

# 時 系 列 解 析

慶應義塾大学理工学部・理工学研究科講義録

柴田里程

2012年

# 1 弱定常時系列のフーリエ表現 (Fourier Representation of Stationary Time Series)

$\{Z_t\}$  で時間  $t$  は連続時間  $-\infty < t < \infty$  あるいは離散時間  $t = \dots, -1, 0, 1, \dots$  であるものを考える. なお実数値あるいは複素数値で,  $E|Z_t|^2 < \infty$  とする

$EZ_t = 0$  (一般には定数であるが, 0 としても一般性を失わない),  $E(Z_{t+h}\bar{Z}_t) = \gamma_h$  ( $-\infty < h < \infty$  あるいは  $h = \dots, -1, 0, 1, \dots$ ) が満たされるときの弱定常時系列 (Weakly Stationary Time Series) という.

$$L(Z) = \left\{ \sum_{k=1}^n c_k Z_{t_k} \text{ の形の線形結合 } \right\}; \quad \text{線形空間}$$

$$\text{つまり} \quad z \in L(Z) \rightarrow \alpha z \in L(Z)$$

$$z, w \in L(Z) \rightarrow z + w \in L(Z)$$

$$0 \in L(Z)$$

内積を  $\langle z, w \rangle = E(z\bar{w})$ ,  $z, w \in L(Z)$  で定義すれば内積空間となる. ノルムは  $\|z\|^2 = E|z|^2$  である.  $z_n \in L(Z)$  に対し  $\lim_{n \rightarrow \infty} z_n = z$  を  $\lim_{n \rightarrow \infty} \|z_n - z\| = 0$  で定義する.  $L(Z)$  をコーシー列が収束してその極限を空間内に持つように完備化

$$L_2(Z) \supset L(Z)$$

すれば  $L_2(Z)$  は Hilbert 空間となる. また線形作用素  $T_t$  (shift operator) を  $z = \sum_{k=1}^n c_k Z_{t_k} \in L(Z)$  に対して  $T_t z = \sum_{k=1}^n c_k Z_{t_k+t}$  で定義する.  $\{T_t\}$  は次の性質をもつ.

a)  $\|T_t z\| = \|z\|$ ,  $z \in L(Z)$  (Unitary)

b) 各  $t$  について  $T_t$  は 1 対 1, 全射で,  $T_t^{-1} = T_{-t}$

c)  $T_0 = I$

d)  $T_t T_s = T_{t+s}$  (Group).

更に,  $T_t$  は a) ~ d) の性質を保ったまま  $L_2(Z)$  の上に一意的に拡張できる. つまり任意の  $z \in L_2(Z)$  に対して  $\lim_{n \rightarrow \infty} z_n = z$  なる  $z_n \in L(Z)$  が存在するので  $T_t z = \lim_{n \rightarrow \infty} T_t z_n$  と定義すればよい.

群:  
結合律,  
逆元,  
単位元

## Stone の定理

(参照：竹之内脩「函数解析」朝倉 p149～, Functional Analysis by F. Riesz and B. SZ.-Nagy (1955, 1972), Chapter X)

ヒルベルト空間  $H$  上で定義された線形作用素  $T_t$  が a)～d) の性質に加えて (ただし  $T_t:1$  対  $1$  は必要ではない)

e)  $\lim_{s \rightarrow t} T_s z = T_t z$  ( $t$  に関する連続性)

の性質を持てば

$$T_t = \int_{-\infty}^{\infty} e^{2\pi i t \lambda} dW_\lambda$$

と一意的に表せる. ただし,  $W_\lambda$  は次の性質をもつ (射影) 作用素である.

- (1)  $W_\lambda(\alpha z_1 + \beta z_2) = \alpha W_\lambda z_1 + \beta W_\lambda z_2$  (線形性)  
 $\lim_{z_n \rightarrow z} W_\lambda z_n = W_\lambda z$  (連続性)
- (2)  $W_\lambda^2 = W_\lambda$  (射影)
- (3)  $\langle W_\lambda z_1, z_2 \rangle = \langle z_1, W_\lambda z_2 \rangle$  (エルミート性)
- (4)  $\lambda < \mu$  について  $W_\lambda W_\mu = W_\lambda$
- (5)  $W_{\lambda+0} = W_\lambda$  ( $\lambda$  に関する右連続性)
- (6)  $W_{-\infty} = 0, \quad W_\infty = I$

別証明

今後, shift operator  $T_t$  に e) の性質を持たせるために, 次の仮定をおく.

**仮定** ( $Z_t$  の  $t$  に関する連続性)

$$\lim_{s \rightarrow t} Z_s = Z_t.$$

注) 時間が  $t = \dots, -1, 0, 1, \dots$  のように離散の場合にはこの仮定は自動的に満たされる.

**Lemma** 上の仮定から性質 e) が従う.

**Proof**

まず,  $z \in L(Z)$  について e) が成り立つことを示す.  $z = \sum_{k=1}^m c_k Z_{t_k}$  と書けるので,

$$\begin{aligned} \|T_s z - T_t z\|^2 &= \left\| \sum_{k=1}^m c_k (T_s - T_t) Z_{t_k} \right\|^2 \\ &\leq \left( \sum_{k=1}^m |c_k|^2 \right) \sum_{k=1}^m \|(T_s - T_t) Z_{t_k}\|^2 \\ &= \left( \sum_{k=1}^m |c_k|^2 \right) m \|Z_s - Z_t\|^2 \end{aligned}$$

より明らか.  $z \in L_2(Z)$  については  $z_n \in L(Z)$  s.t.  $\lim_{n \rightarrow \infty} z_n = z$  が存在することから, 不等式

$$\begin{aligned} \|T_s z - T_t z\| &\leq \|T_s(z - z_n)\| + \|T_s z_n - T_t z_n\| + \|T_t(z - z_n)\| \\ &= 2\|z - z_n\| + \|(T_s - T_t)z_n\| \end{aligned}$$

の右辺で,  $n$  を十分大きくとれば  $2\|z - z_n\|^2 \leq \frac{\varepsilon}{2}$  とでき, その  $n$  に対し,  $s$  を  $t$  に十分近くとれば  $\|(T_s - T_t)z_n\| \leq \frac{\varepsilon}{2}$  とできる. つまり,

$$\|T_s z - T_t z\| \leq \varepsilon$$

にできるので, 証明された.

この仮定のもとで Stone の定理を適用すれば,

$$Z_t = T_t Z_0 = \int_{-\infty}^{\infty} e^{2\pi i t \lambda} d(W_\lambda Z_0)$$

と定常時系列  $\{Z_t\}$  は表現できることになる (Cramér, 1942). そこで, 演算子  $W_\lambda$  と同じ記号を用いて  $W_\lambda = W_\lambda Z_0$  と定義すれば  $W_\lambda$  はふたたび 1 つの確率過程で, 次の性質をもつ.

- (i)  $\lim_{\lambda \rightarrow -\infty} W_\lambda = 0, \quad \lim_{\lambda \rightarrow \infty} W_\lambda = Z_0$  ((6) より)
- (ii)  $\lim_{\lambda \rightarrow \lambda_0 + 0} W_\lambda = W_{\lambda_0}$  (右連続性, (5) より)
- (iii)  $\lambda_1 < \lambda_2 < \lambda_3 < \lambda_4$  について  $\langle W_{\lambda_4} - W_{\lambda_3}, W_{\lambda_2} - W_{\lambda_1} \rangle = 0$  すなわち  $E(W_{\lambda_4} - W_{\lambda_3})(\overline{W_{\lambda_2} - W_{\lambda_1}}) = 0$  (直交増分性)

このような  $W_\lambda$  を  $\lambda$  をインデックスとする確率過程  $\{W_\lambda\}$  と考えたとき、直交増分過程という。直交増分過程の例として有名なものに、次のようなものがある。

**例 1.** ランダムウォーク (Random Walk)

$\{X_i\}_{i=1}^n$  を独立同分布な確率変数の集まりとしたとき、 $S_m = \sum_{i=1}^m X_i$ ,  $m = 1, \dots, n$  を考えれば  $\{S_m\}_{m=1}^n$  が直交増分 (独立増分) 過程であることは明らか。

**例 2.** ブラウン運動 (Brownian Motion)

例 1 の連続時間版で、分布を正規分布に限定したものがブラウン運動である。ウィナー過程とも呼ばれる。ブラウン運動  $\{B_t\}$  とは、直交増分過程で、 $B_t - B_s$ ,  $s < t$  が  $N(0, (t-s)\sigma^2)$  に従うものをいう。特に  $\sigma = 1$  のときを標準ブラウン運動という。もちろんブラウン運動は分布の正規性から独立増分にもなる。

—— 定常時系列のスペクトル表現 ——

弱定常時系列  $\{Z_t\}$  は

$$Z_t = \int_{-\infty}^{\infty} e^{2\pi it\lambda} dW_\lambda$$

と表現できる。ただし  $E(dW_\lambda)(\overline{dW_\mu}) = 0$ ,  $\lambda \neq \mu$  である。

— Karhunen-Loeve の定理 —

定常時系列のスペクトル表現は非定常の場合にも次のように拡張できる。自己共分散関数  $\gamma(s, t) = E(Z_s \bar{Z}_t)$  が

$$\gamma(s, t) = \int_{\Lambda} f(s, \lambda) \bar{f}(t, \lambda) d\mu(\lambda), f(t, \cdot) \in L_2(\Lambda, \mu)$$

と表現できれば、直交増分過程  $\{W_{\lambda}, \lambda \in \Lambda\}$  が存在して

$$Z_t = \int_{\Lambda} f(t, \lambda) dW_{\lambda}$$

と表現できる (Abstract Inference by U. Grenendar (1981), Wiley).  
 なお,  $f(s, \lambda)$  は瞬間スペクトル (instantaneous spectral) と呼ばれる。  
 この表現の例としては, ウェーブレット (Wavelet) がある。  $f(t, \lambda) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$  にとれば, 信号  $Z_t$  のウェーブレット表現

$$Z_t = \int_{\Lambda} \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dW_{\lambda}$$

が得られる。ただし  $\lambda = (a, b)$  である (Kawasaki and Shibata (1995)).

更に,  $\{Z_t\}$  がこのような表現をもてば自己共分散関数

$$\gamma_h = \langle Z_{t+h}, Z_t \rangle = E(Z_{t+h} \bar{Z}_t)$$

も同じような表現をもつことも示そう。

$$\begin{aligned} \gamma_h &= \left\langle \int_{-\infty}^{\infty} e^{2\pi i(t+h)\lambda} dW_{\lambda}, \int_{-\infty}^{\infty} e^{2\pi i t \mu} dW_{\mu} \right\rangle \\ &= \int_{-\infty}^{\infty} e^{2\pi i(t+h)\lambda} d_{\lambda} \left( \langle W_{\lambda}, \int_{-\infty}^{\infty} e^{2\pi i t \mu} dW_{\mu} \rangle \right) \\ &= \int_{-\infty}^{\infty} e^{2\pi i(t+h)\lambda} d_{\lambda} \left( \int_{-\infty}^{\infty} e^{-2\pi i t \mu} d_{\mu} \langle W_{\lambda}, W_{\mu} \rangle \right) \end{aligned}$$

ここで

$$\begin{aligned} &\int_{-\infty}^{\infty} e^{-2\pi i t \mu} d_{\mu} \langle W_{\lambda}, W_{\mu} \rangle \\ &= \int_{-\infty}^{\lambda} e^{-2\pi i t \mu} d_{\mu} \langle W_{\mu}, W_{\mu} \rangle + \int_{\lambda}^{\infty} e^{-2\pi i t \mu} d_{\mu} \langle W_{\lambda}, W_{\lambda} \rangle \\ &= \int_{-\infty}^{\lambda} e^{-2\pi i t \mu} d_{\mu} \|W_{\mu}\|^2 \end{aligned}$$

従って

$$\begin{aligned}\gamma_h &= \int_{-\infty}^{\infty} e^{2\pi i(t+h)\lambda} e^{-2\pi i t \lambda} d_\lambda \|W_\lambda\|^2 \\ &= \int_{-\infty}^{\infty} e^{2\pi i h \lambda} \|dW_\lambda\|^2\end{aligned}$$

と表現できる.

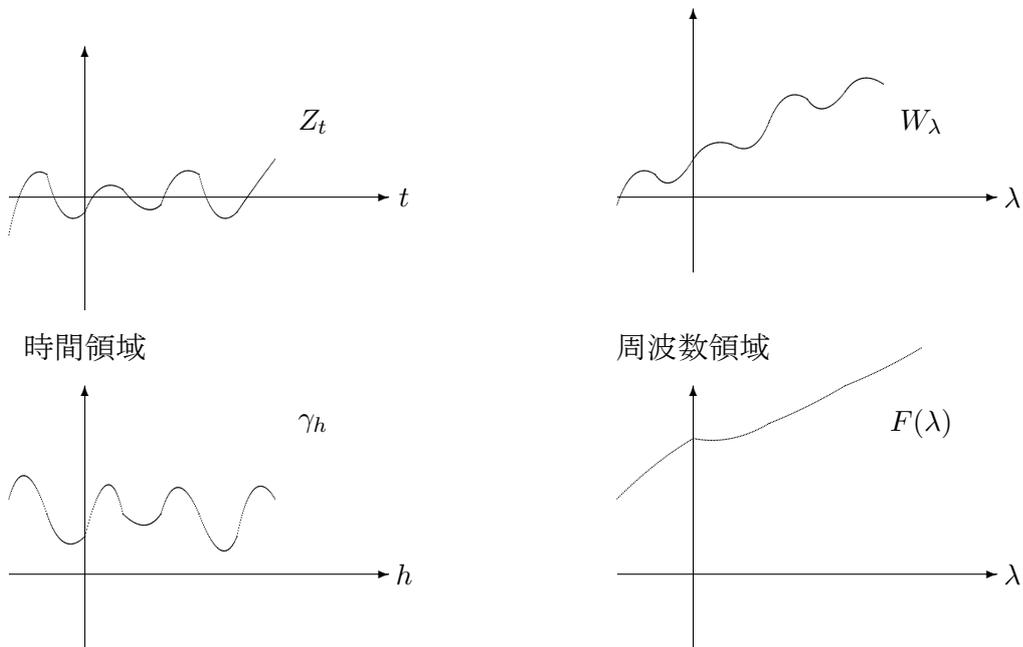
普通

$$dF(\lambda) = \|dW_\lambda\|^2$$

とにおいて,

$$\gamma_h = \int_{-\infty}^{\infty} e^{2\pi i h \lambda} dF(\lambda)$$

と書く.  $dF(\lambda)$  は  $\{Z_t\}$  を  $\{W_\lambda\}$  に変換 (分解) したときの  $\{W_\lambda\}$  の増分  $dW_\lambda$  の大きさ (power) である. そこで  $F(\lambda)$  はスペクトル分布関数 (power spectral distribution) と呼ばれる. ただし,  $F(\lambda) = \int_{-\infty}^{\lambda} dF(\lambda)$  で,  $F(-\infty) = 0$  と定義しておく. したがって,  $F(\lambda) = \|W_\lambda\|^2$ ,  $F(\infty) = \|Z_0\|^2$  であり, 右連続である.



$\{W_\lambda\}$  は直交増分過程なので、その共分散は  $F(\lambda)$  だけで特定できる。  
すなわち

$$\begin{aligned}\langle W_\lambda, W_\mu \rangle &= \|W_\lambda\|^2 = F(\lambda) \quad \lambda < \mu \\ &= \|W_\mu\|^2 = F(\mu) \quad \lambda \geq \mu.\end{aligned}$$

一方、もとの  $\{Z_t\}$  の共分散は  $\{\gamma_h\}_{-\infty < h < \infty}$  で定まる。そしてその間の関係が

EQ1

$$\gamma_h = \int e^{2\pi i h \lambda} dF(\lambda) \quad (1)$$

である。また、一般に

$$F(\lambda) = F_a(\lambda) + F_d(\lambda) + F_s(\lambda)$$

と分解でき

特異スペクトルをもつ例

$$\begin{cases} F_a(\lambda) = \int_{-\infty}^{\lambda} f(\lambda) d\lambda & (\text{絶対連続, absolutely continuous}) \\ F_d(\lambda) & (\text{離散, ジャンプ, discrete}) \\ F_s(\lambda) & (\text{特異, singular}) \end{cases}$$

であるが、 $F_d(\lambda) = F_s(\lambda) = 0$  ならば (6) は

$$\gamma_h = \int_{-\infty}^{\infty} e^{2\pi i h \lambda} f(\lambda) d\lambda$$

と表せる。

次のような場合にはさらに特殊な形をとる。

(a) 時間  $t$  が整数値。

整数  $t, h$  に対し、 $e^{2\pi i t \lambda}, e^{2\pi i h \lambda}$  は  $\lambda$  に関して周期 1 の周期関数なので

$$\begin{aligned}Z_t &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i t \lambda} dW_\lambda, \\ \gamma_h &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i h \lambda} dF(\lambda)\end{aligned}$$

と書き換えられる。またさらに  $F(\lambda)$  が絶対連続部分だけを持つなら、その密度は

$$f(\lambda) = \sum_{h=-\infty}^{\infty} e^{-2\pi i h \lambda} \gamma_h$$

と (逆) 表現できる。

(b)  $\{Z_t\}$  が実数値しかとらない。

$$dW_\lambda = dW_\lambda^{(1)} + idW_\lambda^{(2)}$$

とおくと,

$$\begin{aligned}\bar{Z}_t &= \int e^{-2\pi it\lambda} d\bar{W}_\lambda = \int e^{2\pi it\lambda} \overline{dW_{-\lambda}} \\ &= \int e^{2\pi it\lambda} dW_\lambda = Z_t\end{aligned}$$

より,  $\overline{dW_{-\lambda}} = dW_\lambda$ . 従って,

$$dW_{-\lambda}^{(1)} = dW_\lambda^{(1)}, \quad dW_{-\lambda}^{(2)} = -dW_\lambda^{(2)}.$$

一方,  $\{W_\lambda\}$  の性質 (iii) より

$$E(W_{\lambda+\Delta} - W_\lambda)(\overline{W_{-\lambda} - W_{-\lambda-\Delta}}) = 0$$

であるから  $dW_\lambda^{(1)}, dW_\lambda^{(2)}$  は次の性質をもつ.

- (i)  $\|dW_\lambda^{(1)}\|^2 = \|dW_\lambda^{(2)}\|^2$
- (ii)  $\|dW_\lambda^{(1)}\|^2 + \|dW_\lambda^{(2)}\|^2 = dF(\lambda), \quad dF(\lambda) = dF(-\lambda),$
- (iii)  $\langle dW_\lambda^{(1)}, dW_\lambda^{(2)} \rangle = 0$

従って,

$$\begin{aligned}Z_t &= \int e^{2\pi it\lambda} (dW_\lambda^{(1)} + idW_\lambda^{(2)}) \\ &= \int \cos 2\pi t\lambda dW_\lambda^{(1)} - \int \sin 2\pi t\lambda dW_\lambda^{(2)} \\ &\quad + i \left( \int \cos 2\pi t\lambda dW_\lambda^{(2)} + \int \sin 2\pi t\lambda dW_\lambda^{(1)} \right) \\ &= 2 \left( \int_0^\infty \cos 2\pi t\lambda dW_\lambda^{(1)} - \int_0^\infty \sin 2\pi t\lambda dW_\lambda^{(2)} \right),\end{aligned}$$

$$\gamma_h = 2 \int_0^\infty \cos(2\pi h\lambda) dF(\lambda) \text{ とも書ける.}$$

((i)~(iii) を示せ. レポート問題)

注) ただし, さらに  $t$  が整数値しかとらないときは  $0 \leq \lambda < \frac{1}{2}$  の範囲だけの積分になるが,  $\lambda = -\frac{1}{2}$  でジャンプがあると (つまり  $F_d(-\frac{1}{2}+0) - F_d(-\frac{1}{2}) > 0$  の場合), その部分だけの修正が必要.

**例 1** 巡回過程 (circular process)

$\{Z_t\}$   $t = \dots, -1, 0, 1, \dots$  (離散時間), 弱定常  
 $Z_{t+r} = Z_t$  : 周期  $r$   
 すると,  $\gamma_k = \gamma_{k+r}$ ,  $k = \dots - 1, 0, 1, \dots$

この例は  $r \rightarrow \infty$  とすることによって任意の弱定常過程を近似できる簡単なモデルを与える点で興味深い.

$\{Z_1, Z_2, \dots, Z_r\}$  の共分散行列は

$$\Gamma = (\gamma_{i-j}) = \begin{bmatrix} \gamma_0 & \bar{\gamma}_1 & \cdots & \bar{\gamma}_{r-1} \\ \vdots & & & \vdots \\ \gamma_{r-1} & \cdots & \cdots & \gamma_0 \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{r-1} & \cdots & \gamma_1 \\ \vdots & & & \vdots \\ \gamma_{r-1} & \gamma_{r-2} & \cdots & \gamma_0 \end{bmatrix}$$

で巡回行列 (circular matrix) となる. このような行列の固有値は

$$\nu_j = \sum_{k=0}^{r-1} \gamma_k \exp(-2\pi i j k / r) \geq 0, \quad j = 1, \dots, r \quad (2)$$

と明示的に表せるので

$$\gamma_k = \sum_{j=0}^{r-1} \frac{\nu_j}{r} \exp(2\pi i j k / r) \quad (3)$$

と表現できる. ((2) と (3) を確かめよ. レポート問題).

—— 単位根 ——

$r$  次の単位根  $\omega = \exp(2\pi i / r)$  を導入すれば上の関係は

$$\nu_j = \sum_{k=0}^{r-1} \gamma_k \omega^{-jk}, \quad \gamma_k = \frac{1}{r} \sum_{j=0}^{r-1} \nu_j \omega^{jk}$$

のように表すこともできる.

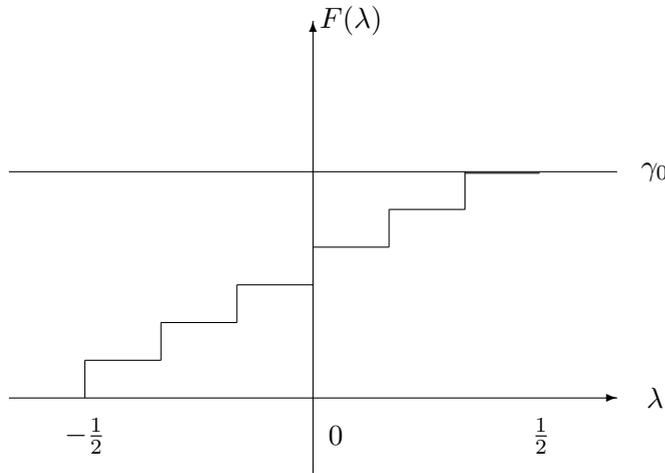
表現

$$\gamma_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i k \lambda} dF(\lambda)$$

と比較すれば,  $\nu_0 = \nu_r$  において

$$dF(\lambda) = \begin{cases} \nu_{-j}/r & \text{at } \lambda_j = -j/r \quad j = -\lfloor \frac{r}{2} \rfloor, \dots, -1, 0 \\ \nu_{r-j}/r & \text{at } \lambda_j = -j/r \quad j = 1, 2, \dots, r - \lfloor \frac{r}{2} \rfloor - 1 \\ 0 & \text{その他} \end{cases}$$

つまり, 次の図のようにジャンプばかりのスペクトル分布  $F(\lambda) = F_d(\lambda)$  である.



このことは, 直接,  $\gamma_h = \gamma_{h-r}$  の性質から導くこともできる. すなわち任意の整数  $h$  について

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i h \lambda} dF(\lambda) \equiv \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i (h-r) \lambda} dF(\lambda)$$

より

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i h \lambda} (1 - e^{-2\pi i r \lambda}) dF(\lambda) \equiv 0$$

が成立するが, フーリエ変換の一意性より

$$(1 - e^{-2\pi i r \lambda}) dF(\lambda) \equiv 0.$$

これが成り立つためには  $1 \neq e^{-2\pi i r \lambda}$  となる  $\lambda$  については  $dF(\lambda) = 0$ , つまり  $r\lambda = \text{整数}$  のところだけ  $dF(\lambda) \neq 0$  となるしかないことがわかる.

従って、 $\lambda_j = j/r$  とし、そこでのジャンプを  $\xi_j$  とすれば

$$\gamma_k = \sum_{j=-\lfloor \frac{r}{2} \rfloor}^{r-\lfloor \frac{r}{2} \rfloor-1} e^{2\pi i k \lambda_j} \xi_j \quad k = 0, 1, 2, \dots, r-1.$$

これを連立方程式として解けば、 $\{\xi_j\}$  は先の  $\{\nu_j/r\}$  と集合として一致することがわかる。

(レポート、 $r = 4$  のとき、上の連立方程式を解き、 $\{\xi_j\}$  が  $\{\nu_j/r\}$  と一致することを確かめよ.)

