CH5350: Applied Time-Series Analysis

Arun K. Tangirala

Department of Chemical Engineering, IIT Madras

Models for Linear Stationary Processes

Classical Time-Series Model

Once a time-series is realized as predictable, the search for a suitable mathematical model is carried out. Classical approaches in the early days rested on the philosophy that a series is made up of three components

$$\boxed{\text{Time-Series} = \text{Trend} + \text{Seasonal Component} + \text{Stationary component}}$$

The trend and seasonal components could be combined into a single component under the banner of *deterministic* component.

Classical Approach ...contd.

Several efficient non-parametric and semi-parametric methods were subsequently developed to realize such a decomposition. The trend usually contains a polynomial type of trend while the seasonal component captures the periodic characteristics, if any.

Extracting the deterministic portions of a series is not trivial, but can be effectively carried out with suitable **regression, smoothing and filtering operations.**

Note: The seasonal component is usually a deterministic periodic signal, and assumed to be uncorrelated with the non-seasonal component.

Modern Approach

In 1970s, a new approach to modelling the seasonal (including the non-stationary and trend components) was introduced.

Unlike the models based on additive approach, **multiplicative models** were postulated. These are more generic in the nature because they take into account the correlation between seasonal and non-seasonal (stationary) components, and also model the integrating (random walk) effects.

The resulting models are known as **seasonal ARIMA (SARIMA) models.**

Models for Stationary Processes

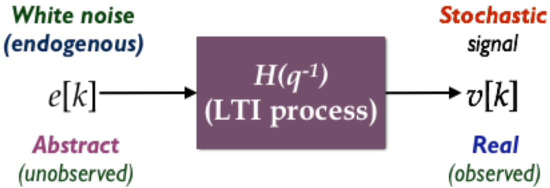
Models for stationary processes

The stationary component, by virtue of definition, cannot be purely explained by a mathematical model but requires the assistance of statistics.

It turns out that a large class of stationary stochastic processes, specifically linear processes, can be explained by mathematical models (convolution / difference equation) driven by forcing functions that are random in nature.

In fact, the forcing function is **the white-noise sequence.**

Spectral Factorization



The existence of such descriptions (for linear random processes) is centered around a milestone result known as the **spectral factorization theorem**.

The ability to represent $v[k]$ as WN passing through a linear filter, is possible **if and only if the spectral density** (TBD later) of $v[k]$ **satisfies certain mild conditions.**

Absolute convergence of ACVF

While we shall discuss the conditions on spectral density later, it is obvious that the spectral density function (SDF), denoted by $\gamma(\omega)$ itself should exist in the first place.

Postponing the formal definition of SDF, at this juncture it is useful to recall from the discussion on non-negative definiteness that the SDF is related to the ACVF through the Fourier transform,

$$\gamma(\omega) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \sigma[l] e^{-j\omega l} \tag{1}$$

ACVF should be absolutely convergent

Clearly, for the SDF to exist, and hence the linear representation of the stationary process, the ACVF should be absolutely convergent

$$\boxed{\sum_{l=-\infty}^{\infty} \sigma[l] < \infty} \tag{2}$$

Interpretation

For a stationary process to possess a linear filter representation, its ACVF should decay with lag.

- A periodic stationary process, i,e., **a harmonic process,** does not lend itself to a linear representation!

Models and Predictors

Developing a model is as good as developing a predictor. There exists a one-on-one equivalence between a model and a predictor. When the predictor is accompanied by a description of the **prediction error** (what has gone unpredicted), the result is a model.

For stochastic signals, there exists no model that can accurately predict them. However, an optimal model is expected to result in a prediction error that offers no further scope for prediction. Therefore, **the prediction errors associated with an optimal model should possess the properties of a white-noise series.**

Basic idea

Any (stationary) random process can be thought of as consisting of a *predictable* portion plus an *unpredictable* component.

$$v[k] = \hat{v}[k] + e[k] \tag{3}$$

where $\hat{v}[k]$ represents the predictable portion and $e[k]$ the unpredictable ideal random process, i.e., the *white-noise* process or in general the *i.i.d.* process.

Remarks

- The second term in (3) is an indispensable component of any random process because when it is absent, $v[k]$ condenses to a deterministic process.
- On the other hand, the first term can be absent, in which case, $v[k]$ has white-noise characteristics.

