土木学会北海道支部　論文報告集　昭和51年度

ARIMAモデルの適用範囲について

北海道大学工学部　正　員　泛中建一郎

1. まえがき

非定常時系列に対する一つの確率モデルとして、Box & Jenkins[1] によりARIMA過程（Autoregressive Integrated Moving Average Process）が提唱され、水文解析においても適用され始めている[2][3]。しかし、このモデルによって、どの様な非定常時系列をモデル化しようとしているのかは必ずしも明確になっていない。ここでは、Box & Jenkins によって与えられたARIMA過程そのものが、どの様な確率的性質を持っているかという点と、逆に、Box & Jenkins がどのような確率的性質を持った時系列をモデル化しようとしているかという二つの点から、ARIMA過程の適用範囲を調べてみる。

2. ARIMA過程の確率的性質

ARIMA過程の一般式は次のように表わされる。

$$\phi(B)\nabla^d z(t) = \theta(B)\eta(t) \qquad \text{---} \quad (1)$$

ただし　$\phi(B)$；自己回帰演算子　$\theta(B)$；移動平均演算子　∇^d；d回の差分演算子
$\eta(t)$；White noise　$\phi(B), \theta(B)$ は共に転換性を持つとする

$\phi(B)$ の転換性の仮定から

$$\psi(B) = \phi^{-1}(B)\,\theta(B) = 1 - \psi_1 B - \psi_2 B^2 - \cdots\cdots \qquad \text{---} \quad (2)$$

の様な無限演算子列で表わされる。又、$\nabla^d z(t)$ が2次過程となる様 $1+\sum_{i=1}^{\infty}\psi_i^2 < \infty$ とする。$\eta(t)$ White noise であるから明らかに $\nabla^d z(t)$ は定常ガウス過程となる。さらに ∇^d は d 回の差分演算子であるから、$z(t)$ は定常ガウス過程の d 回 summation を表わすことになる。ガウス型確率変数の線形結合はガウス分布をなすことから、$z(t)$ はガウス過程となる。しかし定常性は満されない。ガウス過程においては、平均値関数と自己共分散関数が決まれば、過程は決まるという便利な性質がある。$\nabla^d z(t)$ の平均値関数は明らかに、常に零であり、同様に $z(t)$ の平均値関数も常に零になる。したがって $z(t)$ の確率的性質は自己共分散関数によって決まる。直感的な考察から $z(t)$ の共分散関数は時間と共に増大することが予想されるが、次に具体的な例題を2つとって、その自己共分散関数を調べてみる。

連続パラメータを持つ確率過程は、離散パラメータを持つ確率過程の適当な極限として表わされるから、以後計算の容易さを考え、連続パラメータ過程として考察を進める。すなわち、$z(t)$ は適当な定常ガウス過程の d 回の積分過程となる。

3. ウィナー過程の積分過程

ウィナー過程は White noise の形式的な積分過程として表わされ、最も基本的なマルコフ過程である。その平均値関数は零で、自己共分散関数及び分散関数は[4]

$$r_W(s,t) = \begin{cases} t & \text{if } s \geqq t \\ s & \text{if } s < t \end{cases} \qquad \text{---} \quad (3)$$

$$r_W(t,t) = t$$

一方、2次過程においては、その自己共分散関数と、微分過程のそれとの間には次のような関係がある[4]。

$$r_{X'}(s,t) = \frac{\partial^2}{\partial s\,\partial t} r_X(s,t) \qquad \text{---} \quad (4)$$

ウィナー過程を $W(t)$ として、$Z(t)=\int W(t)\,dt$ とするとき（形式的には $\nabla^2 Z(t)=\eta(t)$）、(4)式を用いて $Z(t)$ の自己共分散を求める。$F(s)$, $G(t)$ を各々積分定数とすると、

$$r_Z(s,t)=\begin{cases}\frac{1}{2}st^2-\frac{1}{6}t^3+F(s)+G(t) & \text{if } s\geqq t\\ \frac{1}{2}s^2t-\frac{1}{6}s^3+F(s)+G(t) & \text{if } s<t\end{cases}$$

