Temporal Information Processing

Student

Máster en Ciencia de Datos

Universidad Autónoma de Madrid

tu-web.es

Contents

Part I.
Introduction

1. Time Series

A time series, also known as discrete time signal, is a sequence of observations taken periodically in time. We can use time series to perform many tasks such as predictions of future values, behaviour analysis or information extraction. Examples of time series are audio signals, industrial instrument measures or diary finantial activity.

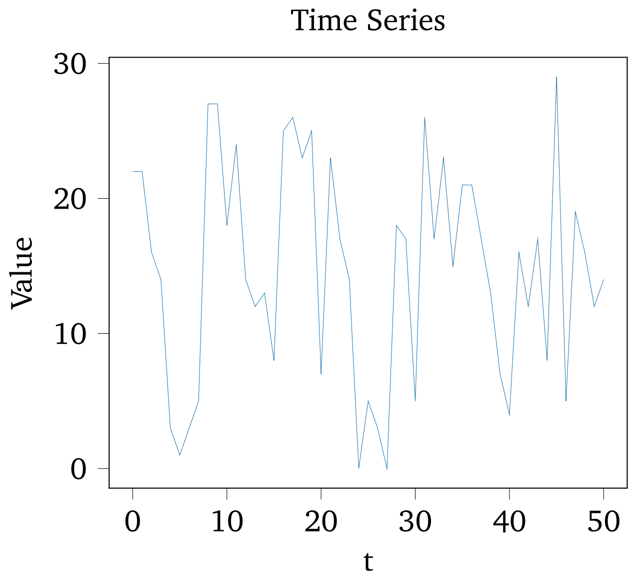


Figure 1: Example of a random time series.

A system can be determined comparing the input and the output. We call the system a filter if it is linear and time invariant. Considering the dynamic system as a black box, we can estimate the transference function or the impulse response to taht filter.

We can also consider **multivariate** time series, where some values of the time series have an influence on the other values in different or the same time instant. We can **classify** the time series in two wide types:

- Determinist: based in dynamic systems, they exploit the phisics of the generation algorithm of the time series.
- Stochastic: where the series are realizations of a stochastic process, which can be modelated.

In this subject, we will focus on stochastic models.

1.1. Stochastic Models

We can make three big considerations on the stochastic models.

- Stationary models.

Definition 1.1. Let $\{X_t\}$ be a stochastic process and let $F_X\left(x_{t_1+\tau},\ldots,x_{t_n+\tau}\right)$ represent the CDF of the **unconditional** joint distribution of $\{X_t\}$ at times $t_1+\tau,\ldots,t_n+\tau$. Then $\{X_t\}$ is strictly stationary if

$$F_X\left(x_{t_1+\tau},\ldots,x_{t_n+\tau}\right)=F_X\left(X_{t_1},\ldots,x_{t_n}\right)$$

However, we will use the case of **weak stationarity**, where we assume that the expectation of the stochastic process and the covariance at times $t, t+\tau$ are constant.

Example 1.1. AR, MA, ARMA

- Non stationary models, where we do not make the assumption that the average of the process is constant in time and that there is seasonality

Example 1.2. ARIMA, SARIMA

- Influenced by exogenous(extern) variables. In this cases, the exogenous variable affects the model, but the model does not affect this variable.

Example 1.3. SARIMAX

Let us introduce some **notation** for the following explanations

Definition 1.2. Let z_t be the value of the time series at instant t.

- The **backward shift** operator is $z_{t-m}=B^m z_t$
- The **forward shift** operator is $z_{t+m}=F^m z_t=B^{-m}z_t$
- The difference or discrete gradient operator is $\nabla z_t=z_t-z_{t-1}=(1-B)z_t$

Recall that, having a time series we can consider its **Z-transform**, that converts the discrete-time signal into a complex frequency-domain representation. In the Z-transform representation, the previously introduced notation is:

- The backward shift is $z_{t-m}=B^m z_t=Z^{-m}z_t$
- The forward shift is $z_{t+m}=B^{-m}z_t=Z^m z_t$
- The difference or discrete gradient is $\nabla z_t=(1-Z^{-1})z_t$

2. Linear filter based models

The stochastic models we use are based on time series z_t in which sucessive values are highly dependent. In these cases, we can see that the time series is generated from a series of independependent "shocks".

