

統計物理と統計数理

統計数理研究所 田村義保

統計数理の研究で扱われる数式・数学公式と統計物理の研究で扱われる数式・数学公式の類似性について講演する目的で題目を決定した。実際に、話した類似性はテプリッツ行列が時系列解析でもでてくると言うことのみであった。時系列データが与えられた時の解析方法を中心話題としての講演であった。表題とは異なるが、統計物理の研究者には、時系列データの解析方法を十分に御存知でない方も多いと思われるので、知識の交流という意味で、自己回帰モデルを用いた解析について説明する。

図1で表わされるような時系列データ

$$x(1), x(2), \dots, x(n) \quad (1)$$

が与えられたとする。このようなデータを解析するために、良く用いられるのが、次に示す自己回帰モデル、いわゆる AR モデルである。

$$x(s) = a(1)x(s-1) + a(2)x(s-2) + \dots + a(m)x(s-m) + \varepsilon(s) \quad (2)$$

但し、 $\varepsilon(s)$ は白色ガウス雑音である。

$$E[\varepsilon(s)] = 0$$

$$E[\varepsilon(s)\varepsilon(s')] = \sigma^2 \delta_{ss'}$$

$$s, s' = 1, 2, \dots, n$$

時系列解析と言われているのは、データに合致するように、AR モデルの係数

$$a_m(1), a_m(2), \dots, a_m(m)$$

と雑音項（イノベーションと呼ばれることが多い）の分散 σ^2 を求めるための方法の総称である。ここでは、2種類の方法について説明する。まず、標本相関関数を用いてできる Yulle-Walker 方程式を解く方法について説明する。簡単化のために、データの標本平均 $\bar{x}$ は零であるとする。この時、標本自己共分散関数は次式で定義することができる。

$$C(t) = (1/n) \sum_{s=1}^{s=n-t} x(s+t)x(s) \quad (3)$$

この時、Yulle-Walker 方程式とは、次式で与えられるような方程式のことをさす。

$$\begin{pmatrix} C(0) & C(1) & \dots & C(M-1) \\ C(1) & C(0) & \dots & C(M-2) \\ C(2) & C(1) & \dots & C(M-3) \\ \vdots & \vdots & \dots & \vdots \\ C(M-1) & C(M-2) & \dots & C(0) \end{pmatrix} \begin{pmatrix} a_M(1) \\ a_M(2) \\ a_M(3) \\ \vdots \\ a_M(M) \end{pmatrix} = \begin{pmatrix} C(1) \\ C(2) \\ C(3) \\ \vdots \\ C(M) \end{pmatrix} \quad (4)$$

ここで、AR モデルでは次数を小文字の m で書き、Yulle-Walker 方程式では、大文字の M で書いているのは、誤植ではなく、自己回帰モデルの解析には、次数 m の決定方法も含まれているということを強調したいためである。すなわち、AR モデルのあてはめを行なう場合、次数の候補の最大値として M を考え、0 から M のどれが最適かの選択基準といっしょになって初めて解析方法として意味を持つことを強調しなかったためである。このための基準として、赤池情報量基準 (AIC) が用いられることが多い。今の場合、AIC は

$$AIC_{YW}(m) = n \log(2\pi\hat{\sigma}_m^2) + n + 2(m+1) \quad (5)$$

で与えられる。ここで、AIC について少しふれておく。AIC は一般に、モデル分布と真の分布との距離を測るための尺度である。物理の言葉で言えば、ボルツマンエントロピの -2 倍の推定値である。統計数理の言葉で言えばカルバック・ライブラの情報量の -2 倍の推定値である。AIC が小さい程、モデルは真のモデルに近いことを意味している。統計数理的に、より適切な表現で言えば、AIC が小さいほど、選ばれたモデルは、今考えている（候補としている）モデル族の中では、最も真のモデルとの距離が遠くないということである。次数 m を 0 から M まで変化させて、AIC が最も小さくなるようなモデルを捜せば良い。次数を変えるごとに、方程式を解くのは時間が非常にかかるので、普通は次の逐時アルゴリズムを用いる。これは、レビンソンアルゴリズムと呼ばれており、初期値として

