Model Identification And Data Analysis I
Theory

Christian Rossi

Academic Year 2023-2024

Abstract

The course topics are:

- Basic concepts of stochastic processes.
- ARMA and ARMAX classes of parametric models for time series and for Input/Output systems.
- Parameter identification of ARMA and ARMAX models.
- Analysis of identification methods.
- Model validation and pre-processing.

Contents

CHAPTER 1

Introduction

1.1 Modeling

Definition (*System*)**.** A system denoted by S refers to a physical entity designed to convert inputs (causes) into outputs (effects).

Definition (*Model*)**.** A model, symbolized as M, constitutes a mathematical description of a system.

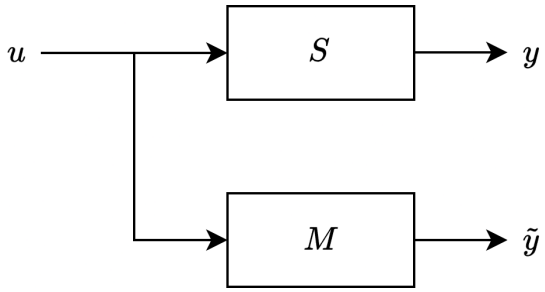


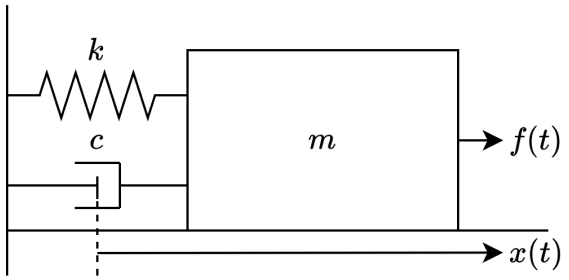
Figure 1.1: Visual representation of system and model

A model can be constructed through various methodologies:

1. *White-box modeling*: this approach relies on established physical laws or existing knowledge. The resultant model is typically generalizable, with clear physical interpretations for each variable. However, precise knowledge of all parameters beforehand is necessary, making it a costly and time-intensive process. Consequently, it's often impractical for complex systems.

2. *Black-box* modeling: this method is based on experimental data. Parameters of the model are estimated using statistical relationships derived from the data. It's feasible even without in-depth knowledge of the underlying processes, and it's comparatively faster and less expensive. However, models generated through this method lack physical interpretability and may not be universally applicable; changes in the system often necessitate repeating the experiment.

Example:
Consider a block with mass m and a spring with an elastic constant k.



In white-box modeling, precise values of parameters such as m, k, and c are required. With this information, the following model can be utilized:

$$m\ddot{x}(t) = f(t) - c\dot{x}(t) - kx(t)$$

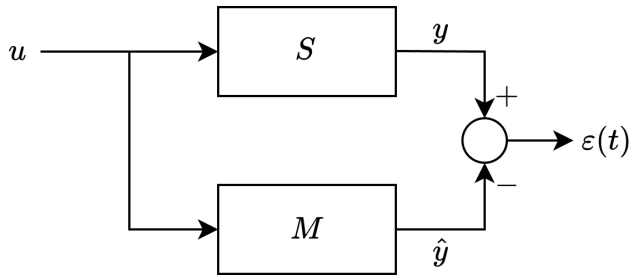
On the other hand, black-box modeling necessitates understanding the relationship between inputs and outputs to derive the model. In this scenario, the model obtained is:

$$x(t) = -a_1x(t-1) - a_2x(t-2) + b_0f(t) + b_1f(t-1) + b_2f(t-2)$$

Here, the parameters are determined from the output-input relationships.

1.1.1 Error

The modeling error, also known as the residual, is calculated as the disparity between the system output and the model output generated with the same input.



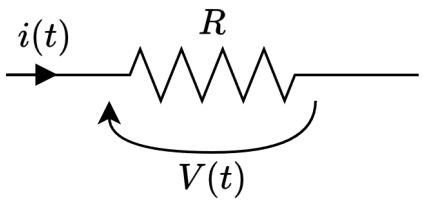
When the outputs exhibit similarity based on certain metrics, it signifies that the model accurately mirrors the dynamics of the system. However, if patterns persist within the error graph, it indicates that not all information has been effectively extracted from the data. Conversely, if the error graph lacks of patterns, it is termed as white noise, suggesting an inability to extract further meaningful information from the data. Consequently, a model is deemed complete only when the error demonstrates a completely unpredictable pattern.

1.1.2 Classification

Static and dynamic A system can be categorized as follows:

- *Static system*: in this type of system, knowledge of the input variables alone is adequate to determine the output value. Classical machine learning primarily addresses the black-box modeling of static systems.
- *Dynamic system*: this refers to a system with memory, wherein the past behavior of the output impacts its current value.

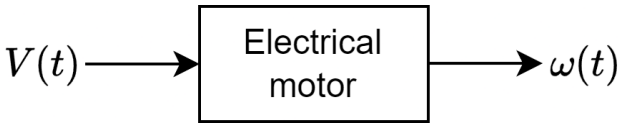
Example:
An illustration of a static system is represented by a circuit containing a resistor.



This system solely relies on the voltage across the resistor at each moment, adhering to Ohm's law:

$$i(t) = \frac{V(t)}{R}$$

On the other hand, an example of a dynamic system is exemplified by an electrical motor, wherein under certain conditions, even if the input remains constant, the output persists in its evolution.



Discrete and continuous Systems can be further categorized based on their time description, which can be either discrete or continuous. Natural and physical phenomena are inherently continuous and are often mathematically described using ordinary differential equations. On the other hand, discrete systems are mathematically described using difference equations.

However, a computer can only handle a limited amount of data. However, computers have limitations in handling data, which necessitates the sampling of signals at discrete intervals with a sampling time T_s. This ensures that only a finite amount of data is stored at discrete time points $t \cdot T_s$, where $t = 1, \ldots, N$:

$$y(t) = y(t \cdot T_s)$$

1.2 Estimation problem

An estimation problem involves quantifying an unknown parameter through estimation. This parameter, denoted as ϑ, can be discrete or continuous, scalar or vectorial, and constant or time-variant. We are provided with a set of observations, d, taken at various time points $t_1, t_2, \ldots, t_N$, formally defined as:

$$d = \{d(t), t \in T\}$$

Our goal is to derive an estimator to obtain an estimate of the unknown variable ϑ, expressed as:

$$\vartheta = f(d)$$

Definition (*Estimate*)**.** An estimate, denoted as $\hat{\vartheta}$, is a value produced by an estimator and depends on the input values of the estimator.

For a parameter ϑ with a constant value, we seek the estimate $\hat{\vartheta}$. For a parameter $\vartheta(t)$ with a dynamic value, we aim to find the estimate $\hat{\vartheta}(t|t_N)$ where the value of t_N is provided. The choice of t determines the nature of the estimation:

- *Prediction*: when $t > t_N$, indicating a time instant beyond t_N, we are forecasting a future event.
- *Filtering*: when $t = t_N$, we are estimating the noise in the estimator.
- *Regularization* or *interpolation* or *smoothing*: when $t < t_N$, representing a time instant before t_N, we are estimating variables that are not directly accessible.

CHAPTER 2

Stochastic processes

2.1 Prediction problem

To predict the value of $v(t)$ given a set of observations $\{v(1), v(2), \ldots, v(t-1)\}$, we can devise a predictor using the formula:

$$\hat{v}(t|t-1) = f\left(v(t-1), v(t-2), \ldots, v(1)\right)$$

In this formulation, we impose certain constraints:

- The function f is linear.
- Older data have diminishing importance compared to recent ones (finite memory predictor).
- The function remains invariant over time.

Under these simplifications, the predictor takes the form:

$$\hat{v}(t|t-1) = a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v(t-n)$$

Here, v is represented as a vector:

$$v = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix}^T$$

A reliable prediction is one that yields estimates closely aligned with the actual values. The accuracy of these estimates hinges on the parameters a_i in the v vector. Determining these parameter values equates to identifying the model that best characterizes the data distribution. This task translates into an optimization problem.

2.1.1 Model quality

Uncertainty represents a critical aspect of noise in prediction problems, yet its precise magnitude cannot be predetermined. The sole method to calculate an estimate of uncertainty is by comparing known values with those provided by the estimator at corresponding instants.

Example:
Consider the predictor given by:

$$\hat{v}(t|t-1) = a_1 v(t-1) + a_2 v(t-2) + a_3 v(t-3)$$

The values to be examined are:

- $\hat{v}(4|3) = v(1)v(2)v(3)$: to be compared with $v(4)$.
- $\hat{v}(5|4) = v(2)v(3)v(4)$: to be compared with $v(5)$.

After comparing all possible sequences, we can generate a sequence of residuals using the formula:

$$\varepsilon(i) = v(i) - \hat{v}(i|i-1) \qquad i = n+1, \ldots, N$$

From this sequence, we seek to find v by minimizing the following function:

$$\mathcal{J}(v) = \sum_{n+1}^{N} \varepsilon(i)^2$$

It's worth noting that the error is squared to ensure it is always considered as positive. A predictor is considered effective if the remaining error exhibits no discernible pattern, indicating that any remaining error is attributable solely to white noise.

Definition (*White noise*)**.** White noise refers to an error characterized by its lack of correlation between values at different points in time.

Consequently, if the residual is white noise, it signifies that there's no meaningful information within it that can enhance predictions.

Finally, to derive the accurate value from an estimate, we must incorporate the residual into the estimate. Thus, the previous formula transforms to:

$$\hat{v}(t) = a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v(t-n) + \varepsilon(t)$$

This implies that addressing the prediction problem involves examining a stochastic system.

2.1.2 Zeta transform

The same system can be reformulated using the Z-transform defined as:

$$V(z) = \mathcal{Z}\left[v(t)\right]$$

When considering the same system at a time $t-1$ in the Z-transform formulation, we have:

$$\underbrace{z^{-1}}_{\text{unity delay operator}} \cdot V(z) = [v(t-1)]$$

By incorporating the system described above with the time-domain equation:

$$\hat{v}(t) = a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v(t-n) + \varepsilon(t)$$

It's feasible to rewrite the same model in the frequency domain with the Z-transform, resulting in:

$$V(z) = \mathcal{Z}\left[a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v(t-n) + \varepsilon(t)\right]$$

This simplifies to:

$$V(z) = a_1 z^{-1} V(z) + a_2 z^{-2} V(z) + \cdots + a_n z^{-n} V(z) + \xi(z)$$

Note that this formula can also be expressed using operatorial notation:

$$V(z) = a_1 z^{-1} v(t-1) + a_2 z^{-2} v(t-2) + \cdots + a_n z^{-n} v(t-n) + \xi(t)$$

Rearranging terms such that all elements multiplied by $V(z)$ are on the left side and the residual $\xi(z)$ is on the right side, we obtain:

$$V(z)\left(1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_n z^{-n}\right) = \xi(z)$$

Finally, from this expression, we can derive the transfer function as the ratio of $V(z)$ to the residual $\xi(z)$:

$$\frac{V(z)}{\xi(z)} = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_n z^{-n}}$$

2.1.3 Summary

The components of an identification problem consist of:

- A system $\mathcal{S}$ requiring modeling.
- A model $\mathcal{M}$ to be ascertained, describing the system.
- An identification algorithm $\mathcal{J}$ governing data processing.
- An identification experiment $\mathcal{E}$ providing the data.

From these elements, it's crucial to emphasize that the model cannot convey more information than what is inherent in the data.

2.2 Stochastic processes

Definition (*Stochastic process*)**.** A stochastic process is an infinite sequence of random variables that is contingent upon the outcomes of a random experiment.

Thus, a stochastic process can be represented by the function:

$$v(t) = \varphi(s, t)$$

Here, s denotes the outcome of a random experiment, and t represents the corresponding time.

Definition (*Realization*)**.** A specific instance $v(t) = \varphi(\bar{s}, t)$ is termed a realization of the process.

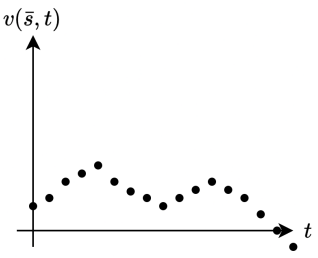


Figure 2.1: Visual representation fixed outcome

If instead of fixing the outcome of the experiment, we constrain the value of time, multiple outcomes can occur simultaneously at that time instant.

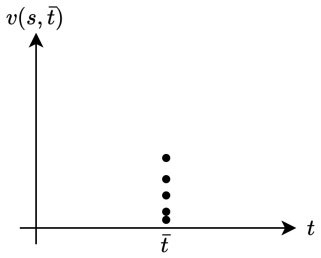


Figure 2.2: Visual representation fixed time

For any positive integer n and any n-tuple of instants $t_1, t_2, \ldots, t_n$, we can characterize $v(t_1), v(t_2), \ldots, v(t_n)$ as follows:

$$F_{t_1,t_2,\ldots,t_n}(q_1, q_2, \ldots, q_n) = \mathrm{P}\left(v(t_1) < q_1, v(t_2) < q_2, \ldots, v(t_n) < q_n\right)$$

However, specifying all the q_i values in practice is often too complex, rendering this method impractical.

2.2.1 Covariance

Consider various realizations $s_1, s_2, \ldots, s_n$ stemming from a specific experiment. The expected value signifies the average across all potential realizations at each time instance $\bar{t}$. Extending this concept across infinite time instants yields the expected value $\mu(t)$:

$$\mu(t) = \mathbb{E}\left[v(t)\right]$$

This framework enables a comparison of μ values across distinct times t_1 and t_2. Such comparisons prove beneficial in predictive scenarios since leveraging historical data enhances accuracy. Comparison across different time instances can be facilitated using the auto-covariance function:

$$\gamma(t_1, t_2) = \mathbb{E}\left[(v(t_1) - \mu(t_1)) \cdot (v(t_2) - \mu(t_2))\right]$$

Here, the terms $v(t_1) - \mu(t_1)$ and $v(t_2) - \mu(t_2)$ represent deviations from the respective mean values $\mu(t_1)$ and $\mu(t_2)$. A large variance suggests significant disparity among realizations. When the two time instances coincide, i.e., $t = t_1 = t_2$, the equation simplifies to:

$$\gamma(t, t) = \mathbb{E}\left[(v(t) - \mu(t))^2\right]$$

This expression corresponds to the variance of the variable $v(t)$ with itself.

