Stochastic Signals and Systems

L.Gerencsér Zs. Vágó

Basic concepts

Lecture 1.

Version 3.

September 22, 2020

Wide sense stationary processes I.

A discrete time stochastic process is a sequence of r.v.-s $y = (y_n)$.

Random variables (r..v) are defined over a probability space $(\Omega, \mathcal{F}, P)$.

The subscript n indicates time, with typical range $-\infty < n < +\infty$.

Random variables are real-valued, unless stated otherwise. But: complex valued r.v.-s such as the Fourier transform

$$\sum_n e^{i\omega y_n}$$

will also play major role.

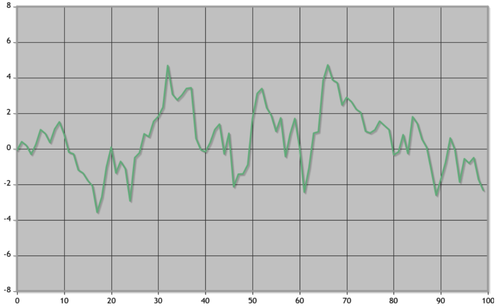
Terminologies: time series in economics and finance

random signals in telecommunication and control.

A basic example

The graph of a simulated AR(1) process defined by $y_n = ay_{n-1} + e_n$.

Here a is called the pole, and (e_n) is i.i.d. It may model a price process.



An AR(1) process with an almost unstable positive pole $a = 0.8$.

Wide sense stationary processes II.

Key features: dependence structure among the r.v.-s y_n.

statistical homogeneity in time.

Our first feature of statistical homogeneity: $\mathbb{E}\, y_n = m$ is constant.

We will in fact assume that $\mathbb{E}\, y_n = 0$ for all n.

Our standing assumption: $\mathbb{E}\,(y_n^2) < +\infty$ or $y_n \in L_2(\Omega, \mathcal{F}, P)$ for all n.

The simplest measure of dependence is the **auto-covariance**:

$$\mathrm{Cov}(y_{n+\tau}, y_n) = \mathbb{E}\,(y_{n+\tau} y_n), \qquad \tau \in \mathbb{Z}.$$

Wide sense stationary processes, III.

Our second feature of statistical homogeneity: for any fixed lag $\tau \in \mathbb{Z}$ the auto-covariances are independent of n. Thus we can write

$$r(\tau) := \mathrm{Cov}(y_{n+\tau}, y_n) = \mathbb{E}\,(y_{n+\tau} y_n), \qquad \tau \in \mathbb{Z}.$$

The function $r(.)$ is called the auto-covariance function.

Definition. A real-valued stochastic process $y = (y_n)$ given on $(\Omega, \mathcal{F}, P)$ is a **wide sense stationary** process, (w.s.st. for short), if

$$\mathbb{E}\,y_n = 0 \quad \text{and} \quad \mathbb{E}\,(y_n^2) < +\infty \quad \text{for all} \quad n,$$

and for any fixed lag $\tau \in \mathbb{Z}$ the auto-covariances $\mathrm{Cov}(y_{n+\tau}, y_n)$ are independent of n.

Complex-valued processes

For complex-valued processes (y_n) our standing assumption is:

$$\mathbb{E}\,|y_n|^2 < +\infty \quad \text{or} \quad y_n \in L_2^c(\Omega, \mathcal{F}, P) \quad \text{for all} \quad n.$$

Assuming $\mathbb{E}\, y_n = 0$ for all n, the auto-covariances are defined as:

$$\mathrm{Cov}(y_{n+\tau}, y_n) = \mathbb{E}\,(y_{n+\tau}\overline{y}_n), \quad \text{for any lag} \quad \tau \in \mathbb{Z}.$$

Definition. A complex-valued stochastic process $y = (y_n)$ given on $(\Omega, \mathcal{F}, P)$ is a **wide sense stationary** process, (w.s.st. for short), if

$$\mathbb{E}\, y_n = 0 \quad \text{and} \quad \mathbb{E}\,|y_n|^2 < +\infty \quad \text{for all} \quad n,$$

and for any fixed lag $\tau \in \mathbb{Z}$ the auto-covariances $\mathrm{Cov}(y_{n+\tau}, y_n)$ are independent of n.

