Wold's Decomposition

Any wss random process X_t can be decomposed in the following form

$$X_t = m + a_t + G_1 a_{t-1} + G_2 a_{t-2} + \dots$$

where

- a_t is a random process satisfying:
  - $E[a_t] = 0$
  - $E[a_t a_{t-\ell}] = \sigma_a^2 \delta_\ell$, where δ is a Kronecker delta function
    $$\delta_\ell = \begin{cases} 1, & \ell = 0 \\ 0, & \text{otherwise} \end{cases}$$

  a_t is an uncorrelated stationary process (white noise process)

- The terms G_ℓ, $\ell \in \{0,1,2,\dots\}$ satisfy
  - $\sum_{\ell=0}^{\infty} G_\ell^2 < \infty$

Coefficients G_ℓ are often referred to as Green's fcn coefficients.

Note: $X_t = \sum_{l=0}^{\infty} G_l \, a_{t-l} = G_t * a_t$

(without a loss of generality - WLOG - we assumed $E[X_t] = \mu = 0$).

$a_t \longrightarrow \boxed{\text{System}} \longrightarrow X_t$

If I replace white noise a_t with an impulse $\delta_t = \begin{cases} 1, & t=0 \\ 0, & \text{othws} \end{cases}$

G_l-s can be seen as impulse response of the system that when driven by white noise gave us the process X_t

$$\sum_{l=0}^{\infty} G_l \, \delta_{t-l} = G_t$$

$\delta_t \longrightarrow \boxed{\text{System}} \longrightarrow G_t$

Wold's decomposition is a very general and important result for describing WSS random processes. We will use it to demonstrate generality of our modeling approaches in the form of ARMA models.

Nevertheless, we can already observe some useful concepts that become (more) easily describable when a random process is decomposed into Wold's decomposition

i) Calculating variance of X_t

$$\mathrm{Var}[X_t] = E[\sum_{l_1} G_{l_1} a_{t-l_1} \sum_{l_2} G_{l_2} a_{t-l_2}] =$$

$$= E[\sum_{l_1} \sum_{l_2} G_{l_1} G_{l_2} a_{t-l_1} a_{t-l_2}] =$$

$$= \sum_{l_1} \sum_{l_2} G_{l_1} G_{l_2} E[a_{t-l_1} a_{t-l_2}] =$$

$$= \sigma_a^2 \sum_{l_1} \sum_{l_2} G_{l_1} G_{l_2} \delta_{l_1 - l_2} =$$

$$= \sigma_a^2 \sum_{l} G_l^2 \quad \text{(which is convergent)}$$

ii) Covariance of the process X_t

$$E[X_t X_{t+l}] = E[\sum_{k=0}^{\infty} G_k a_{t-k} \sum_{n=0}^{\infty} G_n a_{t+l-n}] =$$

$$= E[\sum_{k=0}^{\infty} \sum_{n=0}^{\infty} G_k G_n a_{t-k} a_{t+l-n}] =$$

$$= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} G_k G_n E[a_{t-k} a_{t+l-n}] =$$

$$= \sigma_a^2 \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} G_k G_n \delta_{n-l-k} = \sigma_a^2 \sum_{k=0}^{\infty} G_k G_{k+l}$$

Note that indeed, the covariance function is origin independent (it only depends on the lag ℓ).

iii) Characterizing prediction errors

$$\text{Let } X_t = \sum_{\ell=0}^{\infty} G_\ell a_{t-\ell} = \overset{G_0 = 1}{G_0} a_t + G_1 a_{t-1} + \ldots$$

$$\text{Then } X_{t+\ell} = \underbrace{a_{t+\ell} + G_1 a_{t+\ell-1} + \ldots + G_{\ell-1} a_{t+1}}_{\text{prediction error at time } t} + G_\ell a_t + G_{\ell+1} a_{t-1} + \ldots$$

$$\hat{X}_t(\ell) = E[X_{t+\ell} \mid t] = G_\ell a_t + G_{\ell+1} a_{t-1} + \ldots$$

prediction is conditional expectation given the information

$$\hat{e}_t(\ell) = X_{t+\ell} - \hat{X}_t(\ell) = G_0 a_{t+\ell} + G_1 a_{t+\ell-1} + \ldots + G_{\ell-1} a_{t+1}$$

prediction error at time t

$$\mathrm{Var}\, \hat{e}_t(\ell) = E[\hat{e}_t(\ell)^2] = \sigma_a^2 (G_0^2 + G_1^2 + \ldots + G_{\ell-1}^2)$$

Elegant description of the variance of prediction errors

Wold's decompositions (Green fcn coefficients for some ARMA models)