また、 $F(\lambda)$  の形から  $\{W_\lambda\}$  自身

$$W_\lambda = \sum_{\lambda_j \leq \lambda} X_j \quad \text{つまり} \quad dW_\lambda = \begin{cases} 0 & \lambda \neq \lambda_j \\ X_j & \lambda = \lambda_j \end{cases}, \quad j = -\lfloor \frac{r}{2} \rfloor, \dots, r - \lfloor \frac{r}{2} \rfloor - 1$$

と表せるはずである。ただし、 $\{X_j\}$  は無相関な (複素数値の) 確率変数で  $E|X_j|^2 = dF(\lambda_j)$ 、 $EX_j = 0$  である。したがって  $\{Z_t\}$  は

$$Z_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i t \lambda} dW_\lambda = \sum_{j=-\lfloor \frac{r}{2} \rfloor}^{r-\lfloor \frac{r}{2} \rfloor-1} e^{2\pi i t \lambda_j} X_j$$

と有限和で表現されることになる。

特に、 $\{Z_t\}$  が実数値のときは  $\xi_j = \xi_{-j}$  より、 $r = 2m$  (偶数) ならば自己共分散  $\gamma_k$  は

$$\begin{aligned} \gamma_k &= \sum_{j=-m}^{m-1} (\cos 2\pi k \lambda_j + i \sin 2\pi k \lambda_j) \xi_j \\ &= 2 \sum_{j=1}^{m-1} (\cos 2\pi k \lambda_j) \xi_j + (\cos 2\pi k \lambda_{-m} + i \sin 2\pi k \lambda_{-m}) \xi_{-m} + \xi_0 \\ &= 2 \sum_{j=1}^{m-1} \xi_j \cos 2\pi k \lambda_j + \xi_0 + \xi_{-m} \cos \pi k, \end{aligned}$$

$r = 2m + 1$  (奇数) ならば

$$\begin{aligned} \gamma_k &= \sum_{j=-m}^m (\cos 2\pi k \lambda_j + i \sin 2\pi k \lambda_j) \xi_j \\ &= 2 \sum_{j=1}^m \xi_j \cos 2\pi k \lambda_j + \xi_0 \end{aligned}$$

と表現できる.

この  $\gamma_h$  の表現に対応して,  $Z_t$  自身も

$$X_t = X_j^{(1)} + iX_j^{(2)} \quad \text{より} \quad X_{-j}^{(1)} = X_j^{(1)}, X_{-j}^{(2)} = -X_j^{(2)}$$

となることより, また, 特に  $X_0^{(2)} = 0$  であることより

$$Z_t = \sum_{j=-\lfloor \frac{r}{2} \rfloor}^{r-\lfloor \frac{r}{2} \rfloor-1} \left( X_j^{(1)} \cos 2\pi t \lambda_j - X_j^{(2)} \sin 2\pi t \lambda_j \right)$$

と表現でき,  $r = 2m$  なら

$$Z_t = 2 \sum_{j=1}^{m-1} X_j^{(1)} \cos 2\pi t \lambda_j + X_0^{(1)} + X_{-m}^{(1)} \cos \pi t \quad \left( \lambda_j = \frac{j}{2m} \right)$$

$r = 2m + 1$  なら

$$Z_t = 2 \sum_{j=1}^m X_j^{(1)} \cos 2\pi t \lambda_j + X_0^{(1)} \quad \left( \lambda_j = \frac{j}{2m+1} \right)$$

となることがわかる. この表現から,  $\{Z_t\}$  は周波数  $\lambda_j$  の余弦波で強さが  $2X_j^{(1)}$  の波の線形結合であらわせることがわかる.

$$r = 2m = 2 \text{ のとき } Z_t = X_0^{(1)} - X_{-1}^{(1)} \cos \pi t$$

$$r = 2m + 1 = 3 \text{ のとき } Z_t = 2X_1^{(1)} \cos(2\pi t \frac{1}{3}) + X_0^{(1)}$$

$$r = 2m = 4 \text{ のとき } Z_t = 2X_1^{(1)} \cos(2\pi t \frac{1}{4}) + X_0^{(1)} - X_{-2}^{(1)} \cos \pi t$$

一般に, このような離散時間の定常時系列の表現には, 周波数  $1/2$  (Nyquist 周波数) 以上の波は含まれない. これは上の例でもわかるが, より一般的にはスペクトル表現から明らかである. つまり, 時間間隔が  $1$  の観測では, 周波数  $1/2$  以上の波は同定できない. 当然, 時間間隔が  $h$  ならば  $h/2$  が Nyquist 周波数である.

注) 位相のずれた波でも, 三角関数の和の公式から, 正弦波も用いれば位相  $0$  の波の合成で表される. (レポート,  $c_1 \cos(2\pi(\frac{t}{4} + \lambda_0)) + c_2 \cos(2\pi\frac{t}{2}) + c_3$  を  $\cos \frac{\pi}{2}t, \cos \pi t, \sin \frac{\pi}{2}t, 1$  の線形結合で書き直せ.)

## 例 2 ランダムな振動 (連続時間)

$$Z_t = r \cos 2\pi(t + \theta) \quad -\infty < t < \infty, \text{ 実数値}$$

$$r : 0 \leq r < \infty, \text{Er}^2 = \sigma^2 \text{ である確率変数}$$

$$\theta : \left[-\frac{1}{2}, \frac{1}{2}\right] \text{ 上の一様分布に従う確率変数}$$

とすれば  $r$  は振幅,  $\theta$  は振動を定める. ただし  $r$  と  $\theta$  は独立と仮定する

$$\begin{aligned} \gamma_h &= \text{E}(r \cos(2\pi(t+h+\theta)) \cdot r \cos(2\pi(t+\theta))) \\ &= \text{E}(r^2) \text{E}(\cos 2\pi(t+h+\theta) \cdot \cos 2\pi(t+\theta)) \\ &= \frac{\sigma^2}{2} \text{E}(\cos 2\pi(2t+h+2\theta) + \cos 2\pi h) \\ &= \frac{\sigma^2}{2} \text{E}(\cos 2\pi h) = \frac{\sigma^2}{2} \cos 2\pi h \end{aligned}$$

従って,  $\{\gamma_h\}$  は  $\gamma_{h+1} = \gamma_h$  で, 例 1 と同じ連続時間の巡回過程の例である. これは,  $Z_t = Z_{t+1}$  であることから明らかである.

これより

$$dF(\lambda) = \begin{cases} \frac{\sigma^2}{4} & \lambda = \pm 1 \\ 0 & \text{その他} \end{cases}$$

を得る. 確かに,  $\gamma_h = 2 \int_0^\infty \cos(2\pi h \lambda) dF(\lambda) = \frac{\sigma^2}{2} \cos 2\pi h$  となっている.

また,  $\{Z_t\}$  自身は直交増分過程

$$dW_\lambda = \begin{cases} \frac{r}{2} e^{2\pi i \theta} & \lambda = 1 \\ \frac{r}{2} e^{-2\pi i \theta} & \lambda = -1 \\ 0 & \text{その他} \end{cases}$$

を用いて,

$$\begin{aligned} Z_t &= \int_{-\infty}^{\infty} e^{2\pi i t \lambda} dW_\lambda \\ &= \frac{r}{2} e^{-2\pi i \theta} e^{-2\pi i t} + \frac{r}{2} e^{2\pi i \theta} e^{2\pi i t} \end{aligned}$$

とあらわせ, 周波数 1 と -1 の余弦波のランダムな和となっていることがわかる. この時系列は 2 点観測できればすべての値がわかるはずである.

**例 3** 白色雑音 (white noise)

$$Z_t = a_t \quad t = \dots, -1, 0, 1, \dots$$

ただし,  $\{a_t\}$  は実数値をとる確率変数で, 直交し  $E(a_t) = 0, \text{Var}(a_t) = \sigma^2 \neq 0$ . 正規分布を仮定することもあるがその場合は正規白色雑音という.

$$\gamma_h = \begin{cases} \sigma^2 & h = 0 \\ 0 & \text{その他} \end{cases}$$

このような  $\gamma_h$  を与えるスペクトル密度は  $f(\lambda) = dF(\lambda)/d\lambda = \sigma^2$  しかないことはすぐわかる. 従って

$$Z_t = \int_0^{1/2} 2 \cos(2\pi t\lambda) dW_\lambda^{(1)} - \int_0^{1/2} 2 \sin(2\pi t\lambda) dW_\lambda^{(2)}$$

と表せることになる. ただし,  $dW_\lambda^{(1)}, dW_\lambda^{(2)}$  はそれぞれ平均 0, 分散  $\frac{\sigma^2}{2} d\lambda$  の直交増分過程.

$\{a_t\}$  が正規白色雑音ならば対応する  $\{W_\lambda^{(1)}\}$  と  $\{W_\lambda^{(2)}\}$  はブラウン運動にほかならない.

複素数  
値正規  
白色  
雑音

**例 4** 移動平均過程 (Moving Average Process)

例 3 の白色雑音  $\{a_t\}$  を用いて

$$Z_t = \sum_{j=0}^q \psi_j a_{t-j}$$

と表せる  $\{Z_t\}$  を次数  $q$  の移動平均過程といい,  $MA(q)$  で表す. (レポート:  $\{Z_t\}$  のスペクトル表現とスペクトル密度を求めよ.) さて

(i)  $Z_t = a_t + \frac{1}{2}a_{t-1}$ ,  $a_t$  の平均は 0, 分散は 1 とする.

(ii)  $Z_t = a_t + 2a_{t-1}$ ,  $a_t$  の平均は 0, 分散は 1/4 とする.

を考えると (i), (ii) とともに同じスペクトル密度  $f(\lambda)$  をもつことはすぐわかる. 言い換えれば同じ  $\{\gamma_h\}$  をもつ確率過程である. 同じ  $\{\gamma_h\}$  を持った  $\{Z_t\}$  なのに 2通りの表現ができることになる. もちろん, (i) での  $\{a_t\}$  と (ii) での  $\{a_t\}$  は, 定数倍の違いだけでなく, 次のような違いがある.

(i)  $a_t = Z_t - \frac{1}{2}Z_{t-1} + \frac{1}{4}Z_{t-2} + \dots$

(ii)  $a_t = \frac{1}{2}Z_{t+1} - \frac{1}{4}Z_{t+2} + \frac{1}{8}Z_{t+3} + \dots$  すなわち  $2a_t = Z_{t+1} - \frac{1}{2}Z_{t+2} + \frac{1}{4}Z_{t+3} + \dots$

すなわち，過去の  $\{Z_t\}$  で表されるか，将来の  $\{Z_t\}$  で表されるかの違いである。

bispectrum

## 2 弱定常時系列の幾何学

$\{Z_t\}$ ,  $t = \dots, -1, 0, 1, \dots$  (離散時間, 複素数値) なる弱定常時系列を考える.

$$H_t = L_2(Z_s : s \leq t) \subset L_2(Z).$$

$\dots \subset H_{-1} \subset H_0 \subset H_1 \subset \dots \subset L_2(Z)$  である.  $H_t - H_{t-1} = H_t \cap H_{t-1}^\perp$  とすれば

$$\dim(H_t - H_{t-1}) \leq 1.$$

もし, ある  $t$  で  $\dim(H_t - H_{t-1}) = 0$  ならすべての  $t$  で  $\dim(H_t - H_{t-1}) = 0$  (レポート: Shift operator  $T_t$  を用いてこれを示せ). さらに一般的にある  $s < t$  で  $\dim(H_t - H_s) < t - s$  ならば, すべての  $s < t$  で  $\dim(H_t - H_s) = 0$ . このような場合には,  $\lim_{t \rightarrow -\infty} H_t = H_{-\infty}$  が,  $H_{-\infty} = H_t = L_2(Z)$  となるのでこの弱定常過程は純決定的と呼ばれる. このような特殊な場合以外は  $\dim(H_t - H_{t-1}) = 1$  である. なお,  $H_t = \cup_{s \leq t} (H_s - H_{s-1}) \cup H_{-\infty}$  である.

**定理** (Wold の分解定理)

$\{Z_t\}$  が平均 0 の弱定常過程ならば,

$$Z_t = \sum_{s=0}^{\infty} \psi_s u_{t-s} + v_t$$

と一意的に分解される. ただし,  $\psi_s$  は定数で  $\sum |\psi_s|^2 < \infty$ .  $\{u_t\}$  は平均 0, 分散 1 の直交過程で,  $u_t \in H_t - H_{t-1}$ ,  $v_t \in H_{-\infty}$ ,  $\{v_t\} \perp \{u_t\}$ .

**証明**

$Z_s^*$  を  $Z_s$  の  $H_{s-1}$  への projection, i.e.  $\|Z_s - z\| \rightarrow \min$ ,  $z \in H_{s-1}$  として,  $u_s = (Z_s - Z_s^*) / \|Z_s - Z_s^*\|$  と定義すれば  $\|u_s\| = 1$  で  $u_s \in H_s - H_{s-1}$ . しかも,  $\langle u_s, u_{s'} \rangle = 0$   $s \neq s'$  は明らか. すなわち,  $\{u_s\}$  は直交過程である. また, 作り方によらない (レポート).  $\psi_s = \langle Z_t, u_{t-s} \rangle$  と定義すれば,

$$0 \leq E \left| Z_t - \sum_{s=0}^m \psi_s u_{t-s} \right|^2 = E|Z_t|^2 - \sum_{s=0}^m |\psi_s|^2$$

従って,  $\sum_{s=0}^m |\psi_s|^2 \leq E|Z_t|^2 = \gamma_0$  で,  $\sum_{s=0}^{\infty} |\psi_s|^2 < \infty$  であることがわかる.

$z$  が存在すること

これより

$$\left\| \sum_{s=0}^m \psi_s u_{t-s} - \sum_{s=0}^n \psi_s u_{t-s} \right\|^2 = \sum_{s=m+1}^n |\psi_s|^2$$

は,  $m < n$  を十分大きくとることにより, いくらでも小さくなる. すなわち  $\sum_{s=0}^m \psi_s u_{t-s}$  は Cauchy 列. ヒルベルト空間  $H_t$  の完備性より  $\sum_{s=0}^{\infty} \psi_s u_{t-s}$  が存在し  $H_t$  の要素となる.  $v_t = Z_t - \sum_{s=0}^{\infty} \psi_s u_{t-s}$  と定義すれば  $t \geq s$  のとき

$$\begin{aligned} \langle u_s, v_t \rangle &= \mathbb{E} \left( \overline{u_s (Z_t - \sum_l \psi_l u_{t-l})} \right) \\ &= \mathbb{E} (u_s \overline{Z_t}) - \overline{\psi_{t-s}} \mathbb{E} |u_s|^2 \\ &= \overline{\psi_{t-s}} - \overline{\psi_{t-s}} = 0, \end{aligned}$$

$t < s$  のときは,  $u_s \in H_s - H_t$  より

$$u_s \perp v_t \quad \text{i.e.} \quad \langle u_s, v_t \rangle = 0.$$

いずれにしても  $\{u_s\} \perp \{v_t\}$  が成立する.

$v_t \in H_{-\infty} = \bigcap_{s \leq t} H_s$  を示すために,  $\exists s, v_t \notin H_{t-s}$  と仮定すると,  $v_t = v_t^{(1)} + v_t^{(2)}$ ,  $v_t^{(1)} \in H_{t-s}$ ,  $v_t^{(2)} \neq 0 \in H_t - H_{t-s}$  と分解できる. ところで,

$$H_t - H_{t-s} = \text{Span}\{u_l; t-s+1 \leq l \leq t\}$$

なので,  $v_t^{(2)} = \sum_{l=0}^{s-1} c_l u_{t-l}$  と表せる. 従って

$$\begin{aligned} \langle v_t, u_{t-l} \rangle &= \langle v_t^{(1)}, u_{t-l} \rangle + \langle v_t^{(2)}, u_{t-l} \rangle \\ &= 0 + c_l \mathbb{E} |u_{t-l}|^2 \\ &= c_l. \end{aligned}$$

ところが,  $\{v_t\} \perp \{u_s\}$  より  $c_l = 0$ . 結局,  $v_t^{(2)} = 0$  で矛盾.

一意性は, もし  $Z_t = \sum_{s=0}^{\infty} \psi'_s u_{t-s} + v_t$  と表せたとすると,  $\sum_s (\psi_s - \psi'_s) u_{t-s} = 0$ .

従って,

$$0 = \left\| \sum_s (\psi_s - \psi'_s) u_{t-s} \right\|^2 = \sum_s |\psi_s - \psi'_s|^2$$

すなわち,  $\psi_s \equiv \psi'_s$  よりいえる.

$\{v_t\}$  は  $\{Z_t\}$  の決定的部分と呼ばれる.  $Z_t = v_t$  なら  $\{Z_t\}$  は純決定的 (purely deterministic) と呼ばれ,  $v_t = 0$  なら純非決定的 (purely non-deterministic) と呼ばれる. もちろん  $\dim(H_t - H_{t-1}) = 0$  なら  $Z_t \equiv v_t$  である. また一般的に  $Z_t - Z_t^*$  をイノベーション (innovation) と呼ぶ.  $\{u_t\}$  のように正規化されたイノベーションを標準イノベーションと呼ぶ.

たとえば, 例 4 の (i) は  $v_t \equiv 0$  の例になっている.

(i)  $a_t = u_t$ ,  $\psi_0 = 1, \psi_1 = \frac{1}{2}, \psi_2 = \dots = 0$ . なぜなら

$$a_t = \sum_{l=0}^{\infty} \left(-\frac{1}{2}\right)^l Z_{t-l} \quad \text{で} \quad \in H_t - H_{t-1}$$

となるので Wold 分解の一意性より従う. 一方, 例 4 の (ii) は

(ii)  $a_t \neq u_t$  で, 表現

$$2a_t = \sum_{l=0}^{\infty} \left(-\frac{1}{2}\right)^l Z_{t+l+1}$$

を持つが, これは時間を逆方向に考えて  $H^t = L_2\{Z_s, s > t\}$  を  $H_t$  の代わりに考えたときの表現である. つまり,  $a_t \in H^t - H^{t+1}$ . 一方,  $a_t + 2a_{t-1} \in H_t - H_{t-2}$ . しかし,  $H_t - H_{t-2} = (H_t - H_{t-1}) + (H_{t-1} - H_{t-2})$  が直交分解であることに注意すれば,  $a_t \in H_t - H_{t-1}$ . 結局  $H^t - H^{t+1} = \text{Span}(a_t) = H_t - H_{t-1}$  となる. すると, Wold 分解

$$Z_t = \psi_0 a_t + \psi_1 a_{t-1} + \dots + v_t$$

に加えて, 時間を逆にした

$$Z_t = \psi^0 a_t + \psi^1 a_{t+1} + \dots + v^t$$

も成立することになるが,  $Z_t \neq v_t$  とすると  $Z_t = a_t$  でなければならず, もとの表現に矛盾する. したがって  $Z_t = v_t$  で純決定的な時系列になっている.

この分解定理をスペクトルの言葉で解釈し直すと,  $u_t, v_t$  はそれぞれ

$$\begin{aligned} u_t &= \int e^{2\pi i t \lambda} dU_\lambda, \\ v_t &= \int e^{2\pi i t \lambda} dV_\lambda \end{aligned}$$

と書けるはずで、

$$\begin{aligned}
Z_t &= \sum_l \psi_l u_{t-l} + v_t \\
&= \int e^{2\pi i t \lambda} \left( \sum_l \psi_l e^{-2\pi i l \lambda} \right) dU_\lambda + \int e^{2\pi i t \lambda} dV_\lambda \\
&= \int e^{2\pi i t \lambda} \left\{ \left( \sum_l \psi_l e^{-2\pi i l \lambda} \right) dU_\lambda + dV_\lambda \right\} .
\end{aligned}$$

従って、 $\{U_\lambda\} \perp \{V_\lambda\}$  より

$$\begin{aligned}
dF(\lambda) &= \left\| \sum_l \psi_l e^{-2\pi i l \lambda} dU_\lambda \right\|^2 + \|dV_\lambda\|^2 \\
&= \left| \sum_l \psi_l e^{-2\pi i l \lambda} \right|^2 \|dU_\lambda\|^2 + \|dV_\lambda\|^2 \\
&= \left| \sum_l \psi_l e^{-2\pi i l \lambda} \right|^2 d\lambda + \|dV_\lambda\|^2 .
\end{aligned}$$

この形をみれば、純非決定的な部分がスペクトルの絶対連続部分に対応し、決定的が残りのスペクトルを定めることがわかる。

なお、例1, 例2での  $\{Z_t\}$  は Wold の分解定理での純決定的な  $v_t$  の部分からだけなる例であり、例3, 例4の (i) は純非決定的な  $\sum \psi_l u_{t-l}$  の部分だけの例、例4の (ii) は絶対連続でない部分だけの例であることになる。

なお、 $\Psi(z) = \sum_{l=0}^{\infty} \psi_l z^l$  は複素数  $z$  を変数とする複素関数とみたとき伝達関数 (transfer function) と呼ばれる。

Wold の分解定理の意味

この定理は、数学としては、ヒルベルト空間上の直交分解にすぎない。しかし、 $\{u_t\}$  は直交過程であって、独立過程とは限らないことに注意が必要である。正規性を仮定すれば独立になるが、そうでない限り  $\{u_t\}$  は独立ではなく、互いにある程度の情報を共有している。したがって、ヒルベルト空間  $H_{t-1}$  への射影  $Z_t^*$  と  $\sigma$ -集合体  $\mathcal{F}_{t-1} = \mathcal{F}\{Z_s : s \leq t-1\}$  に関する条件付平均値

$$Z_{t|t-1} = E(Z_t | \mathcal{F}_{t-1})$$

が必ずしも一致しない。実際、Wold の分解定理を用いれば  $Z_{t|t-1}$  と  $Z_t^*$  は

$$Z_{t|t-1} = Z_t^* + \psi_0 E(u_t | \mathcal{F}_{t-1})$$

の関係にある.  $u_t$  の平均は 0 であるが, 条件付平均は必ずしも 0 ではない.

条件付平均値が一般的に最良予測を与えることは, 次のことから容易にわかる.  $Y$  と  $Z$  を 2 つの確率変数,  $g$  を任意の可測な関数としたとき,

$$\begin{aligned} E(Z - g(Y))^2 &= E(Z - E(Z|Y) + E(Z|Y) - g(Y))^2 \\ &= E(Z - E(Z|Y))^2 + 2E((Z - E(Z|Y))(E(Z|Y) - g(Y))) \\ &\quad + E(E(Z|Y) - g(Y))^2 \\ &= E(Z - E(Z|Y))^2 + E(E(Z|Y) - g(Y))^2 \\ &\geq E(Z - E(Z|Y))^2. \end{aligned}$$

もちろん正規性の仮定があれば一致し,  $Z_t^* = Z_{t|t-1}$  である. ここで正規分布と線形性が結び付くのは,  $(X, Y)$  が 2 変量正規分布に従うとしたとき,

$$E(Y|X) = \alpha X + \beta$$

の形になることを思い出せば, 理解できるであろう (レポート:  $(X, Y) \sim N(\mu, \Sigma)$  のとき上の  $\alpha, \beta$  を求めよ). 一般的に  $Z_t^*$  を線形最良予測,  $Z_{t|t-1}$  を最良予測と呼ぶことが多い. 上の不等式から

$$\|Z_t^* - Z_t\|^2 \geq \|Z_{t|t-1} - Z_t\|^2$$

である.

### 3 弱定常時系列の予測

Wold の分解定理で,  $v_t = 0$  と仮定すれば

$$Z_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}$$

なる表現をもつので,  $u_{t-1}, u_{t-2}, \dots$  が与えられたときの  $Z_t$  の最良線形予測は

$$Z_t^* = \sum_{j=1}^{\infty} \psi_j u_{t-j}$$

で与えられ, その予測誤差は

$$E|Z_t - Z_t^*|^2 = |\psi_0|^2$$

$v_t \neq 0$   
のとき  
の予測

となる。以下この章では常に  $v_t = 0$  と仮定する。しかし,  $u_{t-1}, u_{t-2}, \dots$  は直接観測できるわけではないので,  $Z_{t-1}, Z_{t-2}, \dots$  が与えられたときの最良線形予測を求めるには,  $\{Z_t\}$  の  $AR(\infty)$  表現 (Infinite Order Autoregressive Representation)

$$\sum_{j=0}^{\infty} \phi_j Z_{t-j} = u_t$$

が必要となる。ただし,  $\phi_0 = \psi_0 = 1$  で,  $\{u_t\}$  は Wold 分解のときと同様に  $\{Z_t\}$  のイノベーションである。このとき,  $Z_t$  の最良線形予測つまり  $E|Z_t - \sum_{l=1}^{\infty} a_l Z_{t-l}|^2$  を最小にする  $Z_t^* = \sum_{l=1}^{\infty} a_l Z_{t-l}$  は

$$Z_t^* = - \sum_{j=1}^{\infty} \phi_j Z_{t-j}$$

で与えられる。

この  $AR(\infty)$  表現に対応して, 先の Wold 分解で  $v_t = 0$  と仮定したときの

$$Z_t = \sum_{l=0}^{\infty} \psi_l u_{t-l}$$

なる表現 (例 4) を  $MA(\infty)$  表現という。また,  $B = T_{-1}$  を Backward Shift operator として,

$$Z_t = \Psi(B)u_t$$

とも表す。  $AR(\infty)$  表現は  $\Phi(z) = \sum_{j=0}^{\infty} \phi_j z^j$  として

$$\Phi(B)Z_t = u_t$$

である。しかし,  $AR(\infty)$  表現が無条件に可能であるとは限らない。そのための 1 つの条件が

#### 条件 (可逆性)

複素関数  $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$  が  $|z| \leq 1$  で正則で 0 とならない。

である。

### 定理

可逆性の条件が満たされれば,  $\{Z_t\}$  は  $AR(\infty)$  表現をもつ.

### 証明

$\Psi(z)$  は  $|z| \leq 1$  で正則であるので,

$$\exists \varepsilon > 0 \quad 1/\Psi(z) = \sum_{j=0}^{\infty} \phi_j z^j, \quad |z| < 1 + \varepsilon.$$

従って,  $\phi_j(1 + \frac{\varepsilon}{2})^j$  は  $j \rightarrow \infty$  としたとき 0 に収束する. このことは,  $\exists K > 0$

$$|\phi_j| < K \left(1 + \frac{\varepsilon}{2}\right)^{-j}, \quad j = 0, 1, 2, \dots$$

を意味する. このことから  $\sum_{j=0}^{\infty} |\phi_j| < \infty$  であり,  $\Phi(z)$  は  $|z| \leq 1$  で正則で  $\Phi(z)\Psi(z) \equiv 1$  が成り立つことになる. そこで,

$$\begin{aligned} \mathbb{E} \left| \sum_{j=m+1}^n \phi_j Z_{t-j} \right|^2 &= \sum_{j=m+1}^n \sum_{k=m+1}^n \phi_j \bar{\phi}_k \gamma_{j-k} \\ &\leq \gamma_0 \left( \sum_{j=m+1}^n |\phi_j| \right)^2 \end{aligned}$$

に注意すれば,  $S_n = \sum_{j=0}^n \phi_j Z_{t-j}$  は  $L^2(Z)$  でのコーシー列となり, 極限  $S \in L^2(Z)$  をもつ. この極限が実際  $u_t$  と一致していることは Fatou の Lemma より

$$\begin{aligned} \mathbb{E}|S - u_t|^2 &= \mathbb{E} \liminf_{k \rightarrow \infty} |S_{n_k} - u_t|^2 \\ &\leq \liminf_{k \rightarrow \infty} \mathbb{E}|S_{n_k} - u_t|^2 = 0 \end{aligned}$$

より明らかである (レポート:  $\Phi(B)\Psi(B)u_t = u_t$  を用いて  $\lim_{n \rightarrow \infty} \mathbb{E}|S_n - u_t|^2 = 0$  を示せ).

