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.

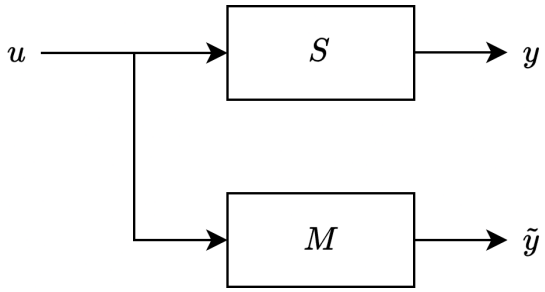


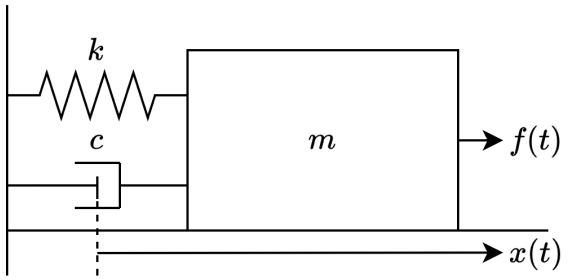
Figure 1.1: Visual representation of system and model

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:

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

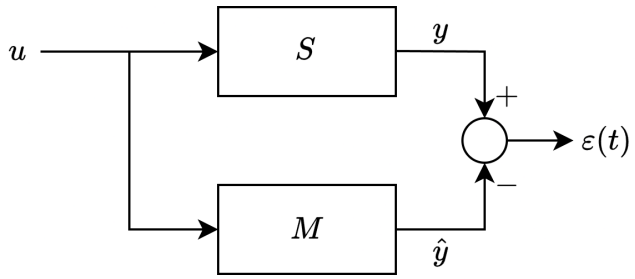
On the other hand, black-box modeling necessitates understanding the relationship between inputs and outputs to derive the model. In this scenario, the model obtained is:

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

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.



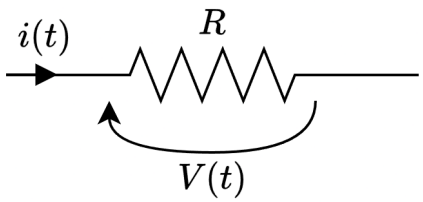
When the outputs exhibit similarity based on certain metrics, it signifies that the model accurately mirrors the dynamics of the system. However, if patterns persist within the error graph, it indicates that not all information has been effectively extracted from the data. Conversely, if the error graph 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.

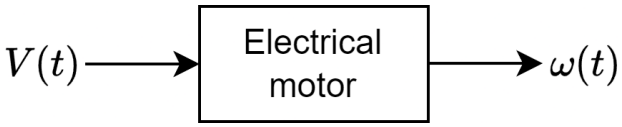
Example:
An illustration of a static system is represented by a circuit containing a resistor.



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 $\widehat{\vartheta}$, 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 $\widehat{\vartheta}$. For a parameter $\vartheta(t)$ with a dynamic value, we aim to find the estimate $\widehat{\vartheta}(t|t_N)$ 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:

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

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:

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

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:

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

The values to be examined are:

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

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

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

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:

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

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:

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

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.

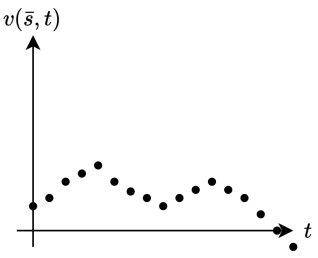


Figure 2.1: Visual representation fixed outcome

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

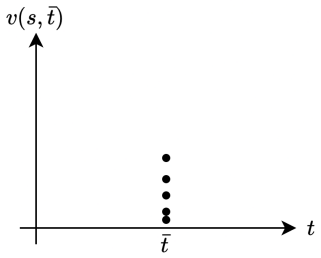


Figure 2.2: Visual representation fixed time

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:

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

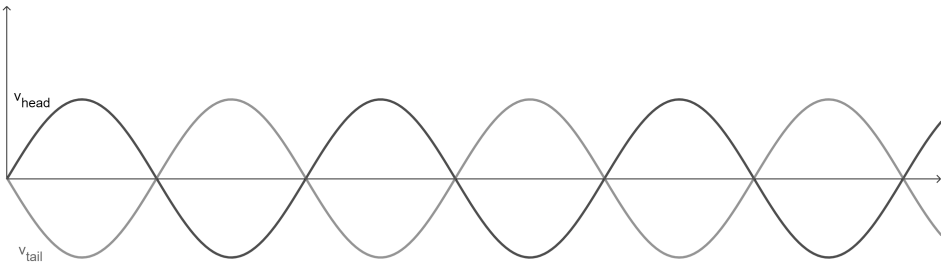


Figure 2.3: Graphs of the functions $v_{head}(t)$ and $v_{tail}(t)$

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 $\bar{v} \sim G(1, 3)$.