The auto-covariance function

Thus we can define

$$r(\tau) := \mathrm{Cov}(y_{n+\tau}, y_n) = \mathbb{E}\,(y_{n+\tau}\overline{y}_n).$$

The function $r(.)$ is called the auto-covariance function.

For real-valued processes the auto-covariance function is symmetric:

$$r(\tau) = r(-\tau).$$

Note also that $\mathbb{E}\,(y_n^2) = r(0)$.

For complex-valued processes the auto-covariance function $r(.)$ satisfies

$$r(\tau) = \overline{r(-\tau)}.$$

In this case $\mathbb{E}\,|y_n|^2 = r(0)$.

Modelling a damped oscillator, I.

The differential equation for a damped oscillator with external input f is:

$$y'' + 2\zeta\omega_0 y' + \omega_0^2 y = f.$$

Assume for the damping ratio $0 < \zeta < 1$ (underdamping) and $f = 0$.

Then we get an oscillation gradually decreasing to 0.

For $f \neq 0$ we may get oscillation around 0.

Discretization over time yields as a so-called AR(2) process defined by

$$y_n + a_1 y_{n-1} + a_2 y_{n-2} = e_n :$$

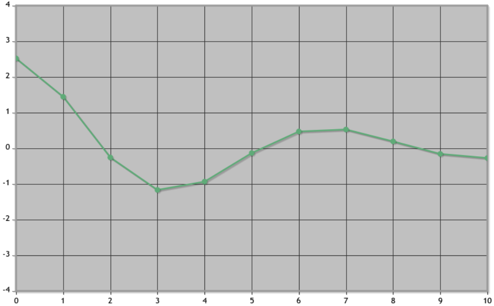
Let us associate with the above dynamics the following polynomial:

$$A(z^{-1}) := 1 + a_1 z^{-1} + a_2 z^{-2}.$$

For $0 < \zeta < 1$ we have complex roots: $A(z^{-1}) = (1 - \alpha z^{-1})(1 - \overline{\alpha} z^{-1})$.

Modelling a damped oscillator, II.

Assuming that (e_n) is i.i.d., the auto-covariance function looks like this:



Auto-covariance of an AR(2) process with complex poles $\alpha = 0.8e^{\pm i0.3\pi}$

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} = (1 - \alpha z^{-1})(1 - \overline{\alpha} z^{-1}).$$

LINEAR OPERATIONS

Linear operations, I.

Let (y_n) be a real-valude w.s.st. process and let us define

$$u_n = a_1 y_{n-1} + \cdots + a_p y_{n-p}, \qquad \text{with } a_k \text{ real.}$$

The process (u_n) is called a moving average of (y_n).

Exercise 1.1. Show that (u_n) is a w.s.st. process. Try first $p = 1$.

Let us compute $\mathbb{E}\, u_n^2$. Squaring and taking expectation we get:

$$\mathbb{E}\, u_n^2 = \sum_{k=1}^{p} \sum_{j=1}^{p} a_k a_j r(j-k).$$

Exercise 1.2. Verify the above equality.

Linear operations, II.

Let $R := (R_{j,k}) = r(k-j)$ and $a = (a_1, ...a_p)^T$

denote a $p \times p$ matrix and a p-vector, resp. Then we can write

$$\mathbb{E}\, u_n^2 = a^T R a.$$

The matrix R is symmetric and positive semi-definite, and has the form:

$$R = \begin{pmatrix} r(0) & r(1) & r(2) & \dots & r(p-1) \\ r(-1) & r(0) & r(1) & \dots & r(p-2) \\ r(-2) & r(-1) & r(0) & \dots & r(p-`3) \\ \vdots & \vdots & \ddots & \ddots & r(1) \\ r(-p+1) & r(-p+2) & \dots & r(-1) & r(0) \end{pmatrix}$$

Note that the elements of R are constant along any sub-diagonal ! Such a matrix is called a **Toeplitz matrix**.