この定理から, 逆に  $AR(\infty)$  表現をもつときに  $\Phi(z) \neq 0$ ,  $|z| \leq 1$  ならば同じ  $u_t$  を用いた  $MA(\infty)$  表現を持つこともいえる. さらに  $AR(p)$  表現や  $MA(q)$  表現を持つときの可逆性についても同様である.

**例**  $Z_t = u_t - 0.8u_{t-1}$  は  $Z_t + 0.8Z_{t-1} + 0.8^2Z_{t-2} + \dots = u_t$  なる  $AR(\infty)$  表現を持つが,  $Z_t = u_t - u_{t-1}$  は  $AR(\infty)$  表現できない.

繰り返しになるが、 $v_t = 0$ である限り、Wold分解から弱定常時系列は常に $MA(\infty)$ 表現を持つ。しかし、 $MA(\infty)$ 表現だけだと $u_t$ が現在と過去の $Z_t, Z_{t-1}, \dots$ で表現できなければ、時間の流れにそった因果関係があやしくなる。そこで、 $AR(\infty)$ 表現もできるような $\{Z_t\}$ のことを因果関係(Causality)のある弱定常過程という。ここには先に指摘した時間の流れだけでなく、無限級数の収束性の問題もからんでいる。

ここで予測の問題に戻ると、 $AR(\infty)$ にもとづいた予測 $Z_t^*$ は実際には使えないことは明らかである。それは過去のすべての観測値が必要となるからである。そこで実用的には $AR(p)$ 表現

$$\sum_{l=0}^p \phi_l Z_{t-l} = u_t$$

で近似したときの予測が用いられることが多い。

### 定理 (Kolmogorov)

$\{Z_t\}$ が離散時間実数値弱定常過程で連続なスペクトル密度 $f(\lambda)$ だけを持ち、 $f(-\frac{1}{2}) = f(\frac{1}{2}-)$ であるとする。このとき、 $f(\lambda) > 0$ 、 $-\frac{1}{2} \leq \lambda < \frac{1}{2}$ ならば、 $Z_{t-1}, Z_{t-2}, \dots$ が与えられたときの $Z_t$ の最良線形予測の予測誤差は

$$\sigma^2(f) = \exp \left( \int_{-\frac{1}{2}}^{\frac{1}{2}} \log f(\lambda) d\lambda \right)$$

で与えられる。

### 証明

まず、複素単位円内 $|z| < 1$ で

$$\log(1 - z) = \sum_{j=1}^{\infty} \frac{z^j}{j}$$

なるテイラー展開が成り立つことに注意する。従って、 $|\alpha| < 1$ なる複素数 $\alpha$ に対して

$$\begin{aligned} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log |1 - \alpha e^{-2\pi i \lambda}|^2 d\lambda &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \log (1 - \alpha e^{-2\pi i \lambda}) (1 - \bar{\alpha} e^{2\pi i \lambda}) d\lambda \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left( \sum_{j=1}^{\infty} \frac{\alpha^j e^{-2\pi i j \lambda}}{j} + \sum_{j=1}^{\infty} \frac{\bar{\alpha}^j e^{2\pi i j \lambda}}{j} \right) d\lambda \\ &= 0 \end{aligned}$$

となる．まず，定理を  $\{Z_t\}$  が次のような  $AR(p)$  過程であるときに示す．つまり，

$$\Phi(z) = 1 + \phi_1 z + \cdots + \phi_p z^p$$

として

$$\Phi(B)Z_t = a_t$$

を満たすとする．ただし， $\{a_t\}$  は  $Ea_t = 0$ ， $Ea_t^2 = \sigma^2$  である直交過程，つまり白色雑音で， $a_t \perp H_{t-1}$  である（従って  $\{a_t\}$  は  $\{Z_t\}$  のイノベーションとなる）．さらに  $\Phi(z) \neq 0$ ， $|z| \leq 1$  であると仮定すれば  $\{Z_t\}$  のスペクトル密度は

$$f(\lambda) = \frac{\sigma^2}{|\Phi(e^{2\pi i \lambda})|^2}$$

で与えられる（レポート問題）．ここで， $\Phi(z) = 0$  の根を  $\alpha_1^{-1}, \dots, \alpha_p^{-1}$  とすれば仮定より， $|\alpha_j| < 1$ ， $j = 1, \dots, p$  であり

$$\Phi(z) = \prod_{j=1}^p (1 - \alpha_j z)$$

と分解できる．結局

$$f(\lambda) = \frac{\sigma^2}{\prod_{j=1}^p |1 - \alpha_j e^{2\pi i \lambda}|^2}$$

と表せることになり，(16) を用いれば

$$\begin{aligned} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log f(\lambda) d\lambda &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \sigma^2 d\lambda - \sum_{j=1}^p \int_{-\frac{1}{2}}^{\frac{1}{2}} \log |1 - \alpha_j e^{2\pi i \lambda}|^2 d\lambda \\ &= \log \sigma^2 . \end{aligned}$$

ここで， $\sigma^2$  は確かに上のような  $AR(p)$  過程であるときの最良線形予測の予測誤差になっているので，定理は成立している．

一般的には，以下の補題とその系で表示されるように  $f(\lambda)$  が定理の条件を満たし，対称（これは実数値であることより従う）ならば任意の  $\varepsilon > 0$  に対して，ある  $AR(p)$  過程が存在して，そのスペクトル密度  $g(\lambda)$  が

EQ3.2

$$|f(\lambda) - g(\lambda)| < \varepsilon \tag{5}$$

のように  $f(\lambda)$  を近似できる．もちろん  $|g(\lambda)| < \infty$  である．したがって，

$$\begin{aligned}\sigma_p^2(f) &= \min_{c_1 \cdots c_p} \mathbb{E}(Z_t - c_1 Z_{t-1} - \cdots - c_p Z_{t-p})^2 \\ &= \min_{c_1 \cdots c_p} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left| 1 - \sum_{j=1}^p c_j e^{2\pi i \lambda} \right|^2 f(\lambda) d\lambda\end{aligned}$$

と定義すれば、 $\sigma_p^2(f)$  は  $p$  について単調減少であるので収束し、

$$\hat{Z}_{t,p} = \sum_{j=1}^p c_{j,p} Z_{t-j}$$

も  $Z_t$  の  $H_{t-1}$  への射影へ収束する。つまり、 $\sigma_p^2(f)$  の極限  $\sigma^2(f)$  は  $Z_{t-1}, Z_{t-2}, \dots$  にもとづく  $Z_t$  の最良線形予測である  $H_{t-1}$  への射影の予測誤差にほかならない。一方、(5) より

$$|\sigma^2(f) - \sigma^2(g)| < C \cdot \varepsilon$$

が従うので、

$$\sigma^2(g) = \exp \left( \int_{-\frac{1}{2}}^{\frac{1}{2}} \log g(\lambda) d\lambda \right)$$

に注意すれば目的の

$$\sigma^2(f) = \exp \left( \int_{-\frac{1}{2}}^{\frac{1}{2}} \log f(\lambda) d\lambda \right)$$

を得る。

注 1)  $\sigma^2(f) - \sigma^2(g) = \sigma^2(f) - \sigma_p^2(f) + \sigma_p^2(f) - \sigma_p^2(g) + \sigma_p^2(g) - \sigma^2(g)$ .

注 2) Jansen の不等式より、

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \log f(\lambda) d\lambda < \log \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\lambda) d\lambda = \log \gamma_0$$

従って、 $\int_{-\frac{1}{2}}^{\frac{1}{2}} \log f(\lambda) d\lambda < \infty$  が常に成立する。また上の定理は

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \log f(\lambda) d\lambda = -\infty$$

のときでも成立し,  $\sigma^2(f) = 0$  となるので,  $\sigma^2(f) > 0$  と  $\int_{-\frac{1}{2}}^{\frac{1}{2}} \log f(\lambda) d\lambda > -\infty$  は同値で, このとき  $f(\lambda) > 0$  a.e. でもあることがわかる.

**補題 (チェザロ総和法, Fejér の定理)**

$f(\lambda)$  が  $-\frac{1}{2} \leq \lambda < \frac{1}{2}$  で連続で  $f(-\frac{1}{2}) = f(\frac{1}{2}-)$  であれば, 部分フーリエ級数

$$S_n f(\lambda) = \sum_{j=-n}^n a_j e^{2\pi i j \lambda}$$

のチェザロ和

$$\frac{1}{n} (S_0 f + S_1 f + \cdots + S_{n-1} f)$$

は  $f$  に一様収束する. ただし,

$$a_j = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\lambda) e^{-2\pi i j \lambda} d\lambda$$

である.

一般に,  $f \in L^2$  である限り  $S_n f$  は  $L^2$  の意味では  $f$  に収束する. つまり,

$$\lim_{n \rightarrow \infty} \|S_n f - f\|^2 = \lim_{n \rightarrow \infty} \int |S_n f(\lambda) - f(\lambda)|^2 d\lambda = 0$$

が成立する. しかし,  $S_n f(\lambda)$  が  $f(\lambda)$  に各点収束するかどうかまでは保証されない. もちろん,  $f(\lambda)$  が無限回連続微分可能であれば, 対応する伝達関数が単位円上で正則であるので, 可逆性の定理の証明と同じようにして  $\sum_{j=-\infty}^{\infty} |a_j| < \infty$  がいえ,  $S_n f(\lambda)$  は  $f(\lambda)$  に各点収束するが, このような条件がない限り, 各点収束するとは限らない. そこで代わりにするのが上のチェザロ和である. 収束が一様であることまでいえているので, きわめて強力な補題である. フーリエ解析の教科書には必ず載っている補題 (定理) であるので証明は省略する (レポート問題).

**補題**

$f(\lambda)$ ,  $-\frac{1}{2} \leq \lambda < \frac{1}{2}$  が連続で  $f(\lambda) = f(-\lambda)$ ,  $f(-\frac{1}{2}) = f(\frac{1}{2}-)$  ならば, 任意の  $\varepsilon > 0$  に対して

$$\left| f(\lambda) - \sigma^2 \left| g_\varepsilon(e^{2\pi i \lambda}) \right|^2 \right| < \varepsilon$$

となるような実係数の  $p$  次多項式

$$g_\varepsilon(z) = 1 + g_1 z + \cdots + g_p z^p$$

が存在する. ただし,  $g_\varepsilon(z) \neq 0, |z| \leq 1$ ,

$$\sigma^2 = \left( \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\lambda) d\lambda \right) / (1 + g_1^2 + \cdots + g_p^2)$$

である.

### 証明

$f(\lambda) \equiv 0$  なら  $\sigma^2 = 0$  であり, どんな  $g_\varepsilon$  でもよい. そこで  $M = \sup f(\lambda) > 0$  のときに補題を示せばよい.  $\delta > 0$  に対して  $f^\delta(\lambda) = \max\{f(\lambda), \delta\}$  を定義すれば,  $f^\delta(\lambda) \geq \delta$  で  $0 \leq f^\delta(\lambda) - f(\lambda) \leq \delta$  である. チェザロ総和定理を用いれば, 十分大きな  $n$  に対して一様に

$$|f_n^\delta(\lambda) - f^\delta(\lambda)| < \epsilon'$$

となる. ただし,  $f_n^\delta(\lambda) = \frac{1}{n} \sum_{j=0}^{n-1} S_j f^\delta(\lambda) > 0$  である.  $f(\lambda)$  の対称性よ

り  $f^\delta(\lambda)$  も対称となり, フーリエ係数  $a_j = \int_{-\frac{1}{2}}^{\frac{1}{2}} f^\delta(\lambda) e^{-2\pi i j \lambda} d\lambda$  は実数で,  $a_j = a_{-j}$  である. これから多項式

$$\begin{aligned} P(z) &= \frac{1}{n} \sum_{k=0}^{n-1} \sum_{j=-k}^k a_j z^j \\ &= \sum_{k=-n}^n \left( 1 - \frac{|k|}{n} \right) a_k z^k \end{aligned}$$

を作れば,  $f_n^\delta(\lambda) = P(e^{2\pi i \lambda})$  である.  $p = \max\{k : a_l = 0, l \geq k\}$  とすれば,  $a_{p+1} = a_{p+2} = \cdots = 0$  であり,  $P(z)$  はその形から根  $\alpha_1, \cdots, \alpha_p, \alpha_1^{-1}, \cdots, \alpha_p^{-1}$  をもつ. ただし,  $|\alpha_j| > 1, j = 1, \cdots, p$  にとっておく. ここで  $P(e^{2\pi i \lambda}) > 0$  であることを用いた. つまり, 単位円上に根をもたない. このことから, ある定数  $c_1$  を用いて

$$z^p P(z) = c_1 \prod_{j=1}^p (1 - \alpha_j^{-1} z)(1 - \alpha_j z)$$

と表せるはずである。そこで

$$\begin{aligned} g_\varepsilon(z) &= \prod_{j=1}^p (1 - \alpha_j^{-1}z) \\ &= 1 + g_1z + \cdots + g_pz^p \end{aligned}$$

と定義すれば

$$\begin{aligned} g_\varepsilon(z)g_\varepsilon(z^{-1}) &= \prod_{j=1}^p (1 - \alpha_j^{-1}z)(1 - \alpha_j^{-1}z^{-1}) \\ &= (-1)^p \prod_{j=1}^p (\alpha_j^{-1})z^{-p} \prod_{j=1}^p (1 - \alpha_j^{-1}z)(1 - \alpha_jz). \end{aligned}$$

つまり、 $P(z)$  と  $g_\varepsilon(z)g_\varepsilon(z^{-1})$  は定数倍の違いを除いて一致する。そこで  $z^0$  の係数を比較することによって、

$$P(z) = c_2 g_\varepsilon(z)g_\varepsilon(z^{-1})$$

を得る。ただし、

$$c_2 = \frac{\int_{-\frac{1}{2}}^{\frac{1}{2}} f^\delta(\lambda) d\lambda}{1 + g_1^2 + \cdots + g_p^2}.$$

以上より、

$$\begin{aligned} \left| f(\lambda) - \sigma^2 |g_\varepsilon(e^{2\pi i\lambda})|^2 \right| &\leq |f(\lambda) - f^\delta(\lambda)| + \left| f^\delta(\lambda) - c_2 |g_\varepsilon(e^{2\pi i\lambda})|^2 \right| + |c_2 - \sigma^2| |g_\varepsilon(e^{2\pi i\lambda})|^2 \\ &= |f(\lambda) - f^\delta(\lambda)| + |f^\delta(\lambda) - f_n^\delta(\lambda)| \\ &\quad + \frac{|f(f^\delta(\lambda) - f(\lambda))d\lambda|}{1 + g_1^2 + \cdots + g_p^2} |g_\varepsilon(e^{2\pi i\lambda})|^2 \end{aligned}$$

ここで、

$$\frac{|g_\varepsilon(e^{2\pi i\lambda})|^2}{1 + g_1^2 + \cdots + g_p^2} = \frac{f_n^\delta(\lambda)}{\int_{-\frac{1}{2}}^{\frac{1}{2}} f^\delta(\lambda) d\lambda} \leq \frac{f^\delta(\lambda) + \epsilon'}{\int_{-\frac{1}{2}}^{\frac{1}{2}} f^\delta(\lambda) d\lambda}$$

であるので、 $\delta$  を十分小さくとれば

$$\left| f(\lambda) - \sigma^2 |g_\varepsilon(e^{2\pi i\lambda})|^2 \right| < \varepsilon$$

とできる。 $g_\varepsilon(z) \neq 0$ ,  $|z| \leq 1$  は作り方より明らかである。

### 系

補題と同じ仮定のもとで、 $f(\lambda)$  をいくらでも近似するスペクトルをもつ  $MA$  過程が存在する。

これは、補題の  $g_\varepsilon$  から  $MA(p)$  過程

$$Z_t = g_\varepsilon(B)a_t, \quad E a_t^2 = \sigma^2$$

を定義すればよいだけである。もっとも Wold の分離定理からもこの系はあきらかである。同じように  $AR$  過程でも近似できることを示しているのが次の系である。

### 系

補題と同じ仮定のもとで、 $f(\lambda)$  をいくらでも近似する  $AR$  過程が存在する。

### 証明

$f^\delta(\lambda) = \max\{f(\lambda), \frac{\delta}{2}\}$  とすれば、 $f^\delta(\lambda) \geq \frac{\delta}{2}$  で、補題を  $\frac{1}{f^\delta(\lambda)}$  に適用すれば

$$\left| \frac{1}{f^\delta(\lambda)} - \sigma^2 |g_\varepsilon(e^{-2\pi i\lambda})|^2 \right| < \varepsilon$$

なる  $g_\varepsilon$  が存在する。  $g(\lambda) = \sigma^2 |g_\varepsilon(e^{-2\pi i\lambda})|^2$  とおけば

$$\begin{aligned} \left| f^\delta(\lambda) - \frac{1}{g(\lambda)} \right| &= \frac{f^\delta(\lambda)}{g(\lambda)} \left| \frac{1}{f^\delta(\lambda)} - g(\lambda) \right| \\ &\leq \frac{f^\delta(\lambda)^2}{1 - \varepsilon f^\delta(\lambda)} \cdot \varepsilon \end{aligned}$$

となるので、 $\varepsilon$  を十分小さくとれば

$$\begin{aligned} \left| f(\lambda) - \frac{1}{g(\lambda)} \right| &\leq |f(\lambda) - f^\delta(\lambda)| + \left| f^\delta(\lambda) - \frac{1}{g(\lambda)} \right| \\ &\leq \frac{\delta}{2} + \frac{\delta}{2} = \delta. \end{aligned}$$

従って、 $AR(p)$  過程

$$g_\varepsilon(B)Z_t = a_t, \quad E a_t^2 = 1/\sigma^2$$

を考えれば  $\{Z_t\}$  は  $f(\lambda)$  を  $\delta$  の誤差で近似するスペクトル  $g(\lambda)$  をもつ弱定常過程である。

## 4 基本的な時系列モデル

ここで,  $AR(p)$  モデルと  $MA(q)$  モデルについてまとめておく. 以下で  $\{a_t\}$  は直交過程である.

### $AR(p)$ モデル

$$\Phi(B)Z_t = a_t$$

を満たす  $\{Z_t\}$  である.

- a)  $\{Z_t\}$  が弱定常であるための必要十分条件は  $\Phi(z) \neq 0$  for  $|z| = 1$ . なぜなら

$$\sigma^2 d\lambda = |\Phi(e^{2\pi i\lambda})|^2 dF(\lambda)$$

に矛盾する.

- b)  $\{a_t\}$  が  $\{Z_t\}$  のイノベーションであるためには更に  $\Phi(z) \neq 0$  for  $|z| \leq 1$  でなければならない. なぜなら弱定常性より,  $\Phi(z) \neq 0$  on  $|z| = 1$  であるので  $\Phi(z) \neq 0$  for  $1 - \varepsilon < |z| < 1 + \varepsilon$ , したがってこの範囲で

$$\Phi(z)^{-1} = \sum_{l=-\infty}^{\infty} \psi_l z^l$$

のようにローラン展開できる. 可逆性の証明と同じように  $\sum_{l=-\infty}^{\infty} |\psi_l| < \infty$  がいえるので

$$Z_t = \sum_{l=-\infty}^{\infty} \psi_l a_{t-l}$$

なる表現 (両側  $MA$  表現) が従う.  $\{a_t\}$  がイノベーションであるためには

$\psi_{-1} = \psi_{-2} = \dots = 0$  であることが必要十分で, これは  $\Phi(z) \neq 0$  for  $|z| \leq 1$  を意味する.

両側  
 $MA$   
表現

- c) 更に, b) の条件があれば  $\{Z_t\}$  は  $MA(\infty)$  表現も持つ.

通常は  $AR(p)$  モデルといえば  $\Phi(z) \neq 0$  for  $|z| \leq 1$  の条件を満たすものとする.

### $MA(q)$ モデル

$$Z_t = \Psi(B)a_t$$

を満たす  $\{Z_t\}$  であるが,  $AR$  と異なり, 常に弱定常であり自動的に  $\{a_t\}$  は  $\{Z_t\}$  のイノベーションとなる. また

$$\Psi(z) \neq 0 \quad \text{for } |z| \leq 1$$

ならば  $AR(\infty)$  表現をもつ.

— Ornstein-Uhlenbeck 過程と  $AR$  モデル —

確率微分方程式

$$dZ_t = \theta(\mu - Z_t)dt + \sigma dB(t)$$

を満たす連続時間の確率過程は Ornstein-Uhlenbeck 過程と呼ばれ, さまざまな場面で登場する. ただし,  $B(t)$  はブラウン運動である. この確率過程からの離散時間サンプリング  $\{\dots, Z_1, Z_2, Z_3, \dots\}$  は  $AR(1)$  モデルにしたがう. 実際, この確率微分方程式の解は

$$Z_t = \mu + \frac{\sigma}{\sqrt{2\theta}} B(e^{2\theta t}) e^{-\theta t}$$

と表せるので,

$$(Z_t - \mu) - e^{-\theta}(Z_{t-1} - \mu) = \frac{\sigma}{\sqrt{2\theta}} (B(e^{2\theta t}) - B(e^{2\theta(t-1)})) e^{-\theta t}$$

となる, 右辺を  $a_t$  とおけば,  $Var(a_t) = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta})$  で, 時間によらず一定の分散で, ブラウン運動の定義よりイノベーションになっていることがわかる.

確率微分方程式

確率微分方程式

$$dX_t = a(t, \omega)dt + \nu(t, \omega)dB(t)$$

に従う確率過程  $\{X_t\}$  を微分可能な 2 変数関数  $f(x, t)$  で変換した  $Y_t = f(X_t, t)$  は確率微分方程式

$$\begin{aligned} dY_t &= \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dX_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (dX_t)^2 \\ &= \left( \frac{\partial f}{\partial t} + a \frac{\partial f}{\partial x} + \frac{\nu^2}{2} \frac{\partial^2 f}{\partial x^2} \right) dt + \nu \frac{\partial f}{\partial x} dB(t) \end{aligned}$$

にしたがう (Ito の公式). この公式を利用すれば, Ornstein-Uhlenbeck 過程の解も求まる, 実際,  $Z_t$  に  $f(x, t) = xe^{\theta t}$  なる変換を施せば

$$df(Z_t, t) = e^{\theta t} \theta \mu dt + \sigma e^{\theta t} dB(t)$$

これを  $t_0$  から  $t$  まで積分すれば

$$Z_t e^{\theta t} - Z_{t_0} e^{\theta t_0} = \int_{t_0}^t e^{\theta s} \theta \mu ds + \int_{t_0}^t \sigma e^{\theta s} dB(s)$$

ここで,

$$\int_{t_0}^{t_1} e^{\theta s} dB(s) = \frac{1}{\sqrt{2\theta}} B(e^{2\theta t_1} - e^{2\theta t_0})$$

となることに注意すれば (ただし, 右辺のブラウン運動は左辺のブラウン運動と同じものではない),  $t_0 \rightarrow -\infty$  として

$$Z_t = \mu + \frac{\sigma}{\sqrt{2\theta}} B(e^{2\theta t}) e^{-\theta t}$$

を得る. 同じように, 対数正規過程

$$dZ_t = \mu Z_t dt + \sigma Z_t dB(t)$$

の解も

$$\log(Z_t) = \log(Z_{t_0}) + (\mu - \sigma^2/2)(t - t_0) + \sigma(B(t) - B(t_0))$$

のように得られる.

## ARMA モデル

$AR(p)$  と  $MA(q)$  を合わせたモデルが  $ARMA(p, q)$  モデルで

$$\Phi(B)Z_t = \Psi(B)a_t$$

を満たす  $\{Z_t\}$  として定義される.

a) 弱定常となるための必要十分条件は

$$\Phi(z) \neq 0 \quad \text{on} \quad |z| = 1.$$

b) さらに  $\{a_t\}$  が  $\{Z_t\}$  のイノベーションであるためには

$$\Phi(z) \neq 0 \quad \text{for} \quad |z| \leq 1$$

が必要である.

c)  $AR(\infty)$ ,  $MA(\infty)$  表現を持つかどうかは,  $MA$ ,  $AR$  のときと同様である.

$ARMA$  過程のスペクトル密度は

$$f(\lambda) = \sigma^2 \frac{|\Psi(e^{2\pi i\lambda})|^2}{|\Phi(e^{2\pi i\lambda})|^2}$$

で与えられる. なお  $ARMA$  モデル特有の問題として, 同定可能性の問題がある. つまり,  $\Phi(z)$  と  $\Psi(z)$  が共通根 (common roots) をもつと冗長な表現となり, パラメータが一意的に定まらない.

**例**

$$\begin{aligned} Z_t &= a_t && ARMA(0, 0) \\ Z_t - \alpha Z_{t-1} &= a_t - \alpha a_{t-1} && ARMA(1, 1) \end{aligned}$$

後者は任意の  $\alpha$  ( $|\alpha| < 1$ ) について前者の方程式と同等である. 共通根  $\alpha^{-1}$  を持つからである.