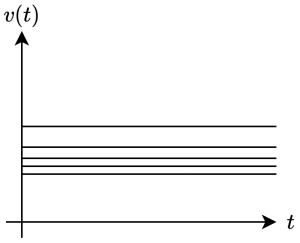


Figure 2.4: Some possible realizations

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 $\bar{v} \sim G(1, 3)$.

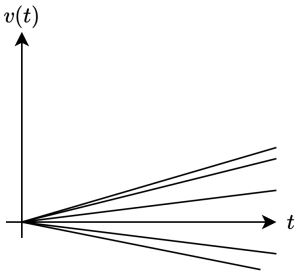


Figure 2.5: Some possible realizations

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:

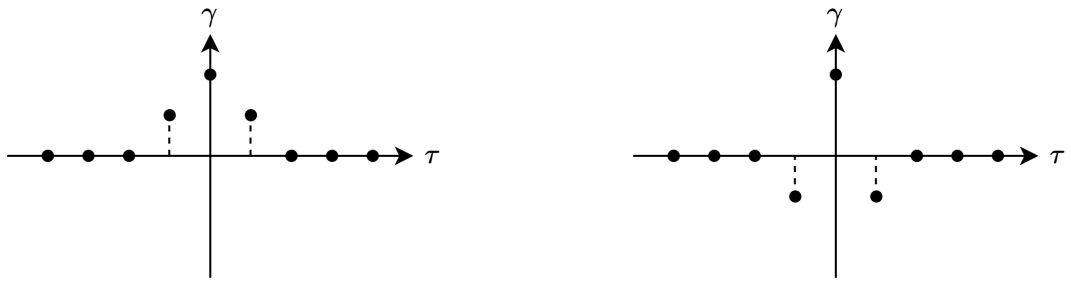
$$\begin{aligned}
\gamma(t,t+1) &= \mathbb{E}\left[(v(t)v(t+1))^2\right] \\
&= \mathbb{E}\left[(\eta(t)+c\eta(t-1))\,(\eta(t+1)+c\eta(t))\right] \\
&= \mathbb{E}\left[\eta(t)\eta(t+1)+c\eta(t-1)\eta(t+1)+c\eta(t)^2+c^2\eta(t-1)\eta(t)\right] \\
&= \underbrace{\mathbb{E}\left[\eta(t)\eta(t+1)\right]}_{0}+c\underbrace{\mathbb{E}\left[\eta(t-1)\eta(t+1)\right]}_{0}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

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

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:

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

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 $\widehat{v}(t)$ are independent:

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

The correlation is given by:

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

The purely nondeterministic component can be expressed as:

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

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

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

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

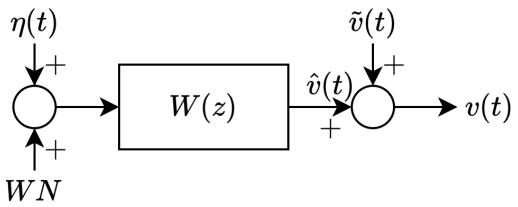


Figure 2.7: Stochastic process composition

2.5.1 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:

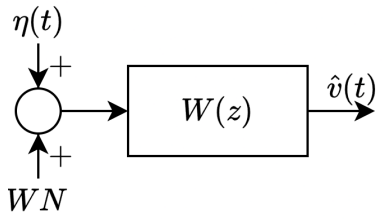
$$v(t) = \alpha_1\lambda_1^t + \alpha_2\lambda_2^t$$

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

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

2.6 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:

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

2.6.1 Moving average 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:

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

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 $\gamma(\tau)$.

2.6.2 Autoregressive 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:

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

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.