Prediction approach to developing models

Given P past observations of the process, the modeling objective is to develop a predictor $\hat{v}^\star[k]$ that leaves nothing predictable in the prediction error. In other words, the prediction error

$$\epsilon[k] = v[k] - \hat{v}^\star[k] \tag{4}$$

should have either white-noise or i.i.d. characteristics.

Conditional expectation: the optimal predictor

Recalling a key result, the best predictor of $v[k]$ is its conditional expectation given its past,

$$\hat{v}^\star[k] = E(v[k] | \{v[k-1], v[k-2], \cdots, v[k-P]\}) \tag{5}$$

- In general, the conditional expectation is **a non-linear function** of the past observations.
- Further, the conditional expectation in (5) is quite difficult to evaluate since the joint p.d.f. of past observations needs to be known.

Linear-time invariant models

If the observations follow a joint Gaussian distribution, the conditional expectation in (5) can be replaced by a **linear model.** (why?)

Replacing the RHS of (5) with a linear function yields,

$$\hat{v}^\star[k] = \sum_{i=1}^{P} (-d_i) v[k-i] \tag{6}$$

$$\implies v[k] + \sum_{i=1}^{P} d_i v[k-i] = e[k] \tag{7}$$

The model in (7) is known as the **auto-regressive model** of order P.

Remarks

- The negative sign on the coefficients is introduced to have a positive sign on the coefficients of the difference equation for $v[k]$.
- When the true process does not satisfy the (joint) Gaussianity assumption, the linear predictor is sub-optimal, but is at least mathematically tractable and implementable.
- **The difference equation in (6) shares strong similarities with that of a deterministic LTI system - the key difference is that the forcing function is stochastic.**

LTI representation

Equation (7) can be re-written in terms of the transfer function operator.

Example

Suppose that $v[k]$ can be modeled as an AR process of first-order, i.e., by (7) with $P = 1$ (as determined by the sharp cut-off in PACF at lag $l = 1$),

$$v[k] + d_1 v[k-1] = e[k] \tag{8}$$

Bringing in the shift-operator, we can express (8) using the transfer function operator

$$v[k] = H(q^{-1})e[k] \qquad \text{where} \qquad H(q^{-1}) = \frac{1}{1 + d_1 q^{-1}} \tag{9}$$

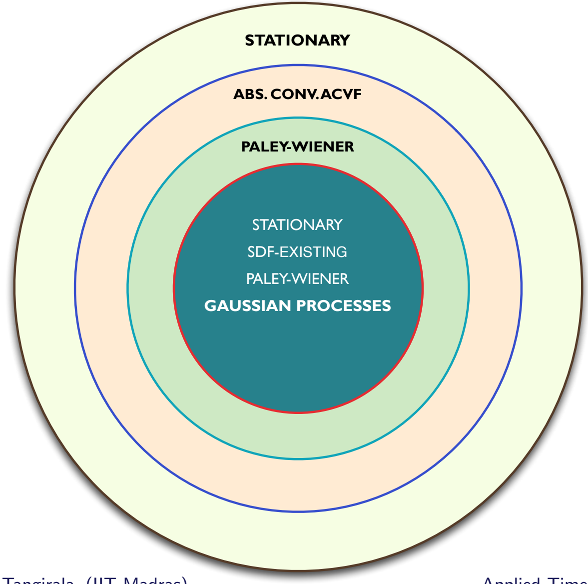
LTI representations and white-noise

The white noise, thus, acts as a fictitious input that drives the random process. It is, however, **endogenous**, i.e., intrinsic to the process.

Recall the two other roles of WN in time-series analysis:

- **As a benchmark for testing predictability:** The series (prior to model development) and the residuals (post modelling) are tested for "whiteness."
- **As an integral element of time-series models:** Every random process contains an element of unpredictability, i.e., the WN as its basic element.

Linear representations



Linear representations exist only for **stationary, SDF-existent, Paley-Wiener condition satisfying** and yield optimal predictions for **Gaussian** processes.

Linear Random Processes

Linear Random Process

Any stationary process is said to be linear if and only if it can be represented as

$$v[k] = \sum_{n=-\infty}^{\infty} h_n e[k-n] \quad \forall k \quad \text{where} \quad e[k] \sim \text{WN}(0, \sigma_e^2) \tag{10a}$$

$$\text{and} \quad \sum_{n=-\infty}^{\infty} |h_n| < \infty \tag{10b}$$

Implication: Any (weakly) stationary process can be represented as the weighted influences of the past & present ($n \geq 0$) and future ($n < 0$) shock waves.