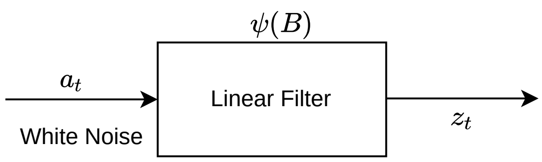
Definition 2.1. Let $a_t \sim \mathcal{N}\left(0, \sigma_a^2\right)$ be *white noise* (where each *shock* is related to a_t) which is not observed. Consider a linear filter that transforms the unobserved a_t to a observed time series z_t. We say that a **linear filter model is**

$$z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \mu + \psi(B) a_t,$$

where

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots$$

is called the **transfer function** of the filter.



As we can see, we are expressing the filter in terms of a infinite sum of the coefficients ψ_i. If there are finite coefficients of the sum is *absolutely summable*, that is: $\sum_{j=0}^{\infty} \left|\psi_j\right| < \infty$ or the vector of coefficients has finite ℓ^1 norm, we say that the filter is **stable** and the process z_t is **stationary**.

In the case where the ℓ^1 norm is not finite, our filter are non-stable and produce non-stationary series.

2.1. Autoregressive Models (AR)

Let us firstly consider the simplest case of linear filter. An **autoregressive model** is a linear filter where the current value of the process $\tilde{z}_t$ is expressed as a finite sum of the previous values and a random shock a_t.

Definition 2.2. Let us denote the values of a process af equally spaced times $t, t-1, \ldots$ by $z_t, z_{t-1}, \ldots$. Consider that the values are centered, that is $\tilde{z}_t = z_t - \mu$. Then, the **autoregressive (AR) process** of **order p** is

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} + a_t \tag{1}$$

Note that it is called autoregressive since, if you consider $\tilde{z}_{i-k}$ for $k = 1, \ldots, p$ as points, you are doing a *linear regression* over the past values.

Now, if we define the **autoregressive operator of order p** using the backward shift operator B as:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p,$$

we can economically write the autoregressive model in (1) as

$$\phi(B)\tilde{z}_t = a_t \tag{2}$$

In practice, this model has $p+2$ unknown parameters $\mu, \phi_1, \ldots, \phi_p, \sigma_a^2$ which have to be estimated from the data.

Proposition 2.1. The autoregressive model is a particula case of a linear filter

Proof. Although we will not be estrictly formal in this proof, we will give an intuition on the iterative process that has to be done.

Consider the term $\tilde{z}_{t-1}$, let us eliminate it. Recall that

$$\tilde{z}_{t-1} = \phi_1 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p-1} + a_{t-1}.$$

We can substitute this term in the expression of the AR model given in Equation (1). The same can be done for $\tilde{z}_{t-2}$ and so on, to yield eventually an infinite series in the a terms.

$\square$

In the case where $p = 1$, we have the AR process $\tilde{z}_t = \phi \tilde{z}_{t-1} + a_t$. After m sucessive substitutions of $\tilde{z}_{t-j} = \phi \tilde{z}_{t-j-1} + a_{t-j}$, with $j = 1, \ldots, m$, we obtain

$$\tilde{z}_t = \phi^{m+1} \tilde{z}_{t-m-1} + a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots + \phi^m a_{t-m}$$

