Specialist Courses

Daniele Zago

December 6, 2021

Contents

Lecture 2

2021-11-29

2.1 White noise process

It serves as a building block in defining more complex linear time series processes and reflects information that is not directly observable. I It is easy to see that a sequence of independent and identically distributed (i.i.d.) random variables with mean 0 and finite variance σ^2 is a special white noise process.
In general it's convenient to write a stochastic process as a sum of white noise terms, since it's easier to prove theorems related to them.
The probability behavior (law) of a stochastic process is completely determined by all of its finite-dimensional distributions.
When all of the finite-dimensional distributions are Gaussian (normal), the process is called a Gaussian process.
Since uncorrelated normal random variables are also independent, a Gaussian WN process is, in fact, a sequence of i.i.d. normal random variables.

2.2 Random walk

Whereas the white noise is a simple process with no memory, the random walk has infinite memory and is nonstationary,

$$X_t = \mu + X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2),$$

where $X_0 = 0$ by convention. The process is such that by recursion,

$$\begin{aligned} X_t &= \mu + (\mu + X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \dots \\ &= \underbrace{t\mu}_{\text{drift}} + \underbrace{\sum_{i=1}^{t} \varepsilon_i}_{\text{stoch. trend}}. \end{aligned}$$

The last sum is called ***stochastic trend***, since every error ε enters with the same error even from past observations. Moreover, $\mathbb{E}[X_t] = t\mu$ and $\mathbb{V}[X_t] = t\sigma^2$. Applying the first-difference operator to the process yields

$$(1 - B)X_t = X_t - X_{t-1} = \mu + \varepsilon_t,$$

which is a stationary model.

2.3 Linear time series

We introduce the ARMA model, the most famous type of linear time-series model which is used even for non-linear data. Forecasts from these models in these case have been empirically shown to be more accurate than forecasts from more complicated models.

A general linear process is of the form

$$X_t = \varepsilon_t + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i},$$

where $\sum_{i=1}^{\infty} \psi_i^2 < \infty$. This type of process is such that

i. $\mathbb{E}[X_t] = 0$ for each t

ii. $\mathrm{Cov}(X_t, t_{t-k}) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$ for $k \geq 0$ and $\psi_0 = 1$.

An example is when the weights are an exponentially decaying sequence $\psi_j = \varphi^j$ with $|\varphi| < 1$, and in this case

$$X_t = \varepsilon_t + \varphi \varepsilon_{t-1} + \varphi^2 \varepsilon_{t-2} + \dots$$

The variance of this process can be written as a geometric series

$$\mathbb{V}[X_t] = \sigma_\varepsilon^2 \cdot \sum_{i=0}^{\infty} \varphi^k = \frac{\sigma_\varepsilon^2}{1-\varphi^2},$$

moreover, the covariance functions are

$$\mathrm{Cov}(X_t, X_{t-k}) = \frac{\varphi^k \sigma_\varepsilon^2}{1-\varphi^2}$$

$$\mathrm{Cor}(X_t, X_{t-k}) = \varphi^k$$

A moving average model of the form MA(q) $=$

$$X_t = \vartheta_1 \varepsilon_{t-1} + \dots + \vartheta_q \varepsilon_{t-q} + \varepsilon_t.$$

These models are easily tractable since they are stationary by definition and
An autoregressive model AR is such that

$$X_t = c + \varphi_1 X_1 + \dots + \varphi_p X_{t-p} + \varepsilon_t,$$

where X_t's could also be random variables each uncorrelated with the next value X_{t+1}. The expected value of the process can be calculated in terms of the autoregressive coefficients, by assuming the process to be stationary

$$\mathbb{E}[X_t] = \mathbb{E}[c + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t] \implies \mathbb{E}[X_t] = \frac{c}{1 - \sum_{i=1}^{p} X_{t-i}}.$$

Again, by assuming the process to be stationary we observe an autocovariance of the form

$$\gamma_k = \begin{cases} \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + \dots + \varphi \gamma_p + \sigma_\varepsilon^2 & k = 0 \\ \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + \dots + \varphi \gamma_p & k > 0 \end{cases}$$

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p}, \quad k > 0,$$

which yield the Yule-Walker equations when considering them for $k = 1, \dots, p$. These equations can be used to compute the model coefficients when solving them in terms of the unknown $\boldsymbol{\varphi}$ and sample autocorrelation $\widehat{\rho}_k$.
We have that the sample autocorrelation is exponentially decaying in k, depending on the model parameters.