ウィナー過程は普通 $t\geqq 0$ で考え $W(0)=0$ と与えることから、$W(t)$の見本関数の連続性より $Z(0)=0$ したがって 初期条件として

$$r_Z(0,0)=F(0)+G(0)=0$$
$$r_Z(s,0)=F(s)+G(0)=0 \qquad \text{if } s\geqq t$$
$$r_Z(0,t)=F(0)+G(t)=0 \qquad \text{if } s<t$$

$r_Z(s,t)$ が s と t に関する対称性を考えると $F(s)=G(t)=0$ したがって

$$r_Z(s,t)=\begin{cases}\frac{1}{2}st^2-\frac{1}{6}t^3 & \text{if } s\geqq t\\ \frac{1}{2}s^2t-\frac{1}{6}s^3 & \text{if } s<t\end{cases} \qquad \text{--- (5)}$$

$$r_Z(t,t)=\frac{1}{3}t^3 \qquad \text{--- (6)}$$

4. O−U過程の積分過程

より一般的な定常ガウス過程の一つ例として、Orenstein = Uhlenbeck過程をとりあげる。[4] O−U過程は

$$X(t)=\int_{-\infty}^{t}e^{-\alpha(t-s)}\,dW_s \qquad \text{--- (7)}$$

で表わされ、その平均値関数は零、自己共分散関数と分散関数は

$$r_X(s,t)=\begin{cases}\frac{\sigma^2}{2\alpha}e^{-\alpha(s-t)} & \text{if } s\geqq t\\ \frac{\sigma^2}{2\alpha}e^{\alpha(s-t)} & \text{if } s<t\end{cases} \qquad \text{--- (8)}$$

$$r_X(t,t)=\frac{\sigma^2}{2\alpha} \qquad \text{--- (9)}$$

前節と同様、$Z(t)=\int X(t)\,dt$ とすると（形式的には $\nabla Z(t)=\psi(B)\eta(t)$）

$$r_Z(s,t)=\begin{cases}\frac{\sigma^2}{2\alpha^3}\left(2\alpha t-e^{-\alpha(s-t)}\right)+F(s)+G(t) & \text{if } s\geqq t\\ \frac{\sigma^2}{2\alpha^3}\left(2\alpha s-e^{\alpha(s-t)}\right)+F(s)+G(t) & \text{if } s<t\end{cases}$$

ここでウィナー過程と同様、$t\geqq 0$ $Z(0)=0$ なるものだけを考えると、前節と同様な手順によって

$$F(t)=G(t)=\frac{\sigma^2}{2\alpha^3}\left(e^{-\alpha t}-\frac{1}{2}\right)$$

よって

$$r_Z(s,t)=\begin{cases}\frac{\sigma^2}{2\alpha^3}\left(2\alpha t-e^{-\alpha(s-t)}+e^{-\alpha s}+e^{-\alpha t}-1\right) & \text{if } s\geqq t\\ \frac{\sigma^2}{2\alpha^3}\left(2\alpha s-e^{\alpha(s-t)}+e^{-\alpha s}+e^{-\alpha t}-1\right) & \text{if } s<t\end{cases} \qquad \text{--- (10)}$$

$$r_Z(t,t)=\frac{\sigma^2}{\alpha^3}\left(t+e^{-\alpha t}-1\right) \qquad \text{--- (11)}$$