It's worth noting that variance is always non-negative, and its square root is referred to as the standard deviation σ:

$$\sigma(t) = \sqrt{\mu(t)}$$

Additionally, the following relation holds:

$$\left|\mathrm{Cov}\left[v(t_1), v(t_2)\right]\right| \leq \sqrt{\mathrm{Var}\left[v(t_1)\right]} \cdot \sqrt{\mathrm{Var}\left[v(t_2)\right]}$$

Proof. Let's consider the realization of the function $v(t)$ at two time instants t_1 and t_2 as a vector $v(t) = \begin{bmatrix} v(t_1) & v(t_2) \end{bmatrix}^T$. The variance of this vector is defined through a vector-vector product as follows:

$$\operatorname{Var}[v] = \mathbb{E}\left[\begin{bmatrix} (v(t_1) - \mu(t_1)) \\ (v(t_2) - \mu(t_2)) \end{bmatrix} \begin{bmatrix} (v(t_1) - \mu(t_1)) & (v(t_2) - \mu(t_2)) \end{bmatrix}\right]$$

Upon performing the multiplication, we obtain:

$$\operatorname{Var}[v] = \begin{bmatrix} \mathbb{E}\left[(v(t_1) - \mu(t_1))^2\right] & \mathbb{E}\left[(v(t_1) - \mu(t_1))\,(v(t_2) - \mu(t_2))\right] \\ \mathbb{E}\left[(v(t_1) - \mu(t_1))\,(v(t_2) - \mu(t_2))\right] & \mathbb{E}\left[(v(t_2) - \mu(t_2))^2\right] \end{bmatrix}$$

In other words:

$$\operatorname{Var}[v] = \begin{bmatrix} \operatorname{Var}(v(t_1)) & \operatorname{Cov}[v(t_1), v(t_2)] \\ \operatorname{Cov}[v(t_1), v(t_2)] & \operatorname{Var}(v(t_2)) \end{bmatrix}$$

The determinant of this matrix must be non-negative, implying:

$$\operatorname{Var}(v(t_1)) \cdot \operatorname{Var}(v(t_2)) - \operatorname{Cov}[v(t_1), v(t_2)]^2 \geq 0$$

From this inequality, we can derive the initial formula. □

The covariance is a valuable tool for comparing two distinct processes, leading to the concept of cross-covariance.

Normalized covariance function The normalized covariance function, also known as the Pearson correlation coefficient, is defined as:

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\gamma(t_1, t_1)} \cdot \sqrt{\gamma(t_2, t_2)}}$$

The absolute value of this function is always less than or equal to one:

$$|\rho(t_1, t_2)| \leq 1$$

Moreover, when considering the same time instant, i.e., $t = t_1 = t_2$, the normalized covariance function equals one.

2.2.2 Correlation

We can also compare values at different time instants using the auto-correlation function:

$$\tilde{\gamma}(t_1, t_2) = \operatorname{Corr}[v(t_1), v(t_2)] = \mathbb{E}[v(t_1) \cdot v(t_2)]$$

The correlation bears a relation to the covariance:

$$\gamma(t_1, t_2) = \tilde{\gamma}(t_1, t_2) - \mu(t_1)\mu(t_2)$$

Proof. The formula can be derived as follows:

$$\begin{aligned} \gamma(t_1, t_2) &= \mathbb{E}\left[(v(t_1) - \mu(t_1))\,(v(t_2) - \mu(t_2))\right] \\ &= \mathbb{E}\left[v(t_1)v(t_2) - \mu(t_1)v(t_2) - v(t_1)\mu(t_2) + \mu(t_1)\mu(t_2)\right] \\ &= \mathbb{E}\left[v(t_1)v(t_2)\right] - \mu(t_1)\mathbb{E}\left[v(t_2)\right] - \mu(t_2)\mathbb{E}\left[v(t_1)\right]\mu(t_2) + \mu(t_1)\mu(t_2) \\ &= \tilde{\gamma}(t_1, t_2) - \mu(t_1)\mu(t_2) - \mu(t_1)\mu(t_2) + \mu(t_1)\mu(t_2) \\ &= \tilde{\gamma}(t_1, t_2) - \mu(t_1)\mu(t_2) \end{aligned}$$

In the second step, we removed the mean values $\mu(t)$ since they are not random variables. Also note that by definition, $\mathbb{E}[v(t_1)] = \mu(t_1)$ and $\mathbb{E}[v(t_2)] = \mu(t_2)$. □

The correlation can also be used to compare two different processes, resulting in a cross-correlation.

2.3 Stationary processes

Definition (*Strongly stationary process*)**.** A process is called strongly stationary if:

$$F_{t_1,t_2,\dots,t_n}(\dots) = F_{t_1+\tau,t_2+\tau,\dots,t_n+\tau}(\dots)$$

Definition (*Weakly stationary process*)**.** A process is called weakly stationary if both the mean and the covariance are stationary:

$$\begin{cases} \mu(t) = \mu(t+\tau) & \forall\tau \\ \gamma(t_1,t_2) = \gamma(t_1+\tau, t_2+\tau) & \forall\tau \end{cases}$$

The first condition ensures that the value of $\mu(t)$ remains constant, while the second condition ensures that $\gamma(\tau)$, where $\tau = t_2 - t_1$, remains constant.

Example:
Consider the experiment of flipping a coin. The outcome of the experiment can be either heads or tails. This can be viewed as the combination of two processes:

$$v_{head}(t) = \sin\left(\frac{2\pi t}{N}\right) \qquad v_{tail}(t) = -\sin\left(\frac{2\pi t}{N}\right)$$

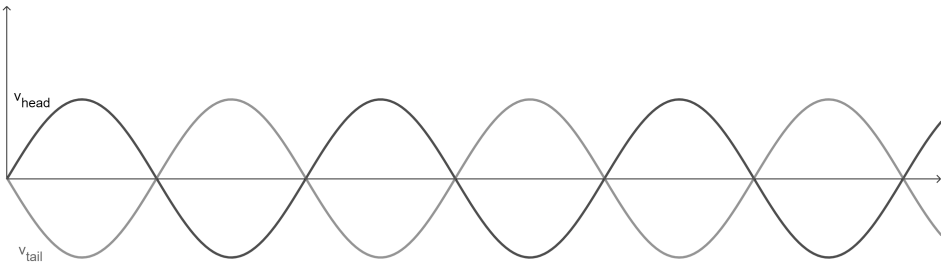


Figure 2.3: Graphs of the functions $v_{head}(t)$ and $v_{tail}(t)$

The expected mean $\mu(t)$ is zero for each time instant, making it constant. However, the variance is zero only when both functions intersect the t-axis, and it is greater than zero at all other instances. Consequently, the variance is not constant, indicating that the process is not stationary.

Example:
Consider the process $v(t) = \bar{v}$, where $\bar{v} \sim G(1, 3)$.

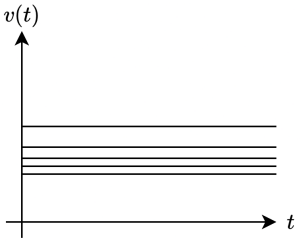


Figure 2.4: Some possible realizations

We have that:

$$\mu(t) = \mathbb{E}[v(t)] = \mathbb{E}[\bar{v}] = 1$$

Thus, the expected mean value is constant.

For the covariance, we have:

$$\begin{aligned}\gamma(t, t+\tau) &= \mathbb{E}[(v(t) - 1)(v(t+\tau) - 1)] \\ &= \mathbb{E}[(\bar{v} - 1)(\bar{v} - 1)] \\ &= \mathbb{E}\left[(\bar{v} - 1)^2\right] \\ &= \mathrm{Var}[\bar{v}] \\ &= 3\end{aligned}$$

Furthermore, the γ function is constant, indicating that the process is weakly stationary.

Example:
Consider the process $v(t) = t\bar{v} - t$, where $\bar{v} \sim G(1, 3)$.

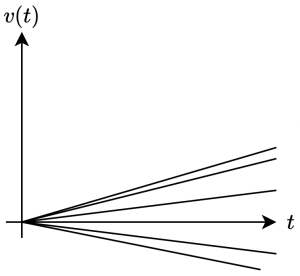


Figure 2.5: Some possible realizations

We have:

$$\mu(t) = \mathbb{E}[v(t)] = \mathbb{E}[t\bar{v} - t] = t\mathbb{E}[\bar{v}] - t = t \cdot 1 - t = 0$$

Thus, the expected mean value is constant.

For the covariance, we have:

$$\begin{aligned}\gamma(t, t+\tau) &= \mathbb{E}\left[(v(t)-1)\left(v(t+\tau)-1\right)\right] \\ &= \mathbb{E}\left[(t\bar{v}-t-0)\left((t+\tau)\,\bar{v}-(t+\tau)-0\right)\right] \\ &= \mathbb{E}\left[t\left(\bar{v}-1\right)(t+\tau)\left(\bar{v}-1\right)\right] \\ &= t\,(t+\tau)\,\mathbb{E}\left[\left(\bar{v}-1\right)^2\right] \\ &= 3t\,(t+\tau)\end{aligned}$$

Since the γ function is not constant, the process is not stationary.

2.3.1 Properties of weakly stationary stochastic processes

For a weakly stationary stochastic process, the following properties hold:

- $\mathbb{E}\left[v(t)\right] = \mu$
- $\gamma(\tau) = \mathbb{E}\left[(v(t)-\mu)(v(t+\tau)-\mu)\right]$
- $\tilde{\gamma} = \mathbb{E}\left[v(t)v(t+\tau)\right]$

These properties exhibit the following characteristics:

- $\gamma(0) = \mathbb{E}\left[(v(t)-\mu)^2\right] = \mathrm{Var}\left[v(t)\right]$
- $|\gamma(\tau)| \leq \gamma(0) \qquad \forall\tau$
- $\gamma(\tau) = \gamma(-\tau)$ (even function).
- The Toepliz matrix:

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(N-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(N-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \dots & \gamma(N-3) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(N-1) & \gamma(N-2) & \gamma(N-3) & \dots & \gamma(0) \end{bmatrix}$$

is a semi-definite matrix, requiring its determinant to be greater than or equal to zero.

2.3.2 Gaussian processes

A Gaussian process is defined by the property that its probability distribution function follows a joint Gaussian distribution:

$$F_{t_1,t_2,\dots,t_n}(\dots) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$

Property 2.3.1. If a Gaussian process is weakly stationary, it implies that it is also strongly stationary.

2.3.3 Ergodic processes

In an ergodic process, the statistical properties can be accurately derived from the analysis of a single realization, with a probability approaching one as the number of observations tends to infinity:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \cdot = \mathbb{E}\left[\cdot\right]$$

2.4 White noise

White noise is characterized as a sequence of independent identically distributed random variables, exhibiting the following properties:

- The probability distribution remains constant across all times, indicating stationarity.
- Independence among variables implies zero correlation at different times.

It is commonly defined with an expected value of zero:

$$\mathbb{E}\left[v(t)\right] = 0$$

And a covariance structured as:

$$\gamma(\tau) = \begin{cases} 0 & \text{for } \tau \neq 0 \\ \lambda^2 & \text{for } = 0 \end{cases}$$

This formulation indicates that white noise is entirely unpredictable. Typically, white noise is expressed as:

$$v(\cdot) \sim WN(0, \lambda^2)$$

Or as white noise with a Gaussian distribution:

$$v(\cdot) \sim WGN(0, \lambda^2)$$

Example:
Let's analyze the process $v(t) = \eta(t) + c\eta(t-1)$, where $\eta(\cdot) \sim WN(0, \lambda^2)$. First, let's calculate the expected mean value $\mu(t)$:

$$\mu(t) = \mathbb{E}\left[v(t)\right] = \mathbb{E}\left[v(t) = \eta(t) + c\eta(t-1)\right] = \underbrace{\mathbb{E}\left[\eta(t)\right]}_{0} + c\underbrace{\mathbb{E}\left[\eta(t-1)\right]}_{0} = 0$$

Next, we find the covariance $\gamma(0)$:

$$\begin{aligned}
\gamma(0) &= \mathrm{Var}\left[v(t)\right] && = \mathbb{E}\left[\left(v(t) - 0\right)^2\right] \\
&= \mathbb{E}\left[\left(\eta(t) + c\eta(t-1)\right)^2\right] \\
&= \mathbb{E}\left[\eta(t)^2 + 2c\eta(t)\eta(t-1) + c^2\eta(t-1)^2\right] \\
&= \mathbb{E}\left[\eta(t)^2\right] + \mathbb{E}\left[2c\eta(t)\eta(t-1)\right] + \mathbb{E}\left[c^2\eta(t-1)^2\right] \\
&= \underbrace{\mathbb{E}\left[\eta(t)^2\right]}_{\lambda^2} + 2c\underbrace{\mathbb{E}\left[\eta(t)\eta(t-1)\right]}_{0} + c^2\underbrace{\mathbb{E}\left[\eta(t-1)^2\right]}_{\lambda^2} \\
&= \left(1 + c^2\right)\lambda^2
\end{aligned}$$

It's important to note that since $\eta(\cdot)$ has a mean of zero, the expected mean between two different time instants is null. The $\gamma(0)$ function is constant. Now, let's consider other time instants:

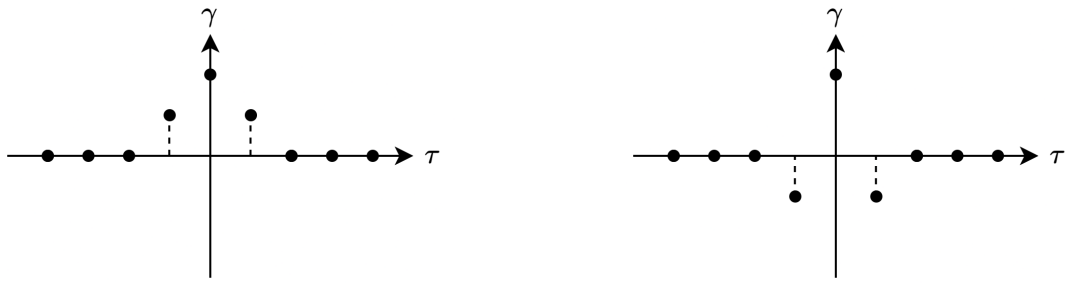
$$\begin{aligned}
\gamma(t,t+1) &= \mathbb{E}\left[(v(t)v(t+1))^2\right] \\
&= \mathbb{E}\left[(\eta(t)+c\eta(t-1))\,(\eta(t+1)+c\eta(t))\right] \\
&= \mathbb{E}\left[\eta(t)\eta(t+1)+c\eta(t-1)\eta(t+1)+c\eta(t)^2+c^2\eta(t-1)\eta(t)\right] \\
&= \underbrace{\mathbb{E}\left[\eta(t)\eta(t+1)\right]}_{0}+c\underbrace{\mathbb{E}\left[\eta(t-1)\eta(t+1)\right]}_{0}+c\underbrace{\mathbb{E}\left[\eta(t)^2\right]}_{\lambda^2}+c^2\underbrace{\mathbb{E}\left[\eta(t-1)\eta(t)\right]}_{0} \\
&= c\lambda^2
\end{aligned}$$

$$\begin{aligned}
\gamma(t,t+2) &= \mathbb{E}\left[(v(t)v(t+2))^2\right] \\
&= \mathbb{E}\left[(\eta(t)+c\eta(t-1))\,(\eta(t+2)+c\eta(t+1))\right] \\
&= \mathbb{E}\left[\eta(t)\eta(t+2)+c\eta(t-1)\eta(t+2)+c\eta(t)\eta(t+1)+c^2\eta(t-1)\eta(t+1)\right] \\
&= \underbrace{\mathbb{E}\left[\eta(t)\eta(t+2)\right]}_{0}+c\underbrace{\mathbb{E}\left[\eta(t-1)\eta(t+2)\right]}_{0}+c\underbrace{\mathbb{E}\left[\eta(t)\eta(t+1)\right]}_{0}+c^2\underbrace{\mathbb{E}\left[\eta(t-1)\eta(t+1)\right]}_{0} \\
&= 0
\end{aligned}$$

We find that both $\gamma(t,t+1)$ and $\gamma(t,t+2)$ are constants. Also, all time instants after $t+2$ will have a covariance with t equal to zero. Thus, in formula, we have:

$$\begin{cases}
\gamma(0) = (1+c^2)\lambda^2 \\
\gamma(1) = c\lambda \\
\gamma(\tau) = 0 \qquad \tau \geq 2
\end{cases}$$

Graphically, we can illustrate two cases based on the sign of the constant c:



(a) Positive constant c (b) Negative constant c

Figure 2.6: Possible values for the covariance

2.5 Dynamic representation

Let's consider the constant process $v(t,s) = v(s)$. In this scenario, once we determine the value of the experiment $v(s)$, we can derive the value of $v(t,s)$ for all time instants since its value remains unchanged. This process is perfectly predictable.

In contrast, white noise is the opposite: it is a completely unpredictable process.

Other processes lie between these two extremes. Consequently, we can decompose a process into two parts: one purely deterministic and one purely non-deterministic:

$$v(t) = \tilde{v}(t) + \hat{v}(t)$$

For $\tilde{v}(t)$, the knowledge of the past is sufficient to predict a value at a certain instant. Note that $\tilde{v}(t)$ and $\hat{v}(t)$ are independent:

$$\mathbb{E}\left[\tilde{v}(t_1)\hat{v}(t_2)\right] = \mathbb{E}\left[\tilde{v}(t_1)\right]\underbrace{\mathbb{E}\left[\hat{v}(t_2)\right]}_{0} = 0$$

The correlation is given by:

$$\begin{aligned}\tilde{\gamma}_v(\tau) &= \mathbb{E}\left[v(t)v(t+\tau)\right] \\ &= \mathbb{E}\left[\left(\tilde{v}(t)+\hat{v}(t)\right)\left(\tilde{v}(t+\tau)+\hat{v}(t+\tau)\right)\right] \\ &= \mathbb{E}\left[\tilde{v}(t)+\tilde{v}(t+\tau)\right]\mathbb{E}\left[\hat{v}(t)+\hat{v}(t+\tau)\right] \\ &= \tilde{\gamma}_{\tilde{v}}(\tau) + \tilde{\gamma}_{\hat{v}}(\tau)\end{aligned}$$

The purely nondeterministic component can be expressed as:

$$\hat{v}(t) = \sum_{i=-\infty}^{t} W(t,1)\eta(i) = \sum_{i=-\infty}^{t} W(t-1)\eta(i) = \sum_{k=0}^{\infty} W(k)\eta(t-k)$$

Here, $\eta(\cdot) \sim WN(0, \lambda^2)$. This formula can be expanded as:

$$\begin{aligned}\hat{v}(t) &= W(0)\eta(t) + W(1)\eta(t-1) + W(2)\eta(t-2) + \dots \\ &= W(0)\eta(t) + W(1)z^{-1}\eta(t) + W(2)z^{-2}\eta(t) + \dots \\ &= \underbrace{\left(W(0) + W(1)z^{-1} + W(2)z^{-2} + \dots\right)}_{\text{transfer function } W(z)}\eta(t)\end{aligned}$$

The stability of this function $W(z)$ implies the stationarity of the process $\hat{v}(t)$.

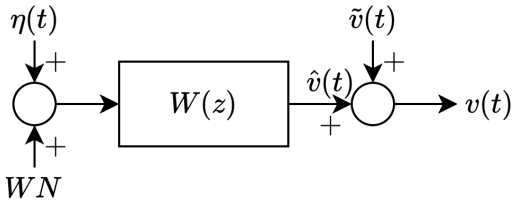


Figure 2.7: Stochastic process composition

2.6 Purely deterministic processes

A purely deterministic process lacks any white noise component and can be expressed as:

$$\tilde{v}(t) = a_1\tilde{v}(t-1) + a_2\tilde{v}(t-2) + \dots + a_n\tilde{v}(t-n)$$

In operator notation, this becomes:

$$\begin{aligned}\tilde{v}(t) &= a_1 z^{-1}\tilde{v}(t) + a_2 z^{-2}\tilde{v}(t) + \cdots + a_n z^{-n}\tilde{v}(t)\\ \tilde{v}(t) &= \underbrace{\left(a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}\right)}_{P(z)}\tilde{v}(t)\\ 0 &= P(z)\tilde{v}(t)\end{aligned}$$

This implies that $P(z)$ acts as a filter for $\tilde{v}(t)$. Consequently, we can predict all future values exactly. The only restriction is that if such filtering is possible, then the function $\tilde{v}(t)$ has the following form:

$$\tilde{v}(t) = \alpha_1\lambda_1^t + \alpha_2\lambda_2^t + \cdots + \alpha_n\lambda_n^t$$

Here, λ_i are the zeros of $P(z)$, with $|\lambda_i| < 1$ for all i.

Example:
Consider a constant process $v(t) = v(t-1)$, which can be rewritten as $v(t) = \alpha 1^t$. To extract the $P(z)$ function, we proceed as follows:

$$\begin{aligned}v(t) &= z^{-1}v(t)\\ 0 &= v(t) - z^{-1}v(t)\\ 0 &= \left(1 - z^{-1}\right)v(t)\end{aligned}$$

Hence, $P(z) = (1 - z^{-1})$.

Example:
Consider a constant alternated process $v(t) = -v(t-1)$, which can be rewritten as $v(t) = \alpha(-1)^t$. To extract the $P(z)$ function, we proceed as follows:

$$\begin{aligned}v(t) &= -z^{-1}v(t)\\ 0 &= z^{-1}v(t) + v(t)\\ 0 &= \left(1 + z^{-1}\right)v(t)\end{aligned}$$

Hence, $P(z) = (1 + z^{-1})$.

Example:
Consider a constant sinusoidal process $v(t) = A\cos(\omega_0 t)$. We can derive $P(z)$ as follows:

$$\begin{aligned}P(z) &= \left(z - e^{j\omega_0}\right)\left(z - e^{-j\omega_0}\right)\\ &= z^2 + 1 - z\left(e^{j\omega_0} + e^{-j\omega_0}\right)\\ &= z^2 + 1 - 2z\left(\frac{e^{j\omega_0} + e^{-j\omega_0}}{2}\right)\\ &= z^2 + 1 - 2\cos(\omega_0)z\end{aligned}$$

To verify if $P(z)v(t) = 0$, we perform a time shift of two:

$$\begin{aligned} P(z)v(t) &= \left(1 - 2\cos(\omega_0)z^{-2}z^{-2}\right) v(t) \\ &= A\cos(\omega_0 t) - 2\cos(\omega_0)A\cos(\omega_0(t-1)) + A\cos(\omega_0(t-2)) \\ &= A\cos(\omega_0 t) - 2A\left[\frac{1}{2}\cos(\omega_0 t) + \cos(\omega_0(t-2))\right] + A\cos(\omega_0(t-2)) \\ &= A\cos(\omega_0 t) - A\cos(\omega_0 t) - A\cos(\omega_0(t-2)) + A\cos(\omega_0(t-2)) \\ &= 0 \end{aligned}$$

Thus, the process can also be written as a sum of exponential:

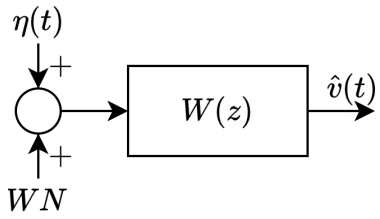
$$v(t) = \alpha_1\lambda_1^t + \alpha_2\lambda_2^t$$

Here, λ_1 and λ_2 needs to be replaced with the zeros of $P(z)$:

$$v(t) = \frac{A}{2}e^{j\omega_0 t} + \frac{A}{2}e^{-j\omega_0 t}$$

2.7 Purely non-deterministic processes

A significant category of stochastic processes involves stationary processes generated by passing white noise through a dynamical system governed by an asymptotically stable rational transfer function.



Let's visualize this with a linear time-invariant system featuring an input signal $u(\cdot)$ and an output signal $y(\cdot)$, characterized by the transfer function $W(z)$. According to the superposition principle, the output comprises two components: the free motion $y_{free}(\cdot)$ and the forced motion $y_{forced}(\cdot)$:

$$y(t) = y_{free}(t) + y_{forced}(t)$$

Here, the forced motion is expressed as:

$$y_{forced}(t) = \sum_{j=t_0}^{t} w(t-j)u(j)$$

In this equation, $w(i)$ represents the impulse response samples of the system. In the context of an asymptotically stable system, the free motion tends towards zero over time:

$$\lim_{t_0 \to -\infty} y(t) = y_{forced}(t) = \sum_{-\infty}^{t} w(t-j)u(j)$$

his implies that regardless of the initial conditions, as t_0 approaches negative infinity, the output converges to:

$$y(t) = \sum_{i=0}^{\infty} w(i)\eta(t-i)$$

2.7.1 MA process

We define the moving average process MA(n) of order n as the stochastic process represented by the equation:

$$v(t) = c_0\eta(t) + c_1\eta(t-1) + \cdots + c_n\eta(t-n)$$

Here, $\eta \sim WN(0, \lambda^2)$, and $c_0, c_1, \ldots, c_n$ are real numbers. In simpler terms, $v(t)$ is obtained by taking a weighted average of the current and past values of the white noise process η over a time window from t to $t-n$. As time progresses, this window shifts, hence the term moving average.

For this process, the mean value is calculated as:

$$\begin{aligned}\mathbb{E}\left[v(t)\right] &= \mathbb{E}\left[c_0\eta(t) + c_1\eta(t-1) + \cdots + c_n\eta(t-n)\right] \\ &= c_0\mathbb{E}\left[\eta(t)\right] + c_1\mathbb{E}\left[\eta(t-1)\right] + \cdots + c_n\mathbb{E}\left[\eta(t-n)\right] \\ &= 0\end{aligned}$$

And the variance is given by:

$$\begin{aligned}\mathrm{Var}\left[v(t)\right] &= \mathbb{E}\left[\left(v(t)-\mu(t)\right)^2\right] \\ &= \mathbb{E}\left[\left(c_0\eta(t) + c_1\eta(t-1) + \cdots + c_n\eta(t-n)\right)^2\right] \\ &= \mathbb{E}\left[\left(c_0\eta(t)\right)^2 + \left(c_1\eta(t-1)\right)^2 + \cdots + \left(c_n\eta(t-n)\right)^2 + \cdots + \underbrace{\text{cross terms}}_{0}\right] \\ &= c_0^2\mathbb{E}\left[\left(\eta(t)\right)^2\right] + c_1^2\mathbb{E}\left[\left(c_1\eta(t-1)\right)^2\right] + \cdots + c_n^2\mathbb{E}\left[\left(c_n\eta(t-n)\right)^2\right] \\ &= \left(c_0^2 + c_1^2 + \cdots + c_n^2\right)\lambda^2\end{aligned}$$

Both the mean value and the variance remain constant, indicating that the process is stationary.