Its partial autocorrelation function is null for $k > p$.

2.4 ARMA model

We introduce the combined ARMA model in order to model more complicated dynamics of time series, yielding the ARMA(p, q) defined as

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \vartheta_1 \varepsilon_{t-1} + \dots + \vartheta_q \varepsilon_{t-q} + \varepsilon_t,$$

which usually allows us to model correlation structures with a smaller number of parameters.

Using a backshift operator $B^k X_t = X_{t-k}$ we can write this model as

$$\varphi(B) X_t = \vartheta(B) \varepsilon_t,$$

where the polynomials in B are defined as

$$\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$$

$$\vartheta(B) = 1 + \vartheta B + \dots + \vartheta_q B^q$$

and are extremely important since we can determine the properties of an ARMA model in terms of $\vartheta(\cdot)$ and $\varphi(\cdot)$. Moreover, if the ARMA model is stationary, we can write an AR(∞) representation as

$$\vartheta(B)^{-1} \varphi(B) X_t = \varepsilon_t,$$

and a $MA(\infty)$ representation as

$$X_t = \varphi(B)^{-1} \vartheta(B) \varepsilon_t.$$

Example (AR(1))
Consider the AR(1) model, then

$$\begin{aligned} Y_t &= \varphi Y_{t-1} + \varepsilon_t \\ &= \varphi(\varphi Y_{t-2} + \varepsilon_{t+1}) + \varepsilon_t \\ &= \dots \\ &= \varepsilon_t + \varphi \varepsilon_{t-1} + \varphi^2 \varepsilon_{t-2} + \dots \end{aligned}$$

Example (General procedure)
We write the relationship
$$(1-\varphi B)y_t = (1-\vartheta B)\varepsilon_t,$$
for which we can write
$$y_t = \frac{1-\vartheta B}{1-\varphi B}\varepsilon_t.$$
Our goal now is to obtain a relationship of the form
$$Y_t = \Psi(B)\varepsilon_t = \sum_{i=0}^{\infty}\psi_i B^i,$$
and therefore $\Psi(B) = \frac{1-\vartheta B}{1-\varphi B}$, from which
$$(1-\varphi)\Psi(B) = 1-\vartheta B$$
$$\Updownarrow$$
$$(1-\varphi B)(1+\psi_2 B+\psi_2 B^2+\ldots) = 1-\vartheta B$$
$$\Updownarrow$$

Example (General AR(∞))
The procedure is the same, except we now have to find a relationship of the form
$$\varphi(B)Y_t = \vartheta(B)\varepsilon_t \longrightarrow$$
$$\frac{\varphi(B)}{\vartheta(B)} = \Xi(B),$$
and find the parameters in terms of φ and ϑ.

In order to check for invertibility of the process, we need to invert the MA operator which is doable if the characteristic equation $\vartheta(B) = 0$ has solutions $|B_i| > 1$.
In order to check for of the process, we need to invert the AR operator which is doable if the characteristic equation $\vartheta(B) = 0$ has solutions $|B_i| > 1$.

Time series models only work if the data is stationary, therefore in general it's recommended to check for the evidence of trend or seasonality before applying an ARMA model. We can remove nonstationarity either via regression or via simple differentiation.

Even though the model might not be invertible, it's still better to have it stationary and not invertible. In general, it's advised to differentiate the series rather than risking for the time series to be nonstationary.

Testing whether the trend is deterministic or stochastic can be performed via a unit root test, which is usually not very powerful.
We obtain the ARIMA(p, d, q) class of models by applying a d-order to X_t and modeling the result as an ARMA(p, q) model, i.e.

$$\varphi(B)(1-B)^d Y_t = \vartheta(B)\varepsilon_t.$$

- In general, this process is simply an ARMA$(p + d, q)$ with d unit roots in the autoregressive polynomial.
- In general, we don't see time series such that $d > 2$.

Lecture 3

2021-12-06

Theorem 1 (Wold decomposition)
Let $(X_t)_t$ be a non-deterministic stationary time series iwth $\mathbb{E}[X_t] = 0$, then

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} + V_t,$$

where V_t is deterministic and

1. $\psi_0 = 1$ *and* $\sum_{j=1}^{\infty} \psi_j^2 < \infty$.
2. $a_t = WN(0, \sigma^2)$.
3. $\mathbb{E}[a_t V_s]$ *for all* $s, t = 0, \pm 1, \pm 2, \dots$

With this decomposition we can approximate any stationary time series using a linear process of the form

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j},$$

where $\sum_{j=1}^{\infty} |\psi_j| < \infty$.