逆に  $ARMA$  モデルの利点は, スペクトルが有理関数の形をしており, 連続スペクトル, 特に高い山や谷を持つスペクトルを近似しやすいことがあげられる. また, 次の定理で示されているように, 和の演算に関し

て閉じていることも大きな利点である。また、たとえば2階の確率微分方程式を満たす  $d^2 Z_t + \alpha_1 dZ_t + \alpha_0 Z_t dt = dB(t)$  を満たす確率過程からの離散サンプリングは ARMA(2,1) モデルに従う。

**定理** (Granger and Morris, 1976, J.Roy. Statist. Soc. A,139,246-257)

$\{Z_t^{(1)}\}$  と  $\{Z_t^{(2)}\}$  が直交していて、それぞれ ARMA( $p_1, q_1$ ), ARMA( $p_2, q_2$ ) 過程ならば、和  $Z_t = Z_t^{(1)} + Z_t^{(2)}$  も ARMA( $p_1 + p_2, \max(p_1 + q_2, p_2 + q_1)$ ) 過程となる。

**証明**

$$\begin{aligned}\Phi_1(B)Z_t^{(1)} &= \Psi_1(B)a_t^{(1)} \\ \Phi_2(B)Z_t^{(2)} &= \Psi_2(B)a_t^{(2)}\end{aligned}$$

より

$$(\Phi_1(B)\Phi_2(B))Z_t = \Phi_2(B)\Psi_1(B)a_t^{(1)} + \Phi_1(B)\Psi_2(B)a_t^{(2)}$$

を得る。ここで右辺の自己相関係数  $\{\rho_k\}$  は

$$\rho_k = 0 \quad \text{for } k > \max(p_1 + q_2, p_2 + q_1)$$

となっていることに注意する。次の補題での  $\Phi_2(B)\Psi_1(B)a_t^{(1)}$  に対応する関数  $v$  を  $v_1(y)$ ,  $\Phi_1(B)\Psi_2(B)a_t^{(2)}$  に対応する関数を  $v_2(y)$  とすれば右辺に対応する  $v$  は

$$v(y) = \frac{\gamma_0^{(1)}}{\gamma_0^{(1)} + \gamma_0^{(2)}} v_1(y) + \frac{\gamma_0^{(2)}}{\gamma_0^{(1)} + \gamma_0^{(2)}} v_2(y)$$

である。  $-2 < y < 2$  に対して  $v_1(y) \geq 0$ ,  $v_2(y) \geq 0$  であることから  $v(y) \geq 0$ 。これはふたたび補題を用いれば、右辺が高々  $\max(p_1 + q_2, p_2 + q_1)$  次の MA 過程で表現できることになる。

**補題 (Wold, 1953)**

与えられた自己相関係数  $\{\rho_k\}$  をもつ MA( $q$ ) 過程が存在するための必要十分条件は、

$$\rho_k = 0 \quad \text{for } k > q$$

であり, かつ

$$u(x) = 1 + \sum_{j=1}^q \rho_j (x^j + x^{-j})$$

から変数変換  $y = x + x^{-1}$  で作った  $q$  次の多項式  $v(y)$  が  $-2 < y < 2$  で奇数次の根を持たない, つまり偶数次の重根のみである (レポート:  $v(y)$  が  $q$  次の多項式であることを示せ) ことである.

ここで

$$\begin{aligned} v(2) &= u(1) = f(0)/\gamma_0 \geq 0 \\ v(-2) &= u(-1) = f\left(-\frac{1}{2}\right) / \gamma_0 \geq 0 \end{aligned}$$

に注意すれば (レポート), 奇数次の根を持たないことは  $-2 < y < 2$  に対して  $v(y) \geq 0$ , つまり符号が変わらないこと同等である.

### 証明

まず, ある定数  $C$  と  $\Psi$  が存在して, 等式

$$u(x) = C\Psi(x)\Psi\left(\frac{1}{x}\right)$$

が任意の実数  $x \neq 0$  について成立することに注意する (レポート:  $MA(q)$  過程の  $\rho_k$  を  $\Psi(B)$  の係数で表すことにより上の等式を示せ).  $y = x + x^{-1}$  は逆に

$$x = \frac{y}{2} \pm \sqrt{\left(\frac{y}{2}\right)^2 - 1}$$

と書き換えられ,  $v(y) = 0$  の根はいずれも  $u(x) = 0$  の根に対応する.  $v(y) = 0$  の根  $y_0$  について場合分けをする.

#### 場合 1. $|y_0| \geq 2$

対応する  $x_1, x_2$  は実数で  $x_1 = \frac{1}{x_2}$  の関係を持つので, どちらかが  $\Psi(x) = 0$  の根となる.

#### 場合 2. $|y_0| < 2$

対応する  $x_1, x_2$  は絶対値 1 で  $x_1 = \bar{x}_2$  の関係にある. この場合, 実係数多項式の複素根は共役も必ず根になるので,  $x_1, x_2$  ともに  $u(x)$  の根つまり  $\Psi(x)$  の根でなければならなくなる.

$v(y)$  も  $\Psi(x)$  も  $q$  次の多項式であるので  $|y_0| < 2$  の根は必ず 2 重根でない  
と根の個数が合わなくなる. さらに対応する  $x_1, x_2 = \bar{x}_1$  が重根である場  
合なども考えれば, 必ず偶数次の重根でなければならないことになる.

逆は,  $v(y) = 0$  の根から  $\Psi(x) = 0$  の根がすべて定まるので,  $\{\rho_k\}$  から  
 $u(x), v(x)$  を経由して  $MA(q)$  の係数をすべて実数として定められること  
に注意すればよい.

注) この定理からわかるように

$$MA(q_1) + MA(q_2) = MA(\max(q_1, q_2))$$

しかし

$$AR(p_1) + AR(p_2) = ARMA(p_1 + p_2, \max(p_1, p_2)).$$

また, 次の例からわかるように, 右辺の次数は最大の次数を与えている  
だけで, 実際の次数はこれ以下.

例

$$\begin{aligned} (1 + \phi B)Z_t^{(1)} &= a_t^{(1)}, & \text{Var}(a_t^{(1)}) &= \sigma^2 \\ (1 - \phi B)Z_t^{(2)} &= a_t^{(2)}, & \text{Var}(a_t^{(2)}) &= \sigma^2 \end{aligned}$$

確率微分  
方程式の  
解が  
ARMA

これから

$$(1 - \phi B)(1 + \phi B)(Z_t^{(1)} + Z_t^{(2)}) = (1 - \phi B)a_t^{(1)} + (1 + \phi B)a_t^{(2)}$$

となるが, 右辺はすでに直交過程 (レポート) になっているので  $ARMA(2, 0)$   
過程である.

## 5 その他のモデル

### ARIMA(p, d, q) モデル

$$\Phi(B)(1 - B)^d Z_t = \Psi(B)a_t$$

$d$  が正整数ならば  $\{Z_t\}$  は非定常であるが,  $d$  階差分をとった  $(1 - B)^d Z_t$  が定  
常な  $ARMA$  モデルに従うというモデル (Autoregressive Integrated Mov-  
ing Average Model) である. 経済時系列では  $d = 1$  とした  $ARIMA(p, 1, q)$   
のほか, 月別時系列のときは

$$\Phi(B)(1 - B^{12})Z_t = \Psi(B)a_t$$

のような非定常モデルもよく用いられる。前年比が重要だからである。一方、 $-\frac{1}{2} < d < \frac{1}{2}$  ならば  $ARIMA(p, d, q)$  は弱定常になり、Fractional  $ARMAX$   $ARIMA$  モデルとも呼ばれる。この場合、 $d$  階差分は

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k$$

なる二項展開として考える。特に  $0 < d < \frac{1}{2}$  ならばスペクトルが原点付近で  $\lambda^{-2d}$  のオーダーで発散し、自己相関が

$$\rho_k = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} \approx \frac{\Gamma(1-d)}{\Gamma(d)} |k|^{2d-1}$$

のように  $k$  が増加するにつれかなりゆっくり減少するので長期記憶性 (long memory) をもつ確率過程の例となる。また  $-\frac{1}{2} < d < 0$  のときは  $f(0) = 0$  であることより  $\sum_{k=-\infty}^{\infty} \rho_k = 0$  という特殊な性質をもつ確率過程となる。

### GARMA(p, q) モデル

$ARMA$  モデルの  $\Phi(B)$ ,  $\Psi(B)$  が過去の観測値に依存してもよいという形での一般化である。

$$\Phi_t(B)Z_t = \Psi_t(B)a_t$$

$$\Phi_t(B) = 1 + \phi_1(t, Z_t, \dots)B + \dots + \phi_p(t, Z_t, \dots)B^p$$

$$\Psi_t(B) = 1 + \psi_1(t, Z_t, \dots)B + \dots + \psi_q(t, Z_t, \dots)B^q$$

ただし、定常になるとは限らない。

例  $Z_t = e^{|Z_{t-1}|}Z_{t-1} + a_t$  (Ozaki, JAP18, 443-451)

また、 $AR(p)$  の係数  $\phi_1, \dots, \phi_p$  を  $\phi_i + B_i(t)$ ,  $i = 1, \dots, p$  のように  $\{a_t\}$  とは独立で時間  $t$  によらず一定の分布をもつ誤差  $(B_1(t), \dots, B_p(t))$  で揺動した Random Coefficient  $AR$  モデルも研究されている (Andel(1996), Math. Operationsforsch Statist., 7, 735-741, Nicholls and Quinn(1981), J.Multiv Anal., 185-198) .

### EAR, EMA, EARMA モデル

$AR, MA, ARMA$  であることには変わりがないが、雨量時系列のように周辺分布が指数分布であるときには、特別な表現をもつ。

例  $Z_t = \rho Z_{t-1} + a_t$

ここで、 $\phi_{Z_t}(s) = E(e^{-Z_t s})$  : モーメント母関数  
とすれば、

$$\phi_{Z_t}(s) = \phi_{Z_{t-1}}(\rho s) \phi_{a_t}(s)$$

が成立し、 $Z_t$  の周辺分布が指数分布であることから  $\phi_{Z_t}(s) = \phi_{Z_{t-1}}(s) = \frac{\lambda}{\lambda+s}$ . 従って、

$$\phi_{a_t}(s) = \frac{\lambda + \rho s}{\lambda + s} = \rho + (1 - \rho) \frac{\lambda}{\lambda + s}.$$

この形から

$$a_t = \begin{cases} 0 & \text{with prob. } \rho \\ e_t & \text{with prob. } 1 - \rho \end{cases}$$

ただし、 $\{e_t\}$  は独立で指数分布する変数である。これをもとへ戻せば

$$Z_t = \begin{cases} \rho Z_{t-1} & \text{with prob. } \rho \\ \rho Z_{t-1} + e_t & \text{with prob. } 1 - \rho \end{cases}$$

という  $\{Z_t\}$  の表現が得られる。

Lawrence & Lewis(1980) J.R.S.S., B, 42, 150-161

Cover & Lewis(1980) Adv. Appl. Prob.,12, No.3.

### Bilinear(p, q, s) モデル

$$\Phi(B)Z_t = \Psi(B)a_t + \sum_{l,m=1}^s \xi_{lm} Z_{t-l} a_{t-m}$$

を満たす  $\{Z_t\}$  を双線形モデルという。この名前は右辺第2項が、

$$\mathbf{Z}_t = (Z_t, \dots, Z_{t-s})^T, \quad \mathbf{a}_t = (a_t, \dots, a_{t-s})^T$$

に対して

$$\begin{aligned} f(\alpha \mathbf{Z}_t, \mathbf{a}_t) &= \alpha f(\mathbf{Z}_t, \mathbf{a}_t) \\ f(\mathbf{Z}_t, \beta \mathbf{a}_t) &= \beta f(\mathbf{Z}_t, \mathbf{a}_t) \end{aligned}$$

のように双線形性を持つことによる。右辺第2項の係数  $\{\xi_{lm}\}$  がすべて0ならば

$ARMA(p, q)$  モデルとなる。この意味で、 $ARMA$  モデルの1つの拡張とみることもできる。

Granger and Anderson(1978) "Introduction to Bilinear Time Series Models".

T.S.Rao(1981) J.R. Statist. Soc. 1343, 244-255.

### ARCH, GARCH モデル

これは  $\{Z_t\}$  自身のモデルではなく、 $AR$  モデルや  $ARMA$  モデルに現れる  $\{a_t\}$  の部分の1つのモデルである。 $\{a_t\}$  が直交過程と仮定する代わりに、

$$Ea_t = 0$$

$$E(a_t^2 | a_{t-1} \cdots) = \sigma_t^2$$

のように相関を持ち条件付分散も変化するモデルで、経済時系列のモデルとして現在1つの標準的なモデルとして用いられている。 $\sigma_t^2$  が

$$\sigma_t^2 = c + \sum_{i=1}^p \alpha_i a_{t-i}^2$$

の方程式に従うとき  $ARCH(p)$  (Autoregressive Conditional Heteroskedastic Process) と呼ぶ。理由は、背後に

$$a_t^2 = c + \sum_{i=1}^p \alpha_i a_{t-i}^2 + w_t$$

なる  $\{a_t^2\}$  に対する  $AR$  モデルを考えているからである。 $E(w_t | a_{t-1}, a_{t-2}, \cdots) = 0$  を仮定すれば両辺の条件付平均をとることにより上の方程式を得る。ただし、 $a_{t-i}^2, i \geq 1$  が  $\sigma(a_{t-1}, a_{t-2}, \cdots)$  に関して可測であるという条件は必要である。なお、 $\xi_t = \frac{a_t}{\sigma_t}$  とおけば

$$a_t = \xi_t \sigma_t$$

と表現でき、ランダムウォーク  $\{\xi_t\}$  のスケールを  $\{\sigma_t\}$  で変化させたものと考えられる。また

$$\sigma_t^2 = c + \sum_{i=1}^p \alpha_i a_{t-i}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2$$

の方程式に従うとき  $GARCH(p, q)$  (Generalized ARCH) と呼ぶ.

### 局所 AR モデル

全体としては非定常であるが, 局所的には定常とみなせる場合も多い. このような場合に, いくつかの AR モデルを短期間でスイッチして当てはめることが考えられる.

#### ・ 時間に関する局所 AR

$$Z_t = Z_t^{(i)} + m_i, \quad t_{i-1} \leq t \leq t_i, \quad i = 1, \dots, k$$

ただし,  $Z_t^{(i)} \sim AR(p_i)$

#### ・ 値に関する局所 AR

値がある範囲を越えたら別の AR モデルにスイッチするモデルで  $TAR$  (Threshold AR) と略称される.

$$Z_t + \sum_{i=1}^p \phi_i^{(J)} Z_{t-i} = a_t^{(J)}$$

$J$  は過去の観測値で定まる確率変数.

例  $\mathbf{R} = \bigcup_{j=1}^k \mathbf{R}_j$  と分割し,  $Z_{t-d} \in \mathbf{R}_j$  ならば  $J = j$  にスイッチする.

より具体的には, たとえば

$$Z_t = \begin{cases} \frac{3}{4}Z_{t-1} + a_t^{(1)} & \text{if } |Z_{t-1}| \leq \frac{1}{2} \\ \frac{1}{2}Z_{t-1} + a_t^{(2)} & \text{otherwise} \end{cases}$$

Tong(1983), "Threshold Models in Non-linear Time Series Analysis",  
Lecture Notes in Statistics, 21, Springer.

## 6 ARモデルにもとづく推測

まず,

$$\Phi(B)Z_t = a_t$$

を満たす  $AR(p)$  過程の自己共分散の性質を調べる.  $AR(p)$  だから当然  $\{Z_t\}$  は定常で,  $\{a_t\}$  はそのイノベーションである.

$$E(a_{t+k} Z_t) = 0, \quad k \geq 1$$

より

$$\gamma_k + \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p} = 0, \quad k \geq 1$$

つまり

$$\Phi(B)\gamma_k = 0, \quad k \geq 1$$

が成り立つ. これは  $p$  次の差分方程式で, その解は次のようにして求まる.

まず2次の差分方程式

$$h_t + \alpha_1 h_{t-1} + \alpha_2 h_{t-2} = 0$$

を例として考えよう.  $\alpha(B) = 1 + \alpha_1 B + \alpha_2 B^2 = (1 - \xi_1^{-1} B)(1 - \xi_2^{-1} B)$  とする. まず,  $h_t = \xi_1^{-t}, h_t = \xi_2^{-t}$  が解であることは直接計算によりすぐわかる. 従って, 線形結合  $h_t = c_1 \xi_1^{-t} + c_2 \xi_2^{-t}$  も解となる. 差分方程式で相続く2つの初期値, たとえば  $h_0$  と  $h_1$  を与えればすべての解の値は定まってしまうので,  $\xi_1 \neq \xi_2$  ならば  $c_1$  と  $c_2$  を選ぶことにより任意の解を  $h_t = c_1 \xi_1^{-t} + c_2 \xi_2^{-t}$  の形で表現できる. また,  $\xi_1^{-t}$  と  $\xi_2^{-1}$  が線形独立であることから,  $c_1, c_2$  も一意的に定まる. また,  $\xi_1 \neq \xi_2$  が複素根ならば  $\xi_1 = \bar{\xi}_2$  で

$$h_t = c \xi_1^{-1} + \bar{c} \bar{\xi}_1^{-t} = 2ar^{-t} \cos(t\theta - b)$$

となる. ただし,  $c = ae^{bi}, \xi_1 = re^{i\theta}$  とおいた.

$\xi_1 = \xi_2$  ならば, 同様の議論から  $h_t = \xi_1^{-t}$  と  $h_t = t\xi_1^{-t}$  が解となり, すべての解は

$$h_t = c_1 \xi_1^{-t} + c_2 t \xi_1^{-t}$$

と一意的に表現できる.

以上を一般化することによって次の定理を得る.

**定理**

差分方程式  $\alpha(B)h_t = 0$  の解は,  $\alpha(B) = \prod_{j=1}^m (1 - \xi_j^{-1}B)^{r_j}$  と分解できたと仮定して

$$h_t = \sum_{j=1}^m \sum_{l=0}^{r_j-1} c_{jl} t^l \xi_j^{-t}$$

で与えられる. ただし,  $\{c_{jl}\}$  は (複素) 定数である.

この定理から,  $\Phi(B)$  が定理の  $\alpha(B)$  と同じように分解できたとして,  $\gamma_k$  は

$$\gamma_k = \sum_{j=1}^m \sum_{l=0}^{r_j-1} c_{jl} k^l \xi_j^{-k}$$

と表現できる. ただし定常性から  $\Phi(B)$  の根はすべて単位円の外, つまり  $|\xi_j| > 1, j = 1, \dots, m$  であるので,  $\gamma_k$  は  $k$  の増加にともなって, 指数的に減少することがわかる. これは  $MA$  モデルの場合の  $\gamma_k$  の cut off 性と対照的である.

## 6.1 $\gamma_k$ の推定

1つの自然な推定量は

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} Z_t Z_{t+k}$$

である. もちろん  $\{Z_t\}$  の弱定常性が前提である.  $n$  で割る代わりに  $n-k$  で割れば不偏つまり  $E\hat{\gamma}_k = \gamma_k$  になるが,  $\gamma_k$  が  $k \rightarrow \infty$  で 0 に収束することを考慮すると,  $n-k$  で割って, 大きな  $k$  に対して不安定な推定になるよりは  $n$  で割った方が誤解を招かなくてよい.

$\hat{\gamma}_k$  の共分散, 信頼幅, 漸近分布

## 6.2 $\phi_1, \dots, \phi_p$ の推定

1つの方法は

$$\gamma_k + \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} = 0, \quad k \geq 1$$

を利用して連立方程式

$$\hat{\gamma}_k + \phi_1 \hat{\gamma}_{k-1} + \dots + \phi_p \hat{\gamma}_{k-p} = 0, \quad k = 1, \dots, p$$

を解くことであるが、推定量のよさなどを考えると本質的には同じであるが、2乗誤差

$$\sum_{t=p+1}^n (Z_t + \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p})^2$$

を  $\phi_1, \phi_2, \dots, \phi_p$  に関して最小化する方法の方がよい。この方法は近似的には  $\{a_t\}$  に正規性を仮定したときの最尤推定量にもなっている。具体的には連立方程式

$$\sum_t Z_t Z_{t-k} + \phi_1 \sum_t Z_{t-1} Z_{t-k} + \cdots + \phi_p \sum_t Z_{t-p} Z_{t-k} = 0, \quad k = 1, 2, \dots, p$$

を解くことになる。つまり行列

$$\hat{\Gamma}_p = \left( \frac{1}{n} \sum_{t=p+1}^n Z_{t-i} Z_{t-j}, \quad 1 \leq i, j \leq p \right)$$

を用いて書き直せば,

$$\hat{\Gamma}_p \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} = - \begin{pmatrix} \tilde{\gamma}_1 \\ \vdots \\ \tilde{\gamma}_p \end{pmatrix}$$

を解くことになる。ただし,

$$\tilde{\gamma}_k = \frac{1}{n} \sum_{t=p+1}^n Z_t Z_{t-k}$$

である。上の方程式は Yule Walker の方程式と呼ばれる。

注) 理論やソフトウェアによっては  $AR(p)$  モデルを

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + a_t$$

と表現していることもある。この場合には  $\phi_1, \dots, \phi_p$  の符号が逆転するので注意が必要である。

### 6.3 スペクトル密度の推定

$AR(p)$  モデルに対するスペクトル密度が

$$f(\lambda) = \frac{\sigma^2}{|\Phi(e^{2\pi i\lambda})|^2}$$

で与えられることに注意すれば,  $\phi_1, \dots, \phi_p$  の推定量  $\hat{\phi}_1, \dots, \hat{\phi}_p$  を用いて, 推定量

$$\hat{f}(\lambda) = \frac{\hat{\sigma}^2}{|\sum_{l=0}^p \hat{\phi}_l e^{2\pi i l \lambda}|^2}$$

が構成できる. ただし,  $\hat{\phi}_0 = 1$  であり,

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=p+1}^n \left( Z_t + \hat{\phi}_1 Z_{t-1} + \dots + \hat{\phi}_p Z_{t-p} \right)^2$$

である.

### 6.4 偏自己相関

偏自己相関は, 通常 of 相関が対象とする変量以外の変量を介した間接的な相関も含んでいるのに対し, そのような間接的な相関を射影により除去した2つの変量間の直接的な相関である. ただし, 2次モーメントにだけ依存した操作であり, 正規分布以外でどれだけ意味があるかどうかはわからない. 時系列では,  $Z_t$  と  $Z_{t+k}$  の偏自己相関 (partial autocorrelation) は  $Z_{t+1}, \dots, Z_{t+k-1}$  を介した間接的な相関を除いた  $Z_t$  と  $Z_{t+k}$  の直接的な相関として定義される. つまり,

$$E \left( Z_{t+k} - \sum_{l=1}^{k-1} \alpha_l Z_{t+k-l} \right)^2$$

を最小化する  $\alpha_1, \dots, \alpha_{k-1}$  を  $\alpha_1^*, \dots, \alpha_{k-1}^*$  と表し,

$$E \left( Z_t - \sum_{l=1}^{k-1} \beta_l Z_{t+l} \right)^2$$

を最小化する  $\beta_1, \dots, \beta_{k-1}$  を  $\beta_1^*, \dots, \beta_{k-1}^*$  と表したとき,

$$\varepsilon_{t+k} = Z_{t+k} - \sum_{l=1}^{k-1} \alpha_l^* Z_{t+k-l}$$

は  $Z_{t+k}$  から  $Z_{t+1}, \dots, Z_{t+k-1}$  の影響を除いた変量,

$$e_t = Z_t - \sum_{l=1}^{k-1} \beta_l^* Z_{t+l}$$

は  $Z_t$  から  $Z_{t+1}, \dots, Z_{t+k-1}$  の影響を除いた変量である。そこで,

$$R_k = E(\varepsilon_{t+k} e_t)$$

を時系列  $\{Z_t\}$  のラグ  $k$  の偏自己共分散,

$$r_k = \frac{R_k}{\sqrt{E\varepsilon_{t+k}^2} \sqrt{Ee_t^2}}$$

をラグ  $k$  の偏自己相関と呼ぶ。ただし,  $r_0 = 1$  と定義しておく。明らかに  $|r_k| \leq 1$  である。ところで  $\varepsilon_{t+k}$  が  $Z_{t+1}, \dots, Z_{t+k-1}$  で張る空間に直交していることを利用すれば

$$\begin{aligned} R_k &= E(\varepsilon_{t+k})(Z_t - \sum_l \beta_l^* Z_{t+l}) \\ &= E(\varepsilon_{t+k} Z_t) \\ &= \gamma_k - \sum_{l=1}^{k-1} \alpha_l^* \gamma_{k-l} \end{aligned}$$

と書くことができる。また,

$$E\varepsilon_{t+k}^2 = Ee_t^2 = \gamma_0 - \sum_{l=1}^{k-1} \alpha_l^* \gamma_l$$

と書くこともできるので,

$$r_k = \frac{\gamma_k - \sum_l \alpha_l^* \gamma_{k-l}}{\gamma_0 - \sum_l \alpha_l^* \gamma_l}$$

なる表現を得る。ここで  $\alpha_1^*, \dots, \alpha_{k-1}^*$  は  $\varepsilon_{t+k}$  が  $Z_{t+k-1}, \dots, Z_{t+1}$  と直交することより,

$$\Gamma_{k-1} \begin{pmatrix} \alpha_1^* \\ \vdots \\ \alpha_{k-1}^* \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_{k-1} \end{pmatrix}$$

の解として求まる. 上の表現を利用すれば観測値  $Z_1, \dots, Z_n$  にもとづく  $r_k$  の推定量も容易に得られる. ところで,  $\{Z_t\}$  が  $AR(p)$  過程ならば  $\varepsilon_{t+k}$  の定義より  $\varepsilon_{t+k} = a_{t+k}$ ,  $k > p$  である. 従って,

$$R_k = 0, \quad k > p$$

つまり

$$r_k = 0, \quad k > p$$

となる. 一方,  $\phi_p \neq 0$  である限り  $r_p \neq 0$  であることが示せるので,  $\{r_k\}$  は cut off の性質をもつ. これは  $MA(q)$  について  $\{\rho_k\}$  が cut off の性質をもつことと対照的である. 従って,  $\{\rho_k\}$  と  $\{r_k\}$  の挙動を調べることにより,  $AR$  か  $MA$  の区別が容易にできることになる.

