークリッド空間になっている．したがって，$\alpha=0$ の距離(これは Hellinger 距離に他ならない)を用いて誤差の分解定理が成立するが，これはピタゴラスの定理そのものである．この他 ARMA モデルの空間を考察すると，これは複雑な大域的トポロジーを持つ空間であることがわかる．

古典的統計学は統計的モデルを与えられたものとして，その枠組の内で理論を展開した．しかし一度モデルを相対化し客体化してみれば，モデルそのものが研究の対象となるべきことがわかる．赤池の理論はこれを追求したものである．モデルそのものの研究にあたっては，幾何学的考察が重要でありまた有用である．筆者は情報幾何学という新しい研究方法を提唱するものであり，その一端をここで述べた．統計数理研究所の 40 周年を記念するシンポジウム「エントロピー最大化原理と統計モデル」においてこのような発表を行なう機会を与えられたことに感謝したい．

参 考 文 献

S. Amari. (1982). Differential geometry of curved exponential families-curvature and information loss, *Ann. Statist.*, **10**, 357-387.

S. Amari and M. Kumon. (1983). Differential geometry of Edgeworth expansions in curved exponential family, *Ann. Inst. Statist. Math.*, **35**A, 345-365.

S. Amari. (1984). Differential geometry of statistics—towards new developments, METR 83-1 (London Workshop on Differential Geometry in Statistical Inference), Univ. Tokyo.

S. Amari. (1985). Differential-Geometrical Methods in Statistics, Springer Lecture Notes in Statistics, vol. **28**.

ベイズ型重回帰モデル

統計数理研究所 石 黒 真木夫

1. 問 題 意 識

最近，Akaike (1980) の考えにもとづいた ABIC 最小化法によるベイズモデルの利用が活発に行われるようになってきた．

ベイズモデルは性格を異にする2つのモデル，データ分布のモデルとこのモデルのパラメータの分布（事前分布）のモデル，で構成されるが，事前分布のモデルとして現実に利用されているのはいわゆる smoothness prior と呼ばれる形の，パラメータの間に自然な順序がつけられて，その順序にパラメータの値が滑らかに変化する場合のものだけである．

赤池の方法は，すくなくとも形式的には，いかなる事前分布を持って来ても使える一般的な形をしている．従って，様々なデータ分布と事前分布を組合わせることによって多種多様の統計的方法を構成することができる．しかし，そうして導かれる統計的方法のすべてが実用的なものであるとは限らない．

ABIC 最小化法はこれまで成功をおさめてきているが，これは単にベイズモデルを利用したからそうなったということではなく，smoothness prior という極めて特殊な形をした事前分布が何らかの意味で自然なものであったからだ，と考えられる．では，smoothness prior ではない事前分布であって，同じように自然で実用的なものは在りうるか．もし在るものならそのような事前分布を構成してみたい．

2. ベイズモデルと AIC 最小化法

smoothness prior をもちいたベイズモデルによる解析が有効な典型的な例として回帰分析がある（石黒・荒畑，1982）．ベイズモデルを利用することによってデータの大局的な動きを捕えた適度に滑らかな回帰曲線が得られる．

従来の AIC 最小化法によって滑らかな回帰曲線を得ることは，たとえば，多項式回帰における次数選択によって実現されていた(坂元他，1983)．一般に，何等かの意味であまり激しい変動をしない滑らかな推定を得ようとしている場合に「次数選択」という手段が使われるといってよい．逆にある場面で AIC が「次数選択」に使われているなら，そこではベイズモデル，