Homework. Set $Y = (y_{n-1}, \dots, y_{n-p})^T$. Show that $R = \mathbb{E}\,(YY^T) \geq 0$.

Orthogonal processes

How do we get a wide sense stationary process?

The simplest example is the w.s.st. **orthogonal process**, say, $e = (e_n)$:

$$\mathbb{E}\; e_n\, e_m = \sigma^2 \delta_{n,m} \qquad \text{for all} \quad n,\; m. \tag{1}$$

In the geometry of the Hilbert space $L_2(\Omega, \mathcal{F}, P)$ we can say:

e_n and e_m are orthogonal for $n \neq m$.

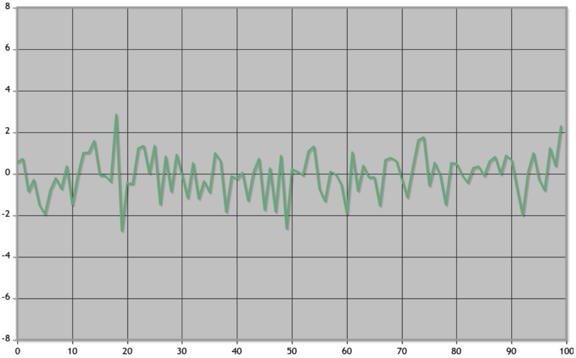
An alternative terminology is that $e = (e_n)$ is a **white noise process**.

Example. An i.i.d. process (e_n) with 0-mean and finite second moment.

In terms of the auto-covariance function we may say that

$$r(\tau) = r^e(\tau) = 0 \qquad \text{for} \quad \tau \neq 0.$$

A simulated orthogonal process



The rough path of an orthogonal process

Example: financial time series

Let S_n denote the price of a stock at time n. Its return is given by

$$y_n = \frac{S_n - S_{n-1}}{S_{n-1}}.$$

A standard approximation to this is its log-return defined by

$$y_n' = \log \frac{S_n}{S_{n-1}} =: e_n.$$

A basic assumption in finance, based on empirical evidence, is that the

log-returns are i.i.d. Gaussian.

We may also assume, that they have 0 mean after discounting. See:

Bachelier, L.: Théorie de la spéculation, 1900.

Samuelson, P. A.: Foundations of Economic Analysis, 1947.

Active suspension

Objective: attenuate the vibration of a car caused by uneven roads.

Modelling: let us take an equidistant spatial mesh, and let (e_n) denote the normalized vertical displacement of the road surface at position n.

We thus assume that (e_n) is a bounded, 0-mean i.i.d. process.

The vertical displacement of a car rolling along this road is (y_n).

Elementary physics implies that the past of e may effect the present of y:

$$y_n = \sum_{k=0}^{\infty} h_k e_{n-k}.$$

Here the h_k-s are the **impulse responses** of the system.

Linear transformations revisited

Focus on linear transformations of a w.s.st. orthogonal process (e_n). Let

$$y_n = \sum_{k=0}^{\infty} h_k e_{n-k}. \qquad (2)$$

To ensure that (y_n) is well defined we use Hilbert space theory.

If $\sum_{k=0}^{\infty} h_k^2 < +\infty$, then the right hand side of (2) converges in $L_2(\Omega, \mathcal{F}, P)$.

Exercise 1.4. Show that $y = (y_n)$ given in (2) is a w.s.st. process.

Moving Average processes

Consider a finite (!) linear combination of an orthogonal process $e = (e_n)$:

$$y_n = \sum_{k=0}^{r} c_k e_{n-k}. \tag{3}$$

The w.s.st. process (y_n) is called a **moving average** os MA process.