As for the auto-covariance, we have:

$$\begin{aligned}\gamma\left[t, t+1\right] &= \mathbb{E}\left[v(t)v(t+1)\right] \\ &= \mathbb{E}\left[\left(c_0\eta(t) + \cdots + c_n\eta(t-n)\right)\left(c_0\eta(t+1) + \cdots + c_n\eta(t-n+1)\right)\right] \\ &= \mathbb{E}\left[c_0c_1\eta(t)^2\right] + \mathbb{E}\left[c_1c_2\eta(t-1)^2\right] + \mathbb{E}\left[c_{n-1}c_n\eta(t-n+1)^2\right] + \cdots + \underbrace{\text{cross terms}}_{0} \\ &= c_0c_1\mathbb{E}\left[\eta(t)^2\right] + c_1c_2\mathbb{E}\left[\eta(t-1)^2\right] + c_{n-1}c_n\mathbb{E}\left[\eta(t-n+1)^2\right] \\ &= \left(c_0c_1 + c_1c_2 + \cdots + c_{n-1}c_n\right)\lambda^2\end{aligned}$$

In the general case, we find:

$$\gamma(t, t+\tau) = \mathbb{E}\left[v(t)v(t+\tau)\right] = \begin{cases}\left(c_0c_\tau + c_1c_{\tau+1} + \cdots + c_{n-\tau}c_n\right)\lambda^2 & \tau \le n \\ 0 & \tau > n\end{cases}$$

The covariance depends solely on the difference between the two time indices. Since the mean value and the variance are constant, and the auto-covariance depends only on the time index difference, the MA process of any finite order n is weakly stationary. Moreover, if $\eta(\cdot)$ is a Gaussian process, then $v(\cdot)$ is also a Gaussian process since it arises from a linear combination of Gaussian variables.

Transfer function We can derive the transfer function as follows:

$$\begin{aligned} v(t) &= c_0\eta(t) + c_1\eta(t-1) + \cdots + c_n\eta(t-n) \\ &= c_0\eta(t) + c_1 z^{-1}\eta(t) + \cdots + c_n z^{-n}\eta(t) \\ &= \left(c_0 + c_1 z^{-1} + \cdots + c_n z^{-n}\right)\eta(t) \end{aligned}$$

This simplifies to:

$$W(z) = c_0 + c_1 z^{-1} + \cdots + c_n z^{-n} = \frac{c_0 z^n + c_1 z^{n-1} + \cdots + c_n}{z^n}$$

All poles of this function are at $z = 0$, indicating that the system is asymptotically stable.

The MA(n) process is characterized by $n + 2$ parameters, but this representation is redundant. To mitigate this redundancy, parameter c_0 is typically set to ($C(z)$ becomes a monic polynomial).

Infinite order When the value of n tends to infinity, the expected value remains zero, but the variance:

$$\mathrm{Var}\left[v(t)\right] = \left(\sum_{i=0}^{\infty} c_i^2\right)\lambda^2$$

may either converge or diverge. The variance is finite if and only if:

$$\sum_{i=0}^{\infty} c_i^2 < \infty$$

This condition not only ensures a finite variance but also guarantees that all elements of $\gamma(\tau)$ are finite, thus establishing the process as stationary.

The MA(∞) process possesses an auto-covariance function of infinite length, allowing it to model any stationary stochastic process. However, its practical utility is hindered by the challenge of handling an infinite number of parameters and the necessity for series calculations to derive $\gamma(\tau)$.

2.7.2 AR process

The autoregressive process AR(n) of order n is represented by the equation:

$$v(t) = a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v(t-n) + \eta(t)$$

Here, $\eta(cdot) \sim WN(0, \lambda^2)$.

Transfer function The transfer function can be derived as follows:

$$\begin{aligned} v(t) &= a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v + \eta(t) \implies \\ \eta(t) &= v(t) - a_1 v(t) z^{-1} - a_2 v(t) z^{-2} - \cdots - a_n v(t) z^{-n} \\ &= v(t)\left(1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_n z^{-n}\right) \end{aligned}$$

From this expression, the transfer function $W(z)$ is derived as:

$$W(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_n z^{-n}}$$

In its standard form, $W(z)$ is represented as:

$$W(z) = \frac{z^n}{z^n - a_1 z^{n-1} - a_2 z^{n-2} - \cdots - a_n}$$

Here, the denominator polynomial is denoted as $A(z)$.

Stability The AR(1) model is described by the equation:

$$v(t) = av(t-1) + \eta(t)$$

Consider a specific time point t_0 where the value of v is known as $v(t_0) = v_0$. Iterating the equation allows computing subsequent values:

- $t_0 \to v(t_0) = v_0$
- $t_0 + 1 \to v(t_0 + 1) = av(t_0) + \eta(t_0 + 1) = av_0 + \eta(t_0 + 1)$
- $t_0 + 2 \to v(t_0 + 2) = av(t_0 + 1) + \eta(t_0 + 2) = a^2 v_0 + a\eta(t_0 + 1) + \eta(t_0 + 2)$

In general, for $t > t_0$:

$$v(t) = \sum_{i=t_0}^{t-1} a^{t-1-i}\eta(i+1) + a^{t-t_0}v_0$$

This system has the transfer function $A(z) = za$. The only root is a thus the condition for asymptotic stability requires $|a| < 1$.

When t_0 tends to $-\infty$ and the system is asymptotically stable, $v(t)$ becomes an infinite linear combination of current and past noise values, forming an MA(∞) process. The expected value is trivially null. The variance of $v(t)$ is:

$$\begin{aligned}\operatorname{Var}[v(t)] &= \left((a^0)^2 + (a^1)^2 + (a^2)^2 + \dots\right)\lambda^2 \\ &= \left(1 + a^2 + a^4 + \dots\right)\lambda^2 \\ &= \frac{1}{1-a^2}\lambda^2\end{aligned}$$

Since $\sum_{j=0}^{\infty} a^{2j} = \dfrac{1}{1-a^2} < \infty$, the variance is finite, and $v(t)$ indeed represents a well-defined stationary MA(∞) process with $c_j = a^j$.

The auto-covariance function can be obtained using the expression for MA processes:

$$\gamma(\tau) = \lambda^2 \sum_{j=0}^{\infty} c_j c_{j+\tau} = \frac{\lambda^2 a^\tau}{1-a^2}$$

AR and MA equivalence Since the AR(1) process is equivalent to an MA(∞) process, its transfer function $W(z)$ can be represented as:

$$W(z) = \frac{z}{z-a} = w_0 + w_1 z^{-1} + w_2 z^{-2} + \dots$$

The coefficients w_i can be computed by dividing the numerator of $W(z)$ by its denominator, yielding:

$$W(z) = \frac{z}{z-a} = 1 + \frac{a}{z-a} = 1 + az^{-1} + \frac{a^2 z^{-1}}{z-a} = \dots$$

Higher order The insights gained from analyzing the AR(1) process extend to autoregressive processes of arbitrary order n. An AR(n) model that is asymptotically stable corresponds to an MA(∞) process, indicating stationarity. As t_0 tends to $-\infty$, the solution of the time domain equation converges asymptotically to a stationary process.

Let's consider the specific case of a second-order autoregressive process:

$$v(t) = a_1 v(t-1) + a_2 v(t-2) + \eta(t)$$

By utilizing Yule-Walker's equations, we can determine the auto covariance function γ of the process.

Multiplying both sides of the equation by $v(t-\tau)$ and taking expectations leads to:

$$\gamma(\tau) = a_1\gamma(\tau-1) + a_2\gamma(\tau-2) + \mathbb{E}\left[\eta(t)v(t-\tau)\right]$$

For the AR(2) process, this results in the following equations:

$$\begin{cases} \gamma(0) = a_1\gamma(1) + a_2\gamma(2) + \lambda^2 \\ \gamma(1) = a_1\gamma(0) + a_2\gamma(1) \\ \gamma(2) = a_1\gamma(1) + a_2\gamma(0) \end{cases}$$

Solving this system of equations allows determining $\gamma(0)$, $\gamma(1)$, $\gamma(2)$ given the parameters a_1, a_2, λ^2. Similar equations can be derived for autoregressive processes of higher orders, such as the generic AR(n) process.

2.7.3 ARMA process

The series combination of an autoregressive and a moving average model is described by the equation ARMA(n_a, n_c), given by:

$$v(t) = a_1 v(t-1) + a_2 v(t-2) + \cdots + a_{n_a} v(t-n_a) + c_0\eta(t) + c_1\eta(t-1) + \cdots + c_{n_c}\eta(t-n_c)$$

Here, $\eta(\cdot) \sim WN(0, \lambda^2)$. This can be visually represented as:

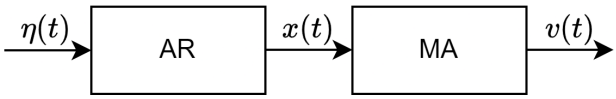


Figure 2.8: ARMA process representation

Considering the series connection, we have:

- Autoregressive process: $x(t) = a_1 x(t-1) + a_2 x(t-2) + \cdots + a_{n_a} x(t-n_a) + \eta(t)$. Rewritten as a transfer function:
$$A(z) = 1 - az^{-1} - \cdots - a_{n_a} z^{-n_a}$$
- Moving average process: $v(t) = c_0 x(t) + c_1 x(t-1) + \cdots + c_{n_c} x(t-n_c)$. Rewritten as a transfer function:
$$C(z) = c_0 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}$$

The transfer function of the two processes in series is given by:

$$v(t) = \frac{C(z)}{A(z)}\eta(t) = \frac{c_0 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}}{1 - az^{-1} - \cdots - a_{n_a} z^{-n_a}}\eta(t)$$

The stability of the process depends solely on the roots of $A(z)$. If the transfer function is asymptotically stable, then the process generated by this equation is stationary.

Example:
Consider an ARMA(1,1) process described by the equation $v(t) = av(t-1) + \eta(t) + c\eta(t-1)$. The corresponding transfer function is given by:

$$W(z) = \frac{1 + cz^{-1}}{1 - az^{-1}}$$

If $|a| < 1$, the solution of this equation converges to a stationary process. In this case, the expected value must be constant:

$$\mathbb{E}\left[v(t-1)\right] = \mu$$

Its value can be calculated as follows:

$$\begin{aligned}\mathbb{E}\left[v(t-1)\right] &= \mathbb{E}\left[av(t-1) + \eta(t) + c\eta(t-1)\right] \\ &= a\mathbb{E}\left[v(t-1)\right] + \mathbb{E}\left[\eta(t)\right] + c\mathbb{E}\left[\eta(t-1)\right]\end{aligned}$$

Substituting, we have:

$$\mu = a\mu + 0 + c \cdot 0 \rightarrow \mu = 0$$

For the variance, we compute:

$$\mathrm{Var}\left[v(t)\right] = \mathbb{E}\left[\left(av(t-1) + \eta(t) + c\eta(t-1)\right)^2\right]$$

By developing the squares we obtain:

$$\begin{aligned}\mathrm{Var}\left[v(t)\right] = \mathbb{E}\left[\left(av(t-1)\right)^2\right] + \mathbb{E}\left[\left(\eta(t)\right)^2\right] + \mathbb{E}\left[\left(c\eta(t-1)\right)^2\right] + 2a\mathbb{E}\left[v(t-1)\eta(t)\right] \\ + 2ac\mathbb{E}\left[v(t-1)\eta(t-1)\right] + 2c\mathbb{E}\left[\eta(t)\eta(t-1)\right]\end{aligned}$$

From which we obtain:

$$\begin{aligned}\mathrm{Var}\left[v(t)\right] = a^2 \underbrace{\mathbb{E}\left[v(t-1)^2\right]}_{\mathrm{Var}[v(t)]} + \underbrace{\mathbb{E}\left[\eta(t)^2\right]}_{\lambda^2} + c^2 \underbrace{\mathbb{E}\left[\eta(t-1)^2\right]}_{\lambda^2} + 2a \underbrace{\mathbb{E}\left[v(t-1)\eta(t)\right]}_{0} \\ + 2ac \underbrace{\mathbb{E}\left[v(t-1)\eta(t-1)\right]}_{\lambda^2} + 2c \underbrace{\mathbb{E}\left[\eta(t)\eta(t-1)\right]}_{0}\end{aligned}$$

Finally, we obtain:

$$\mathrm{Var}\left[v(t)\right] = \frac{1 + c^2 + 2ac}{1 - a^2}\lambda^2$$

Furthermore, we have:

$$\gamma(1) = \frac{a + ac^2 + a^2c + c}{1 - a^2}\lambda^2$$

The vanishing covariance property The AR, MA, and ARMA processes share the property that as the lag τ approaches infinity, the auto covariance function $\gamma(\tau)$ tends to zero:

$$\lim_{\tau \to \infty} \gamma(\tau) = 0$$

Here are the specific cases:

- For MA(n) processes, $\gamma(\tau) = 0$ for $\tau > n$.
- For an AR(1) process, it's known that $\gamma(\tau) = a^{\tau}\gamma(0)$, implying the vanishing property due to $|a| < 1$.
- For an AR(n) process, we can derive the Yule-Walker equation:

$$A(z)\gamma(\tau) = 0$$

  This equation suggests that $\gamma(\cdot)$ can be interpreted as the output of a system with transfer function $\frac{1}{A(z)}$ and null input. Under the stability condition, the output of such a system must tend to zero.

- For ARMA(n_a, n_c) processes, the Yule-Walker equation remains valid for sufficiently large values of τ ($\tau > n_c$), leading to the same consequence as the AR process.

2.7.4 ARMAX process

The models we've explored thus far are effective for describing time series data but lack the capability to represent phenomena influenced by external variables (exogenous variables).

An extension to these models is the ARMAX model, where we denote the process as $y(\cdot)$ to align with input-output system notation:

$$A(z)y(t) = B(z)u(t-k) + C(z)\eta(t)$$

Here, $u(t)$ represents the input variable, k is the input-output delay, and $B(z) = b_0 + b_1 z^{-1} + \cdots + b_{n_b} z^{n_b}$.

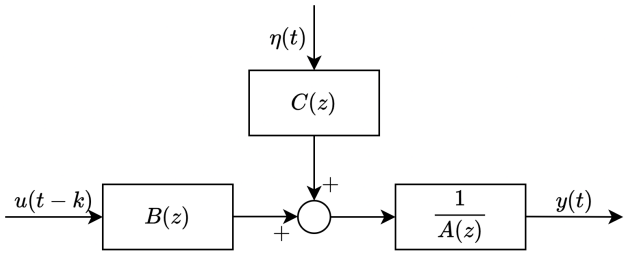


Figure 2.9: ARMAX process representation

The core concept is to model the process through an input-output relationship while accommodating model uncertainty through the influence of white noise input.