$$\hat{\sigma}_0^2 = C(0)$$

$$\Delta_0 = C(1)$$

$$a_0(0) = 0$$

を考え、次の漸化式を用いて順次計算して行けば良い。

$$\hat{\sigma}_m^2 = C(0) - C(1)a_m(1) - C(2)a_m(2) - \dots - C(m)a_m(m) \quad (6)$$

$$= \hat{\sigma}_{m-1}^2(1 - \Delta_{m-1}^2/\hat{\sigma}_{m-1}^4) \quad (7)$$

$$\Delta_m = C(m+1) - C(1)a_m(m) - C(2)a_m(m-1) - \dots - C(m)a_m(1) \quad (8)$$

$$a_m(m) = \Delta_{m-1}/\hat{\sigma}_{m-1}^2 \quad (9)$$

$$a_m(l) = a_{m-1}(l) - a_m(m)a_{m-1}(m-l) \quad l = 1, 2, \dots, m-1 \quad (10)$$

図1のデータにARモデルをあてはめた例を述べる。1次のARモデルの場合、 $a_1(1)$ 、及び σ_1^2 の推定値は次のようになる。

$$a_1(1) = -0.3553658$$

$$\hat{\sigma}_1^2 = 1.750279$$

係数 $a_1(1)$ を用いて1ステップ先の予測を行うと図2.1のようになる。AR係数がわかれば、予測が可能になるだけでなく、パワースペクトルも次式を用いて求めることができる。

$$p(g) = \hat{\sigma}_m^2 / |1 - \sum_{k=1}^m a_m(k) \exp(-i2\pi gk)|^2 \quad (11)$$

但し、サンプリング時間 Δt は1であるとしているので、周波数 g は、 $-1/2 \leq g \leq 1/2$ である。1次のARモデルの場合、パワースペクトルは、図3.1のようになる。真のモデルのパワースペクトルを図3.4に描いている。かなり違うものになっているが、ここで注意しておきたいことは、単にFFTを用いただけの推定と違って、滑らかなパワースペクトルの推定を行なうことができるということである。図3.2、図3.3には2次のARモデル、3次のARモデルの推定結果

$$a_2(1) = -0.6198579$$

$$a_2(2) = -0.2222021$$

$$\hat{\sigma}_2^2 = 1.396609$$

$$\begin{aligned}
a_3(1) &= -0.3157684 \\
a_3(2) &= -0.3327667 \\
a_3(3) &= 0.4248625 \\
\hat{\sigma}_3^2 &= 0.968692
\end{aligned}$$

を用いて求めた、パワースペクトルを図示しておいた。この例の場合、AIC は 3 次のモデルで最小になるが、図 3.3 が図 3.4 の真のモデルのパワースペクトルを良く再現していることがわかると思う。また、図 2.2、図 2.3 に 2 次、3 次のモデルを用いた 1 ステップ先の予測値を図示しておいた。

次に最小二乗法を用いた AR 係数の推定方法について説明する。次数 m の AR モデルをあてはめる場合、次の残差二乗和を最小にするように係数を決めるのが最小二乗法であることは良く知られていると思う。

$$S_m = \sum_{s=M+1}^n (x(s) - a_m(1)x(s-1) - a_m(2)x(s-2) - \cdots - a_m(m)x(s-m))^2 \quad (12)$$