3.1 Transfer function models

Transfer function models are models where an output series is related to one or more input series. We link the output series y_t to the predictor series x_t via

$$y_i = \nu(B) x_t + \eta_t,$$

where $\nu(B) = \sum_{j=-\infty}^{\infty} \nu_j B^j$ is the *transfer function of the linear filter* that transform x_t into y_t, and η_t is a noise series independent of x_t. The weights ν_j are called ***impulse response weights*** and the TFM is called ***stable*** if

$$\sum_{j=-\infty}^{\infty} |\nu_j| < \infty,$$

and in particular we are in a BIBO setting (Bounded Input Bounded Output). The TFM is said to be ***causal*** if $\nu_j = 0$ for $j < 0$, since the present output is affected only by the system current and past values,

$$y_t = \sum_{j=0}^{\infty} \nu_j B^j.$$

The purpose of TF models is to identify the TF $\nu(B)$ and the noise model, possibly using a simpler representation, similar to an ARIMA model

$$\delta(B) y_t = \omega(B) B^b x_t,$$

where

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \dots \delta_r B^r$$

$$\omega(B) = \omega_0 - \omega_1 B - \dots - \omega_s B^s$$

and b is a delay parameter that tells us the lag that elapses before the impulse of the input variable produces an effect on the output variable.
The transfer function can be written as

$$\nu(B) = \frac{\omega(B)B^b}{\delta(B)}, \tag{1}$$

and the model has an explicit representation in terms of x_t,

$$y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + \eta_t,$$

and the orders s, r, b can be found by equating the coefficients of B^j to both sides in Equation (1)

$$\delta(B)\nu(B) = \omega(B)B^b,$$

which yields the following equation

$$(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r)(\nu_0 + \nu_1 B + \dots) = (\omega - \omega_1 B - \dots - \omega_s B^s)B^b,$$

and we obtain the following set of equations

$$\begin{aligned}
&\nu_j = 0 && \text{if } j < b\\
&\nu_j = \delta_1 \nu_{j-1} + \delta_2 \nu_{j-2} + \dots + \delta_r \nu_{j-r} + \omega_0 && \text{if } j = b\\
&\nu_j = \delta_1 \nu_{j-1} + \delta_2 \nu_{j-2} + \dots + \delta_r \nu_{j-r} - \omega_{j-b} && \text{if } j = b+1, \dots, b+s\\
&\nu_j = \delta_1 \nu_{j-1} + \delta_2 \nu_{j-2} + \dots + \delta_r \nu_{j-r} && \text{if } j > b+s
\end{aligned}$$

Observing the behaviour of this transfer function (similarly to what we do with ACF and PACF for ARIMA models, we can find the appropriate values of s, r, b.

Def. (Cross-correlation function)
We say that X_t and Y_t are ***jointly stationary*** if they are univariate stationary and $\text{Cov}(X_t, Y_s) = f(|s-t|)$, and in this case we define the ***cross-correlation function*** between X_t and Y_t as the function

$$\gamma_{XY}(k) = \mathbb{E}[(X_t - \mu_X)(Y_{t+k} - \mu_Y)],$$

Marginals We have that $\rho_{XX}(k) = \rho_X(k)$.

Symmetry It's relevant the order in which we compute the cross-correlation function, since unlike the ACF the CCF is not symmetric around the origin,

$$\rho_{XY}(k) \neq \rho_{XY}(-k),$$

instead we have that

$$\rho_{XY}(k) \neq \rho_{YX}(-k).$$

However, we have a way of obtaining the direction of association between the time series by inspecting the graph of the ACF.

Example (AR(1) model)
Let $Y_t \sim \text{AR}(1)$, then we have $(1-\varphi B)Y_t = X_t$ and for time $t+k$ we can write

$$Y_{t+k} = \frac{1}{1-\varphi B} X_{t+k} = X_{t+k} + \varphi X_{t+k-1} + \varphi^2 X_{t+k-2} + \ldots,$$

therefore the cross-covariance function between X_t and Y_t are

$$\gamma_{XY}(k) = \mathbb{E}[X_t Y_{t+k}] = \begin{cases} \varphi^k \sigma_k^2 & \text{if } k \geq 0 \\ 0 & \text{if } k \leq 0 \end{cases}$$

In general the ARMA(p,q) model can be written as a transfer function model without the white noise term, and where X_t is a white noise.