2.7.5 NARMAX process

A further extension leads us to the nonlinear domain with the NARMAX model:

$$y(t) = f(y(t-1), \ldots, y(t-n_a), u(t-k), \ldots, u(t-k-n_b), \eta(t), \ldots, \eta(t-n_c c))$$

Here, $f(\cdot)$ represents a nonlinear parametric function.

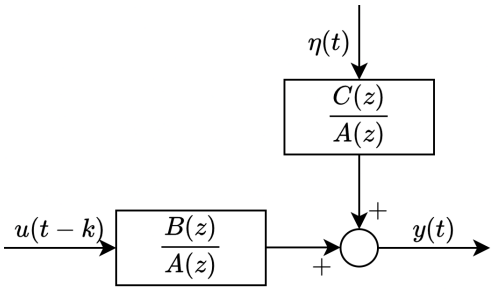


Figure 2.10: NARMAX process representation

CHAPTER 3

Frequency analysis

3.1 Power spectral density

The power spectral density (or spectrum) of a stationary stochastic process is defined as the Fourier transform of its correlation function:

$$\Gamma(\omega) = \mathcal{F}\left[\tilde{\gamma}(\tau)\right] = \sum_{\tau=-\infty}^{\infty} \tilde{\gamma}(\tau) e^{-j\omega\tau}$$

Here, ω represents the angular frequency measured in rad/s. The Fourier series sum exists for stationary processes, specifically those with a correlation function $\tilde{\gamma}(\tau)$ that rapidly tends to 0 as τ approaches infinity.

A sufficient condition for the existence of the Fourier transform is the absolute convergence of $\tilde{\gamma}(\tau)$:

$$\sum_{\tau=-\infty}^{\infty} |\tilde{\gamma}(\tau)| < \infty$$

The anti-transformation formula establishes a one-to-one correspondence between $\tilde{\gamma}(\tau)$ and $\Gamma(\omega)$:

$$\tilde{\gamma}(\tau) = \mathcal{F}^{-1}\left[\Gamma(\omega)\right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma(\omega) e^{j\omega\tau} d\omega$$

From the definition, we derive that:

$$\begin{aligned}
\Gamma(\omega) &= \cdots + \tilde{\gamma}(-2)e^{j2\omega} + \tilde{\gamma}(-1)e^{j\omega} + \tilde{\gamma}(0) + \tilde{\gamma}(1)e^{-j\omega} + \tilde{\gamma}(2)e^{-j2\omega} + \ldots \\
&= \tilde{\gamma}(0) + \tilde{\gamma}(1)(e^{j\omega} + e^{-j\omega}) + \tilde{\gamma}(2)(e^{j2\omega} + e^{-j2\omega}) \\
&= \tilde{\gamma}(0) + 2\tilde{\gamma}(1)\cos\omega + 2\tilde{\gamma}(2)\cos 2\omega
\end{aligned}$$

Therefore, the spectrum $\Gamma(\omega)$ is a non-negative, real, even, and periodic function with a period of 2π.

The maximum angular frequency for sinusoidal discrete-time signals is $\omega_{max} = \pi$. This is due to the minimum period of a discrete signal, $T = 2\left(f = \frac{1}{2}\right)$, corresponding to $\omega = \frac{2\pi}{2} = \pi = \omega_{max}$.

For processes with zero expected value, the variance of the process equals the area below the spectral density curve:

$$\gamma(0) = \tilde{\gamma}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma(\omega) d\omega$$

This can be directly derived from the anti-transformation formula with $\tau = 0$.

The area below the spectrum in a given angular frequency range represents the contribution to the overall variability of the process at those angular frequencies.

Example:
Consider a white noise with zero mean and variance equal to λ^2. Both the covariance and correlation functions are zero everywhere except for $\tau = 0$, where $\gamma(0) = \lambda^2$. Therefore, $\Gamma(\omega) = \lambda^2$ for all ω:

$$\tilde{\gamma}(\tau) = \gamma(\tau) = \begin{cases} 0 & \tau \neq 0 \\ \lambda^2 & \tau = 0 \end{cases}$$

In the frequency domain, this translates to:

$$\Gamma(\omega) = \tilde{\gamma}(0) e^{-j\omega 0} = \lambda^2$$

This indicates that the power spectral density of a white noise is constant across all frequencies. In other words, all frequencies contribute equally to the process variability, affirming the absolute unpredictability of a white noise.

The spectrum can be equivalently defined in two steps:

1. Take the (bilateral) Z-transform of the correlation function:

$$\Phi(z) = \sum_{\tau=-\infty}^{+\infty} \tilde{\gamma}(\tau) z^{-\tau}$$

2. Evaluate $\Phi(z)$ at $z = e^{j\omega}$:

$$\Gamma(\omega) = \Phi(e^{j\omega})$$

3.2 Stable systems

For stationary processes resulting from filtering white noise through a stable dynamical system, the spectrum can be calculated using the transfer function. Let's consider the process $y(\cdot)$ obtained by filtering an input process $u(\cdot)$ through an asymptotically stable dynamical system described by transfer function $W(z)$. The spectra of the two processes are related by the equation:

$$\Gamma_{yy}(\omega) = \left| W(e^{j\omega}) \right|^2 \Gamma_{uu}(\omega)$$

Given that the expected value of $u(\cdot)$ is zero, applying the convolution formula reveals that $y(\cdot)$ also has a null expectation. To compute the input-output cross-covariance function, we start with the convolution formula:

$$y(t_2) = \sum_{i=0}^{\infty} w(i) u(t_2 - i)$$

Multiplying both sides by $u(t_1)$ yields:

$$u(t_1)y(t_2) = u(t_1)\sum_{i=0}^{\infty} w(i)u(t_2-i) = \sum_{i=0}^{\infty} w(i)u(t_1)u(t_2-i)$$

Taking the expected value of both sides, we obtain:

$$\mathbb{E}\left[u(t_1)y(t_2)\right] = \mathbb{E}\left[u(t_1)\sum_{i=0}^{\infty} w(i)u(t_2-i)\right] = \sum_{i=0}^{\infty} w(i)\mathbb{E}\left[u(t_1)u(t_2-i)\right]$$

This leads to:

$$\gamma_{uy}(t_1,t_2) = \sum_{i=0}^{\infty} w(i)\gamma_{uu}(t_1,t_2-i)$$

Similarly, multiplying both sides of the convolution formula by $y(t_1)$ gives:

$$y(t_1)y(t_2) = y(t_1)\sum_{i=0}^{\infty} w(i)u(t_2-i) = \sum_{i=0}^{\infty} w(i)y(t_1)u(t_2-i)$$

Taking the expectation leads to:

$$\gamma_{yy}(t_1,t_2) = \sum_{i=0}^{\infty} w(i)\gamma_{yu}(t_1,t_2-i)$$

For stationary processes, these auto-covariance and cross-covariance functions depend only on the difference between the two temporal indices $\tau = t_2 - t_1$:

$$\begin{cases}\gamma_{uy}(\tau) = \sum_{i=0}^{\infty} w(i)\gamma_{uu}(\tau-i)\\ \gamma_{yy}(\tau) = \sum_{i=0}^{\infty} w(i)\gamma_{yu}(\tau-i)\end{cases}$$

Let's introduce the (bilateral) Z-transforms of these covariance functions:

$$\Phi_{uu}(z) = \sum_{-\infty}^{+\infty}\gamma_{uu}(\tau)z^{-\tau} \qquad \Phi_{yy}(z) = \sum_{-\infty}^{+\infty}\gamma_{yy}(\tau)z^{-\tau} \qquad \Phi_{uy}(z) = \sum_{-\infty}^{+\infty}\gamma_{uy}(\tau)z^{-\tau}$$

These transforms are useful because they allow us to compute the power spectral densities of the input and output, as well as the cross-spectrum, by simply applying $z = e^{j\omega}$:

$$\Gamma_{uu}(\omega) = \Phi_{uu}(e^{j\omega}) \qquad \Gamma_{yy}(\omega) = \Phi_{yy}(e^{j\omega}) \qquad \Gamma_{uy}(\omega) = \Phi_{uy}(e^{j\omega})$$

It's worth noting that while $\Gamma_{uu}(\omega) = \Phi_{uu}(e^{j\omega})$ and $\Gamma_{yy}(\omega) = \Phi_{yy}(e^{j\omega})$ are real functions, the input-output cross spectrum can be complex. Additionally, since $\gamma_{uy}(\tau) = \gamma_{yu}(-\tau)$, we have:

$$\Phi_{yu}(z) = \Phi(z^{-1}) \rightarrow \Gamma_{yu}(\omega) = \Gamma_{uy}(\omega)^*$$

Here, $*$ denotes the complex conjugate. By performing Z-transforms on the expressions for $\gamma_{uy}(\cdot)$ and $\gamma_{yy}(\cdot)$, and recalling the properties of Z-transforms, specifically:

- The Z-transform of the convolution product is equal to the product of the transforms of the two processes.

- The Z-transform of the impulse response $w(\cdot)$ is the transfer function of the system.

We arrive at:

$$\gamma_{uy}(\tau) = \sum_{i=0}^{\infty} w(i)\gamma_{uu}(\tau - i) \rightarrow \Phi_{uy}(z) = W(z)\Phi_{uu}(z)$$

$$\gamma_{yy}(\tau) = \sum_{i=0}^{\infty} w(i)\gamma_{yu}(\tau - i) \rightarrow \Phi_{yy}(z) = W(z)\Phi_{yu}(z)$$

Expanding these expressions yields:

$$\Phi_{yy}(z) = W(z)\Phi_{yu}(z) = W(z)\Phi_{uy}(z^{-1}) = W(z)W(z^{-1})\Phi_{uu}(z^{-1})$$

Finally, since $\Phi_{uu}(z^{-1}) = \Phi_{uu}(z)$, we also have:

$$\Phi_{yy}(z) = W(z)W(z^{-1})\Phi_{uu}(z)$$

To obtain the power spectral density, we simply substitute $z = e^{j\omega}$:

$$\Gamma_{yy}(\omega) = \left|W(e^{j\omega})\right|^2 \lambda^2$$

3.2.1 MA(1) spectrum

To compute the spectrum of an MA(1) process, we utilize the direct definition, leveraging the property that the auto-correlation function is zero for lags greater than 1:

$$\begin{aligned}
\Gamma(\omega) &= \sum_{\tau=-\infty}^{\infty} \gamma(\tilde{\tau})e^{-j\omega\tau} \\
&= \tilde{\gamma}(0) + \tilde{\gamma}(1)e^{-j\omega} + \tilde{\gamma}(-1)e^{j\omega} \\
&= \tilde{\gamma}(0) + 2\tilde{\gamma}(1)\frac{e^{j\omega} + e^{-j\omega}}{2} \\
&= \tilde{\gamma}(0) + 2\tilde{\gamma}(1)\cos(\omega) \\
&= \left(1 + c^2 + 2c\cos(\omega)\right)\lambda^2
\end{aligned}$$

Alternatively, we can employ the transfer function $W(z) = 1 + cz^{-1}$ to compute the complex spectrum:

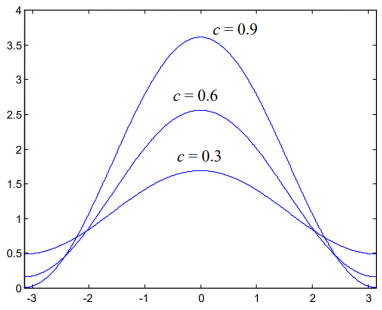
$$\Phi_{vv}(z) = W(z)W(z^{-1})\Phi_{\eta\eta}(z) = (1 + cz^{-1})(1 + cz)\lambda^2 = \left(1 + c^2 + 2c\cos(\omega)\right)\lambda^2$$

As a result, we have:

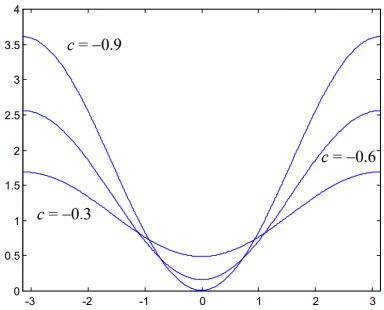
$$\Gamma(\omega) = \Phi_{vv}(e^{j\omega}) = \left(1 + c^2 + 2c\cos(\omega)\right)\lambda^2$$

The spectrum shape varies with the value of parameter c:

- $c > 0$: realizations tend to maintain the sign between consecutive times, with low frequencies dominating.
- $c < 0$: opposite behavior occurs, with high frequencies dominating.