ここで、 M は Yulle-Walker 法の時と同様に候補となる AR モデルの最大次数である。分散の推定値は次式で与えられる。

$$\sigma_m^2 = S_m / (n - M) \quad (13)$$

また、AIC は次のようになる。

$$AIC_{LS}(m) = (n - M) \log(2\pi\sigma_m^2) + (n - M) + 2(m + 1) \quad (14)$$

ここで、Yulle-Walker 法の場合の AIC と、最小二乗法の場合の AIC が違うことを注意しておく。この違いは、二つの推定法で、使用するデータ数が違うことによっている。Yulle-Walker 法の場合は n 個のデータ全てを使うのに対し、最小二乗法では、先頭の M 個のデータを使わないためである。データ数が多い場合、この違いは無視でき、どちらの方法を用いても、AIC が最小になる次数は等しくなるが、データ数が少ない場合は、異なる結果を与えることもある。どちらの推定法が良いかは、計算精度を重視するなら、最小二乗法の方が良いように思えるが、小さいメモリーの計算機で計算する場合は、Yulle-Walker 法の方が良いことを述べておく。実際の計算は次のような手順で行なう。データから、次式の左辺の行列及び右辺のベクトルを求める。係数は、この式を解くことによって求めるわけである。

$$\begin{pmatrix}
x(M) & x(M-1) & \cdots & x(M-m+1) \\
x(M+1) & x(M) & \cdots & x(M-m) \\
\vdots & \vdots & & \vdots \\
\vdots & \vdots & & \vdots \\
x(n-1) & x(n-2) & \cdots & x(n-m)
\end{pmatrix}
\begin{pmatrix}
a_m(1) \\
a_m(2) \\
\vdots \\
\vdots \\
a_m(m)
\end{pmatrix}
=
\begin{pmatrix}
x(M+1) \\
x(M+2) \\
\vdots \\
\vdots \\
x(n)
\end{pmatrix} \quad (15)$$

次にハウスホルダー変換等の方法により、行列を上三角行列に変換する。この時、右辺のベクトルにも同様の変換を施す。

$$\begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1m} & \cdots & s_{1M} \\ & s_{22} & \cdots & s_{2m} & \cdots & s_{2M} \\ & & \ddots & \vdots & \cdots & \vdots \\ & & & s_{mm} & \cdots & s_{mM} \\ & & & & \ddots & \vdots \\ & & & & & s_{MM} \end{pmatrix} \begin{pmatrix} a_m(1) \\ a_m(2) \\ \vdots \\ a_m(m) \end{pmatrix} = \begin{pmatrix} s_{M+1,1} \\ s_{M+1,2} \\ \vdots \\ s_{M+1,m} \\ s_{M+1,m+1} \\ \vdots \\ s_{M+1,M+1} \end{pmatrix} \quad (16)$$

m 次 AR モデルの係数を求めるには、右辺の三角行列の $s_{ij}, i, j = 1, 2, \dots, m$ の部分のみ用いて求めれば良い。また、分散の推定値は右辺のベクトルを用いて次のように求めることができる。

$$\sigma_m^2 = (s_{M+1,m+1}^2 + s_{M+1,m+2}^2 + \cdots + s_{M+1,M+1}^2) / (n - M) \quad (17)$$

最小二乗法を用いる計算の良い点は、上三角行列への変換を一回行なうだけで、0 次のモデルから次数 M のモデルまで、全ての AR モデルの係数、分散を計算できることである。

簡単に自己回帰モデル (AR モデル) の説明を行ってきた。これだけでは不十分かも知れないが、何かの参考になれば幸いである。なお、時系列解析のためのプログラムライブラリとして、TIMSAC が統計数理研究所長赤池弘次を中心として開発されている。利用を希望される方は、田村に連絡して欲しい。教育・研究のための使用に限定しているが、無料で配付している。