Now, if we take the limit $m \to \infty$ this leads to the *convergent inifinite series representation* $\tilde{z}_t = \sum_{j=0}^{\infty} \phi^j a_{t-j}$, with $\psi_j = \phi^j, j \geq 1$, provided that $|\phi| < 1$. In the general AR case,

$$\phi(B)\tilde{z}_t = a_t$$

is equivalent to

$$\tilde{z}_t = \phi^{-1}(B) a_t = \psi(B) a_t, \qquad \psi(B) = \phi^{-1}(B) = \sum_{j=0}^{\infty} \psi_j B^j.$$

AR processes can be stationary or nonstationary. From the definition, it is clear that for a AR process to be stationary, the coefficients ϕ must be such that the weights $\psi_1, \psi_2, \ldots$ in $\psi(B) = \phi^{-1}(B)$ form a convergent series. A **necessary requirement** for stationarity is that the autoregressive operator $\phi(B)$, considered a polynomial in B of degree p, must have all roots greater than 1 in absolute value.

2.2. Application: Linear Prediction Coefficients in Speech Coding

Let us now set in the case of the **Speech Coding** topic. It is considered that a speech sample can be approximated as a linear combination of the past samples, which is how an AR model behaves. We have to find the coefficients that best suit our problem, using for instance the mean squared error prediction. We use the obtained **Linear Prediction Coefficients (LPCs)** to represent the signal frame.

Using this technique, we would be **reducing** the signal size significantly. However, since we are only approximating the signal, we would be most probably losing information. Two examples of codification of the audio signals are:

- MP3: which produces a different audio signal, involving loss of information
- FLAC: where the output is almost equal to the input, no loss of information

Signals are digitalized using a coding system.

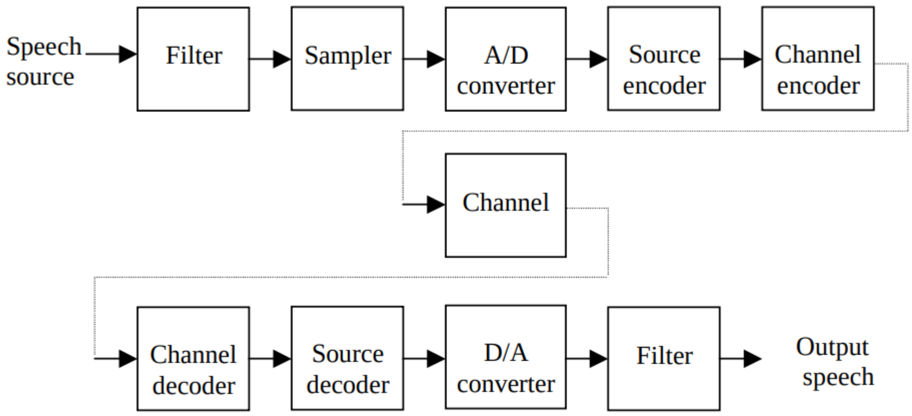


Figure 2: Block diagram of a speech coding system.

The filter eliminates aliasing and the sampler makes the continuous to discrete time conversion.

Example 2.1. In this example, we present the digital CD audio signal and why we would like to reduce its size without losing information. This signal has the following properties:

1. Sample rate $\Omega_s = 44.1kHz$
2. Bits per sample: 16
3. 2 channels (although sometimes 3 are used)

With this properties, the input bit rate is

$$R = \Omega_s \cdot \text{Bits/sample} \cdot \text{Channels} = 44.1 * 10^3 * 16 * 2 = 14112000 \frac{\text{bits}}{s} = 1.41 \frac{\text{Mb}}{s}$$

Which implies that, in a single minute we would need

$$60s\;*\;1,4112\frac{\mathrm{Mb}}{s}\;*\;\frac{1\,\mathrm{byte}}{8\,\mathrm{bits}} = 10.09\,\mathrm{MB},$$

which is a high size for a single minute audio.

Example 2.2. In this example, we will present the input bit rate for the speech digital signal. Its common properties are:

1. Sample rate $\Omega_s = 8\;\mathrm{kHz}$
2. Bits per sample: 16
3. 1 channel

With this properties, the input bit rate is 128 Kb per second.