統計数理 第35巻 第1号（1987）

公開講演会要旨

統計数理研究の特質

―― その抽象性と具体性 ――

統計数理研究所　赤　池　弘　次

（昭和61年11月5日，統計数理研究所 講堂）

この講演では，統計数理の研究の特徴を具体的な例によって説明してみたい．まず，統計数理とは何かという問題を考えてみよう．統計数理という言葉は，他の学問分野の場合によく見られるような，諸外国で確立された概念の訳語ではない．研究所創設当時に，研究所がこれから取り扱うべき研究分野を規定する上で最も適切と考えられて採用されたものと思われる．したがって，統計数理とは何かという問いに対しては，過去の42年間におけるこの研究所の活動の実態をふまえて，われわれが答えを決定しなくてはならない．

そこでひとまず個人的な定義を試みると

統計数理 ≡ 統計的概念の発展を目的とする数学的理論

となる．ここで，

統計的概念 ≡ 統計的な物の見方に基礎を置く概念

である．こうしてみると，統計数理という言葉の理解の上で中心的なことは

〝統計的な物の見方〟

ということになる．ここに統計数理の研究の抽象性がはっきりと顔を出している．同時にそれは統計数理の研究の対象が，「あのもの」あるいは「この事柄」というように単一的，客観的に指示できるものではなく，人間における知識の利用という，極めて主観的な活動に関わるものであることを示している．

ここで，「統計的な物の見方」などという観念的なものは科学的な研究の対象にはなり得ないのではないか，という疑念が生じるかも知れない．ところがそれに対して数学を用いて議論を進めようというのが統計数理の研究である．

そこで「統計的な物の見方」とは何かについて極く簡単に説明してみよう．統計的な見方が必要とされる場合というのは，あるひとつの事物が提供する情報が，それだけではわれわれがある種の決定を行なうのに十分なものではないという状況のもとで発生する．このようなときわれわれは自然に

〝これまでの種々な経験との対比のもとに特定の事物を考察する〟

という態度をとる．これこそが統計的な見方であり，ここに

統計的な物の見方 ≡ ひとつの事物をある集団の一員としてとらえる見方

という定義が成立する．ここに現われる「ある集団」は，その時その事物を見るわれわれの心の枠組みを示しているわけで，その設定の仕方はわれわれのこれまでの経験に基づく知識に

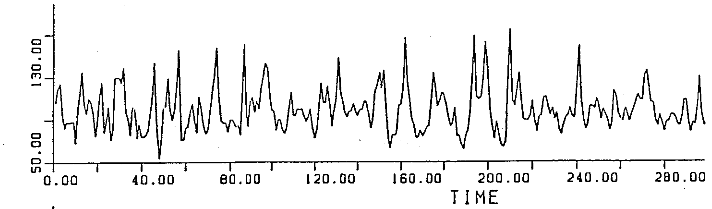

図1. 小麦価格

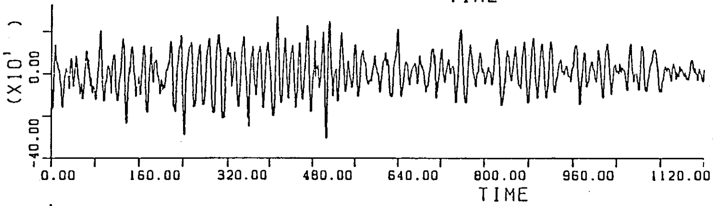

図2. ?

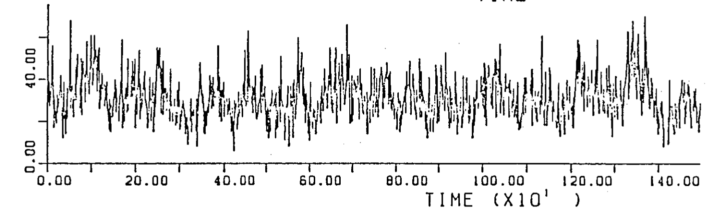

図3. ??

よって規定されることになる.

われわれの経験や知識に基づいて統計的な物の見方の枠組みが設定されるということは，統計数理の研究対象が，それぞれの事物という極めて具体的なものと切り離しては存在し得ないものであることを示す．ここに統計数理研究の具体性の根拠がある．またこのことは，統計数理の研究成果が，具体的な適用分野の研究の発展を通じて初めて具体化されることを示すものである.

さて以上の議論は抽象的であり，分り難い点が多いかと思われるので，これから具体的な例を通じて，統計数理の研究の必要性と有効性とを示すことにする.

図1, 2, 3は，時間の経過に伴い不規則に変動する現象の記録の例である．このような観測データに対して，統計数理がどのように関係するかを見ることにしよう．図1のデータは，ヨーロッパにおける小麦の価格の年々の変動を示す歴史的な時系列である．このデータに関する経済学的関心は，小麦価格の変動に周期性があるかというものであった．周期性が捉えられれば，それによって予測が可能になるか，このような変動を生み出す経済的な構造の理解の手掛かりを得ることが期待される.