図−1，図−2は、各々 3節、4節で求めた自己共分散関数の概略図である。これらから、2節で述べたように共分散が時間と共に増加している様子が分る。

5. トレンドの多項式表示

前節までは非定常確率過程の初期値を $Z(0)=0$ としていたが、ここで零以外の値を持つ場合を考える。d回の積分によって、

$$Y(t) = Z(t) + C_0 + C_1 t + \cdots + C_d t^d \quad \cdots \quad (12)$$

ただし、$Z(t)$ は前節までで述べた平均値零の積分過程
$C_0, C_1, \cdots, C_d$ は t に無関係な確率変数

さて、トレンドというものを考えるとき、その定義は必ずしも明確ではない。ここでは、次の様に定義を与える。

$Z(t)$ を上式と同じ確率過程とするとき、ある確率過程 $Y(t)$ が次の様に表わされるとき、$P(\cdot)$ をトレンドという。

$$Y(t) = Z(t) + P(C_0, C_1, \cdots C_n ; t) \quad \cdots \quad (13)$$

ただし、$C_0, C_1, \cdots, C_n$ は t に無関係な確率変数
$P(\cdot)$ は $\{C_i\}$ の見本値と t に関する実関数

この様に考えると、$Y(t)$ の平均値関数 $\mu_Y(t)$ とトレンド $P(\cdot)$ との関係は、$E\{\ \}$ を期待値とすると

$$\mu_Y(t) = E\{Y(t)\} = E\{P(C_0, C_1, \cdots C_n ; t)\} \quad \cdots \quad (14)$$

(12)式における C_0 を含めた t の多項式は、上の意味でトレンドである。より一般的には $P(\cdot)$ として任意の関数を考えることが出来る。たとえば、三角多項式等。

たしかに(12)式の様に表わされる確率過程は、d回の微分により定常過程となり、ARIMA過程のより一般的な表現となる。

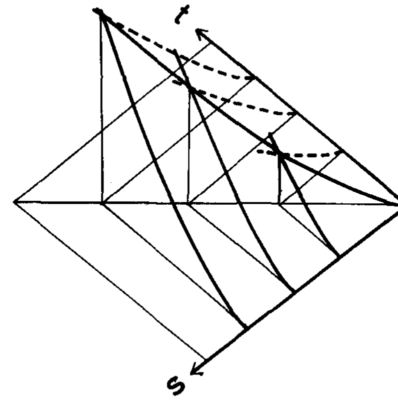


図-1 ウィナー過程の積分過程の自己共分散関数

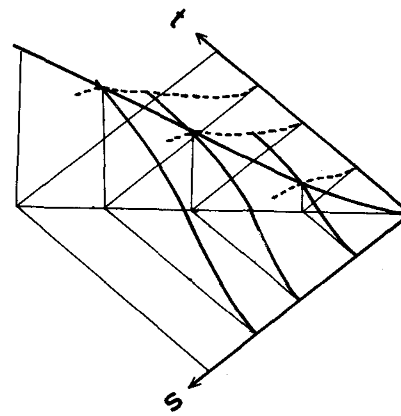


図-2 O-U過程の積分過程の自己共分散関数

6. Box & Jenkins のモデル

これまでの考察で、(1)式で与えられたARIMA過程の確率的性質はほぼ明らかになったと思われるので、次に Box & Jenkins がどの様な時系列をモデル化しようとしたかを考えてみる。図-3は Box & Jenkins がモデル化しようとした時系列の模式図である。(a)は lebel の変化を伴うもの、(b)は slope の変化を伴うものである。これらは明らかに前節までで調べてきたARIMA過程の見本関数とは言い難い。

次に lebel とか slope の概念は不明確なまゝ使われいるので、(13)式による定義を用いて、一つのトレンドとして次の様に定義する。すなわち、次の様な t に無関係な確率変数列を考える。

$C_0, C_1, \cdots, C_n \qquad D_0, D_1, \cdots, D_n$

さらに、今着目している観測区間を次の様に分割する。

$t_0 < t_1 < \cdots < t_{n+1}$ 　　ただし $T = t_{n+1} - t_0 = \sum_{i=0}^{n} D_i \quad i=0,1,\cdots,n$

これに対し、次の様なステップ関数を考える

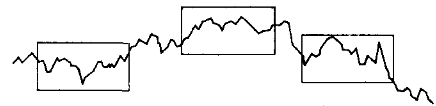
$$P(t) = C_i \quad \text{if} \quad t_i \le t < t_{i+1} \quad i=0,1,\cdots,n \quad \cdots \quad (15)$$