If $c_r \neq 0$, then r is **the order**, and (y_n) is called an MA(r) process. Let

$$A(z^{-1}) = \sum_{k=0}^{r} c_k z^{-k}.$$

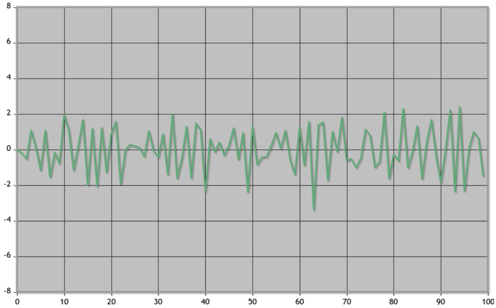
The roots of the equation $A(z^{-1}) = 0$ are called the zeros of the process.

Note on scaling: we may and will assume that $c_0 = 1$.

Example for a MA(1) process, I.

Consider the process $y_n = e_n + c_1 e_{n-1}$. Its zero is then $\gamma = -1/c_1$.

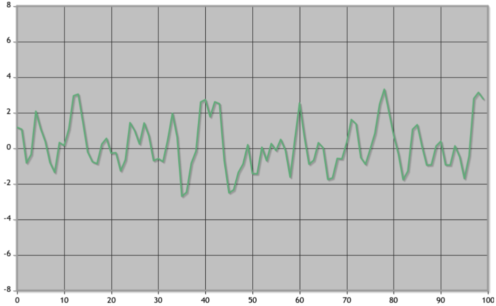
For $c_1 < 0$ we get the weighted difference of the white noise process:



An MA(1) process with a large positive zero $\gamma = 0.9$.

Example for a MA(1) process, II.

For $c_1 > 0$, we get the smoothed average of the white noise process:



An MA(1) process with a large negative zero, $\gamma = 0.9$.

Exercise 1.5 Compute the auto-covariance functions for both examples.

PREDICTION

The problem of prediction

A fundamental problem of the theory of time series:

predict future values of a w.s.st. stochastic process (y_n).

Example: prediction of temperature, precipitation or wind speed;

1 to 10 days ahead. See:

OMSZ, https://www.met.hu/ or

European Centre for Medium-Range Weather Forecasts (ECMWF)

Prediction based on finite past, I.

In practice we have to work with a finite segment of data. The problem: Predict y_n knowing past values $y_{n-1}, y_{n-2}, \ldots, y_{n-p}$. (One-step ahead).

We restrict ourselves to linear predictions of the form:

$$\widehat{y}_n = \sum_{k=1}^{p} \alpha_k y_{n-k}. \qquad (4)$$

The quality of prediction is measured by its mean square error (MSE):

$$J(\alpha) = \mathbb{E}\,(\widehat{y}_n - y_n)^2.$$

Prediction based on finite past, II.

The mean square error (MSE) in detail:

$$J(\alpha) = \mathbb{E}\left(y_n - \sum_{k=1}^{p} \alpha_k y_{n-k}\right)^2 .$$

Minimize $J(\alpha)$ with respect to $\alpha := (\alpha_1, \ldots, \alpha_p)^T$!

Thus we get the least squares or LSQ predictor.

Differentiating $J(\alpha)$ is w.r.t. α_j we get the equation

$$\mathbb{E}\left(y_n - \sum_{k=1}^{p} \alpha_k y_{n-k}\right) y_{n-j} = 0.$$

Prediction based on finite past, III.

The equation above repeated:

$$\mathbb{E}\left(y_n - \sum_{k=1}^{p} \alpha_k y_{n-k}\right) y_{n-j} \;=\; 0.$$

It can be written in the following form:

$$r(j) - \sum_{k=1}^{p} r(j-k)\,\alpha_k = 0.$$

Recall that $r(j-k) = r(k-j) =: (R_{j,k})$. Let $r := (r(1), \ldots, r(p))^T$.

Then we get the following normal equation for $\alpha := (\alpha_1, \ldots, \alpha_p)^T$:

$$R\,\alpha = r.$$

Prediction based on finite past, IV.