Ramsey,  
1974  
Durbin-  
Levin-  
son  
Berg の  
方法

### 演習問題

$\{Z_t\}$  が  $AR(1)$  に従い, 正規過程のとき, つまり

$$Z_t = \phi Z_{t-1} + a_t, \quad t = 1, \dots, n$$

ただし,  $a_1, \dots, a_n$  は独立で  $N(0, \sigma^2)$  に従うときを考える.  $Z_i$  以外を条件として与えたときの  $Z_i$  の条件付分布が,  $i = 1$  のときは  $N(Z_2/\phi, \sigma^2/\phi^2)$ ,  $i = n$  のときは  $N(\phi Z_{n-1}, \sigma^2)$ , それ以外のときは  $N(\phi(Z_{n-1} + Z_{n+1})/(1 + \phi^2), \sigma^2/(1 + \phi^2))$  であることを証明せよ. この結果を用いれば,  $(Z_1, \dots, Z_n)$  の同時密度も導くことができる.

## 7 多変量弱定常時系列 (Multiple Time Series, Multivariate Time Series)

$$Z(t) = \begin{pmatrix} Z_1(t) \\ \vdots \\ Z_m(t) \end{pmatrix} \in \mathbb{C}^m, \quad t = \dots, -1, 0, 1, \dots$$

弱定常性:

$$EZ(t) = 0, \quad \gamma(h) = E(Z(t+h)\overline{Z(t)}^T) < \infty.$$

$\gamma(h)$  : 自己共分散系列

$$\gamma(h) = (\gamma_{jk}(h), 1 \leq j, k \leq m)$$

$\gamma_{jk}(h) = E(Z_j(t+h)\overline{Z_k(t)})$  : 時間差  $h$  の要素  $Z_j(t+h)$  と  $Z_k(t)$  の共分散

ここで,  $\gamma_{jk}(h)$  と  $\gamma_{kj}(h)$  は必ずしも一致しないことに注意. どちらの要素の時間を進めて考えるかで, 当然共分散は異なる. 実際  $\gamma_{jk}(h) = \overline{\gamma_{kj}(-h)}$  の関係にある. 一変量のとおり同じように,  $\gamma(h)$  から自己相関係列

$$\rho(h) = (\rho_{jk}(h)), \quad h = \dots, -1, 0, 1, \dots$$

を

$$\rho_{jk}(h) = \frac{\gamma_{jk}(h)}{\sqrt{|\gamma_{jj}(h)|} \sqrt{|\gamma_{kk}(h)|}}$$

で定義できる. また,  $\{\gamma(h), h = \dots, -1, 0, 1, \dots\}$  の各要素をフーリエ変換することによりスペクトル密度

$$f(\lambda) = \sum_{h=-\infty}^{\infty} e^{-2\pi i h \lambda} \gamma(h)$$

も定義できる. ただし,  $f(\lambda) = (f_{jk}(\lambda))$  は  $m \times m$  行列となる. 対角線以外は実数値になるとは限らない. 実際,

$$f_{jk}(\lambda) = \overline{f_{kj}(\lambda)}$$

が任意の  $k, j$  について成立する. 実は, これは行列  $f(\lambda)$  がエルミート行列であることを示していることに他ならない. したがって,  $f(\lambda)$  の固有値は実数.  $f(\lambda)$  はさらに非負定符号行列でもある. したがって, 固有値は実数であるだけでなく非負. 実際  $\xi = (\xi_1, \dots, \xi_m)^T$  を任意の複素ベクトルとしたとき

$$y_t = \xi^T Z(t)$$

を考えれば  $\{y_t\}$  は自己共分散系列

$$\begin{aligned} \gamma_h &= E(y_{t+h}\overline{y_t}) \\ &= \xi^T E(Z(t+h)\overline{Z(t)}) \xi \\ &= \xi^T \gamma(h) \xi \end{aligned}$$

をもつ一変量弱定常時系列である. したがって  $\{y_t\}$  のスペクトル密度を  $f^y(\lambda)$  とすれば

$$0 \leq f^y(\lambda) = \xi^T f(\lambda) \xi$$

が成立する。これは  $f(\lambda)$  が非負定符号行列であることにほかならない。また、

$$\gamma(h) = \int e^{2\pi ih\lambda} f(\lambda) d\lambda$$

で  $\gamma(h)$  に戻る。

なお、Splus で  $\gamma(h)$  を求める関数は、acf,  $f(\lambda)$  を求める関数は spectrum である。ただし、 $f(\lambda)$  を求めるのに method="pgram" と method="ar" の選択ができるが、前者は特にモデルを仮定しないで求め、後者は後述の AR モデルを仮定して求める。省略時は前者である。関数 spectrum は実際には関数 pgram か関数 ar を呼び出して  $\gamma(h)$  を求め、それにもとづいて  $f(\lambda)$  を求めている。ただし、返す値の構造は幾分複雑である。いずれにしても、省略時は線形なトレンドはあらかじめ除かれる (detrend=T)。

自己共分散  $\gamma(h)$  から自己相関  $\rho(h)$  を作ったのと同じように、スペクトル密度  $f(\lambda)$  から複素コヒーレンシー (complex coherency)  $w(\lambda) = (w_{jk}(\lambda))$  を  $f_{jj}(\lambda) \neq 0$ ,  $f_{kk}(\lambda) \neq 0$  である限り次のよう定義できる。

$$w_{jk}(\lambda) = \frac{f_{jk}(\lambda)}{\sqrt{f_{jj}(\lambda)}\sqrt{f_{kk}(\lambda)}}.$$

スペクトル密度  $f(\lambda)$  がもとの時系列に含まれる周波数  $\lambda$  の波の強さの 2 乗平均、言い替えれば強さの分散を表すものであることを思い出せば、 $f_{jk}(\lambda)$  は  $\{Z_j(t)\}$  と  $\{Z_k(t)\}$  に含まれる周波数  $\lambda$  の波の強さの共分散を表し、 $w_{jk}(\lambda)$  は相関を表すものであることがわかる。実際、1 変量時系列のときと同じように弱定常時系列  $\{Z(t)\}$  は

$$Z(t) = \int e^{2\pi it\lambda} dW(\lambda)$$

なるフーリエ表現をもつ。ただし、

$$W(\lambda) = \begin{pmatrix} W_1(\lambda) \\ \vdots \\ W_m(\lambda) \end{pmatrix}$$

は直交増分過程である。従って、

$$f_{jk}(\lambda) d\lambda = E(dW_j(\lambda) \overline{dW_k(\lambda)})$$

が成立し、確かに周波数  $\lambda$  の波のランダムな強さの共分散 (密度) を表すものであることがわかる。なお、 $|w_{jk}(\lambda)|$  を周波数  $\lambda$  のコヒーレンシー

(coherency) といい,  $\arg(w_{jk}(\lambda)) = \arg(f_{jk}(\lambda))$  を周波数  $\lambda$  のフェーズ (phase) という.

$f_{jk}(\lambda)$  は次の性質をもつ.  $Z(t)$  が実数値ならば

1.  $f_{jk}(\lambda) = f_{jk}(\lambda + 1)$
2.  $f_{jk}(\lambda) = f_{kj}(-\lambda)$ , 特に  $j = k$  なら  $f_{jj}(-\lambda) = f_{jj}(\lambda)$
3.  $f_{jk}(-\lambda) = \overline{f_{jk}(\lambda)}$

従って,

$$\operatorname{Re}(f_{jk}(1 - \lambda)) = \operatorname{Re}(f_{jk}(\lambda))$$

と

$$\operatorname{Im}(f_{jk}(1 - \lambda)) = -\operatorname{Im}(f_{jk}(\lambda))$$

が成立し, 行列  $(f_{jk}(\lambda))$  の上三角部分, しかも  $0 \leq \lambda \leq \frac{1}{2}$  に対する値だけで, 他はすべて定まる.

一方,  $Z(t)$  が複素数値ならば 1. は成立するが, 2., 3. は成立しない. 代わりにエルミート性

$$4. f_{jk}(\lambda) = \overline{f_{kj}(\lambda)}$$

のみが成立する. 従って, 複素数値の場合は,  $0 \leq \lambda \leq 1$  に対する値がすべて必要となる.

コヒーレンシーは  $Z(t)$  の要素ごとの線形変換に関して不変である. たとえば, 2変量時系列  $Z(t) = (Z_1(t), Z_2(t))^T$  から実数係数  $\{c_k\}$ ,  $\{d_k\}$  によって作られた

$$Y_1(t) = \sum_{k=-\infty}^{\infty} c_k Z_1(t - k)$$

$$Y_2(t) = \sum_{k=-\infty}^{\infty} d_k Z_2(t - k) ,$$

$Y(t) = (Y_1(t), Y_2(t))^T$  のコヒーレンシーは  $Z(t)$  のコヒーレンシーと同じである. 実際,  $f(\lambda)$  を  $\{Z(t)\}$  のスペクトル,  $g(\lambda)$  を  $\{Y(t)\}$  のスペクトル

ルとすれば

$$\begin{aligned} g_{11}(\lambda) &= |c(\lambda)|^2 f_{11}(\lambda) \\ g_{22}(\lambda) &= |d(\lambda)|^2 f_{22}(\lambda) \\ g_{12}(\lambda) &= c(\lambda) \overline{d(\lambda)} f_{12}(\lambda), \\ \text{ただし, } c(\lambda) &= \sum_k c_k e^{-2\pi i \lambda k}, \quad d(\lambda) = \sum_k d_k e^{-2\pi i \lambda k} \end{aligned}$$

であることからコヒーレンシーが不変であることがすぐ確かめられる。

演習問題  $g_{12}(\lambda)$  に関する上記の表現を確かめよ。

ただし、フェーズは  $\arg(g_{12}(\lambda)) = \arg(c(\lambda)) - \arg(d(\lambda)) + \arg(f_{12}(\lambda))$  の関係で変化する。

$Z_1(t)$  をシステムの入力、 $Z_2(t)$  をシステムの入力とし、システムが

$$Z_2(t) = \sum_{k=1}^p c_k Z_1(t-k) + \varepsilon_t$$

の形に表せるとする。ただし、 $\{\varepsilon_t\}$  は弱定常なノイズで  $\{Z_1(t)\}$  と  $\{\varepsilon_t\}$  は直交しているとする。ここで、伝達関数  $c(\lambda) = \sum_{k=1}^p c_k e^{-2\pi i k \lambda}$  を導入し、 $f_\varepsilon(\lambda)$  を  $\{\varepsilon_t\}$  のスペクトル密度とすれば、2変量時系列  $Z(t) = (Z_1(t), Z_2(t))^T$  に対し、

$$f_{22}(\lambda) = |c(\lambda)|^2 f_{11}(\lambda) + f_\varepsilon(\lambda)$$

が成立する。これは

EQ1

$$1 = \frac{|c(\lambda)|^2 f_{11}(\lambda)}{f_{22}(\lambda)} + \frac{f_\varepsilon(\lambda)}{f_{22}(\lambda)} \quad (6)$$

と書き換えられるが、 $f_{12}(\lambda) = \overline{c(\lambda)} f_{11}(\lambda)$  であるので、(6)の右辺の第1項はコヒーレンシーの2乗  $|w_{12}(\lambda)|^2$  にほかならない。つまり、第1項はこのシステムの入力と出力の周波数ごとの関係の強さを表し、第2項は出力に占めるノイズの大きさを表していることになる。特に、係数  $\{c_k\}$  が  $c_d = c$  以外は0である場合を考えれば、

$$Z_2(t) = c Z_1(t-d) + \varepsilon_t$$

で、スペクトルは

$$\begin{aligned} f_{22}(\lambda) &= c^2 f_{11}(\lambda) + f_\varepsilon(\lambda) \\ f_{12}(\lambda) &= c e^{2\pi i \lambda d} f_{11}(\lambda) \end{aligned}$$

となり,

$$w_{12}(\lambda) = \frac{cf_{11}(\lambda)^{\frac{1}{2}}e^{2\pi i\lambda d}}{(c^2f_{11}(\lambda) + f_\varepsilon(\lambda))^{\frac{1}{2}}}$$

つまり

$$\begin{aligned} |w_{12}(\lambda)| &= \frac{1}{\left(1 + \frac{f_\varepsilon(\lambda)}{c^2f_{11}(\lambda)}\right)^{\frac{1}{2}}} \\ \arg(w_{12}(\lambda)) &= 2\pi\lambda d \end{aligned}$$

の関係が成立する.

### 例 1

$Z_1(t)$  をある商品の需要,  $Z_2(t)$  をその価格を表すものとしたとき, 1 つの最も簡単なモデルは

$$\begin{aligned} Z_1(t) &= -\alpha Z_2(t) + \varepsilon_1(t) \\ Z_2(t) &= \beta Z_1(t-1) + \varepsilon_2(t) \end{aligned}$$

ただし,  $\alpha > 0$  であるが,  $\beta$  はどちらの符号をとる場合もある ( $\beta > 0$  であれば品不足,  $\beta < 0$  であれば十分な供給ができることを示す). このモデルは  $Z(t) = (Z_1(t), Z_2(t))^T$ ,  $\varepsilon(t) = (\varepsilon_1(t), \varepsilon_2(t))^T$  とすれば行列方程式

$$Z(t) = \begin{pmatrix} 0 & -\alpha \\ 0 & 0 \end{pmatrix} Z(t) + \begin{pmatrix} 0 & 0 \\ \beta & 0 \end{pmatrix} Z(t-1) + \varepsilon(t)$$

つまり,

$$(I - F)Z(t) = GZ(t-1) + \varepsilon(t)$$

の形に書ける. ただし,  $\varepsilon_1(t), \varepsilon_2(t)$  は直交し,  $\{\varepsilon(t)\}$  自体も直交過程で  $\varepsilon(t)$  は  $Z(t-1), Z(t-2)$  と直交していると仮定するのが 1 つの自然な仮定である. また  $|\alpha\beta| \neq 1$  とする

演習問題 例 1 の  $\{Z(t)\}$  の  $f(\lambda), w(\lambda)$  を  $\alpha, \beta$  で表せ. ただし  $Cov(\varepsilon(t)) = \sigma^2 I$  とする

上の例からもわかるように,  $\{Z(t)\}$  に対する線形演算

EQ2

$$Y(t) = \sum_{h=-\infty}^{\infty} g(h)Z(t-h) \quad (7)$$

によってできる  $\{Y(t)\}$  の性質を調べておく必要がある。ここで各  $g(h)$  は  $n \times m$  行列であり、 $Y(t)$  は  $n$  次元の値をとる。(7) のような線形演算は (ノイズレス) フィルターとも呼ばれ、 $g(h)$  はインパルスレスポンス行列と呼ばれる。 $\{Z(t)\}$  が弱定常ならば  $\{Y(t)\}$  も弱定常時系列となるので、それぞれ

$$\begin{aligned} Z(t) &= \int e^{2\pi it\lambda} dW^{(Z)}(\lambda) \\ Y(t) &= \int e^{2\pi it\lambda} dW^{(Y)}(\lambda) \end{aligned}$$

なるフーリエ表現をもつ。したがって、伝達関数

$$\Gamma(\lambda) = \sum_{h=-\infty}^{\infty} g(h)e^{-2\pi ih\lambda}$$

を用いて、直交増分過程  $\{W^{(Z)}(\lambda)\}$  と  $\{W^{(Y)}(\lambda)\}$  の関係

$$dW^{(Y)}(\lambda) = \Gamma(\lambda)dW^{(Z)}(\lambda)$$

を得る。これを

$$dW_j^{(Y)}(\lambda) = \Gamma_{j1}(\lambda)dW_1^{(Z)}(\lambda) + \cdots + \Gamma_{jm}(\lambda)dW_m^{(Z)}(\lambda), \quad j = 1, \dots, n$$

の形に書き下してみれば、(7) の時間領域 (Time Domain) での関係よりもずっときれいな線形関係が成立していることがわかる。また、この関係から、 $\{Y(t)\}$  のスペクトル密度  $f^{YY}(\lambda)$  と  $\{Z(t)\}$  のスペクトル密度  $f^{ZZ}(\lambda)$  の間には

EQ2-A

$$f^{YY}(\lambda) = \Gamma(\lambda)f^{ZZ}(\lambda)\Gamma(\lambda)^* \quad (8)$$

の関係があることがわかる。また、 $f^{YZ}(\lambda) = \Gamma(\lambda)f^{ZZ}(\lambda)$  も成立している。ただし、 $\Gamma(\lambda)^*$  は  $\bar{\Gamma}(\lambda)^T$  を表す。なお、一般に、2つの多変量時系列  $\{Y(t)\}$  と  $\{Z(t)\}$  を考えたとき、

$$\begin{aligned} \gamma^{ZZ}(h) &= E(Z(t+h)Z(t)^*) \\ \gamma^{YY}(h) &= E(Y(t+h)Y(t)^*) \\ \gamma^{YZ}(h) &= E(Y(t+h)Z(t)^*) \\ \gamma^{ZY}(h) &= E(Z(t+h)Y(t)^*) \end{aligned}$$

の4種類の共分散系列が考えられ、これらに対応して4種類のスペクトル  $f^{ZZ}(\lambda)$ ,  $f^{YY}(\lambda)$ ,  $f^{YZ}(\lambda)$ ,  $f^{ZY}(\lambda)$  が考えられる。

演習問題  $Y(t), Z(t)$  が実数値のとき,  $f^{YZ}(\lambda)^* = f^{ZY}(\lambda)$  が成り立つことを示せ。

さて, フィルター (7) の一般化であるノイズのあるフィルター

EQ3

$$Y(t) = \sum_{h=-\infty}^{\infty} g(h)Z(t-h) + \varepsilon(t) \quad (9)$$

を考えよう。ただし, 弱定常なノイズ  $\{\varepsilon(t)\}$  は  $\{Z(t)\}$  と直交すると仮定する。つまり  $\gamma^{Z\varepsilon}(h) = E(Z(t+h)\varepsilon(t)^*) = 0, h = \dots, -1, 0, 1, \dots$  とする。そうすると

$$\varepsilon(t) = \int e^{2\pi i t \lambda} dW^{(\varepsilon)}(\lambda)$$

と表せば,

$$dW^{(Y)}(\lambda) = \Gamma(\lambda)dW^{(Z)}(\lambda) + dW^{(\varepsilon)}(\lambda)$$

なる関係から,

$$\begin{aligned} f^{YZ}(\lambda) &= \Gamma(\lambda)f^{ZZ}(\lambda) \\ f^{YY}(\lambda) &= \Gamma(\lambda)f^{ZZ}(\lambda)\Gamma(\lambda)^* + f^{\varepsilon\varepsilon}(\lambda) \end{aligned}$$

を得る。これは (8) の拡張である。これらを組み合わせれば  $f^{ZZ}(\lambda)^{-1}$  が存在する限り

EQ3'

$$f^{YY}(\lambda) = f^{YZ}(\lambda)f^{ZZ}(\lambda)^{-1}f^{YZ}(\lambda)^* + f^{\varepsilon\varepsilon}(\lambda) \quad (10)$$

が成立することもわかる。

さて, 2つの時系列  $\{Y(t)\}$  と  $\{Z(t)\}$  の総合的な相関の指標はどう定義したらよいだろうか。まず,  $Y(t)$  がスカラー値のときを考える。このとき (9) を  $\{Y(t)\}$  を被説明変数  $\{Z(t)\}$  を説明変数とする一種の回帰モデルとしてみれば,  $\{\varepsilon(t)\}$  の分散が相対的に小さければ小さいほど  $\{Y(t)\}$  と  $\{Z(t)\}$  の総合的な相関は高いと考えられる。ここで, 上の結果を用い

れば,

$$\begin{aligned}
\sigma_\varepsilon^2 &= \text{Var}(\varepsilon(t)) = \int e^{2\pi i \lambda t} f^{\varepsilon\varepsilon}(\lambda) d\lambda \\
&= \int (f^{YY}(\lambda) - f^{YZ}(\lambda) f^{ZZ}(\lambda)^{-1} f^{YZ}(\lambda)^*) d\lambda \\
&= \int \left( 1 - \frac{f^{YZ}(\lambda) f^{ZZ}(\lambda)^{-1} f^{YZ}(\lambda)^*}{f^{YY}(\lambda)} \right) f^{YY}(\lambda) d\lambda
\end{aligned}$$

であるので,

$$w_{YZ}(\lambda) = \left( \frac{f^{YZ}(\lambda) f^{ZZ}(\lambda)^{-1} f^{YZ}(\lambda)^*}{f^{YY}(\lambda)} \right)^{\frac{1}{2}}$$

を定義して多重コヒーレンシーと呼び、 $\{Y(t)\}$  と  $\{Z(t)\}$  の総合的な相関の指標として用いることが多い。  $0 \leq w_{YZ}(\lambda) \leq 1$  は (10) より明らかである (演習問題)。 なお

$$\frac{\text{Var}(Y(t)) - \text{Var}(\varepsilon(t))}{\text{Var}(Y(t))} = \frac{\int w_{YZ}(\lambda)^2 f^{YY}(\lambda) d\lambda}{\int f^{YY}(\lambda) d\lambda}$$

の関係に注意すれば、 $w_{YZ}(\lambda)$  は周波数ごとの線形関係の強さを表していると見ることができる。 考え方としては重回帰における重相関と同じである。 なお、特に  $m = 1$  つまり  $Z(t)$  もスカラー値ならば  $w_{YZ}(\lambda)$  は 2 変量時系列  $X(t) = (Y(t), Z(t))^T$  のコヒーレンシー  $|w_{12}(\lambda)|$  と一致する。 これが  $w_{WZ}(\lambda)$  が多重コヒーレンシー (multiple coherency) と呼ばれる理由でもある。

なお、 $Z(t)$  だけでなく  $Y(t)$  も多変量の場合は、各要素  $\{Y_j(t)\}$  と  $\{Z(t)\}$  の多重コヒーレンシーを考えるか、

$$\left( \frac{\det(f^{YZ}(\lambda) f^{ZZ}(\lambda)^{-1} f^{YZ}(\lambda)^*)}{\det(f^{YY}(\lambda))} \right)^{\frac{1}{2}}$$

を考える。

場合によっては、 $\{Y_j(t)\}$  と  $\{Z(t)\}$  との総合的な関連の強さではなく、 $\{Y_j(t)\}$  と特定の要素  $\{Z_k(t)\}$  との直接的な関連の強さを調べる必要が生じる。 以下では、 $\{Z(t)\}$  が 2 次元の値をとるときに、 $Z_2(t)$  の影響を除いた  $\{Y_1(t)\}$  と  $\{Z_1(t)\}$  の直接的な関連の強さの指標を導いてみよう。 まず、

$\{Y_1(t)\}$  と  $\{Z_1(t)\}$  から

$$\begin{aligned}\varepsilon(t) &= Y_1(t) - \sum_{h=-\infty}^{\infty} g^{(1)}(h)Z_2(t-h) \\ e(t) &= Z_1(t) - \sum_{h=-\infty}^{\infty} g^{(2)}(h)Z_2(t-h)\end{aligned}$$

として,  $\|\varepsilon(t)\|$  と  $\|e(t)\|$  がともに最小になるように,  $\{g^{(1)}(h)\}$  と  $\{g^{(2)}(h)\}$  を定めれば, そのときの  $\{\gamma^{\varepsilon e}(h)\}$  を偏自己共分散系列という. なお, このとき明らかに  $\varepsilon(t)$ ,  $e(t)$  とともに  $\{Z_2(t)\}$  の張る線形空間に直交する. さらに,  $G^{(1)}(\lambda) = \sum_h g^{(1)}(h)e^{-2\pi i\lambda h}$ ,  $G^{(2)}(\lambda) = \sum_h g^{(2)}(h)e^{-2\pi i\lambda h}$  とおけば,

$$\begin{aligned}dW^\varepsilon(\lambda) &= dW^{Y_1}(\lambda) - G^{(1)}(\lambda)dW^{Z_2}(\lambda) \\ dW^e(\lambda) &= dW^{Z_1}(\lambda) - G^{(2)}(\lambda)dW^{Z_2}(\lambda)\end{aligned}$$

の関係となるので, 偏自己共分散系列に対応したスペクトル密度, つまり  $\gamma^{\varepsilon e}(h)$  のフーリエ変換,

$$\begin{aligned}f^{\varepsilon e}(\lambda) &= f^{Y_1 Z_1}(\lambda) - G^{(1)}(\lambda)f^{Z_2 Z_1}(\lambda) \\ &\quad - G^{(2)}(\lambda)^* f^{Y_1 Z_2}(\lambda) + G^{(1)}(\lambda)G^{(2)}(\lambda)^* f^{Z_2 Z_2}(\lambda)\end{aligned}$$

を得る. これはさらに

$$f^{\varepsilon e}(\lambda) = f^{Y_1 Z_1}(\lambda) - \frac{f^{Y_1 Z_2}(\lambda)f^{Z_1 Z_2}(\lambda)}{f^{Z_2 Z_2}(\lambda)}$$

と書き換えられ (演習問題), 偏クロススペクトル密度 (partial cross-spectral density) と呼ばれる.  $f^{\varepsilon e}(\lambda)$  を  $f_{Y_1 Z_1 \cdot Z_2}(\lambda)$  と表記することもある. また,

$$w_{Y_1 Z_1 \cdot Z_2}(\lambda) = \frac{f^{\varepsilon e}(\lambda)}{(f^{\varepsilon \varepsilon}(\lambda)f^{e e}(\lambda))^{\frac{1}{2}}}$$