Remarks

- The condition of absolute convergence (of $\{h_n\}$) is required for the convergence of the infinite sum in (10a) with probability one.
- A weaker requirement is that

$$\sum_n |h_n|^2 < \infty \tag{11}$$

which guarantees that the sum converges **in the mean square sense.**

- The WN process driving force for the linear process is replaced by an I.I.D. process in certain schools of thought, especially, in non-linear time-series analysis.

Further remarks

- We usually restrict ourselves to causal processes, *i.e.*, series with $n \geq 0$
- With the backshift operator notation, we have

$$\boxed{v[k] = H(q^{-1})e[k]} \quad \text{where} \quad \boxed{H(q^{-1}) = \sum_{n=-\infty}^{\infty} h_n q^{-n}} \tag{12}$$

- The transfer function operator $H(q^{-1}$ can be thought of as a "linear filter", which filters the shock waves to produce the series $x[k]$.

Comparison with convolution form

The model in equation (10) has a striking similarity with the convolution form for deterministic processes. Comparing

$$y[k] = \sum_{n=-\infty}^{n=\infty} g[n]u[k-n] \qquad\qquad v[k] = \sum_{n=-\infty}^{\infty} h_n e[k-n]$$

we can draw a few useful analogies

Comparison with LTI deterministic processes

- The coefficient h_n has the some role to play as $g[n]$
- **Stability** of an LTI system requires **absolute convergence of** $g[n]$ **Stationarity** requires **absolute convergence of the coefficients** h_n
- Thus, h_n can be thooght of as the **impulse response coefficient** of $H(q^{-1})$

A marked difference between these two forms is that while the input for deterministic systems is known to the user, the input in the stochastic case is fictitious, unknown, random but with known ACF.

Comparison with LTI descriptions ...contd.

- Just as $G(q^{-1}) = \sum_{n=-\infty}^{\infty} g[n]q^{-n}$ acts as a filter, so does $H(q^{-1}) = \sum_{n=-\infty}^{\infty} h_n q^{-n}$
- It is possible to re-write the infinitely long convolution form for deterministic LTI systems in two different forms:
  - **FIR form** (the impulse response dies down exactly after M instants)
  - **Difference equation** (regressive) form (whenever the IR can be parametrized)
- The **FIR form** is known as the **Moving Average** form (of finite-order) while the **difference equation form** is known as the **Auto-Regressive** form

Modelling viewpoints

1. **Time-series modelling is concerned with the estimation of $H(q^{-1})$ and variance of $e[k]$ given the series $v[k]$**

   - Note that $e[k]$ is fictitious input with specific, but unknown, statistical properties
   - This is one of the prime differences between modelling of linear stationary stochastic and LTI deterministic processes.

Modelling viewpoints ... contd.

2. The model in (12) is **not unique** in the sense that $\alpha H(q^{-1})$ and σ_e^2/α^2 also offer an equally suitable description of to this process.

An easy way to resolve this non-uniqueness is by fixing the leading coefficient of the polynomial $H(q^{-1})$ be unity. Stated otherwise,

$$\boxed{h_0 = 1} \tag{13}$$

Modelling viewpoints ... contd.

3. Even with the fixation of $h_0 = 1$, we are presented with the problem of estimating **infinite unknowns** $h_n, \ n = 1, 2, \cdots$.

There are two different ways of overcoming this issue:

i. **Assume** that the process is described by a **finite number of IR coefficients**, i.e.,

$$h[n] = \begin{cases} c_n, & 1 \leq n \leq M, \quad c_n < \infty \\ 0, & n > M \end{cases} \tag{14}$$

leading to the class of **Moving Average** models.

Dealing with infinite unknowns

ii. **Parametrize** $h[.]$ in terms of **finite number of unknowns (parameters).** For e.g.,

$$h[n] = \alpha_1 \lambda_1^n \qquad \text{OR} \qquad h[n] = \sum_{i=1}^{P} \alpha_i \lambda_i^P \tag{15}$$

leading to the class of **difference equation form** or the **Auto-Regressive** models.