(a) Positive



(b) Negative

Figure 3.1: Possible sign for the constant c

We can also deduce the process variance from the spectrum:

$$\begin{aligned}
\gamma(0) &= \tilde{\gamma}(0) \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma(\omega)d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(1 + c^2 + 2c\cos(\omega)\right)\lambda^2 d\omega \\
&= \left(1 + c^2\right)\lambda^2 \frac{1}{2\pi}\int_{-\pi}^{\pi} d\omega + 2c\lambda^2 \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(\omega)d\omega \\
&= \left(1 + c^2\right)\lambda^2 + 2c\lambda^2 \frac{1}{2\pi}\left[\sin(\omega)\right]_{-\pi}^{\pi} \\
&= \left(1 + c^2\right)\lambda^2
\end{aligned}$$

3.2.2 AR(1) spectrum

Now, let's examine an AR(1) process:

$$v(t) = av(t-1) + \eta(t)$$

Here, $\eta(\cdot)$ follows a white noise distribution $WN(0, \lambda^2)$ with $|a| < 1$. The transfer function of this process is:

$$W(z) = \frac{1}{1 - az^{-1}}$$

The power spectral density can be computed as follows:

$$\Phi_{vv}(z) = W(z)W(z^{-1})\Phi_{\eta\eta}(z) = \frac{1}{(1 - az^{-1})(1 - az)}\lambda^2 = \frac{1}{1 + a^2 - a(z + z^{-1})}\lambda^2$$

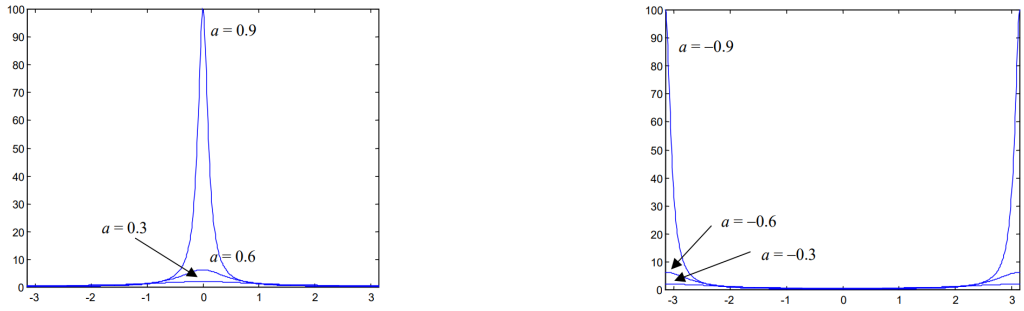
And the spectral density at frequency ω is given by:

$$\Gamma(\omega) = \Phi_{vv}(e^{j\omega}) = \frac{1}{1 + a^2 - a\cos(\omega)}\lambda^2$$

The spectrum exhibits a markedly different shape depending on the sign of a. Specifically:

- For $a > 0$, most of the frequency content concentrates at low frequencies.
- Conversely, for $a < 0$, the opposite trend occurs.

This difference is visually represented in the following figure.



(a) Positive (b) Negative

Figure 3.2: Possible sign for the constant a

3.2.3 ARMA(1,1) spectrum

Let's now examine the ARMA(1, 1) process with $|a| < 1$:

$$v(t) = av(t-1) + \eta(t) + c\eta(t-1)$$

Here, $\eta(\cdot)$ follows a white noise distribution $WN(0, \lambda^2)$. The transfer function is given by:

$$W(z) = \frac{1 + cz^{-1}}{1 - az^{-1}}$$

The power spectral density can then be computed as:

$$\Phi(z) = W(z)W(z^{-1})\lambda^2 = \frac{(1 + cz^{-1})(1 + cz)}{(1 - az^{-1})(1 - az)}\lambda^2 = \frac{1 + c^2 + c(z + z^{-1})}{1 + a^2 - a(z + z^{-1})}\lambda^2$$

And the spectral density at frequency ω is expressed as:

$$\Gamma(\omega) = \Phi(e^{j\omega}) = \frac{1 + c^2 + 2c\cos(\omega)}{1 + a^2 - 2a\cos(\omega)}\lambda^2$$

3.3 Purely deterministic processes

Wold's decomposition conceptualizes a stochastic process as the sum of a purely nondeterministic component and a purely deterministic one:

$$v(t) = \tilde{v}(t) + \hat{v}(t)$$

he purely deterministic part of the process $v(t)$ can be represented as:

$$v(t) = \alpha_1 v(t-1) + \alpha_2 v(t-2) + \cdots + \alpha_n v(t-n)$$

3.3.1 Constant process

Let's examine the constant process $v(t) = v(t-1) = v$ with a correlation function $\tilde{\gamma}(\tau) = v^2$. Consequently, the power spectral density is:

$$\Gamma(\omega) = v^2\delta(\omega)$$

3.3.2 Alternated process

Now, consider the alternating process $v(t) = -v(t-1) = v$ with a correlation function $\tilde{\gamma}(\tau) = (-1)^\tau v^2$. This leads to the power spectral density:

$$\Gamma(\omega) = v^2 \delta(\omega - \pi)$$

3.3.3 Sinusoidal process

Lastly, let's consider the sinusoidal process $v(t) = A\cos(\omega_0 t)$ with a correlation function $\tilde{\gamma}(\tau) = \frac{A^2}{2}\mathbb{E}\left[\cos(\omega_0 \tau)\right]$. This yields the power spectral density:

$$\Gamma(\omega) = \frac{A^2}{4}\left[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)\right]$$

3.4 Mixed processes

Let's now consider a mixed process represented as:

$$v(t) = \tilde{v}(t) + \hat{v}(t)$$

Here, $\tilde{v}(t)$ is a deterministic signal with spectrum $\gamma_{\tilde{v}}(\omega)$, while $\hat{v}(t)$ is a stationary stochastic process with zero expected value and spectrum $\gamma_{\hat{v}}(\omega)$. As we've observed earlier, since $\tilde{v}(t)$ and $\hat{v}(t)$ are independent, it follows that:

$$\tilde{\gamma}_v(\tau) = \tilde{\gamma}_{\tilde{v}}(\tau) + \tilde{\gamma}_{\hat{v}}(\tau)$$

This relationship also implies:

$$\Gamma_v(\omega) = \Gamma_{\tilde{v}}(\omega) + \Gamma_{\hat{v}}(\omega)$$

Example:
Now, let's consider an example with a white noise process $\eta(\cdot)$ having a non-zero expected value $\bar{\eta}$ and variance λ^2. Since the process $\eta(t) - \bar{\eta}$ is a standard white noise with zero expected value, its auto-covariance function is:

$$\gamma(\tau) \begin{cases} 0 & \tau \neq 0 \\ \lambda^2 \tau = 0 \end{cases}$$

From this, we can compute the auto-correlation function, recalling that:

$$\tilde{\gamma}(\tau) = \gamma(\tau) + \bar{\eta}^2$$

Hence, the power spectral density is:

$$\Gamma(\omega) = \lambda^2 + \bar{\eta}^2 \delta(\omega)$$

This result aligns with the interpretation of $\eta(t)$ as a mixed process with:

- A purely deterministic component equivalent to a constant process (which exhibits an impulsive spectrum).

- A purely nondeterministic component equivalent to a white noise with zero expected value (which has a constant spectrum).

Since these two components are independent, the spectrum of $\eta(t)$ is the sum of their spectra:

$$\Gamma_v(\omega) = \Gamma_{\tilde{v}}(\omega) + \Gamma_{\hat{v}}(\omega)$$

CHAPTER 4

Process characteristics estimation

4.1 Estimation of mean and covariance function

Let $v(t)$ represent a stationary stochastic process with unknown probabilistic properties, for which we have a finite realization $\{v(1), v(2), \ldots, v(n)\}$. Our objective is to estimate the probabilistic properties of $v(t)$ from this data:

- *Mean*: $\mu = \mathbb{E}[v(t)] \rightarrow \hat{\mu} = \hat{\mu}_N(v(1), v(2), \ldots, v(N))$
- *Autocovariance*: $\gamma(\tau) = \mathbb{E}[(v(t) - \mu)(v(t-\tau) - \mu)] \rightarrow \hat{\gamma}(\tau) = \hat{\gamma}_N(\tau; v(1), v(2), \ldots, v(n))$

The functions $\hat{\mu}$ and $\hat{\gamma}(\tau)$ are derived from the available data and are referred to as sample estimators. Their primary objective is to provide estimates that closely match the underlying probabilistic properties being estimated, regardless of the specific data sequence available.

4.1.1 Mean function

A natural sample estimator for the mean is:

$$\hat{\mu}_N = \frac{1}{N}\sum_{t=1}^{N} v(t)$$

It's important to note that $\hat{\mu}_N$ is a random variable because it depends on the realization of $v(t)$.

Definition (*Unbiased estimator*)**.** An estimator is unbiased if its expected value (taken over all possible finite realizations of $v(t)$) equals the property being estimated.

For our estimator to be unbiased, we require $\mathbb{E}[\hat{\mu}_N] = \mu$. In our case:

$$\mathbb{E}[\hat{\mu}_N] = \mathbb{E}\left[\frac{1}{N}\sum_{t=1}^{N} v(t)\right] = \frac{1}{N}\sum_{t=1}^{N}\mathbb{E}[v(t)] = \frac{1}{N}\sum_{t=1}^{N}\mu = \mu$$

Consider a fixed $s = \bar{s}$, where the process realization is constant over time:

- $\hat{\mu}_N = \frac{1}{N}\sum_{t=1}^{N} v(t, \bar{s}) = \frac{1}{N}\sum_{t=1}^{N} v(\bar{s}) = v(\bar{s})$

- $\mu = \mathbb{E}[v(t,s)] = \mathbb{E}[v(s)] = 0$

While $\hat{\mu}_N$ is unbiased, it may not provide a good estimate of the expected value of $v(t)$, even with infinite data.

We need another property to ensure a good estimate.

Definition (*Consistent estimator*)**.** An estimator is consistent if the variance of its estimation error tends to zero as the data size tends to infinity:

$$\mathbb{E}\left[(\hat{\mu}_N - \mu)^2\right] \overset{N\to\infty}{\longrightarrow} 0$$

This property doesn't hold for all stationary processes. A good estimator satisfies both correctness and consistency.

Example:
Consider the constant process $v(t,s) = v(s)$, where $v(s) \sim N(0,1)$:

$$\mathbb{E}\left[(\hat{\mu}_N - \mu)^2\right] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{t=1}^{N} v(s) - 0\right)^2\right] = \mathbb{E}\left[v(s)^2\right] = 1 \neq 0$$

For stationary ARMA processes, the mean estimator is consistent.

4.1.2 Covariance function

Let $v(t)$ be a stationary stochastic process with a zero mean. In this scenario:

$$\gamma(\tau) = \tilde{\gamma}(\tau) = \mathbb{E}[v(t)v(t-\tau)] = \mathbb{E}[v(t)v(t+\tau)]$$

The sample estimator of $\gamma(\tau)$ is given by:

$$\hat{\gamma}_N(\tau) = \frac{1}{N-\tau}\sum_{t=1}^{N-\tau} v(t)v(t+\tau) \qquad 0 \leq \tau N - 1$$

It's noteworthy that given $\{v(1), v(2), \ldots, v(N)\}$, we have:

- $\hat{\gamma}_N(0)$ is computed from a sum of N elements.
- $\hat{\gamma}_N(1)$ is computed from a sum of $N-1$ elements.
- $\hat{\gamma}_N(N-1)$ is computed from just one element.

The accuracy of $\hat{\gamma}_N(\tau)$ decreases as τ (the approximation is reliable only for $\tau \ll N$). The provided sample estimator $\hat{\gamma}_N(\tau)$ is both correct and consistent for ARMA processes.

Another estimator of $\gamma(\tau)$ can be considered:

$$\hat{\gamma}'_N(\tau) = \frac{1}{N}\sum_{t=1}^{N-\tau} v(t)v(t+\tau) \qquad 0 \leq \tau N - 1$$

This estimator is biased for finite N, but it becomes unbiased as $N \to \infty$ (asymptotically correct). Only $\hat{\gamma}'_N(\tau)$ satisfies the positive semi-definiteness property of a covariance function.

4.2 Spectral density estimator

We aim to estimate the spectral density $\Gamma(\omega)$ of a stationary process based on a set of samples $\{v(1), v(2), ..., v(N)\}$. We assume that the process has a zero mean, implying $\tilde{\gamma}(\tau) = \gamma(\tau)$. By definition, the spectrum is expressed as an infinite sum:

$$\Gamma(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-j\omega\tau}$$

In practical applications, we approximate it using only a finite number of terms:

$$\hat{\Gamma}_N(\omega) = \sum_{\tau=1-N}^{N-1} \hat{\gamma}_N(\tau) e^{-j\omega\tau}$$

It's essential to note two sources of approximation in this definition:

- The sample estimator $\hat{\gamma}_N(\tau)$ is utilized instead of $\gamma(\tau)$.
- The summation is restricted to $\pm(N-1)$ terms.

While this estimator is correct in the sense that it's derived from the available data, it's not consistent due to the finite number of terms used in the estimation process.

4.3 Spectral factorization

Here are four distinct representations of a stochastic process resulting from filtering white noise through a stable digital filter:

- Time-domain representation:

$$V(T) = a_1 v(t-1) + \cdots + a_{n_a}(t-n_a) + c_o\eta(t-1) + \cdots + c_{n_c}\eta(t-n_c)$$

- Operatorial representation:

$$v(t) = \frac{C(z)}{A(z)}\eta(t)$$

- Probabilistic representation:

$$\mu, \gamma(\tau)$$

- Frequency domain representation:

$$\mu, \Gamma(\omega)$$

These representations are interchangeable, allowing for translation between them. While calculating the spectrum of a process from the transfer function has been addressed, in practice, it's often more useful to do the reverse. Given data with specific spectral characteristics, the challenge lies in describing the generating process, or finding the pair $(W(z), \eta(\cdot))$.

However, not every spectrum corresponds to a transfer function due to the constraints on spectrum shapes. Furthermore, even if a transfer function exists, it may not be unique. Multiple combinations of transfer functions and white noise can represent the same process.