上式を含め、さらに適当回の積分も併せ考えると、順次 lebel, slope, etc. が定義されよう。逆に、これらの適当回の微分は、任意の重みを持ち、任意の間隔で発生するパルス列を表わすことになる。これらの確率的性質を一般的に述べることは出来ないが、例えば、交番ポアソンパルス列は、次の自己共分散関数を持つ[4]。

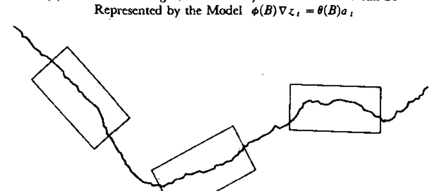
$$r(t) = \lambda\delta(t) - \lambda^2 e^{-2\lambda|t|}$$

これは、この様なパルス列に基づく lebel や slope は、適当回の微分によっても、自己共分散関数の原点以外にも値を持ち、Box & Jenkins による、lebel や slope からの自由という保障は持ち得ないことを意味する。

さて次に、実際の時系列解析における種々のパラメータの推定手順を考えると、これまでの議論では平均操作がすべて確率空間で行えていたのに対し、時間空間で行われる。この二つの平均値が一致するのは定常エルゴード過程においてだけであることは、充分注意しておかなければならないが、逆に、時間空間で平均操作を行うことによって上で述べたARIMA過程の適用に対する制限をゆるめることが出来ると思われる。すなわち、適当回の微分（あるいわ 差分）によって、定常な過程とパルス列になったとすると、特にそのパルス間隔の広い部分だけをとり出して、時間空間での自己共分散を考えれば、その値は小さなものになる。定常な過程のAR-MA 係数の推定値は必ずある分散を持ち、パルス列による影響がその分散の中に含まれる程小さいならそれは無視しても良いであろう。なを順序が逆になったが、トレンドの自己共分散関数に特に注目している理由は、Box & Jenkins によるパラメータ同定の原理は、差分により定常になった過程の自己共分散関数には トレンド（lebel や slope 等）の影響が 原点以外のところには表われないとしているからである。



(a) A Series Showing Nonstationarity in Level such as can be Represented by the Model $\phi(B)\nabla z_t = \theta(B)a_t$



図－3

(b) A Series Showing Nonstationarity in Level and in Slope such as can be Represented by the Model $\phi(B)\nabla^2 z_t = \theta(B)a_t$
（Box & Jenkins より）

結局、結論として次の様なことが言えるだろう。ARIMAモデルの適用し得る範囲としては、厳密には

イ) 例題3、4の様に、平均値関数が常に零、自己共分散関数は時間と共に増加する様なものとして、すなわち、定常エルゴード過程の任意回積分過程として、

ロ) トレンドを除いた部分のだけのモデル化を行うなら、イ)の過程に大の多項式のトレンドを含むものとして、

さらに近似的には

ハ) パラメータ同定において、その影響が無視し得る程充分に広い間隔で発生するパルス列に基づくトレンドも含めたものとして。

7. あとがき

これまでの議論で、ARIMAモデルがどの様な時系列に対し適用し得るのかを述べてきた。しかし実際に時系列解析に用いるためには、さらにいくつかの問題を解決しなければならない。その一つは予測においてトレンドを無視し得なくなったとき、どの様に推定するのか、又 ハ) の様な場合、トレンドがおよぼす影響が無視し得るかどうかの具体的な評価基準をどう設定するかといった点である。さらに、謂ゆる Seasonal な成分に対する検討も残された問題である。

参考文献

(1) G.Box & G.Jenkins ; Time series analysis — forecasting and control , Holden-Day , 1970
(2) 長谷部 正彦 ; 非定常時系列の解析と予測について , 第20回 水理講演会論文集 , 1976
(3) 〃 ; ARIMA モデルによる年降水量系列の解析について , 第30回年講 , 1976
(4) ホーエル＝ポート＝ストーン ; 確率過程入門 , 東京図書 , 1976
(5) 宮川 洋 他 ; 不規則信号論と動特性推定 , コロナ社 , 1971
(6) 浜中建一郎 ; 時系列における定常性と非定常性について ; 第30回年講 , 1976