Thus we arrive at the following result:

Proposition

Assume that R is non-singular. Then the coefficients of the LSQ linear predictor $\widehat{y}_n = \sum_{k=1}^{p} \alpha_k y_{n-k}$ are obtained as the unique solution of

$$R\,\alpha = r. \tag{5}$$

The case of singular R

Remark: R is nonsingular if and only if R is positive definit. (Why?)

Exercise 1.6 Show that if R is singular (for a given p), then y_n can be predicted with 0 error. (*Hint.* Recall the definition of R.)

A geometric approach, I.

Assume now that the infinite past of (y_n) up to time $n-1$ is known.

This is a matter of convenience for theoretical arguments.

Surprise: this assumption is not as impractical as it looks !

Consider the linear space spanned by $(y_{n-1}, y_{n-2} \ldots)$:

$$\mathcal{L}_{n-1} = \left\{ u : \sum_{k=0}^{p} \alpha_k y_{n-k}, \quad \alpha_k \in \mathbb{R} \quad \text{for some} \quad p \right\}$$

Obviously (?) $\mathcal{L}_{n-1}$ is a linear subspace of the Hilbert-space $L_2(\Omega, \mathcal{F}, P)$.

The latter is equipped with the scalar product $\langle \xi, \eta \rangle = \mathbb{E}\, \xi\, \eta$. Let

$$H_{n-1} := \mathrm{Cl}\, \mathcal{L}_{n-1}$$

be the closure of $\mathcal{L}_{n-1}$ in the metric of $L_2(\Omega, \mathcal{F}, P)$.

Geometric approach, II.

Thus $H_{n-1} = \mathrm{Cl}\,\mathcal{L}_{n-1}$ is a Hilbert subspace of $L_2(\Omega, \mathcal{F}, P)$.

The prediction problem: find the best approximation of y_n in H_{n-1},

in the in the LSQ sense, i.e solve

$$\min_{u \in H_{n-1}} \mathbb{E}\,(y_n - u)^2.$$

It is well-known that the solution $u = \widehat{y}_n$ is given

$$\widehat{y}_n = (y_n | H_{n-1}),$$

the orthogonal projection of y_n on the subspace H_{n-1}.

Geometric approach, III.

In general let H' be closed subspace of a Hilbert space H.

Let $y \in H$ and let $\widehat{y}$ denote its orthogonal projection on H' :

$$\widehat{y} = (y|H').$$

The projection $\widehat{y}$ is uniquely defined by the following two properties:

$$\widehat{y} \in H' \qquad \text{and} \qquad (y - \widehat{y}) \perp u \qquad \forall u \in H',$$

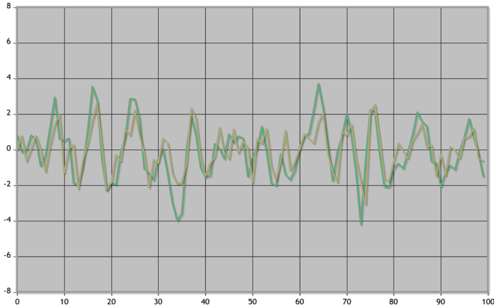
where the symbol $\perp$ stands for orthogonality.

A challenge: can we compute $\widehat{y}_n$ under appropriate conditions ?

We will see: the answer is yes.

Example 1.

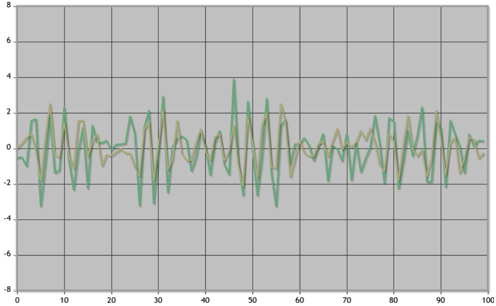
The graphs of an AR(2) process and its LSQ predictor, marked yellow:



LSQ prediction of an AR(2) process with complex poles $0.8e^{\pm 0.3\pi i}$

Example 2.

The graphs of an AR(2) process an its LSQ predictor, marked yellow:



LSQ prediction of an AR(2) process with complex poles $0.8e^{\pm 0.6\pi i}$

End of Lecture 1.