Lecture 3

Special case of an AR(1) model is "Random Walk"

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

Substituting successively, we can see that

$$X_t = \sum_{k=0}^{t} a_t$$

and that

$$\hat{X}_t(1) = X_t$$ (best prediction of the next sample is the previous one)

First model of the stock market

Bachelier's hypothesis - 1900

Also important because it's a limiting case between "stable" & "unstable" time-series (we'll talk more about it when we discuss time-series analysis).

Autoregressive Model of Order 2 - AR(2)

$$X_3 = \phi_1 X_2 + \phi_2 X_1 + a_3$$

$$X_4 = \phi_1 X_3 + \phi_2 X_2 + a_4 \qquad \Rightarrow$$

$$\vdots$$

$$X_N = \phi_1 X_{N-1} + \phi_2 X_{N-2} + a_N$$

$$\overset{\underline{Y}}{\begin{bmatrix} X_3 \\ X_4 \\ \vdots \\ X_N \end{bmatrix}} = \overset{\underline{X}}{\begin{bmatrix} X_2 & X_1 \\ X_3 & X_2 \\ \cdots & \cdots \\ X_{N-1} & X_{N-2} \end{bmatrix}} \overset{\vec{\beta}}{\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}} + \overset{\vec{\varepsilon}}{\begin{bmatrix} a_3 \\ a_4 \\ \vdots \\ a_N \end{bmatrix}}$$

Now, this looks like a multivariate regression

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y}$$

$$\hat{\sigma}_a^2 = \frac{1}{N-2-2} \sum_{t=3}^{N} (X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2})$$

ARMA (2,1) Models

$$X_t = \phi_1 X_{t-1} + a_t - \theta_1 a_{t-1} + \phi_2 X_{t-2}$$

$$= \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t - \theta_1 a_{t-1} \qquad (*)$$

If a_t-s are NIID $a_t \sim N(0, \sigma_a^2)$, then model $(*)$ is an Auto-Regressive Moving Average (ARMA) model of AR order 2 & MA order 1

ARMA(2,1)

Differences from AR(1) or AR(2) models

i) More parameters

ii) Has both AR & MA parts (noise term has some structure)

iii) Need past a_t-s to find the current one
=> must recursivelly look for a_t-s

$$a_t = X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} + \theta_1 a_{t-1}$$

Point iii) leads to a more fundamental difference (and a problem)

$$a_t = X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} + \theta_1 a_{t-1}$$

$$\hat{a}_t = X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2} + \theta_1 \hat{a}_{t-1}$$

$$\hat{a}_{t-1} = X_{t-1} - \hat{\phi}_1 X_{t-2} - \hat{\phi}_2 X_{t-3} + \theta_1 \hat{a}_{t-2}$$

wish to make it look like an AR model so I could estimate ϕ's

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t - \theta_1 (X_{t-1} - \phi_1 X_{t-2} - \phi_2 X_{t-3} + \theta_1 a_{t-2}) =$$

$$= a_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} - \theta_1 X_{t-1} + \theta_1 \phi_1 \overset{?}{X_{t-2}} + \theta_1 \phi_2 \overset{?}{X_{t-3}} + \dots$$

Non-linear estimation problem (not as easy to estimate parameters as it was in the AR case)
(we'll use M-Lab's command armax)

How to check adequacy of ARMA(2,1)?

Again, need to formally check if a_t is a white process

$$\hat{\rho}_k = \frac{\frac{1}{N-\ell}\sum_{t=1}^{N-\ell} a_t a_{t+\ell}}{\frac{1}{N}\sum_{t=1}^{N} a_t^2} \approx \delta_\ell = \begin{cases} 1, & \ell = 0 \\ 0, & \ell \neq 0 \end{cases}$$

$\hat{\rho}_k$ should be less than $\frac{2}{\sqrt{N}}$

Special Case of ARMA(2,1)

→ Auto-regressive model of order 2: AR(2)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t$$

This was already discussed in the last lecture!

What to do if neither AR(1), nor AR(2), nor ARMA(2,1) models are adequate (i.e. if residuals for any of these models are not white)?

Naturally, I'll go to ARMA(3,2)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

which is again a non-linear estimation problem

Then ARMA(4,3), ARMA(5,4) ... ARMA(n, n-1)

Where do I stop? We'll learn in Ch. 4.

For now, we'll focus on analyzing models (assume model is given, what does it mean?)

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

    $$\delta_\ell = \begin{cases} 1, & \ell = 0 \\ 0, & \text{otherwise} \end{cases}$$

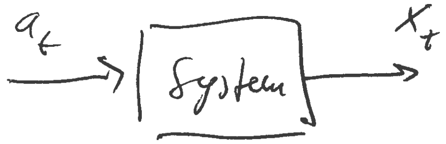
  a_t is an uncorrelated stationary process (white noise process)

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

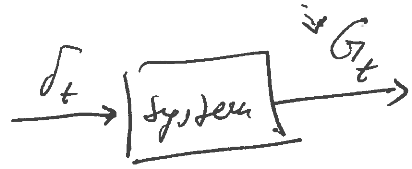
(without a loss of generality - WLOG - we assumed $E[X_t] = \mu = 0$).



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

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

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



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

$$\text{Let} \quad X_t = \sum_{l=0}^{\infty} G_l a_{t-l} = \overset{\swarrow G_0 = 1}{G_0} a_t + G_1 a_{t-1} + \dots$$

$$\text{Then} \quad X_{t+l} = \underbrace{a_{t+l} + G_1 a_{t+l-1} + \dots + G_{l-1} a_{t+1}}_{\text{prediction error at time } t} + G_l a_t + G_{l+1} a_{t-1} + \dots$$

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

prediction is conditional expectation given the information

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

prediction error at time t

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

Elegant description of the variance of prediction errors