3.1.1 CCF and TF

Let x_t and y_t be stationary series with $\mu_x = \mu_y = 0$, then the transfer function at time $t+k$ is

$$y_{t+k} = \nu_0 x_{t+k} + \nu_1 x_{t+k-1} + \nu_2 x_{t+k-2} + \ldots + \eta_{t+k},$$

therefore if we multiply both left and right by x_t and take expectations we have

$$\gamma_{xy}(k) = \nu_0 \gamma_x(k) + \nu_1 \gamma_x(k-1) + \nu_2 \gamma_x(k-2) + \ldots,$$

hence the CCF is

$$\rho_{xy}(k) = \frac{\sigma_x}{\sigma_y} \left[\nu_0 \rho_x(k) + \nu_1 \rho_x(k-1) + \nu_2 \rho_x(k-2) + \ldots \right]. \tag{2}$$

The relationship between CCF and IRF ν_j given in Equation (2) is contaminated by the fact that they are not white noise.

However for a white noise model we would see $\rho_x(k) = 0$ for all $k \neq 0$ and

$$\gamma_{xy}(k) = \nu_k \sigma_k^2,$$

hence we can estimate the covariance function and obtain an impulse response function which is directly proportional to the CCF,

$$\rho_{xy}(k) = \frac{\sigma_x}{\sigma_y} \nu_k \implies \nu_k = \frac{\sigma_y}{\sigma_x} \rho_{xy}(k).$$

In the general TF model,

$$y_t = \nu(B) x_t + \eta_t,$$

if we assume $x_t \sim \text{ARMA}(p, q)$ we can calculate the ***pre-whitened input series***

$$\alpha_t = \frac{\varphi_x(B)}{\vartheta_x(B)} x_t,$$

and applying this transformation to both y_t and η_t we can obtain the ***filtered series***

$$\begin{cases} \beta_t = \frac{\varphi_x(B)}{\vartheta_x(B)} y_t \\ \varepsilon_t = \frac{\varphi_x(B)}{\vartheta_x(B)} \eta_t \end{cases}$$

Finally, the TF model becomes

$$\beta_t = \nu(B)\alpha_t + \varepsilon_t,$$

where the input series $\alpha_t \sim \text{WN}(0, \sigma^2)$ and we can estimate the transfer function using Equation (2) between β_t and α_t.
Given two stationary time series x_t and y_t, we want to identify a transfer function model of the form

$$y_t = \nu(B)x_t + \eta_t = \frac{\omega(B)}{\delta(B)} B^b x_t + \eta_t.$$

General procedure for a TF model

1. Identify an $\text{ARMA}(p, q)$ model for the input x_t,

$$\varphi_x(B)x_t = \vartheta_x(B)\alpha_t$$

2. Prewhiten $x_t \to \alpha_t = \frac{\varphi_x(B)}{\vartheta_x(B)} x_t$ and apply the same filter to $y_t \to \beta_t = \frac{\varphi_x(B)}{\vartheta_x(B)} y_t$.

3. Calculate the CCF between the whitened input series and the residuals of the model for y_t,

$$\widehat{\nu}_k = \frac{\widehat{\sigma}_\beta}{\widehat{\sigma}_\alpha} \ldots,$$

to get a preliminary estimation of the transfer function ν_k.

4. Identify the order b, r, s of the TF model by inspecting the estimated TF (or equivalently the CCF) and estimate the transfer function using the fact that

$$\widehat{\nu}_j = \frac{\widehat{\omega}(B)}{\widehat{\delta}(B)} B^b,$$

which is of course done by nonlinear least squares or other methods.

5. Identify a model for the estimated residuals $\widehat{\eta}_t$ given by

$$\widehat{\eta}_t = y_t - \widehat{\nu}(B)x_t.$$

6. Estimate the model and check goodness-of-fit, generally by checking both $\widehat{\varepsilon}_t$ and $\widehat{\alpha}_t$ are white noise. Moreover, since we assume that $\varepsilon_t \sim \text{WN}$ and $\eta_t \perp\!\!\!\perp x_t$. we need to check that $\widehat{\rho}_{\alpha,\widehat{\varepsilon}}(k)$ is non significant.

For checking the last step, there are test statistics which are based on Portmanteau tests.