を複素偏コヒーレンシー (partial coherency) という. また,  $\arg(w_{Y_1 Z_1 \cdot Z_2}(\lambda)) = \arg(f^{\varepsilon e}(\lambda))$  を偏フェーズ (partial phase) という. さらに  $Z(t)$  が 2 次元でなく, 一般的に  $m$  次元のときの  $Y_j(t)$  と  $Z_k(t)$  の偏コヒーレンシーなども  $Z^{(-k)}(t)$  を  $k$  番目の変数を除いた時系列として

$$\begin{aligned}\varepsilon(t) &= Y_j(t) - \sum_{h=-\infty}^{\infty} g^{(1)}(h)^T Z^{(-k)}(t-h) \\ e(t) &= Z_k(t) - \sum_{h=-\infty}^{\infty} g^{(2)}(h)^T Z^{(-k)}(t-h)\end{aligned}$$

のように定義することによって同様に導くことができる。ただし、 $g^{(1)}(h)$ 、 $g^{(2)}(h)$  は  $m - 1$  次元の係数ベクトルである。

## 8 多変量 ARMA モデル (Multivariate ARMA model)

一変量の際の ARMA (Autoregressive Moving Average) モデルの多変量版を考える。多変量時系列  $\{Z(t)\}$  が方程式

$$\Phi(B)Z(t) = \Theta(B)\varepsilon(t)$$

を満たし、次に述べるような条件が満たされるとき ARMA( $P, Q$ ) モデルに従うという。ただし、 $P = (p_{jk} : 1 \leq j, k \leq m)$ 、 $Q = (q_{jk} : 1 \leq j, k \leq m)$  は次数を定める行列であり、

$$\begin{aligned}\Phi(z) &= (\phi_{jk}(z) : 1 \leq j, k \leq m) \\ \Theta(z) &= (\theta_{jk}(z) : 1 \leq j, k \leq m)\end{aligned}$$

で、

$$\begin{aligned}\phi_{jk}(z) &= \sum_{l=0}^{p_{jk}} \phi_{jkl} z^l, \\ \theta_{jk}(z) &= \sum_{l=0}^{q_{jk}} \theta_{jkl} z^l\end{aligned}$$

はそれぞれ  $p_{jk}$  次、 $q_{jk}$  次の伝達関数である。また、行列

$$\Phi_l = (\phi_{jkl} : 1 \leq j, k \leq m), \quad l = 0, 1, 2, \dots, p = \max(p_{jk})$$

を定義すれば、 $\Phi(z) = \sum_{l=0}^p \Phi_l z^l$  と表すこともでき、同様に

$$\Theta_l = (\theta_{jkl} : 1 \leq j, k \leq m), \quad l = 0, 1, 2, \dots, q = \max(q_{jk})$$

を定義すれば、 $\Theta(z) = \sum_{l=0}^q \Theta_l z^l$  と表すこともできる。ただし、 $\Theta_0 = \Phi_0 = I$  と定義しておくのが普通である。

条件1  $\{\varepsilon(t)\}$  は直交過程である。つまり

$$E(\varepsilon(t+h)\overline{\varepsilon(t)}^T) = \begin{cases} \Sigma & h = 0 \\ 0 & h \neq 0 \end{cases}$$

が成立する。

条件2  $\varepsilon(t)$  は過去の  $Z(t-1), Z(t-1), \dots$  と直交する。つまり

$$E(\varepsilon(t)\overline{Z(t-h)}^T) = 0, \quad h = 1, 2, \dots$$

が成立する。

条件3

$$\begin{aligned} |\Phi(z)| &\neq 0 && \text{for } |z| \leq 1 \\ |\Theta(z)| &\neq 0 && \text{for } |z| \leq 1 \end{aligned}$$

Wold 分  
解の多  
変量版

条件1, 条件2は  $\{\varepsilon(t)\}$  が  $\{Z(t)\}$  のイノベーションであるための条件で、条件3の前半が弱定常性を保証する条件、後半が無限次の AR モデルで表現できるためのいわゆる可逆性 (invertibility) の条件である。

例1の関係式

$$(I - F)Z(t) = GZ(t-1) + \varepsilon(t)$$

も ARMA モデルのように見えるが、 $\Phi_0 \neq I$  なので厳密には ARMA モデルではない。しかし  $(I - F)$  が逆行列をもつので、

$$I - (I - F)^{-1}Gz = \begin{pmatrix} 1 + \alpha\beta z & 0 \\ -\beta z & 1 \end{pmatrix}$$

となり

$$\det(I - (I - F)^{-1}Gz) = 1 + \alpha\beta z .$$

従って、 $|\alpha\beta| < 1$  ならば定常な AR モデルとなる。

スペクトルを求めるには ARMA 方程式の左辺と右辺をそれぞれノイズレスフィルターとみて両辺のスペクトルを求めれば

$$\Phi(e^{-2\pi i\lambda})f^{ZZ}(\lambda)\Phi^*(e^{-2\pi i\lambda}) = \Theta(e^{-2\pi i\lambda})f^{\varepsilon\varepsilon}(\lambda)\Theta^*(e^{-2\pi i\lambda})$$

なる等式を得る. 条件3の前半から  $\Phi(e^{-2\pi i\lambda})$  は任意の  $\lambda$  について逆行列をもつので,  $f^{\varepsilon\varepsilon}(\lambda) = \Sigma$  であることに注意すれば

$$f^{ZZ}(\lambda) = \Phi(e^{-2\pi i\lambda})^{-1} \Theta(e^{-2\pi i\lambda}) \Sigma \Theta^*(e^{-2\pi i\lambda}) \Phi^*(e^{-2\pi i\lambda})^{-1}$$

を得る.

ここで, 多変量 *ARMA* モデルのあてはめを行うソフトウェアについて触れておく. *ARIMA* モデル

$$\Phi(B)(I - B)^d Z(t) = \Theta(B)\varepsilon(t), \quad d \text{ は正整数}$$

のあてはめを最尤法によって行う S 関数 `arima.mle` は一変量時系列に対してだけ有効で, 多変量時系列には適用できない. これは, 統計数理研究所で開発した *TIMSAC* パッケージ (『FORTRAN 77 時系列解析プログラミング』, 北川源四郎著, 岩波書店, 1993) でも同様である. 多変量 *ARIMA* のような複雑なモデルになると非線形最小化アルゴリズムが満足に動かないからである. そこで, 通常は後半で述べる状態空間表現にもとづく方法が用いられる.

## 9 多変量 *AR* モデルにもとづく推測 (Inference on Multivariate AR model)

多変量 *AR* モデルは, 多変量 *ARMA* モデルで  $\Theta(z) = I$  としたモデルであり, 多変量 *ARMA* モデルより特殊なモデルであるが, モデルとしての意味の明快さや予測量の単純さ, 推定の容易さ, 安定性などさまざまな面で, 広く用いられることの多いモデルである. 本章では, 多変量 *AR* モデルに焦点を絞って, そのパラメータ推定の諸問題を論ずる.

### 9.1 多変量 *AR* モデルのパラメータ推定 (parameter estimation)

$$\Phi(B)Z(t) = \varepsilon(t)$$

つまり,

eq-A

$$\sum_{l=0}^p \Phi_l Z(t-l) = \varepsilon(t), \quad \Phi_0 = I \quad (11)$$

のパラメータ行列  $\Phi_1, \dots, \Phi_p$  と共分散行列  $\Sigma = \text{Cov}(\varepsilon(t))$  を  $m$  次元のベクトル観測値  $Z(1), \dots, Z(n)$  にもとづいて最小二乗法で推定するには、次のように書き換えておくとよい。

$$Y = \begin{pmatrix} Z(p+1)^T \\ Z(p+2)^T \\ \vdots \\ Z(n)^T \end{pmatrix} : (n-p) \times m \text{ 行列}$$

$$X = \begin{pmatrix} Z(p)^T & \cdots & Z(1)^T \\ Z(p+1)^T & \cdots & Z(2)^T \\ & \vdots & \\ Z(n-1)^T & \cdots & Z(n-p)^T \end{pmatrix} : (n-p) \times pm \text{ 行列}$$

$$B = \begin{pmatrix} -\Phi_1^T \\ \vdots \\ -\Phi_p^T \end{pmatrix} : pm \times m \text{ 行列}$$

$$E = \begin{pmatrix} \varepsilon(p+1)^T \\ \vdots \\ \varepsilon(n)^T \end{pmatrix} : (n-p) \times m \text{ 行列}$$

とおけば、多変量  $AR$  モデルは

$$Y = XB + E$$

という通常の変量回帰モデルに書き換えられる。定常性から、特殊な場合を除き  $X$  は最大階数となるので、以下では  $X$  は常に最大階数であると仮定する。

### 9.1.1 $\Sigma = \sigma^2 I$ と仮定したとき

$B$  の最小二乗推定量つまり、 $\text{tr}((Y - XB)^T(Y - XB))$  を最小にする  $B$  は

$$\hat{B} = (X^T X)^{-1} X^T Y$$

で与えられ、 $\sigma^2$  の推定量は

$$\hat{\sigma}^2 = \frac{1}{(n-p)m - pm^2} \text{tr} \left( (Y - X\hat{B})^T (Y - X\hat{B}) \right)$$

で与えられるが、これをそのまま解くのは計算効率の面からも、精度の面からも望ましくない。  $X$  の（ハウスホルダー法による）  $QR$  分解  $X = QR$  を用いるのが普通である。  $QR$  分解を用いれば  $\hat{B}$  は上三角方程式

$$R\hat{B} = Q^T Y$$

を解くだけで求まり、  $\hat{\sigma}^2$  は

$$\hat{\sigma}^2 = \frac{1}{(n-p)m - pm^2} \text{tr} (Y^T Y - (Q^T Y)^T (Q^T Y))$$

で求まる。ここでは  $E$  の要素はすべて相関がないと仮定しているのので、結局

$$\text{vec}(Y) = (I \otimes X) \text{vec}(B) + \text{vec}(E)$$

のように、ベクトル観測値の方程式に書き換えて、一変量回帰モデルとしての最小二乗法を考えるのと同じである。ただし、 $\text{vec}$  は行列の各列を積み重ねてベクトルを作る関数であり、 $\otimes$  はクロネッカー積である。  $\text{vec}(E)$  の要素数が  $(n-p)m$ 、  $\text{vec}(B)$  の要素数が  $pm^2$  であることから上の  $\hat{\sigma}^2$  の定義も理解できる。

### 回帰変数の選択

$M$  を要素が 1 か 0 の大きさ  $pm \times pm$  の対角行列とする。対角線上の  $m(i-1) + j$  要素が 0 であることはラグ  $i$  の  $j$  番目の変数  $Z_j(t-i)$  を  $AR$  モデルから除くことを示している。このマスク  $M$  を用いれば回帰変数の選択を含んだ多変量  $AR$  モデルは

$$Y = XMB + E$$

と書き換えられ、  $B$  の推定量は

eq-B

$$M^T R^T R M \hat{B} = M^T R^T Q^T Y \quad (12)$$

の解として求まる。これはさらに

$$(RM)^T (RM) \hat{B} = (RM)^T Q^T Y$$

と書き換えられるが，一般的には  $(RM)^T$  は  $M$  で指定された  $R$  の各行を 0 ベクトルとした退化した行列であり，両辺から簡単にキャンセルするわけにはいかない．しかし， $M$  の対角要素の 0 が後半にかたまっているような特殊な場合には， $(RM)^T$  は

$$\begin{pmatrix} * & \dots & 0 & 0 \\ * & \dots & * & 0 \\ & 0 & & 0 \end{pmatrix}$$

のような形の行列となり，0 でない下三角部分はもとの  $R$  が最大階数である限り，同じように最大階数となるので，両辺の  $(RM)^T$  を  $M$  で置き換えた

$$M(RM\mathcal{B}) = M(Q^T Y)$$

を解くことと同じになる．

もちろん実際には上の方程式系の後半には  $0 = 0$  という自明な方程式が含まれるので，それは除いてから解くことになる．具体的なアルゴリズムとしては， $R, Q^T Y, \mathcal{B}$  の最後の方の行で  $\text{diag}(M)$  の対応する要素が 0 の行を除いた  $\tilde{R}, \tilde{Q}^T Y, \tilde{\mathcal{B}}$  による方程式

$$\tilde{R}\tilde{\mathcal{B}} = \tilde{Q}^T Y$$

を解けばよい．但し， $\tilde{\mathcal{B}}$  には 0 を補って  $\mathcal{B}$  に戻す必要がある．

しかし，このようにきれいなマスクでない場合には (12) 式をそのまま解く必要がある．左辺の係数行列  $M^T R^T R M$  は， $R^T R$  から  $M$  で指定された行と列を除くことと同じで，右辺は  $R^T Q^T Y$  の対応する行を除くことと同じである．ただし，もはや三角方程式とはならないので数値精度の問題を避けるためには，左辺の係数行列をふたたび  $QR$  分解してから解いた方がよい．連立方程式を解くよいルーチンならば，これは自動的に行ってくれるのでそれにまかせてもよい．例えば S 関数 `solve` は  $QR$  分解を利用して解を求めている．

もう一つ考えられる方法としては，最初から  $X$  の対応する要素を取り除いてしまう方法があるが，これはさまざまな変数選択を試したいときなど，きわめて大きな行列  $X$  の  $QR$  分解を何度も実行しなければならないので，あまり勧められない．

### 9.1.2 $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$ と仮定したとき

この場合、最小二乗法で解く限り、ARモデル式を  $m$  個の独立な方程式

$$Y_j = X\mathcal{B}_j + E_j, \quad j = 1, \dots, m$$

に分けて考えても同じである。但し、 $Y_j, \mathcal{B}_j, E_j$  はそれぞれ  $Y, \mathcal{B}, E$  の第  $j$  列である。従って、3.1.1 と同じ  $X$  の  $QR$  分解の結果を利用すれば

$$R\hat{\mathcal{B}}_j = Q^T Y_j, \quad j = 1, \dots, m$$

を解くことによって求まり、 $\sigma_j^2$  の推定量は

$$\hat{\sigma}_j^2 = \frac{1}{n - p - pm} (Y_j^T Y_j - (Q^T Y_j)^T (Q^T Y_j))$$

で求まる。この場合、結果的に  $\mathcal{B}$  の推定量は3.1.1 で求めた推定量と同じである。それは方程式

$$R\hat{\mathcal{B}} = Q^T Y$$

の各列がちょうど上の  $m$  個の方程式になっているからであり、また、

$$\text{tr}((Y - X\mathcal{B})^T (Y - X\mathcal{B})) = \sum_{j=1}^m \|Y_j - X\mathcal{B}_j\|^2$$

と書き換えられるからでもある。違いは  $\sigma^2$  を推定するか、 $\sigma_1^2, \dots, \sigma_m^2$  を推定するかどうかである。この場合に推定を行う S 関数は `ars2` が作ってあるので参考にするとよい。

#### 回帰変数の選択

上の議論からわかるように、3.1.1 と同じ扱いができるので、任意のマスクについて同様である。統計数理研究所の石黒氏作成のプログラム `ARfit` は、 $\Phi_1, \dots, \Phi_p$  について同じマスクという制限はあるものの、この3.1.2の場合についての回帰変数の選択を  $AIC$  にもとづいて行っている。

マスク  
のサー  
チ方法  
を記載  
予定

### 9.1.3 一般の $\Sigma$ のとき

非負定符号行列  $\Sigma$  は  $\Sigma = C^T C$  のように上三角行列  $C$  でコレスキー分解できるが、さらに、 $d = \text{diag}(C)$  を対角要素とする対角行列  $D$  の逆行列を施した

$$B = D^{-1}C$$

を用いれば

eq-C

$$\Sigma = B^T D^2 B \quad (13)$$

と分解できる.  $AR$  モデル式の両辺に左から  $(B^T)^{-1}$  を施せば (11) 式は

$$\sum_{l=0}^p \tilde{\Phi}_l Z(t-l) = \tilde{\varepsilon}(t)$$

と書き換えられる. ただし,  $\tilde{\Phi}_l = (B^T)^{-1} \Phi_l$ ,  $l = 0, \dots, p$ ,  $\tilde{\varepsilon}(t) = (B^T)^{-1} \varepsilon(t)$ . 従って,

$$\text{Cov}(\tilde{\varepsilon}(t)) = D^2$$

のように  $\tilde{\varepsilon}(t)$  の各要素が無相関になるので, 3.1.2 の場合に帰着する.

ただし, 多変量回帰モデルに書き換えるためには  $\tilde{\Phi}_0 \neq I$  であることに注意して,

$$\tilde{\Phi}_0 = (B^T)^{-1} - I$$

とおき直し,

$$Z(t) + \sum_{l=0}^p \tilde{\Phi}_l Z(t-l) = \tilde{\varepsilon}(t)$$

の形に書き直しておいた方がよい.

これを多変量回帰モデルで書き直せば,  $Y$  はこれまでと同じで

$$X = \begin{pmatrix} Z(p)^T & \cdots & Z(1)^T & Z(p+1)^T \\ \vdots & \vdots & & \vdots \\ Z(n-1)^T & \cdots & Z(n-p)^T & Z(n)^T \end{pmatrix} : (n-p) \times (p+1)m \text{ 行列}$$

$$\tilde{B} = \begin{pmatrix} -\tilde{\Phi}_1^T \\ \vdots \\ -\tilde{\Phi}_p^T \\ -\tilde{\Phi}_0^T \end{pmatrix} : (p+1)m \times m \text{ 行列}$$

$$E = \begin{pmatrix} \tilde{\varepsilon}(p+1)^T \\ \vdots \\ \tilde{\varepsilon}(n)^T \end{pmatrix} : (n-p) \times m \text{ 行列}$$

である. なお,  $\tilde{\Phi}_0$  に対応して  $Y$  と同じ列を  $X$  の最初ではなく最後に付け加えたのは  $B$  が対角線が 1 ばかりの上三角行列であるので,  $B^{-1}$  も同

じ上三角行列となり， $\tilde{\Phi}_0 = (B^T)^{-1} - I$ は対角線がすべて0の下三角行列になるという特殊な事情を考慮したためである．これは3.1.1で述べた回帰変数選択と同じ状況である．従って方程式系，

$$R\tilde{\mathcal{B}}_j = Q^T Y_j, \quad j = 1, \dots, m$$

で， $\tilde{\mathcal{B}}_j$ は最初の $pm + (j - 1)$ 個の要素以降は最初から0であり，求める必要はない．つまり， $j$ の値に応じて $R$ から $pm + (j - 1)$ 番以降の行と列を除き，また右辺も対応して除いた方程式に縮小して解いた上で，0を埋めればよい．

また，求めた $\tilde{\mathcal{B}}$ を行列演算 $B$ で $\mathcal{B}$ へ戻すことも必要である．ここで

$$\tilde{\mathcal{B}}B = \begin{pmatrix} -\Phi_1^T \\ \vdots \\ -\Phi_p^T \\ -(I - B) \end{pmatrix}$$

という関係にあることに注意すればよい． $B$ は $B = ((I + \tilde{\Phi}_0)^{-1})^T$ で求まる．さらに， $\Sigma$ は(13)に従って $B$ と $D$ の推定値より求める． $D^2$ の対角要素は基本的には3.1.2で与えた $\hat{\sigma}_j^2$ ， $j = 1, \dots, m$ でよいが，この場合は回帰パラメータが $pm + (j - 1)$ 個あることに注意すれば， $n - p - pm$ ではなく $(n - p) - (pm + (j - 1))$ で割る方が自然である．また， $Q^T Y_j$ は第1要素から第 $pm + j - 1$ 要素までだけを用いる．3.1.3の場合を求めるS関数はars3である．

### 回帰変数の選択

ここでは一般の $\Sigma$ の場合に， $\mathcal{B}$ いくつかの要素が0であるといった制約条件のもとで $\mathcal{B}$ を求める方法を示す．非常に特殊なマスクの場合を除いて， $\Sigma$ を任意とすると簡単な線形演算では解が得られない．TIMSACのmulvarは， $p$ が被説明変量ごとに違ってもよい，という形だけの回帰変数選択を許しているが，このような形，あるいはマスクが被説明変量について同一といった形の場合には，あらかじめ $X$ の列の入れ換えをしておけば変換 $B^{-1}$ で構造が変わらないので，3.1.2までと同様の方法で処理できる．しかし，任意のマスクの場合にはそうはいかない．そこで，推定パラメータ $\mathcal{B}, B, D$ のうち，まず $B$ を既知パラメータとして $\mathcal{B}, D$ を推定し，次に推定された $\mathcal{B}, D$ にもとづいて $B$ を推定し，アップデートする方法によって求めることを考える．まず，

$$Y = XB + E$$

の両辺に右から  $B^{-1}$  を施したものを

$$\tilde{Y} = X\tilde{B} + \tilde{E}$$

とおく ((11) の形の表現で左から  $(B^T)^{-1}$  を施したものと同等である).  
ただし,  $YB^{-1} = \tilde{Y}$ ,  $BB^{-1} = \tilde{B}$ ,  $EB^{-1} = \tilde{E}$ .

まず, ラグランジュ乗数法を用いて  $\tilde{B}$  を求めるときに必要な制約条件として,  $B$  の  $(k, l)$  要素が 0 であるという簡単な制約条件を考えると,

$$V = \sum_{j=1}^m \|\tilde{Y}_j - X\tilde{B}_j\|^2 - 2\lambda e_k^T \tilde{B} B e_l \rightarrow \min$$

となる  $\tilde{B}$  を求めればよいことになる. ここで,

$$e_k^T \tilde{B} B e_l = e_k^T B e_l = B_{kl}$$

より, 制約条件は確かに  $B_{kl} = 0$  と同等になっていることがわかる. なお,  $e_k$  は次元が  $pm$  で  $k$  番目の要素が 1 の単位ベクトル,  $e_l$  は次元が  $m$  で  $l$  番目の要素が 1 の単位ベクトル.

また,

$$\frac{\partial}{\partial \tilde{B}_j} e_k^T \tilde{B} B e_l = \frac{\partial}{\partial \tilde{B}_j} e_k^T \sum_{i=1}^m \tilde{B}_i B_{il}$$

であるから,

$$\frac{\partial V}{\partial \tilde{B}_j} = -2X^T(\tilde{Y}_j - X\tilde{B}_j) - 2\lambda e_k B_{jl} .$$

と書け,  $\frac{\partial V}{\partial \tilde{B}_j} = 0$  の方程式は

$$X^T \tilde{Y}_j + \lambda e_k B_{jl} = X^T X \tilde{B}_j, \quad j = 1, \dots, m$$

つまり,

$$\tilde{B} = (X^T X)^{-1} (X^T \tilde{Y} + \lambda e_k B_l^T)$$

ただし,  $B_l$  は  $B$  の  $l$  列. さらに, 制約条件  $e_k^T \tilde{B} B e_l = 0$  に  $\tilde{B}$  を代入し

$$\lambda = -\frac{(e_k^T R^{-1}) Q^T \tilde{Y} B_l}{(e_k^T R^{-1})(e_k^T R^{-1})^T B_l^T B_l}$$

を得る.