(a) AR(1) model

Let's observe a random process X_t satisfying

$$X_t = \phi_1 X_{t-1} + a_t \qquad a_t \sim NIID\ \mathcal{N}(0, \sigma_a^2)$$

$$|\phi_1| < 1$$

(it's sufficient to just assume a_t is an uncorrelated stationary process and drop Gaussianity

$$X_{t-1} = \phi_1 X_{t-2} + a_{t-1}$$

$$X_{t-2} = \phi_1 X_{t-3} + a_{t-2}$$

$$\vdots$$

$$\Rightarrow X_t = \phi_1 X_{t-1} + a_t = a_t + \phi_1 X_{t-2} + \phi_1 a_{t-1} =$$

$$= a_t + \phi_1 a_{t-1} + \phi_1^2 X_{t-3} + \phi_1^2 a_{t-2} = \dots$$

$$= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \dots = \sum_{l=0}^{\infty} \phi_1^l a_{t-l}$$

$$= \sum_{l=0}^{\infty} G_l a_{t-l} \quad \text{where} \quad G_l = \phi_1^l$$

Note: If X_t had finite support, we wouldn't need $|\phi_1| < 1$ (however, strictly speaking, for stationarity of X_t, we NEED $|\phi_1| < 1$

Variance and covariance fcn of an AR(1) model

$$X_t = \phi_1 X_{t-1} + a_t = \sum_{\ell=0}^{\infty} \phi_1^{\ell} a_{t-\ell} \qquad (|\phi_1| < 1)$$

$$\Rightarrow \operatorname{Var}[X_t] = \sum_{\ell=0}^{\infty} (\phi_1^{\ell})^2 \cdot \sigma_a^2 = \frac{\sigma_a^2}{1-\phi_1^2}$$

$$E[X_t X_{t+\ell}] = \sigma_a^2 \sum_{k=0}^{\infty} \phi_1^{k} \phi_1^{k+\ell} \sigma_a^2 = \frac{\sigma_a^2}{1-\phi_1^2} \phi_1^{\ell}$$

(for $\ell \geq 0$)

Note: $E[X_t X_{t+\ell}] = E[X_t X_{t-\ell}] = \dfrac{\sigma_a^2}{1-\phi_1^2} \phi_1^{|\ell|}$

(b) First order Green's fcn coefficients

Let $X_t = \sum_{l=0}^{\infty} G_l a_{t-l}$ where $G_l = \phi_1^l$, $|\phi_1| < 1$

and a_t is a wss white noise, i.e.

- $E[a_t] = 0$
- $E[a_t a_{t+l}] = \begin{cases} \sigma_a^2, & l=0 \\ 0, & l \neq 0 \end{cases}$

$\Rightarrow \quad X_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \phi_1^3 a_{t-3} + \dots$

$X_{t-1} = \quad a_{t-1} + \phi_1 a_{t-2} + \phi_1^2 a_{t-3} + \dots$

$\Rightarrow \quad X_t = a_t + \phi_1 X_{t-1} \rightarrow$ it's an AR(1) model

Note: Even if $|\phi_1| \geq 1$, if we have that

$$X_t = \sum_{l=0}^{\infty} \phi_1^l a_{t-l}$$

this results in $X_t = \phi_1 X_{t-1} + a_t$, but this process is NOT wss (it has no variance)

Another way of seeing transformations between GF-based and ARMA based representations.

Let's introduce back-shift operator in the space of random processes

$$B X_t = X_{t-1} \qquad (B a_t = a_{t-1})$$

Then AR(1) model is $(1-\phi_1 B) X_t = a_t$

and Wold's decomposition is

$$X_t = \left(\sum_{l=0}^{\infty} G_l B^l\right) a_t$$

it's an operator applied to the time series a_t

We see that for $G_l = \phi_1^l$, $|\phi_1|<1$

$$X_t = \phi_1 X_{t-1} + a_t \iff X_t = \sum_{l=0}^{\infty} \phi_1^l a_{t-l}$$

$$(1-\phi_1 B) X_t = a_t \iff X_t = \left[\sum_{l=0}^{\infty} \phi_1^l B^l\right] a_t$$

$$\Rightarrow X_t = \underbrace{\left[\left(\sum_{l=0}^{\infty} \phi_1^l B^l\right)(1-\phi_1 B)\right]}_{\text{IDENTITY OP.}} X_t$$

$\Rightarrow$ I can see

$$1 + \phi_1 B + \phi_1^2 B^2 + \dots + \phi_1^l B^l + \dots$$

as inverse operator of $(1 - \phi_1 B)$ and label it as

$$1 + \phi_1 B + \dots + \phi_1^l B^l = \frac{1}{1 - \phi_1 B}$$

Another way to see this is that for any random process X_t

$$(1 - \phi_1 B)(1 + \phi_1 B + \dots + \phi_1^n B^n) X_t = X_t - \phi_1^{n+1} X_{t-n-1}$$

$$= (1 - \phi_1^{n+1} B^{n+1}) X_t$$

As we let $n \to \infty$, $\phi_1^n \to 0$ and hence

$$(1 - \phi_1 B)(1 + \phi_1 B + \dots + \phi_1^n B^n + \dots) X_t = X_t$$

or in other words $\underbrace{1 + \phi_1 B + \dots + \phi_1^n B^n + \dots}_{\text{It's inverse operator of the operator } (1 - \phi_1 B)} = \frac{1}{1 - \phi_1 B}$

Actually, if X_t has finite support, then we don't even need $|\phi_1| < 1$.