4.3.1 Representation choices

Consider a generic process:

$$v(t) = W(z)\eta(t)$$

Here, $\eta(\cdot) \sim WN(o, \lambda^2)$ The complex spectrum is given by:

$$\Phi(z) = W(z)W(z^{-1})\lambda^2$$

The following are examples of equivalent representations for $v(t) = \tilde{W}(z), \tilde{\eta}(t)$ where $\tilde{\eta} \sim WN(0, \tilde{\lambda^2})$:

1. Multiplication by a constant α:

$$\tilde{W}(z) = \frac{1}{\alpha}W(z) \qquad \tilde{\eta} \sim WN(0, \alpha^2\lambda^2)$$

   To demonstrate their equivalence, we show they have the same spectrum:

$$\tilde{\Phi}(z) = \tilde{W}(z)\tilde{W}(z^{-1})\tilde{\lambda}^2 = \frac{1}{\alpha}W(z) \cdot \frac{1}{\alpha}W(z^{-1}) \cdot \alpha^2\lambda^2 = W(z)W(z^{-1})\lambda^2 = \Phi(z)$$

2. Multiplication by z^n:

$$\tilde{W}(z) = z^n W(z) \qquad \tilde{\eta} = z^{-n}\eta(t) = \eta(t-n)$$

   To demonstrate their equivalence, we show they have the same spectrum:

$$\tilde{\Phi}(z) = \tilde{W}(z)\tilde{W}(z^{-1})\lambda^2 = z^n W(z)z^{-n}W(z^{-1})\lambda^2 = W(z)W(z^{-1})\lambda^2 = \Phi(z)$$

3. Multiplication of both the numerator and the denominator by the same factor:

$$\tilde{W}(z) = W(z)\frac{z-p}{z-p} \qquad \tilde{\eta} = \eta(t)$$

   To demonstrate their equivalence, we show they have the same spectrum, obtained after simplification:

$$\tilde{\Phi}(z) = \frac{z-p}{z-p}\Phi(z) = \Phi(z)$$

4. Multiplication by an all-pass filter:

$$\tilde{W}(z) = W(z)T(z) \qquad \tilde{\eta} = \eta(t)$$

   Here, $T(z) = \frac{1}{q}\left(\frac{z-q}{z-\frac{1}{q}}\right)$ To demonstrate their equivalence, we show they have the same spectrum:

$$\tilde{\Phi}(z) = \tilde{W}(z)\tilde{W}(z^{-1})\lambda^2 = W(z)T(z)T(z^{-1})W(z^{-1})\lambda^2 = W(z)W(z^{-1})\lambda^2 = \Phi(z)$$

We require a representation that is more suitable than others for solving the prediction problem. The following theorem facilitates the selection of the so-called canonical representation, which eliminates all potential sources of redundancy as outlined.

Theorem 4.3.1. *Let $v(t)$ be a stationary stochastic process with a rational spectrum. There exists a unique pair* $\left\{\hat{W}(z) = \frac{C(z)}{A(z)}, \xi(t)\right\}$ *such that:*

1. *$C(z)$ and $A(z)$ are monic (with the first coefficient being one).*
2. *$C(z)$ and $A(z)$ have the same degree (relative degree equal to zero).*
3. *$C(z)$ and $A(z)$ are co-prime (no common roots).*
4. *$C(z)$ and $A(z)$ have all their roots in the closed and open unit circle, respectively ($|z| \leq 1, \forall z$ s.t. $C(z) = 0$, $|z| < 1, \forall z$ s.t. $A(z) = 0$).*

The function $\hat{W}(z)$ is termed the canonical spectral factor.

Example:
Let's examine the ARMA(1,1) process given by:

$$v(t) = av(t-1) + \eta(t) + c\eta(t-1) \qquad \eta(\cdot) \sim WN(0, \lambda^2)$$

For this process to be stationary, the condition $|a| < 1$ must be satisfied. The spectrum is given by:

$$\Gamma(\omega) = \frac{1 + c^2 + 2c\cos(\omega)}{1 + a^2 - 2a\cos(\omega)}\lambda^2$$

If $|c| \leq 1$, the representation is canonical:

$$\hat{W}(z) = W(z) = \frac{1 + cz^{-1}}{1 - az^{-1}} \qquad \xi(\cdot) = \eta(\cdot)$$

On the other hand, if $|c| > 1$, the canonical representation changes to:

$$\hat{W}(z) = W(z) = \frac{1 + \frac{1}{c}z^{-1}}{1 - az^{-1}} \qquad \xi(\cdot) \sim WN(0, c^2\lambda^2)$$

Example:
Let's examine the given process:

$$v(t) = \eta_1(t-1) + \eta_2(t) - \eta_2(t-1)$$

where $\eta_1 \sim WN(0, lambda_1^2)$ and $\eta_2 \sim WN(0, lambda_2^2)$ are independent white noise processes. This process is constructed as the sum of two independent and stationary processes, thus it is also stationary. The auto-covariance function is null for $|\tau| > 1$. The process can be reformulated as an MA(1):

$$v(t) = \xi(t) + c\xi(t-1)$$

where $\xi(\cdot) \sim WN(0, \lambda^2)$, provided it has the same spectral characteristics, or equivalently, the same auto-covariance function (same covariance implies same spectral representation). Now, considering the auto-covariance function:

$$\gamma(0) = (1 + c^2)\lambda^2 = \lambda_1^2 + 2\lambda_2^2$$

$$\gamma(1) = c\lambda^2 = -\lambda_2^2$$

Solving this system of equations yields only one solution with $|c| < 1$ to this system of equations. For instance, if $\lambda_1^2 = \frac{1}{4}$ and $\lambda_2^2 = \frac{1}{2}$, we have:

$$\begin{cases} (1+c^2)\lambda^2 = \frac{5}{4} \\ c\lambda^2 = -\frac{1}{2} \end{cases} \implies \begin{cases} c = -\frac{1}{2} & \lambda^2 = 1 \\ c = -2 & \lambda^2 = \frac{1}{4} \end{cases}$$

Note that only the first solution is canonical.

CHAPTER 5

Prediction problem

5.1 Introduction

Let's consider a process with a rational spectrum:

$$v(t) = W(z)\eta(t) = \frac{C(z)}{A(z)}\eta(t) \qquad \eta(\cdot) \sim WN(0, \lambda^2)$$

Prediction aims to estimate the future value $v(t+r)$ (with $r > 0$) based on observations of the process up to time t. This estimate, denoted as $\hat{v}(t+r|t)$, depends on the predictor horizon r. Generally, the predictor takes the form:

$$\hat{v}(t+r|t) = f\left(v(t), v(t-1), \dots\right)$$

It's important to note that the prediction error itself is a stochastic process since $\hat{v}(\cdot)$ depends on random variables.

Example:
A naïve predictor could be constructed by averaging the last observed samples:

$$\hat{v}(t+1|t) = \frac{1}{3}v(t) + \frac{1}{3}v(t-1) + \frac{1}{3}v(t-2)$$

Our goal is to find the optimal predictor that effectively combines information from the model and the available past observations.

5.2 Prediction error

To evaluate prediction quality, we define the prediction error (or residual) as:

$$\varepsilon(t+r) = v(t+r) - \hat{v}(t+r|t)$$

The optimal predictor minimizes the mean square prediction error (MSPE), which is the variance of the residual:

$$\min\left(\mathbb{E}\left[\varepsilon(t)^2\right]\right)$$

Considering the expansion of $W(z)$ in negative powers of z:

$$W(z) = w_0 + w_1 z^{-1} + w_2 z^{-2} + \dots$$

We can represent $v(t)$ as an infinite linear combination of past noise values:

$$v(t) = w_0\eta(t) + w_1\eta(t-1) + w_2\eta(t-2) + \dots$$

Assuming we have access to the past of $\eta(\cdot)$, we can estimate $v(t+r)$ for $r \geq 1$:

$$v(t+r) = \underbrace{w_0\eta(t+r) + w_1\eta(t+r-1) + \dots + w_{r-1}\eta(t+1)}_{\alpha(t)} + \underbrace{w_r\eta(t) + w_{r+1}\eta(t-1) + \dots}_{\beta(t)}$$

Here, $\alpha(t)$ and $\beta(t)$ are uncorrelated random variables, computed over non-overlapping time ranges:

- $\beta(t)$ is computable once the past of $\eta(\cdot)$ (up to t) is known.
- $\alpha(t)$ depends on the future of $\eta(\cdot)$ (from $t+1$ to $t+r$).

Thus, $\alpha(t)$ shows no correlation with the past until time t, indicating its unpredictability based on past information. As a result, we can estimate its mean solely, which is zero:

$$\mathbb{E} = w_0\mathbb{E}\left[\eta(t+r)\right] + w_1\mathbb{E}\left[\eta(t+r-1)\right] + \dots + w_{r-1}\mathbb{E}\left[\eta(t+1)\right] = 0$$

This leads to the optimal predictor:

$$\hat{v}(t+r|t) = \beta(t) = w_r\eta(t) + w_{r+1}\eta(t-1) + \dots$$

The prediction error is then:

$$\varepsilon(t+r) = v(t+r) - \hat{v}(t+r|t) = \alpha(t) = w_0\eta(t+r-1) + \dots + w_{r-1}\eta(t+1)$$

It's noteworthy that $\varepsilon(\cdot)$ represents an MA process. Its mean value is 0, and its variance is $(w_0^2 + w_1^2 + \dots + w_{r-1}^2)\lambda^2$.

Furthermore, the variance increases monotonically with r, indicating that prediction uncertainty grows as the prediction horizon extends:

- For $r = 1$, the variance is $w_0^2\lambda^2$. When $w_0 = 1$ (if $W(z)$ is canonical), it matches the noise variance:
$$\mathrm{Var}\left[\varepsilon(t)\right] = \mathrm{Var}\left[\eta(t)\right]$$
- As r approaches infinity, the variance becomes $(w_0^2 + w_1^2 + \dots)\lambda^2$, equivalent to the variance of the entire process $v(t)$:
$$\mathrm{Var}\left[\varepsilon(t)\right] = \mathrm{Var}\left[v(t)\right]$$

Predicting becomes increasingly challenging with larger r as the estimation concerns a distant time point beyond available data. Over time, past data lose utility, and the most reasonable estimate is the variable's mean:

$$\mathbb{E}\left[v(t+r)\right] = 0$$

Consequently, the variance of the prediction error approaches that of the process:

$$\mathrm{Var}\left[\varepsilon(t)\right] = \mathbb{E}\left[\left(v(t+r) - \hat{v}(t+r|t)\right)^2\right] = \mathrm{Var}\left[v(t)\right]$$

The optimal predictor can be expressed using operator notation:

$$\hat{v}(t+r|t) = \left[w_r + w_{r+1}z^{-1} + w_{r+2}z^{-2} + \dots\right]\eta(t) = \hat{W}_r(z)\eta(t)$$

To find $\hat{W}_R(z)$, observe that the transfer function can be expressed as:

$$\begin{aligned} W(z) &= w_0 + w_1 z^{-1} + \cdots + w_{r-1}z^{-(r-1)} + w_r z^{-r} + w_{r+1}z^{-r-1} + \dots \\ &= \underbrace{\left(w_0 + w_1 z^{-1} + \cdots + w_{r-1}z^{-(r-1)}\right)}_{E(z)} + z^{-r}\underbrace{\left(w_r + w_{r+1}z^{-1} + \dots\right)}_{\hat{W}_r(z)} \\ &= E(z) + z^{-r}\hat{W}_r(z) \end{aligned}$$

Where $E(z)$ is a polynomial of degree $r-1$, and $\hat{W}_r(z)$ is a power series in z^{-r} (representing a transfer function). Thus, $\hat{W}_r(z)$ can be determined by performing long division of the numerator of $W(z)$ by its denominator, iterated for r steps:

$$\hat{W}_r(z) = \frac{F_r(z)}{A(z)}$$

Here, $F_r(z)$ is the remainder of the division.

5.2.1 Optimal predictor

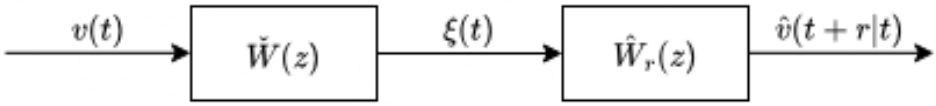
The optimal predictor we derived relies on past values of the white noise process, which may not be readily available in practice. Typically, we only have access to a sequence of data from the $v(\cdot)$ process. Let's make the following assumptions:

- $W(z)$ and $\eta(t)$ represent a canonical representation of $v(t)$.
- $W(z)$ has no zeros on the unit circle boundary ($\Sigma(\omega) > 0 \quad \forall\omega$).

Under these assumptions, we can reconstruct $W(t)$ using a whitening filter with transfer function $W(z)^{-1}$:

$$\eta(t) = \frac{A(z)}{C(z)}v(t)$$

The whitening filter offers a well-defined representation of $\eta(t)$ since polynomials $A(z)$ and $C(z)$ have the same degree. Hence, $\check{W}(z) = W(z)^{-1}$ can be interpreted as a legitimate transfer function. Moreover, $C(z)$ being a Schur-stable polynomial ensures the stationarity of the process.



By combining the whitening filter with the optimal predictor derived from the noise, we obtain the optimal predictor from the process data:

$$W_r(z) = \check{W}(z)\hat{W}_r(z) = \frac{A(z)}{C(z)}\frac{F_r(z)}{A(z)} = \frac{F_r(z)}{C(z)}$$

The optimal predictor from $v(t)$ aligns with that from the noise, with the denominator replaced by the numerator of $\hat{W}(z)$. It's important to note that this predictor is also a stochastic process dependent on $v(\cdot)$, and it remains stationary.

5.2.2 Predictor initialization

Unless $C(z)$ is trivially 1, the optimal predictor from data takes the form of a recursive equation:

$$\hat{v}(t+r|t) = -c_1\hat{v}(t+r-1|t-1) - c_2\hat{v}(t+r-2|t-2) + \dots$$

The challenge arises from having data only up to a certain instant. A heuristic solution is proposed: when data are unavailable, we initialize $\hat{v}$ to $\mathbb{E}[v]$ (trivial predictor). The asymptotic stability of the filter $W_r(z)$ ensures that the effect of initialization diminishes over time.

Example:
Let's examine an MA(1) process:

$$v(t) = \eta(t) + c\eta(t-1) \quad \eta(\cdot) \sim WN(0, \lambda^2)$$

where $|c| < 1$. This form implies the model is in canonical form:

$$W(z) = 1 + cz^{-1}$$

Given that the transfer function is already expressed as an expansion of negative powers of z, we can skip the long division procedure:

$$\begin{cases} \hat{W}_1(z) = c \to \hat{v}(t+1|t) = -c\hat{v}(t|t-1) + cv(t) \\ \hat{W}_r(z) = 0 \to \hat{v}(t+r|t) = 0 \end{cases}$$

As the covariance is null for $|\tau| > 1$, past data up to t do not aid in predicting $v(\cdot)$ two or more steps ahead. The only viable estimate is the mean value, which is 0. Consequently, $\mathrm{Var}[\varepsilon(t)] = \mathrm{Var}[v(t)]$.