Fig.1

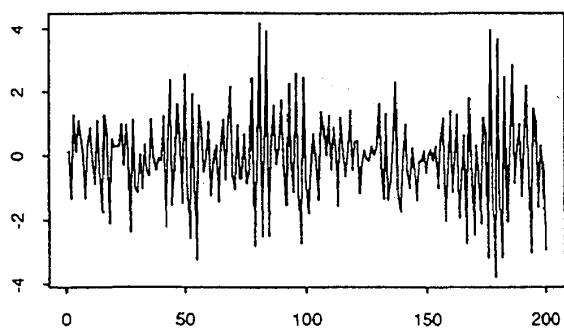


Fig.2.1

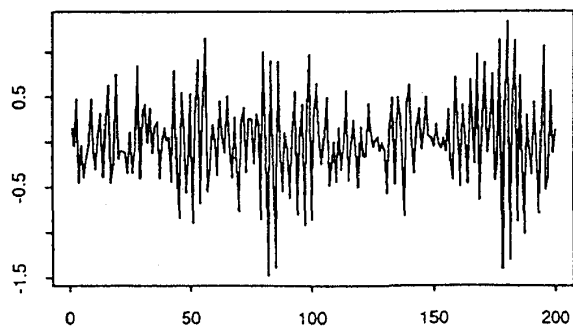


Fig.2.2

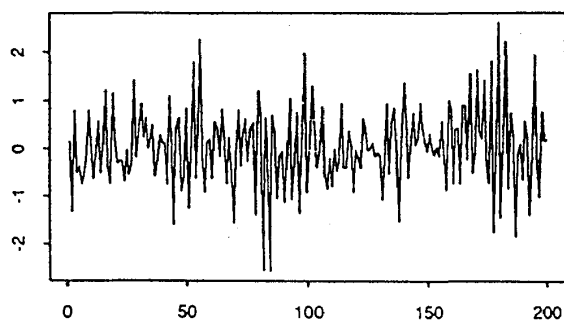


Fig.2.3

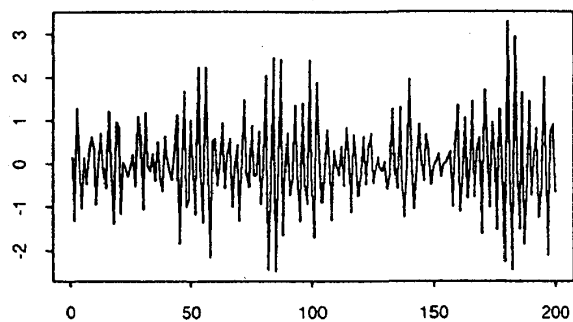


Fig.3.1

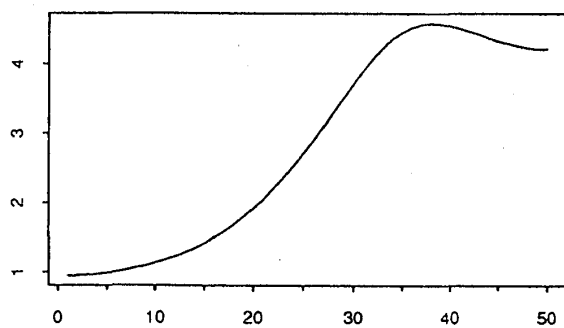


Fig.3.2

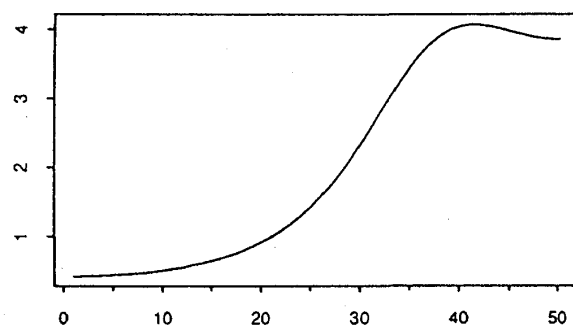


Fig.3.3

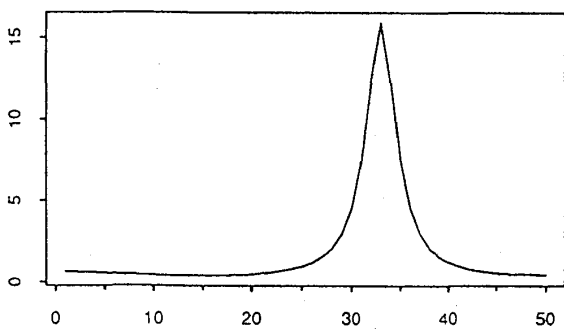


Fig.3.4