同様に、制約条件が  $\mathcal{B}$  の  $(k_1, l_1), \dots, (k_r, l_r)$  要素が 0 という一般的な場合を考える。  $Z = R^T Q^T \tilde{Y}$ ,  $A = R^{-1}(R^T)^{-1} = (R^T R)^{-1}$  とおけば、

$$\tilde{\mathcal{B}} = A(Z + \lambda_1 \mathbf{e}_{k_1} B_{l_1}^T + \dots + \lambda_r \mathbf{e}_{k_r} B_{l_r}^T) .$$

また、制約条件より  $\lambda_1, \dots, \lambda_r$  は

eq-D

$$\begin{pmatrix} \mathbf{e}_{k_1}^T A Z B_{l_1} \\ \vdots \\ \mathbf{e}_{k_r}^T A Z B_{l_r} \end{pmatrix} = - \begin{pmatrix} \mathbf{e}_{k_1}^T A (\lambda_1 \mathbf{e}_{k_1} B_{l_1}^T B_{l_1} + \dots + \lambda_r \mathbf{e}_{k_r} B_{l_r}^T B_{l_1}) \\ \vdots \\ \mathbf{e}_{k_r}^T A (\lambda_1 \mathbf{e}_{k_1} B_{l_1}^T B_{l_r} + \dots + \lambda_r \mathbf{e}_{k_r} B_{l_r}^T B_{l_r}) \end{pmatrix} \quad (14)$$

の解として求まる。  $A_r$  を  $A$  の  $(k_1, \dots, k_r)$  行だけを取り出した行列、  $B^*$  を  $B$  の  $(l_1, \dots, l_r)$  列だけを取り出した行列として (14) を書き直せば、

$$\begin{pmatrix} A_r^{(1)} Z B_1^* \\ \vdots \\ A_r^{(r)} Z B_r^* \end{pmatrix} = -A_{rr} \times ((B^*)^T B^*) \boldsymbol{\lambda}$$

となる。ただし、  $A_{rr}$  は  $A$  の  $(k_1, \dots, k_r)$  行と  $(k_1, \dots, k_r)$  列をそれぞれ組み合わせて取り出した部分行列、  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_r)^T$  であり、  $\times$  は行列演算ではなく、要素ごとの積である。この方程式にもとづいて求めた  $\boldsymbol{\lambda}$  を代入して  $\tilde{\mathcal{B}}$  を求めればよい。もちろん  $\tilde{\mathcal{B}}$  を  $\mathcal{B}$  に戻す必要はある。

一方、  $\mathcal{B}$  を既知としたとき、

$$E = Y - X\mathcal{B}$$

において  $E$  の各行を  $E^{(j)}$ ,  $j = 1, \dots, n-p$  とすれば、モデルの仮定は

$$E^{(j)} \sim N(0, \Sigma), \quad j = 1, \dots, n-p \text{ (独立)}$$

であるから、尤度関数は

$$L = \log \frac{1}{((2\pi)^m |\Sigma|)^{\frac{n-p}{2}}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{n-p} E^{(j)} \Sigma^{-1} (E^{(j)})^T \right\}$$

となり、これを最大にする  $\Sigma$  の最尤推定量は

$$\tilde{\Sigma} = \frac{1}{n-p} \sum_{j=1}^{n-p} (E^{(j)})^T (E^{(j)})$$

となる．実際には  $\hat{\mathcal{B}} = \tilde{\mathcal{B}}\mathcal{B}^T$  によって  $\mathcal{B}$  を推定できるので，残差  $\hat{E} = Y - X\hat{\mathcal{B}}$  を求め，

$$\hat{\Sigma} = \frac{1}{n-p} \sum_{j=1}^{n-p} (\hat{E}^{(j)})^T (\hat{E}^{(j)})$$

で  $\Sigma$  を推定し，

$$\hat{\Sigma} = CC^T = \hat{B}\hat{D}^2\hat{B}^T$$

から  $B$  を  $\hat{B}$  でアップデートし，アップデートされた  $\hat{B}$  を用いて再び  $\tilde{\mathcal{B}}$  を求め直す．これを収束するまで繰り返す．

収束を判定するには対数尤度を基準にするとよい．

eq-E

$$\begin{aligned} & \log \frac{1}{((2\pi)^m |\Sigma|)^{\frac{n-p}{2}}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{n-p} E^{(j)} \Sigma^{-1} (E^{(j)})^T \right\} \\ &= \log \frac{1}{((2\pi)^m |\hat{\Sigma}|)^{\frac{n-p}{2}}} \exp \left\{ -\frac{(n-p)m}{2} \right\} \\ &= \frac{-(n-p)}{2} \left\{ m \log 2\pi + \log |\hat{\Sigma}| + m \right\} \end{aligned} \quad (15)$$

(15) において 1 行めから 2 行めの書換えでは

$$\begin{aligned} & v^T A v \\ &= \\ & \text{tr}(A v v^T) \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^{n-p} \hat{E}^{(j)} \hat{\Sigma}^{-1} (\hat{E}^{(j)})^T &= \sum_{j=1}^{n-p} \text{tr} \left( \hat{\Sigma}^{-1} (\hat{E}^{(j)})^T \hat{E}^{(j)} \right) \\ &= \text{tr} \left( \hat{\Sigma}^{-1} \sum_{j=1}^{n-p} (E^{(j)})^T E^{(j)} \right) \\ &= \text{tr} \left( \hat{\Sigma}^{-1} (n-p) \hat{\Sigma} \right) = (n-p) \text{tr}(\mathbf{I}_m) = (n-p)m \end{aligned}$$

であることを用いている．ここで  $\mathbf{I}_m$  は  $m \times m$  単位行列．

$$|\hat{\Sigma}| = |\hat{B}\hat{D}^2\hat{B}^T| = |\hat{D}^2| = \hat{\sigma}_1^2 \times \cdots \times \hat{\sigma}_m^2$$

なので (15) の対数尤度は

$$-\frac{n-p}{2} \left\{ m \log 2\pi + \sum_{j=1}^m \log \hat{\sigma}_j^2 + m \right\}$$

となり，この値がなるべく大きくなり，しかも収束していると判断できるまで  $\tilde{\mathcal{B}}, \hat{B}, \hat{\Sigma}$  を上記の方法で繰り返し推定する．

さまざまな制約条件のうち、どれが適当かを判断する基準の1つとして、 $AIC$ がある。

$$AIC = -2 \cdot \text{最大対数尤度} + 2 \cdot \text{未知パラメータ数}$$

の定義より、上の場合

$$AIC = (n - p) \left\{ m \log 2\pi + \sum_{j=1}^m \log \hat{\sigma}_j^2 + m \right\} + 2(pm^2 - r + m(m+1)/2)$$

となる。ここで、未知パラメータ数としては、回帰パラメータ数が制約条件の個数  $r$  を考慮すると  $pm^2 - r$ 、分散パラメータが  $m(m+1)/2$  個であることを考慮している。なお、 $p$  も変化させて比較する場合には、最大ラグ  $\bar{p}$  で行列  $X$  を作り、 $p$  の値によらず、同一の  $X$  を用いる必要がある。従って、

$$AIC = (n - \bar{p}) \left\{ m \log 2\pi + \sum_{j=1}^m \log \hat{\sigma}_j^2 + m \right\} + 2(pm^2 - r + m(m+1)/2)$$

となる。いずれの場合も、 $AIC$  の値が最小時の制約条件が一番適していることになる。

## 9.2 多変量 $AR$ モデルのパラメータが弱定常性の条件を満たしているかどうかの判定 (Diagnostics of the stationarity)

弱定常性の必要十分条件

$$|I + \Phi_1 z + \dots + \Phi_p z^p| \neq 0 \quad \text{for } |z| \leq 1$$

を満たしているかどうかを判定するには、多項式行列の単因子 (elementary divisor) 表現を用いる必要がある。

定理 (多項式行列の標準化)



- (2)  $A_{11}(z)$  で  $A_{21}(z), \dots, A_{m1}(z)$  を割ったときの商を  $Q_2(z), \dots, Q_m(z)$  とする.

$$A_{k1}(z) = A_{11}(z)Q_k(z) + R_k(z), \quad k = 2, \dots, m$$

第1行に  $Q_k(z)$  をかけた多項式 (行) ベクトルを  $k$  行から引く,  $k = 2, \dots, m$ .

- (3) 第1列の第2要素以降は  $R_2(z), \dots, R_m(z)$  となっているはずである. これらがすべて0となっていなければ(1)へ戻って操作を繰り返す.
- (4)  $A_{11}(z)$  が  $e_1(z)$  である.
- (5)  $A(z)$  の1行, 1列を除いた部分行列を  $A(z)$  とおき直し, (1)へ戻る. この際, 第1列が0ベクトルとなっていれば列の入れ換えで最終列に移動する (弱定常性の判定ならばこの段階で非定常であることになる).
- (6) (5)でおき直した  $A(z)$  が1行1列の行列になるまで(1)~(5)の過程を繰り返し, 1行1列になればそれが  $e_m(z)$  であり, 終了.

## 10 状態空間モデルと ARMA モデル (State Space Model and ARMA model)

推移方程式 (transition equation)

$$X(t) = FX(t-1) + Gv(t)$$

と観測方程式 (observation equation)

$$Z(t) = HX(t) + w(t)$$

の2つの方程式で表現されるモデルが状態空間モデル (state space model) と呼ばれる.  $X(t)$  は適当な次元のベクトルで状態ベクトル (state vector) と呼ばれる. また,  $F, G, H$  は行列である. さて,  $ARMA(P, Q)$  モデルは  $w(t) = 0$  の状態空間モデルで表現できる. ただし, 表現の仕方は一意的ではなく, 次のような代表的な2つの表現が知られている. なお, 状態空間モデルは一般的には弱定常とはならないが,  $\{v(t)\}$  が平均0, 分散一

定の直交過程で、 $F$  の固有値がすべて1より小で、 $E\|X(t)\|$  が有界ならば、 $\{X(t)\}$  は弱定常過程となり、 $\{Gv(t)\}$  がイノベーションとなる。実際、推移方程式を繰り返し用いることにより、

$$X(t) = F^i X(t-i) + \sum_{l=0}^{i-1} F^l Gv(t-l)$$

のように表現できるが、 $\|F\|$  を  $F$  のオペレータノルムとすれば仮定から  $\|F\| < 1$  であり、

$$E\|F^i X(t-i)\| \leq \|F\|^i E\|X(t-i)\|$$

と評価できるので、 $F^i X(t-i)$  は0に収束し

$$X(t) = \sum_{l=0}^{\infty} F^l Gv(t-l)$$

という表現を得る。これは  $\{Gv(t)\}$  がイノベーションであることを意味している。

## 10.1 直接表現 (direct representation)

$r = \max(p, q + 1)$  として

$$X(t) = \begin{pmatrix} Z(t) \\ Z(t-1) \\ \vdots \\ Z(t-r+1) \end{pmatrix}, \quad v(t) = \begin{pmatrix} \varepsilon(t) \\ \varepsilon(t-1) \\ \vdots \\ \varepsilon(t-q) \end{pmatrix},$$

$$F = \begin{pmatrix} -\Phi_1 & \cdots & -\Phi_p & 0 & \cdots & 0 \\ I & & & & & \mathbf{0} \\ & & \ddots & & & \\ & & & \mathbf{0} & & \\ & & & & I & 0 \end{pmatrix},$$

$$G = \begin{pmatrix} I & \Theta_1 & \cdots & \Theta_q \\ & \mathbf{0} & & \end{pmatrix},$$

$$H = (I \ 0 \ \cdots \ 0)$$

とおけば *ARMA* モデル

$$\Phi(B)Z(t) = \Theta(B)\varepsilon(t)$$

は状態空間モデル

$$\begin{aligned} X(t) &= FX(t-1) + Gv(t) \\ Z(t) &= HX(t) \end{aligned}$$

で表現できる. ただし,  $X(t)$  は  $rm$  次元ベクトル,  $v(t)$  は  $(q+1)m$  次元ベクトル,  $F$  は  $rm \times rm$  行列,  $G$  は  $rm \times (q+1)m$  行列,  $H$  は  $rm \times m$  行列であり, 一般的にかなり大きな行列方程式となる.

## 10.2 マルコフ表現 (Markov representation)

4.1 の直接表現は, 単純に *ARMA* 方程式を書き直したただけであるので,  $\{Gv(t)\}$  は直交過程ではなく, もちろん状態ベクトルの系列  $\{X(t)\}$  に対するイノベーションにもならない. マルコフ表現はこのような問題を解決した表現を与える. また, マルコフ表現することにより, 状態空間の次元の節約も可能となる. まず, 記号として条件付平均値

$$Z(t|s) = E(Z(t)|Z(s), Z(s-1), \dots)$$

を導入する.  $s \geq t$  なら明らかに  $Z(t|s) = Z(t)$  である. 一変量のとくと同じように,  $Z$  を  $m$  次元の値をとる確率変数,  $Y$  を別の確率変数とし  $g(Y)$  を  $m$  次元の値をとる可測関数とすれば

$$E\|Z - g(Y)\|^2 \geq E\|Z - E(Z|Y)\|^2$$

が成立するので,  $E(Z|Y)$  は  $Y$  への射影と考えられる. このことから,  $Z(t|s)$  は  $Z(t)$  の  $Z(s), Z(s-1), \dots$  から生成される  $\sigma$ -集合体への射影である. なお, 正規過程を仮定すればこの条件付平均値  $Z(t|s)$  は実際, 行列を係数とする  $m$  次元の値をとる確率変数の線形空間

$$\text{Span}\{Z(s), Z(s-1), \dots\} = \text{Span}\{\varepsilon(s), \varepsilon(s-1), \dots\}$$

への射影と一致する. ここで  $\text{Span}\{Z(s), Z(s-1), \dots\}$  は  $\text{Span}\{z_1(s), \dots, z_m(s), z_1(s-1), \dots, z_m(s-1), \dots\}$  の意味なので, 射影  $Z(t|s)$  は,  $Z(s), Z(s-1), \dots$

の行列係数の線形結合のうちでもっとも  $Z(t)$  に距離の近いものという意味である。この条件付平均値を用いて、 $rm$  次元の状態ベクトルを

$$X(t) = \begin{pmatrix} Z(t|t) \\ Z(t+1|t) \\ \vdots \\ Z(t+r-1|t) \end{pmatrix}$$

と定義すれば、正規性の仮定のもとで  $ARMA(P, Q)$  モデルのマルコフ表現

$$\begin{aligned} X(t) &= FX(t-1) + G\varepsilon(t) \\ Z(t) &= HX(t) \end{aligned}$$

を得る。ただし、

$$F = \begin{pmatrix} 0 & I & \cdots & 0 \\ \vdots & 0 & \cdots & I \\ 0 & \cdots & 0 & -\Phi_p \cdots -\Phi_1 \end{pmatrix},$$

$$G = \begin{pmatrix} I \\ \Psi_1 \\ \vdots \\ \Psi_{r-2} \\ \Theta_{r-1} \end{pmatrix}, \quad H = (I \ 0 \ \cdots \ 0)$$

である。なお、 $\Psi_1, \dots, \Psi_{r-1}$  は  $\Phi(z)^{-1}\Theta(z)$  を展開したときの係数行列である。つまり、

$$\Phi(z)^{-1}\Theta(z) = \Psi_0 + \Psi_1 z + \Psi_2 z^2 + \cdots$$

の係数行列である。 $ARMA$  モデルが上のようなマルコフ表現をもつこと、逆に上のような表現ができれば  $ARMA$  モデルであることはそれほど明らかでない。以下にそれを証明する。

### 10.2.1 ARMAモデルからマルコフモデル (ARMA model to Markov model)

$\{\varepsilon(t)\}$  は独立系列とする.

$$Z(t) = \sum_{l=0}^{\infty} \Psi_l \varepsilon(t-l)$$

がすべての  $t$  について成立していることより, 任意の非負整数  $j$  について EQ3-1

$$\begin{aligned} Z(t+j|t) &= \sum_{l=0}^{\infty} \Psi_l \varepsilon(t+j-l|t) \\ &= \sum_{l=j}^{\infty} \Psi_l \varepsilon(t+j-l) \end{aligned} \quad (16)$$

ただし,

証明

$$\begin{aligned} \varepsilon(t|s) &= E(\varepsilon(t)|Z(s), Z(s-1), \dots) \\ &= E(\varepsilon(t)|\varepsilon(s), \varepsilon(s-1), \dots) \end{aligned}$$

で  $\varepsilon(t)$  の条件付平均値を表す. また, 任意の  $j \geq 0$  について EQ3-2

$$\begin{aligned} Z(t+j|t-1) &= \sum_{l=0}^{\infty} \Psi_l \varepsilon(t+j-l|t-1) \\ &= \sum_{l=j+1}^{\infty} \Psi_l \varepsilon(t+j-l) \end{aligned} \quad (17)$$

(16) と (17) を比較することにより,

$$Z(t+j|t) = Z(t+j|t-1) + \Psi_j \varepsilon(t), \quad j = 0, 1, 2, \dots$$

を得る. これが推移方程式の最初の  $r-1$  個のブロック行に対応する. 一方, ARMA 方程式

$$\sum_{l=0}^p \Phi_l Z(t-l) = \sum_{l=0}^q \Theta_l \varepsilon(t-l)$$

で,  $t$  を  $t+r-1$  で置き換え, 時間  $t$  での条件付平均をとれば  $r > q$  であることより

$$\sum_{l=0}^p \Phi_l Z(t+r-1-l|t) = \Theta_{r-1} \varepsilon(t)$$

を得る. これが推移方程式の最後のブロックに対応する.

### 10.2.2 マルコフモデルから ARMA モデル (Markov model to ARMA model)

$Z(t)$  が  $m$  次元,  $X(t)$  が  $s = rm$  次元なので,

$$X(t) = F^s X(t-s) + F^{s-1} G \varepsilon(t-s+1) + \cdots + G \varepsilon(t) \quad (18)$$

ARIMA  
でも  
O.K.

EQ3-3

が成立する. さらにこの式を利用すれば任意の  $l < s$  について

EQ3-4

$$\begin{aligned} F^{s-l} X(t-s) &= F^{s-l} X(t-l-(s-l)) \\ &= X(t-l) - \left\{ F^{s-l-1} G \varepsilon(t-l-(s-l-1)) \right. \\ &\quad \left. + F^{s-l-2} G \varepsilon(t-l-(s-l-2)) + \cdots + G \varepsilon(t-l) \right\} \\ &= X(t-l) - \sum_{j=0}^{s-l-1} F^{s-l-1-j} G \varepsilon(t-s+1+j) \end{aligned} \quad (19)$$

が成り立つ. 一方 Cayley - Hamilton の定理から

$$\sum_{l=0}^s \alpha_l F^{s-l} = 0$$

が成立する. ここで,  $\alpha_0 = 1, \alpha_1, \dots, \alpha_s$  は  $F$  の固有多項式

$$|\lambda I - F| = \sum_{l=0}^s \beta_l \lambda^{s-l} \quad (\beta_0 \neq 0)$$

の係数  $\beta_0, \dots, \beta_s$  を  $\beta_0$  で正規化した係数である. 従って, (19) を用いれば

$$\begin{aligned} F^s X(t-s) &= - \sum_{l=1}^s \alpha_l F^{s-l} X(t-s) \\ &= -\alpha_s X(t-s) - \sum_{l=1}^{s-1} \alpha_l \left\{ X(t-l) - \sum_{j=0}^{s-l-1} F^{s-l-1-j} G \varepsilon(t-s+1+j) \right\} \\ &= - \sum_{l=1}^s \alpha_l X(t-l) + \sum_{l=1}^{s-1} \alpha_l \sum_{j=0}^{s-l-1} F^{s-l-1-j} G \varepsilon(t-s+1+j) \end{aligned}$$

これを (18) に代入すると

$$X(t) = - \sum_{l=1}^s \alpha_l X(t-l) + \sum_{l=0}^{s-1} \alpha_l \sum_{j=0}^{s-l-1} F^{s-l-1-j} G \varepsilon(t-s+1+j)$$

となるが, さらに  $K_j = (F^j + \alpha_1 F^{j-1} + \dots + \alpha_j I)G$  とおけば

$$X(t) = - \sum_{l=1}^s \alpha_l X(t-l) + \sum_{j=0}^{s-1} K_j \varepsilon(t-j)$$

となり, 両辺に  $H$  を演算すれば,  $ARMA$  モデル

$$\sum_{l=0}^s \alpha_l Z(t-l) = \sum_{j=0}^{s-1} L_j \varepsilon(t-j)$$

を得る. ただし,  $L_j = HK_j$  である.

ここで注意すべきことは, 得られた  $ARMA$  モデルは, 両辺とも  $s = rm$  次の方程式となっていることで,  $ARMA(P, Q)$  モデルから  $rm$  ( $r = \max(p, q+1)$ ) 次元の状態ベクトルをもつマルコフ表現に直し, 再び  $ARMA$  モデルに戻ってくると, かなり大きなモデル, 言い換えればかなり冗長なモデルになってしまう. このような冗長度をどうやって減らすかを次に議論する.

マルコフ性

## 10.3 同定可能性 (Identifiability)

### 10.3.1 $ARMA$ モデルの同定可能性 (Identifiability of $ARMA$ model)

同定可能性 (identifiability) は, 表現の一意性のことである.  $ARMA(P, Q)$  モデルが同定可能なための十分条件としては, 一変量のときの条件を拡張した Hannan (Annals of Statistics, 975-981, 1975) によるものが知られている.

#### 定理

$Z(t)$  が  $ARMA$  モデル

$$\Phi(B)Z(t) = \Theta(B)\varepsilon(t)$$

に従うとき,  $\Phi(z) = \sum_{l=0}^p \Phi_l z^l$ ,  $\Theta(z) = \sum_{l=0}^q \Theta_l z^l$  の係数行列  $\Phi_1, \dots, \Phi_p, \Theta_1, \dots, \Theta_q$  が一意的に定まるための 1 つの十分条件は次の 4 条件で与えられる.

1.  $\Sigma = \text{Var}(\varepsilon(t))$  が正則
2.  $\Phi(z)^{-1}\Theta(z)$  が  $|z| < 1$  で解析的で各  $z$  ごとに正則行列

3.  $\Phi(z), \Theta(z)$  が Unimodular の common left divisor を持たない

4. 行列  $[\Phi_p, \Theta_q]$  がフルランク

ここで,  $U(z)$  が Unimodular とは

$$\exists V(z) \text{ s.t. } U(z)V(z) = I$$

となる多項式行列のことで, ちょうど正則行列に対応する概念である.

例

$$U(z) = \begin{pmatrix} z-1 & z \\ z & z+1 \end{pmatrix} \text{ は Unimodular で } V(z) = \begin{pmatrix} -z-1 & z \\ z & -z+1 \end{pmatrix}$$

Unimodular 行列の性質としては

$$U(z) : \text{Unimodular} \iff |U(z)| = \text{ゼロでない定数}$$

がよく知られている. また,  $D(z)$  が  $A(z), B(z)$  の common left divisor (c.l.d.) とは,  $\exists A_1(z), \exists B_1(z)$

$$A(z) = D(z)A_1(z)$$

$$B(z) = D(z)B_1(z)$$

と書けることである. ただし, 上の定理の一意性のための条件は実際には強すぎる.

例

$$Z(t) + \Phi_1 Z(t-1) = \varepsilon(t) + \Theta_1 \varepsilon(t-1) \text{ で}$$

$$\Phi_1 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}, \Theta_1 = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & 0 \end{pmatrix}$$

とすれば,

$$\begin{aligned} Z_1(t) + \frac{1}{2}Z_1(t-1) &= \varepsilon_1(t) + \frac{1}{3}\varepsilon_1(t-1) \\ Z_2(t) &= \varepsilon_2(t) \end{aligned}$$

と同等で, この表現は一意である. しかし,

$$[\Phi_1, \Theta_1] = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

の階数は 1 で  $m = 2$  より小さい. より精密な同定可能性の議論にはやはり状態空間で議論した方が簡単である.

### 10.3.2 状態空間モデルの同定可能性 (Identifiability of State Space model)

ARMAモデルのときの同定可能性の条件1, 2のもとで, 同定可能なための必要十分条件は次の(3')である.

(3') マルコフ表現したときの  $F, G$  から作った行列

$$C = [G, FG, \dots, F^{r-1}G]$$

がフルランク. ただし  $r = \max(p, q + 1)$  である.

これは次のようにして示すことができる.

まず, (3') から同定可能性を示す. まず,

$$Z(t+r|t) + \Phi_1 Z(t+r-1|t) + \dots + \Phi_r Z(t|t) = 0 \quad (20)$$

が成立することに注意する. ただし,  $l > p$  に対する  $\Phi_l$  は0と定義する. もし,  $\Phi_1, \dots, \Phi_r$  が一意に定まらないとすると, あるゼロでない  $m \times rm$  行列  $A$  が存在して

$$AX(t) = 0 .$$

ただし,  $X(t)$  はマルコフ表現したときの状態ベクトルである. すると

$$0 = E (AX(t) \varepsilon(t-i)^T) = AF^i G \Sigma, \quad i = 0, 1, 2, \dots$$

つまり  $AC = 0$  となり,  $C$  がフルランクであることに矛盾する. ARMA式の右辺の一意性は  $\varepsilon(t)$  の直交性より明らか.

逆に, 表現が一意であるとすれば(3')が従う. これは, まず  $C$  がフルランクでないとする,  $v^T C = 0$  となるようなゼロでないベクトル  $v$  が存在する. 一方, マルコフ表現より,

$$X(t) = F^{i+1} X(t-i-1) + F^i G \varepsilon(t-i) + \dots + G \varepsilon(t), \quad i = 0, \dots, s-1$$

と表せるので,

$$E (v^T X(t) \varepsilon(t-i)^T) = v^T F^i G \Sigma$$

となる. ここで  $v^T C = 0$  より  $v^T F^i G = 0, i = 0, \dots, s-1$  であるが, Cayley-Hamilton の定理を用いれば

$$v^T F^i G \Sigma = 0, \quad i = 0, 1, 2, \dots$$

EQ3.3-1

も成立することになる。一方、先にも用いたように

$$X(t) = \sum_{l=0}^{\infty} \Psi_l \varepsilon(t-l)$$

と表現できるが、

$$E(v^T X(t) \varepsilon(t-i)^T) = v^T \Psi_i \Sigma, \quad i = 0, 1, 2, \dots$$

より左辺は0なので、 $v^T \Psi_i \Sigma = 0, i = 0, 1, 2, \dots$ 。さらに  $\Sigma$  が正則であることより

$$v^T \Psi_i = 0, \quad i = 0, 1, 2, \dots$$

であることになり、結局

$$v^T X(t) = 0$$

を得る。これは  $V_j$  を第  $j$  行が  $(v_{(j-1)m+1}, \dots, v_{jm})$  で、他はすべて0の行列とすれば

$$\sum_{j=1}^r V_j Z(t+j-1|t) = 0,$$

つまり、 $\Phi_1, \dots, \Phi_r$  の一意性に反する。 $\Phi'_i = \Phi_i + V_i$  という係数を用いた表現も可能となるからである。

もちろん、上の表現は  $AR$  の次数  $p$ ,  $MA$  の次数  $q$  としたときの表現の一意性をいっており、0となる係数の存在までは排除していない。つまり最小表現の保証にはなっていない。

### 10.3.3 状態空間の最小表現 (Minimum Realisation)

冗長な表現を排除するには、まず状態空間表現での状態ベクトル  $X(t)$  の次元縮小を計る必要がある。状態ベクトル

$$X(t) = \begin{pmatrix} Z(t|t) \\ \vdots \\ Z(t+r-1|t) \end{pmatrix} = \begin{pmatrix} x_1(t) \\ \vdots \\ x_s(t) \end{pmatrix}$$

を要素ごとに考えたとき、 $x_1(t), \dots, x_s(t)$  のうちに明らかな線形従属性、たとえば

$$x_j(t) = \sum_{l=1}^h \alpha_l x_{k_l}(t) \quad \text{a.s. for any } t$$

のような関係があれば,  $x_j(t)$  は状態ベクトルから除いてもよいことになる.  $x_1(t), \dots, x_s(t)$  を順に調べて, 線形独立な要素だけを取り出して,  $x_{k_1}(t), \dots, x_{k_h}(t)$  を得たとする. このとき, 添字の集合  $K = \{k_1, \dots, k_h\}$  について次の定理が成立する.

**定理**

$$k \notin K \text{ ならば } k + m \notin K$$

**証明**

$k \notin K$  ならば

$$\sum_{jm+i \leq k} \alpha_{ij} Z_i(t+j|t) = 0$$

となるはずである. 条件付確率の性質

$$E(X|Y) = E(E(X|Y, Z)|Y)$$

に注意すれば

$$\sum_{jm+i \leq k} \alpha_{ij} Z_i(t+j|t-1) = 0$$

が従い, これが任意の  $t$  で成立することから

$$\sum_{jm+i \leq k} \alpha_{ij} Z_i(t+j+1|t) = 0.$$

$j' = j+1$  とおけば, これは

$$\sum_{j'm+i \leq k+m} \alpha_{ij'-1} Z_i(t+j'|t) = 0.$$

つまり,  $k+m \notin K$  がいえた.