Attempting to predict a process with a model that is not in canonical form results in a non-stationary process because the model is unstable. Essentially, we would be trying to predict a stationary process using a non-stationary model, which is not feasible.

5.3 Prediction of AR processes

Let's examine the AR(1) process:

$$v(t) = av(t-1) + \eta(t) \quad \eta(cdot) \sim WN(0, \lambda^2)$$

Performing long division on the transfer function for three steps, we obtain:

$$\hat{W}_1(z) = \frac{a}{1-az^{-1}} \qquad \lambda_\varepsilon^2 = \lambda^2$$

$$\hat{W}_2(z) = \frac{a^2}{1-az^{-1}} \qquad \lambda_\varepsilon^2 = (1+a^2)\lambda^2$$

$$\hat{W}_3(z) = \frac{a^3}{1-az^{-1}} \qquad \lambda_\varepsilon^2 = (1+a^2+a^4)\lambda^2$$

For a generic prediction horizon r, we have:

$$\hat{W}_r(z) = \frac{a^r}{1-az^{-1}} \qquad \lambda_\varepsilon^2 = (1+a^2+a^4+\dots+a^{2(r-1)})\lambda^2$$

As r tends to infinity, a^r approaches zero, and thus $\hat{v}(t+r|t) = 0$. Moreover, since $|a| < 1$, the variance of the residual equals:

$$\lambda_\varepsilon^2 = \sum_{i=0}^{\infty} a^{2i} = \sum_{i=0}^{\infty} (a^2)^i = \frac{1}{1-a^2}$$

This coincides with the variance of the process $\mathrm{Var}\,[v(t)]$.

The most effective one-step ahead predictor is expressed as follows:

$$\hat{v}(t+1|t) = \hat{W}_1\eta(t) = \frac{a}{1-az^{-1}}\eta(t) = a\hat{v}(t|t-1) + a\eta(t)$$

Where the prediction error is given by $\varepsilon(t) = v(t) + \hat{v}(t|t-1)$, and its dynamics are determined by subtracting the predictor equation from the model equation:

$$v(t+1) - \hat{v}(t+1|t) = av(t) + \eta(t+1) - a\hat{v}(t|t-1) - a\eta(t) \qquad \varepsilon(t+1) = a\varepsilon(t) + \eta(t+1) - a\eta(t)$$

Subsequently, the variance of the prediction error can be computed as:

$$\mathrm{Var}\,[\varepsilon(t+1)] = \mathrm{Var}\,[\varepsilon(t)] = \lambda^2$$

It's worth noting that the predictor:

$$\hat{v}(t+1|t) = v(t)$$

is not the optimal choice. In fact,

$$\varepsilon(t+1) = v(t+1) - \hat{v}(t+1|t) = v(t+1) - v(t)$$

Thus, we can deduce that:

$$\mathbb{E}\left[\varepsilon(t+1)^2\right] = \frac{2}{1+a}\lambda^2$$

Given that $|a| < 1$, the mean squared prediction error (MSPE) of the basic predictor exceeds that of the optimal predictor.

Recalling that:

$$\hat{W}_1(z) = \frac{1}{1-az^{-1}} \to F_r(z) = 1$$

This implies $W_R(z) = a$, leading to the optimal one-step ahead predictor derived from data as:

$$\hat{v}(t+1|t) = av(t)$$

This is not surprising, if one considers the model:

$$v(t+1) = av(t) + \eta(t+1) = \hat{v}(t+1|t) + \eta(t+1)$$

This implies that the predicted value differs from the true value by a white noise term, which is inherently unpredictable. Consequently, there's no additional information in the data that could enhance the predictor, validating its optimality. Thus, the prediction error follows a white noise process $\varepsilon(t+1) = \eta(t+1)$ with a variance of $\mathrm{Var}\,[\varepsilon(t)] = \lambda^2$.

5.4 Prediction of ARMA processes

Consider a generic ARMA process represented as:

$$A(z)v(t) = C(z)\eta(t) \quad \eta(\cdot) \sim WN(0, \lambda^2)$$

Here, $A(z)$ and $C(z)$ are monic polynomials of the same degree. Assuming all their roots lie within the unit circle, and they share no common roots, this representation is considered canonical.

To derive the one-step ahead predictor, we employ a single step of the long division procedure on the transfer function. Consequently, the optimal one-step ahead predictor is characterized by the transfer function:

$$W_1(z) = \frac{z\left(C(z) - A(z)\right)}{C(z)}$$

This corresponds to the following time-domain equation:

$$C(z)\hat{v}(t+1|t) = z\left(C(z) - A(z)\right)v(t) = \left(C(z) - A(z)\right)v(t+1)$$

5.5 Prediction of ARMAX processes

Let's consider the generic ARMAX model expressed at time $t + k$:

$$A(z)y(t+k) = B(z)u(t) + C(z)\xi(t+k)$$

The procedure for computing the predictor mirrors that of the ARMA process. Initially, we conduct long division of $C(z)$ by $A(z)$ for k steps, denoting the quotient as $E(z)$ and the remainder as $F_k(z)$:

$$C(z) = A(z)E(z) + z^{-k}F_k(z)$$

Here, $E(z) = e_0 + e_1 z^{-1} + \cdots + e_{k-1}z^{-(k-1)}$, with $e_0 = 1$ due to the monic nature of both $A(z)$ and $C(z)$.

Multiplying both sides of the ARMAX equation by $E(z)$, we obtain:

$$\begin{aligned}
&A(z)E(z)y(t+k) = B(z)E(z)u(t) + C(z)E(z)\xi(t+k) \implies \\
&[C(z) - z^{-k}F_k(z)]y(t+k) = B(z)E(z)u(t) + C(z)E(z)\xi(t+k) \\
&C(z)y(t+k) = z^{-k}F_k(z)y(t+k) + B(z)E(z)u(t) + C(z)E(z)\xi(t+k) \\
&y(t+k) = \frac{F_k(z)}{C(z)}y(t) + \frac{B(z)E(z)}{C(z)}u(t) + E(z)\xi(t+k)
\end{aligned}$$

Breaking down the structure of the equation:

- The first term relies on $y(\cdot)$ up to time t.
- The second term depends on $u(\cdot)$ up to time t.
- The third term, $e_0\xi(t+k) + e_1\xi(t+k-1) + \cdots + e_{k-1}\xi(t+1)$, remains unpredictable from the past.

Hence, the optimal k-steps ahead predictor for $y(t+k)$ is:

$$\hat{y}(t+k|t) = \frac{F_k(z)}{C(z)}y(t) + \frac{B(z)E(z)}{C(z)}u(t)$$

Or equivalently:

$$C(z)\hat{y}(t+k|t) = Fk(z)y(t) + B(z)E(z)u(t)$$

Specifically, for $k = 1$, $E(z) = e_0 = 1$ and $F_1(z) = z(C(z) - A(z))$, resulting in the predictor equation:

$$C(z)\hat{y}(t+1|t) = (C(z) - A(z))y(t+1) + B(z)u(t)$$

CHAPTER 6

Identification problem

6.1 Introduction

System identification is the field dedicated to methodologies for deriving models from data.

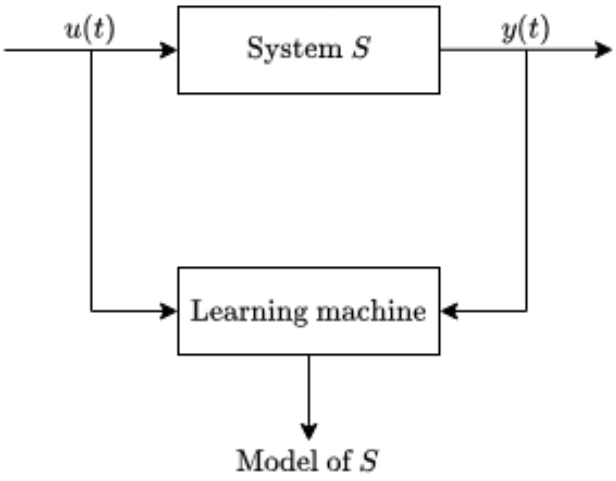


Figure 6.1: Identification process

Key characteristics of derived models include:

- *Limited validity*: the model can't reveal more about the system than what's inherently present in the data used for its derivation.
- *Difficult physical interpretation*: model parameters lack direct physical interpretation; they are conceived to explain data and aren't derived from physical laws.
- *Simple derivation and usage.*

The identification process starts with selecting the model class (linear, nonlinear, AR, MA, ARMA). However, data provided to the system are imperfect (affected by noise and only partially representative of the phenomena involved) and may have other issues like quantization, missing data, outliers and more.

Additionally, the process may vary over time, rendering an approach solely based on time-invariant models insufficient. Furthermore, not all signals and variables relevant to the system are always measurable.

In summary, the fundamental elements of the identification problem are:

1. The system $\mathcal{S}$.
2. The model family $\mathcal{M}$.
3. The identification method $\mathcal{I}$.
4. The identification experiment $\mathcal{E}$.

6.1.1 Parametric system identification

The steps involved in parametric system identification are as follows:

- *Experiment design and data collection*: the first phase involves designing and conducting an identification experiment ($\mathcal{E}$) to gather the necessary data. For input-output systems, the system ($\mathcal{S}$) is subjected to an appropriate input signal, and the inputs and outputs are observed and recorded over a period of time. Factors such as the choice of input signal and the length of the data (N) impact the informativeness and confidence of the resulting dataset. Data preprocessing is typically performed to address outliers, missing data, trends, periodicity, etc. The experiment design ($\mathcal{E}$) depends on factors such as the input signal selection, the presence of feedback loops, sampling time, and data prefiltering. Constraints on the experiment may exist due to security reasons or operational limitations under specific conditions.

- *Choice of model class*: a parametric model family $M(\vartheta) = \{M(\vartheta), \vartheta \in \Theta\}$ is selected, believed to be flexible enough to explain the data. The set Θ defines the admissible parameterization. The choice of model class may be guided by data analysis, prior knowledge of the system, or intended use of the model. The identification problem involves determining the appropriate parameters ϑ. Model types include static/dynamic, discrete/continuous-time, linear/nonlinear, time-invariant/time-varying, and internal/external representation. Non-parametric models are also viable options.

- *Choice of identification criterion and parameter estimation*: a suitable method (identification algorithm, $\mathcal{I}$) is employed to estimate the unknown model parameters. This typically involves selecting an identification criterion $J_N(\theta) \geq 0$ and minimizing it with respect to θ. The choice of identification method ($\mathcal{I}$) depends on the model class. Minimizing $J_N(\vartheta)$ is straightforward for certain classes where it becomes a quadratic function of ϑ, but more challenging for other classes.

- *Model validation*: throughout the identification process, crucial assumptions are made, such as the system ($\mathcal{S}$) belonging to the selected model family $M(\vartheta)$ and fixing model orders. Validation involves conducting a quality check on the obtained model. If the model is deemed unsuccessful, assumptions are reevaluated, and the identification process may need to be revisited from steps one, two, or three.

6.2 Static modeling

The typical scenario in the identification problem involves observing input and output data from a system:

- Input data: $\{u(1), u(2), ..., u(N)\}$.
- Output data: $\{y(1), y(2), ..., y(N)\}$.

The objective is to find a model that elucidates the relationship between input and output variables.

Before delving into the intricate task of identifying dynamic systems, let's first address a relatively simpler scenario, focusing on the static relationship between the variables. In this case, the model can be expressed as:

$$\hat{y}(t) = f\left(u(t), \theta\right)$$

Here, t uniquely identifies an input/output pair, θ represents the model parameters, and $y(t)$ solely depends on $u(t)$ at the corresponding index t.

To tackle this problem, we initially hypothesize about the nature of the relationship and then seek the optimal model within the assumed category. This is achieved through a fitting technique, such as the least squares method, which minimizes the disparity between the model-generated data and the observed data. Two primary classes of problems can be envisioned:

- *Interpolation*: the model is expected to pass through each data point.
- *Fitting*: the model approximates the data points as closely as possible.

However, aiming to find a value for ϑ such that $\hat{y}(t) = y(t)$ (interpolation) is often not advisable due to several reasons:

- Interpolation assumes data are devoid of disturbances. However, in reality, data are often affected by uncertainty, such as measurement errors, distortions, or approximations. Attempting to replicate data exactly through a model could inadvertently encompass noise effects (overfitting), thereby modeling the noise rather than the underlying mechanism generating the data.
- Interpolation typically necessitates a model with a degree of complexity matching the size of the dataset, such as polynomial approximation.
- Extrapolation (generalization) capabilities of the model may be poor and potentially misleading.

For fitting, it is important to consider the following:

- When dealing with noisy data, the function $f(u, \vartheta)$ should accurately fit the data, while avoiding fitting too closely to account for noise effects and prevent overfitting.
- The robustness of a model decreases as the number of parameters (ϑ) increases.
- If a model with fewer degrees of freedom than the dataset size N is used, exact interpolation cannot be achieved, and the data can only be approximated.
- A fitted function will not precisely match the data used in the identification process, but it is expected to possess better generalization properties.
- In many cases, interpolation is not feasible because there may be multiple different values of y for the same u. In such scenarios, only fitting is possible.

6.2.1 Parametric function

One common approach to describe the parametric function $f(u, \vartheta)$ is through a functional expansion:

$$f(u, \vartheta) = \varphi^T \vartheta = \sum_i \vartheta_i \varphi_i(u)$$

Here, the $\varphi_i(u)$ functions are referred to as basis functions. This representation is particularly straightforward and useful because it is linear in the parameters (assuming fixed basis functions), thereby framing the problem as a linear regression for estimating ϑ. Hence, the φ_i terms are sometimes called regressors.

In vector form, the model is expressed as:

$$\mathcal{M} : \hat{y}(t) = \varphi(u(t))^T \vartheta$$

Here, any nonlinearity (if present) is encapsulated within the regressors. A fundamental requirement of this functional expansion is that $f(u, \vartheta)$ serves as a universal approximating function. In other words, it should be capable of approximating any continuous function on a compact domain with arbitrary precision, provided it possesses a sufficiently large number of parameters.