不規則な変動の記録も，これを規則正しく一定の周期的変動をする波（正弦波）に分けて表現することが可能である．このためにフーリエ解析と呼ばれる数学的手法が用いられる．このように分解されたときの各成分の強さを周波数（=1/周期=単位時間の振動数）ごとに表示すると，光を各色成分に分解して得られるスペクトルに対応するものが得られる．これをピリオドグラムと呼ぶ．図4, 5, 6が図1, 2, 3に示されたデータのピリオドグラムを与えている（ただし，図1, 2, 3は原データの一部分しか示していない).

これらのピリオドグラムは何れも激しい上下の動きを示しているが，これらの変動のひとつひとつに意味があるものかどうかは容易には判断できない．そこで統計的な見方として，これらのデータが全く不規則に各周波数成分を一様に含むような時系列から一定の変換によって生じたものと考えることにする．このように考えると，与えられたデータからこの変換の構造を〝推定〟することが可能になる.

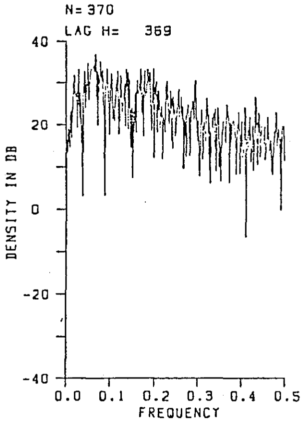

図 4.

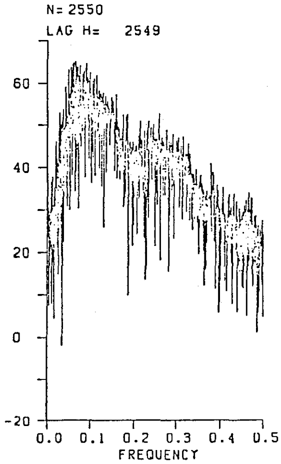

図 5.

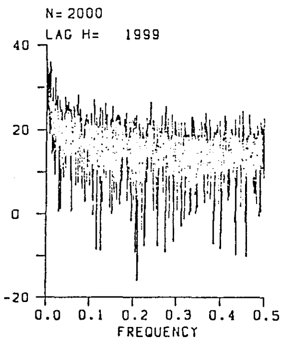

図 6.

こうして "推定" される構造が，全く不規則な系列のそれぞれの周波数成分の強さをどのように変化させるものかを示す図を描くことにより，図 4, 5, 6 よりも，もとのデータの特性を見やすく示すものが得られる．図 7, 8, 9 がそれを示す．

ピリオドグラム（図 4, 5, 6）から見ると，図 7, 8, 9 に示される "推定されたスペクトル" はそれぞれの時系列の特性を良く示しているように見える．特に図 7 は，小麦の価格の時系列に特定の周期的成分は含まれず，広い範囲にわたって成分が分布していることを示している．これを更により詳しい構造を用いて推定した結果が図 10 で，これは全く滑らかな分布を示している．これらの結果は，数多くの構造を想定して得られるものの中から，情報量規準 AIC と呼ばれる評価値によって最良な結果を選び出したものである（参考文献参照）．ただし図 8 だけは別の評価値によっている．

さて，図 5 のピリオドグラムに対応するスペクトルのひとつの推定値が図 8 に与えられているが，これは図 4, 5 のピリオドグラムの類似から当然期待されるような結果に見える．実はこの結果は，推定に用いられたモデルの選択の基準に BIC と呼ばれる評価値を用いて得られたものであるが，もし図 7 の場合と同じように AIC によって結果を求めてみると図 11 のようになる．極めて数多くの細い突起が見られる．これは滑らかな図 8 の結果とは著しい違いである．ではこの何れがより有効な情報を提供していると言えるであろうか．

実はこの場合の原データ（図 2 にその一部が示されている）は，ある地震の記録の一部分で

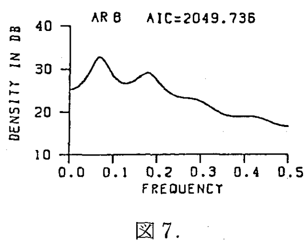

図 7.

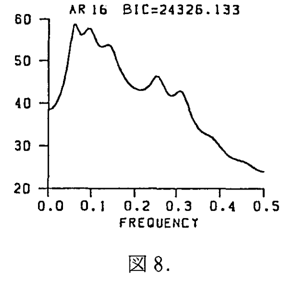

図 8.

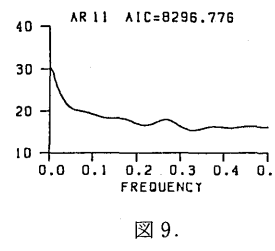

図 9.

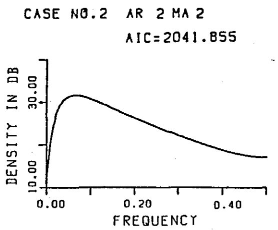

図10.

図11.