この定理を用いれば, 線形独立な要素だけからなる新たな状態ベクトル

$$\tilde{X}(t) = \begin{pmatrix} x_{k_1}(t) \\ \vdots \\ x_{k_h}(t) \end{pmatrix}$$

を用いた具体的な状態空間表現を得る. まず,  $\tilde{X}(t)$  の  $i$  番目の要素の条件付平均  $E(x_{k_i}(t)|t-1)$  が, 次のように  $x_{k_1}(t-1), \dots, x_{k_h}(t-1)$  つまり  $\tilde{X}(t-1)$  の要素の線形結合で表現できることを示そう.

a)  $k_i + m \leq s$  で  $k_i + m \in K$  のとき,

$x_{k_i+m}(t-1) = E(x_{k_i}(t)|t-1)$  をそのまま用いればよい.

b)  $k_i + m \leq s$  で  $k_i + m \notin K$  のとき,

$x_{k_i+m}(t-1) = E(x_{k_i}(t)|t-1)$  は  $\tilde{X}(t-1)$  の要素にはならないが  $X(t-1)$  の  $k_i+m$  番目の要素なので線形従属で  $x_{k_i+m}(t-1) = \sum_{k_j < k_i+m} \alpha_{ij} x_{k_j}(t-1)$

と表せるはずである.

c)  $k_i + m > s$  のとき,

$E(x_{k_i}(t)|t-1)$  は対応する要素が  $\tilde{X}(t-1)$  にも  $X(t-1)$  にも存在しないが, 方程式  $Z(t+r-1|t-1) = -\Phi_1 Z(t+r-2|t-1) - \dots - \Phi_p Z(t+r-p-1|t-1)$  を利用すれば,  $E(x_{k_i}(t)|t-1)$  は  $\tilde{X}(t-1)$  の要素の線形結合で書けるはずである.

これから,

$$E(\tilde{X}(t)|t-1) = \tilde{F}\tilde{X}(t-1)$$

と表現できる.  $G\varepsilon(t)$  が  $X(t)$  のイノベーションであることに注意すれば推移方程式

$$\tilde{X}(t) = \tilde{F}\tilde{X}(t-1) + \tilde{G}\varepsilon(t)$$

を得る. 観測方程式に関しては

$$Z_i(t) = \begin{cases} x_{k_j}(t) & i = k_j \in K \\ \sum_{k_j \in K} c_{ij} x_{k_j}(t) & \text{otherwise} \end{cases}$$

と表せることに注意すれば

$$Z(t) = \tilde{H}\tilde{X}(t)$$

を得る.

さて, 添字集合  $K$  を求めるにはどうしたらよいだろうか. 実際に状態ベクトルの要素の線形独立を調べるのは, 関数の線形独立性を調べることになり, 困難である. また,  $X(t)$  が観測できないといけない. その代わりに, 共分散行列

$C(j) = E(Z(t+j)\overline{Z(t)}^T)$  から作った Hankel 行列

$$\mathcal{H} = \begin{pmatrix} C(0) & C(1) & \cdots & C(r-1) \\ C(1) & C(2) & \cdots & C(r) \\ \vdots & \vdots & \ddots & \vdots \\ C(r-1) & C(r) & \cdots & C(2r-2) \end{pmatrix}$$

の各行の線形独立性を1行目から順に調べ、独立な行の添字だけを取り出せば  $K$  が得られることを以下に示す.

EQ3-A

$$\begin{aligned}
 k \notin K &\iff x_k(t) + \sum_{k_l < k} \alpha_l x_{k_l}(t) = 0 \\
 &\iff Z_i(t+j|t) + \sum_{k_l < k} \alpha_l Z_{i'}(t+j'|t) = 0 \\
 &\quad \text{ただし } i + mj = k, \quad i' + mj' = k_l \\
 &\iff Z_i(t+j) + \sum_{k_l < k} \alpha_l Z_{i'}(t+j') \text{ は任意の } Z_u(t-v) \text{ と直交} \\
 &\quad v = 0, 1, 2, \dots, u = 1, \dots, m \\
 &\iff C_{iu}(j+v) + \sum_{k_l < k} \alpha_l C_{i'u}(j'+v) = 0 \quad (21) \\
 &\quad v = 0, 1, 2, \dots, u = 1, 2, \dots, m
 \end{aligned}$$

(21) は  $v \geq r$  に対しては  $ARMA$  の方程式よりいつでも成立することがわかるので、定理に注意すれば

$$k \notin K \iff \text{Hankel 行列の } k \text{ 行目が } k_l < k \text{ 行の線形結合で表せる.}$$

であることになる.

より具体的に  $K$  を取り出すには、Hankel 行列を推定しその行を1行目から順に取り出して作った部分行列の階数を調べて、階数が増加している限り、その行番号を採用していけばよい. なお  $K$  の要素数  $h$  のことを McMillian degree と呼ぶこともある.

### 10.3.4 $ARMA$ モデルの正準表現 (Canonical Representation of $ARMA$ model)

状態空間での最小表現を利用すれば、冗長性のない、よりわかりやすい  $ARMA$  表現が得られる. この表現は  $ARMA$  の最小表現とも再表現とも呼ばれる. ただし、状態空間表現から  $ARMA$  表現を以下のように注意深く導かないと冗長性を増加させる結果となる.

まず、添字集合  $K$  から、行列

$$S = (s_{ij}; 1 \leq i \leq m, 1 \leq j \leq r)$$

を

$$s_{ij} = \begin{cases} 1 & i + (j - 1)m \in K \\ 0 & \text{otherwise} \end{cases}$$

によって作る. 先の定理から, この行列の各行は, 0 要素の右はすべて 0 要素という特別な性質をもつ行列であることがわかる.

例

$$S = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & \cdots & \cdots & 1 & 0 & \cdots & \cdots & 0 \\ & & & \vdots & & & & \\ 1 & \cdots & \cdots & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}$$

行列  $S$  の添字  $j$  は状態ベクトル  $X(t)$  のブロック番号,  $i$  はブロック内の番号に対応するので,

$$p_i = \sum_j s_{ij}, \quad i = 1, \dots, m$$

を作れば,  $p_i$  は  $Z(t)$  の  $i$  番目の要素が過去にどこまで依存しているかを表す次数である. なお,  $\sum_{ij} s_{ij} = h$  であるので,  $h = \sum_i p_i$  であり,  $(p_1, \dots, p_m)$  は Kronecker indices と呼ばれる.

なお, 以下では  $p_1 \leq p_2 \leq \dots \leq p_m$  であるとする. こうしても,  $Z(t)$  の要素を並べ換えればよいので, 一般性を失わない. このとき,  $S$  は上の例のように階段行列になっていることを注意しておく.

定理

Kronecker indices  $(p_1, \dots, p_m)$  を用いた表現

$$Z_i(t + p_i | t) = \sum_{l=0}^{p_i} d_l(i)^T Z(t + p_i - l | t), \quad i = 1, \dots, m$$

が成立する. ただし,  $d_l(i)$  は  $m$  次元ベクトルであり, 特に  $d_0(i)$  は  $p_i > 0$  ならば 0 ベクトル,  $p_i = 0$  ならば  $i$  番目以降の要素がすべて 0 である  $m$  次元のベクトルである.

**証明**

$p_i = 0$  のとき,  $i \notin K$  であるので,

$$Z_i(t + p_i|t) = Z_i(t|t) = \sum_{k_j < i} c_{ij} x_{k_j}(t)$$

と表せるはずである. しかも  $k_j < i \leq m$  であることより,  $x_{k_j}(t)$  はすべて  $Z(t|t)$  の  $i$  番目の要素より手前の要素に対応する. 従って,  $d_0(i)$  は  $i$  番目以降の要素がすべて 0 のベクトルとなる.

$p_i > 0$  ならば,  $Z_i(t + p_i|t)$  は行列  $S$  の第  $i$  行第  $p_i + 1$  列に対応するので, この要素が 0 であることより, 線形従属となり, しかも,  $S$  が階段行列であるので,  $Z_j(t + p_i|t)$ ,  $j \leq i$  に対応する  $S$  の要素はすべて 0. これは  $Z_i(t + p_i|t)$  は  $Z(t + p_i - l|t)$ ,  $l = 1, \dots, p_i$  の要素の線形結合で表せることを意味する. つまり,

$$Z_i(t + p_i|t) = \sum_{l=1}^{p_i} d_l(i)^T Z(t + p_i - l|t), \quad i = 1, \dots, m$$

のような表現を得る.

この定理から,  $\bar{p} = \max(p_1, \dots, p_m)$ ,

$$D_l = \begin{pmatrix} d_l(1)^T \\ \vdots \\ d_l(m)^T \end{pmatrix}, \quad l = 0, \dots, \bar{p}$$

とおけば,  $D_l$  は階段行列で, これを用いて

$$Z(t + \bar{p}|t) = \sum_{l=0}^{\bar{p}} D_l Z(t + \bar{p} - l|t)$$

と表現できる. ここで,  $D_0$  は下三角行列で, 対角線がすべて 0 であることに注意すれば  $(I - D_0)^{-1}$  が存在し, 下三角行列となるので,

$$\tilde{D}_l = (I - D_0)^{-1} D_l$$

は  $D_l$  と同じ形の階段行列であり,

$$Z(t + \bar{p}|t) = \sum_{l=1}^{\bar{p}} \tilde{D}_l Z(t + \bar{p} - l|t) .$$

これをより詳しく眺めれば

$$Z_i(t + \bar{p}|t) = \sum_{l=1}^{p_i} (\tilde{D}_l)_{ij} Z_j(t + \bar{p} - l|t), \quad i = 1, \dots, m$$

の形となっていることがわかる。さらに  $t$  での条件付をはずせば、正準  $ARMA$  表現

$$Z_i(t + \bar{p}) = \sum_{l=1}^{p_i} (\tilde{D}_l)_{ij} Z_j(t + \bar{p} - l) + \sum_{l=0}^{q_i} (\tilde{E}_l)_{ij} \varepsilon_j(t + \bar{p} - l)$$

を得る。

## 11 状態空間モデルにもとづく予測と推定 (Kalman Filter)

状態空間モデルは、その表現をうまく利用すると逐次的な計算で、状態ベクトルの予測や推定、パラメータの推定が効率よく行える。以下では、一般形

$$\begin{aligned} X(t) &= FX(t-1) + Gv(t) \\ Z(t) &= HX(t) + w(t) \end{aligned}$$

を考える。また、以下では特に定常性は仮定しない。 $F, G, H$  が時間  $t$  に依存してもよい。なお、 $\{v(t)\}, \{w(t)\}$  は独立な系列で、それぞれ  $N(0, Q), N(0, R)$  に従うとする。さらに、 $\forall s < t$  に対して  $v(t)$  と  $X(s)$  は独立、 $w(t)$  と  $X(s)$  も独立であるとする。もちろん、 $v(t), w(t)$  はそれぞれ  $Z(s), s < t$  とも独立である。

### 11.1 一期先予測 (one step ahead prediction)

$t-1$  時点までの観測  $Z(t-1), Z(t-2), \dots$  にもとづく  $X(t)$  の最良予測、つまり、平均2乗誤差

$$E\|X(t) - f(Z(t-1), Z(t-2), \dots)\|^2$$

を最小にする  $f(Z(t-1), Z(t-2), \dots)$  は,  $Z(t-1), Z(t-2), \dots$  から生成された  $\sigma$ -集合体  $\mathcal{F}_{t-1}$  に関する条件付平均値

$$X(t|t-1) = E(X(t)|\mathcal{F}_{t-1})$$

で与えられることは, 以前に述べた通りである. ここで, 条件が  $\{X(t)\}$  に関してではなく  $\{Z(t)\}$  に関してであることに注意されたい. 従って  $Z(t|t) = Z(t)$  であるが,  $X(t|t)$  と  $X(t)$  は一致しない. 今,  $X(t-1|t-1)$  が得られているものとすれば, 推移方程式の両辺の  $\mathcal{F}_{t-1}$  に関する条件付平均値をとって, 予測方程式

$$X(t|t-1) = FX(t-1|t-1) \quad (22)$$

EQ3.4-1

を得る. ただし, ここで

$$E(v(t)|\mathcal{F}_{t-1}) = E(E(v(t)|\mathcal{G}_{t-1})|\mathcal{F}_{t-1}) = 0$$

を用いた. また,  $\mathcal{G}_{t-1}$  は  $X(t-1), X(t-2), \dots$  から生成された  $\sigma$  集合体で  $\mathcal{G}_{t-1} \supset \mathcal{F}_{t-1}$  である. さらに, 条件付共分散は

EQ3.4-2

$$\begin{aligned} V(t|t-1) &= \text{Var}(X(t)|\mathcal{F}_{t-1}) \\ &= E((X(t) - X(t|t-1))(X(t) - X(t|t-1))^T|\mathcal{F}_{t-1}) \\ &= FV(t-1|t-1)F^T + GQG^T \end{aligned} \quad (23)$$

によって予測できる.

## 11.2 カルマンフィルター (Kalman filter)

工学では「現在の観測値から, すでに定まっている値を求める」という意味で, 推定をフィルターと呼ぶことが多い. カルマンフィルターは, この意味で,  $t$  時点で新たな観測  $Z(t)$  が得られたとき, 一期先予測で求まっていた  $X(t|t-1), V(t|t-1)$  を利用して  $X(t|t)$  や  $V(t|t)$  の形での更新を行う効率的な方法と理解できる.

今,  $\xi(t) = Z(t) - Z(t|t-1)$ ,  $\eta(t) = X(t|t) - X(t|t-1)$  とおくと, 正規性の仮定から,  $X(t|t)$  は行列係数を用いた  $Z(t), Z(t-1), \dots$  の線形結合で表現できるはずであるが, これはさらに  $\xi(t), \xi(t-1), \dots$  の線形結合で表せる. つまり,

$$X(t|t) = \sum_{k=0}^{\infty} A_k^{(t)} \xi(t-k)$$

なる表現をもつ（直交分解）．この両辺の  $\mathcal{F}_{t-1}$  での条件付平均をとれば

$$X(t|t-1) = \sum_{k=1}^{\infty} A_k^{(t)} \xi(t-k)$$

となるので

$$X(t|t) - X(t|t-1) = A_0^{(t)} \xi(t)$$

を得る．この係数行列は次のようにして求まる．

$$\eta(t) = A_0^{(t)} \xi(t)$$

より，

$$E(\eta(t)\xi(t)^T | \mathcal{F}_{t-1}) = A_0^{(t)} E(\xi(t)\xi(t)^T | \mathcal{F}_{t-1})$$

左辺は，

$$\begin{aligned} E(\eta(t)\xi(t)^T | \mathcal{F}_{t-1}) &= E(\eta(t)\eta(t)^T | \mathcal{F}_{t-1})H^T + E(\eta(t)w(t)^T | \mathcal{F}_{t-1}) \\ &= V(t|t-1)H^T \end{aligned}$$

となり，同様にして

$$\begin{aligned} E(\xi(t)\xi(t)^T | \mathcal{F}_{t-1}) &= HE(\eta(t)\eta(t)^T | \mathcal{F}_{t-1})H^T + E(w(t)w(t)^T | \mathcal{F}_{t-1}) \\ &= HV(t|t-1)H^T + R \end{aligned}$$

を得る．従って，

$$A_0^{(t)} = V(t|t-1)H^T(HV(t|t-1)H^T + R)^{-1} .$$

この行列はカルマンゲインと呼ばれ， $K_t$  で表されることが多い．結局， $X(t|t-1)$  から  $X(t|t)$  を求める式

$$\begin{aligned} X(t|t) &= X(t|t-1) + K_t \xi(t) \\ &= X(t|t-1) + K_t(Z(t) - HX(t|t-1)) \end{aligned} \tag{24}$$

EQ3.4-3

が得られる．また条件付分散は

$$V(t|t) = (I - K_t H)V(t|t-1) \tag{25}$$

EQ3.4-4

で  $V(t|t-1)$  から求まる．この式 (25) を確かめるには，次の補題を用いる．

## 補題

$X, Y$  をそれぞれベクトル値をとる確率変数とし,  $\begin{pmatrix} X \\ Y \end{pmatrix}$  が多変量正規分布に従うとすると,  $X$  の条件付平均と条件付分散は

$$\begin{aligned} E(X|Y) &= E(X) + \Sigma_{XY}\Sigma_{YY}^{-1}(Y - E(Y)), \\ \text{Var}(X|Y) &= \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \end{aligned}$$

で与えられる. ただし,  $\Sigma_{XX}, \Sigma_{YY}$  はそれぞれ  $X, Y$  の分散行列,  $\Sigma_{XY}$  は  $X, Y$  の共分散行列,  $\Sigma_{YX}$  は  $Y, X$  の共分散行列である. この  $\text{Var}(X|Y)$  はいわゆる  $Y$  を介した間接的な相関を除いた相関を表す  $X$  の偏共分散にほかならない.

この補題で  $X$  を  $X(t)$ ,  $Y$  を  $Z(t)$  と考えれば,  $\text{Var}(X|Y)$  が  $V(t|t)$ ,  $\text{Var}(X)$  が  $V(t|t-1)$  に対応する. 従って,

$$V(t|t) = V(t|t-1) - \text{Cov}(X(t), Z(t)|t-1)\text{Var}(Z(t)|t-1)^{-1}\text{Cov}(Z(t), X(t)|t-1)$$

を得る. ここで,

$$\begin{aligned} \text{Cov}(X(t), Z(t)|t-1) &= \text{Cov}(X(t), X(t)|t-1)H^T \\ &= V(t|t-1)H^T \\ \text{Var}(Z(t)|t-1) &= \text{Var}(\xi(t)|t-1) \\ &= HV(t|t-1)H^T + R \end{aligned}$$

であるので,  $K_t$  の定義を思い出せば (25) が従うことはすぐわかる.

(24) と (25) を合わせてカルマンフィルターと呼ばれる. 3.4.1 の一期先の予測, 推定の式 (22) と (23) と組み合わせれば,  $X(t-1|t-1)$  と  $V(t-1|t-1)$  から, 新たな観測値  $Z(t)$  にもとづいて,  $X(t|t)$ ,  $V(t|t)$  を簡単に更新できることになる.

## 11.3 スムーザー (smoother)

これまでの話は, 観測が行われた最後の時間, あるいはその1つ先の時間の状態ベクトルや条件付分散を求める方法であったが, 場合によっては, それより前の時間に関して同様のことを行いたいこともある. このような演算を工学では平滑化 (スムージング) と呼ぶことが多い.

a) 固定点 smoother

$Z(t), Z(t-1), \dots$  が観測されたとき, 時点  $k \leq t$  に対する  $X(k|t)$  や  $V(k|t)$  を求めるのが固定点 smoother である. カルマンフィルターのときと同じように

$$X(k|t) = \sum_{l=0}^{\infty} B_l^{k,t} \xi(t-l)$$

と表せるので,

$$X(k|t) - X(k|t-1) = B_0^{k,t} \xi(t)$$

となる. 従って,

$$\begin{aligned} \mathbb{E} \left( B_0^{k,t} \xi(t) \xi(t)^T | \mathcal{F}_{t-1} \right) &= \mathbb{E} \left( (X(k|t) - X(k|t-1)) \xi(t)^T \right) = \mathbb{E} \left( X(k|t) \xi(t)^T \right) \\ &= \mathbb{E} \left( X(k) \xi(t)^T \right) = \mathbb{E} \left( X(k) (X(t) - X(t|t-1))^T \right) H^T \end{aligned}$$

となり, この右辺を  $\Omega(t|t-1, k) H^T$  とおけば,

$$B_0^{k,t} = \Omega(t|t-1, k) H^T [H V(t|t-1) H^T + R]^{-1}$$

を得る. これを  $A_t(k)$  とおきなおせば,

$$\begin{aligned} X(k|t) &= X(k|t-1) + A_t(k) \xi(t) \\ &= X(k|t-1) + A_t(k) (Z(t) - H X(t|t-1)) \end{aligned}$$

によって, 観測値  $Z(t)$  が新たに得られるごとに,  $X(k)$  の平滑値  $X(k|t-1)$  を  $X(k|t)$  に更新できる.  $X(t|t-1)$  を求めるには 3.4.1 の結果を用いればよい.  $V(k|t)$  の更新方程式は

$$V(k|t) = V(k|t-1) - \Omega(t|t-1, k) H^T [H V(t|t-1) H^T + R]^{-1} H \Omega(t|t-1, k)^T,$$

となる. ただし,

$$\begin{aligned} \Omega(t|t-1, k) &= \Omega(t-1|t-2, k) [I - K_{t-1} H]^T F^T, \quad t > k \\ \Omega(k|k-1, k) &= V(k|k-1). \end{aligned}$$

の関係があるので,  $t-1$  時点での  $\Omega(t-1|t-2, k)$  から  $\Omega(t|t-1, k)$  が求まる. この証明は演習問題.

b) 固定区間 smoother

一期先予測やカルマンフィルターによって状態ベクトルや条件付分散の推定を逐次的に更新して、時点  $N$  に達した時点で、この smoother を用いれば過去の  $X(t)$  や  $V(t|t), t \leq N$  をさらに更新して精密にできる。

a) の結果を用いれば、

$$\begin{aligned} X(t|N) &= X(t|N-1) + A_N(t)\xi(N) \\ &= X(t|N-2) + A_{N-1}(t)\xi(N-1) + A_N(t)\xi(N) \\ &\quad \vdots \\ &= X(t|t) + A_{t+1}(t)\xi(t+1) + \cdots + A_N(t)\xi(N). \end{aligned}$$

一方、同様にして、

$$X(t+1|N) = X(t+1|t) + A_{t+1}(t+1)\xi(t+1) + \cdots + A_N(t+1)\xi(N)$$

が成立するが、係数行列  $A_l(t)$  と  $A_l(t+1)$  の間には

$$A_l(t) = V(t|t)F^T V(t+1|t)^{-1} A_l(t+1), \quad \forall l > t$$

の関係がある。

(証明は演習問題、ヒント：

$$E\left((X(t) - V(t|t)F^T V(t+1|t)^{-1} X(t+1)) (\xi(t+j))^T\right) = 0, \quad \forall j > 0$$

をいえばよい。これをいうには

$$A(t) = V(t|t)F^T V(t+1|t)^{-1}$$

とにおいて、

$$E\left((X(t) - A(t)X(t+1))(X(t+j) - X(t+j|t+j-1))^T\right) = 0, \quad \forall j > 0$$

が成立することをいえばよいが、まず、 $j=1$  のときを証明し、 $j=2$  の場合は、

$$\begin{aligned} X(t+2) - X(t+2|t+1) &= F(X(t+1) - X(t+1|t)) \\ &\quad - FK_t(w(t+1) + H(X(t+1) - X(t+1|t))) \end{aligned}$$

のように書き換えられることから、 $j = 1$  の場合に帰着すればよい。  
 あとは帰納法である)

これを利用すれば

$$X(t|N) = X(t|t) + A(t)(X(t+1|N) - X(t+1|t)), \quad t \leq N$$

を得る。分散共分散行列に関する更新方程式は

$$V(t|N) = V(t|t) + A(t)(V(t+1|N) - V(t+1|t))A(t)^T, \quad t \leq N$$

で与えられる。この証明も演習問題。

## 11.4 尤度とパラメータ推定 (likelihood and parameter estimation)

これまでの話は、状態空間モデルのパラメータである  $F, G, H, R, Q$  はすべて既知としてきた。しかし、実際にはこれらのパラメータを推定する必要がある。その方法として広く用いられる方法で、しかもそのよさが保証されている方法が最尤法である。

今、観測値  $Z(1), Z(2), \dots, Z(n)$  が得られたとしたとき、尤度はこの観測値に対する密度関数の値として定義される。前節までは、 $Z(t), Z(t-1), \dots$  によって生成される  $\sigma$  集合体を  $\mathcal{F}_t$  としたが、ここでは、 $Z(t), \dots, Z(1)$  で生成される  $\sigma$  集合体を  $\mathcal{F}_t$  とし、 $\xi(t) = Z(t) - E(Z(t)|\mathcal{F}_{t-1})$  とすれば、尤度は  $\xi(t), \xi(t-1), \dots, \xi(1)$  の密度関数の値と同等で、

$$f = \prod_{t=1}^n \frac{1}{\sqrt{2\pi}} |\Sigma_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \xi(t)^T \Sigma_t^{-1} \xi(t)\right)$$

となる。ただし、 $\Sigma_t = \text{Var}(\xi(t)|t-1) = HV(t|t-1)H^T + R$  である。厳密に言えば、この尤度は  $Z(1)$  を与えたときの条件付尤度であり、 $\xi(1) = Z(1)$  と定義しておく。具体的に最尤推定量を得るには、 $f$  あるいは  $\log f$  の未知パラメータ  $F, G, H, R, Q$  に関する最大化問題を解くことになる。もちろん解析的にきれいな形で解くことはできないので、数値計算に頼ることになるが、その段階で、前節までのカルマンフィルター、スムーザを活用することになる。

参考書

Reinsel,G.C., Elements of Multivariate Time Series Analysis, 1993, Springer

北川源四郎,『時系列解析プログラミング』, 1993, 岩波書店

赤池弘次, 北川源四郎編,『時系列解析の実際 I , II』, 1995, 朝倉書店

片山徹,『応用カルマンフィルタ』, 1983, 朝倉書店

Hannan, E.J., Multiple Time Series, 1970, John Wiley

Hannan,E.J., Deistler,M., The Statistical Theory of Linear Systems.  
1988, John Wiley

Priestley, M.B., Spectral Analysis and Time Series Vol.2, 1981, Academic Press.