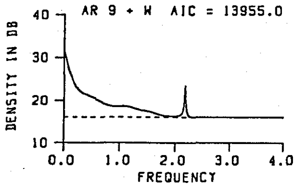

図12.

ある．このような記録は，地球の振動特性を示す信号を含み，これはスペクトルに線のように細い山を多数示すはずであることが知られている．その意味では図11の結果の方がより良いものと判断される．AICの方がBICより良い推定結果を与えたものと言えるであろう．

この例の場合のように，データに関する具体的な知識が得られると，それによって統計的方法で得られた結果の意味が明らかになる．また統計的方法の適，不適も容易に判断できる．ここで注意すべき点は，AICのような客観的な指標を用いることによって初めて図11の結果が得られたということである．小麦の価格のピリオドグラム（図4）とそのスペクトル（図7あるいは図10)のような結果に慣れている人にとっては，図5のピリオドグラムを眺めるだけでは，図8のような滑らかなスペクトルを想定するのが自然であろう．この例は，単なる経験を超えるものを統計数理が生み出していることを示している．

それではここで統計的な物の見方がどこに使われているのかというと，今の場合，われわれが現在眺めているデータが全く不規則な動きをする系列から，一定の変換によって作り出されたとみなすところにある．このような見方を採用すれば，サイコロを投げるような確率的な仕組みを用いて同じようなデータが作られることがわかる．現在のデータをこうして得られる類似のデータの集団の一員であるとみなすことによって，前に述べたAICのような統計数理的な概念を用いたデータの処理方法の展開が可能になる．このようにして，統計数理の研究における基礎概念として，確率の考えが用いられることになる．

ここで図3に示されるデータについてみることにしよう．図6のピリオドグラム，図9のスペクトルがこれに対応するものである．この結果は再び小麦の価格の時系列の場合と類似の結果を与えている．この場合もAICがモデル選択の基準として用いられている．

さて，この結果は満足すべきものであろうか．図9を見ると，他のデータに比べ，このデータのスペクトルは平坦に近いことがわかる．原データの図3を見ると，全く不規則な大きな変動を含んでいることがわかる．

そこでnという時刻におけるデータの値$x(n)$が

$$x(n)=y(n)+z(n)$$

という構造を持っているものと考えることにする．ただしここに，$y(n)$は比較的規則的な動きを示す部分であり，$z(n)$は全く不規則であり平坦なスペクトルを持つものとする．ここで，$y(n)$に前の二つの例の場合の図7, 8, 11を求めたときと同様な確率的な構造を想定して$x(n)$のスペクトルを求めてみると，図12が得られる．もちろん，ここでもAICが用いられてモデルの最終的な決定がなされている．

さて，この図では横軸（周波数軸）上，中央のあたりに鋭いピークを見ることができる．こ

こで改めて図 9 のスペクトルを眺めてみると，確かに同じ位置に極くわずかのゆるやかなスペクトルの膨らみがあることに気づく．しかしこれと図 12 とでは結果の解釈には大きな違いが生じる．音になぞらえていえば，図 9 のようなスペクトルは潮騒のようにごうごうと鳴る場合であり，図 12 はその中に一定のピーと鳴る音が含まれる場合ということになる．ピーと鳴る音が有るか無いかでは，このデータの示す現象に対するわれわれの理解に大きな変化が生じることになる．

図 3 のデータが天空の一角からやってきた電波の記録であったと想定してみよう．その電波の中に，正確な発振器からでなければ得られないような信号が混じっていることが知られたとしたらどうであろう．その信号の源は何であるかという科学的な問いが発せられることになるであろう．科学よりも SF 物語の方が好きな人には，これは異星人からの通信かもという夢と期待を抱かせることになろう．ここでは統計数理という抽象的な枠組みを利用して，小麦価格の変動や地震動の解析に用いられたのと同様な方法が，全く別の分野での新しい科学的な仮説の創出に力を貸すことになるわけである．

さて図 3 のデータは何から得られたものであろうか．これは実際の貴重な科学的観測から得られたものであり，図 12 の結果はより深いデータの理解への探索の第一歩を示すものである．この研究は現在まさに進行中のものであり，その詳細をここで明らかにすることはできない．これに不満を感じる人々は，統計数理の研究がその抽象的な側面だけでは終らないものであることを容易に認めることができるであろう．具体的な応用とは切り離せないという特質を統計数理の研究は持っているのである．一日も早くこのデータの良い解析結果が報告されることを願いつつ，この話しを終ることにしたい．

〔付記〕

この講演で示された数値は，北川源四郎助教授，田村義保助教授，荒畑恵美子助手等の協力を得て求められたものである．一部は昭和 61 年度統計数理研究所共同研究（61-共研-17）の途中結果である．なおこの講演の詳細な記録は講談社ブルーバックス「科学の中の統計学」（赤池編）に収められている．

参考文献

赤池弘次（1976）．情報量規準とは何か，数理科学，**153-3**.

坂元慶行，石黒真木夫，北川源四郎（1983）．情報量統計学，情報科学講座，**A・5・4**，共立出版.