Random Processes and Applications

Jitkomut Songsiri

Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University

CUEE

November 16, 2023

Random Processes and Applications

Jitkomut Songsiri

Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University

CUEE

November 16, 2023

Outline

How to read this handout

1. readers are assumed to have a background on uni-variate random variables and statistics in undergrad level (sophomore year)
2. the note is used with lecture in EE501 (you cannot master this topic just by reading this note) – class lectures include
   - graphical concepts, math derivation of details/steps in between
   - computer codes to illustrate examples
3. pay attention to the symbol ✎; you should be able to prove such ✎ result
4. each chapter has a list of references; find more formal details/proofs from in-text citations
5. almost all results in this note can be Googled; readers are encouraged to 'stimulate neurons' in your brain by proving results without seeking help from the Internet first
6. typos and mistakes can be reported to `jitkomut@gmail.com`

Introduction to Random Processes

Outlines

- definition
- types of random processes
- examples
- statistics
- statistical properties
- analysis of wide-sense stationary process

From RV to RP

extension of how to specify RPs

- definition, elements of RPs
- pdf, cdf, joint pdf
- mean, variance, correlation, other statistics

Types of random processes

- continuous/discrete-valued
- continuous/discrete-time
- stationary/nonstationary

Typical examples

- **Gaussian:** popularly used by its tractability
- **Markov:** population dynamics, market trends, page-rank algorithm
- **Poisson:** number of phone calls in a varying interval
- **White noise:** widely used by its independence property
- **Random walk:** genetic drifts, slowly-varying parameters, neuron firing
- **ARMAX:** time series model in finance, engineering
- **Wiener/Brownian:** movement dynamics of particles

Examples of random signals

- sinusoidal signals with random frequency and phase shift
- a model of signal with additive noise (received = transmitted + noise)
- sum process: $S[n] = X[1] + X[2] + \cdots + X[n]$
- pulse amplitude modulation (PAM)
- random telegraph signal
- electro-cardiogram (ECG, EKG)
- solar/wind power
- stock prices

above examples are used to explain various concepts of RPs

Statistical properties

- stationary processes (strict and wide senses, cyclostationary)
- independent processes
- correlated processes
- ergodic processes

Wide-sense stationary processes

- autocovariance, autocorrelation
- power spectral density
- cross-covariance, cross-correlation
- cross spectrum
- linear system with random inputs
- designing optimal linear filters

Questions involving random processes

- dependency of variables in the random vectors or processes
- probabilities of events in question
- long-term average
- statistical properties of transformed process (under linear system)
- model estimation from data corrupted with noise
- signal/image/video reconstruction from noisy data

and many more questions varied by application of interest

Terminology in Random Processes

- definition and specification of RPs
- statistics: pdf, cdf, mean, variance
- statistical properties: independence, correlation, orthogonal, stationarity

Definition of a random process

elements to be considered:

- let Θ be a random variable (that its outcome, θ is mapped from a sample space S)
- let t be a deterministic value (referred to as 'time') and $t \in T$

definition:

a family (or ensemble) of random variables indexed by t

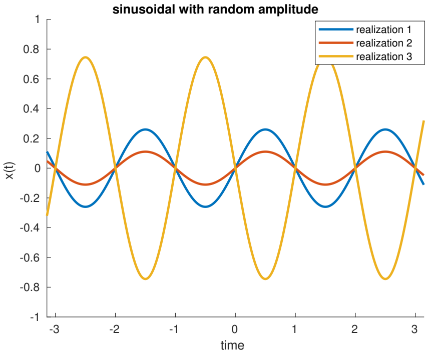
$$\{X(t, \Theta), t \in I\}$$

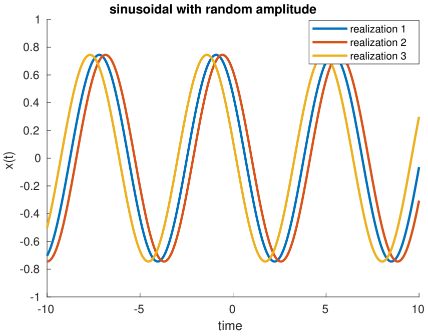
is called a **random (or stochastic) process**

$X(t, \Theta)$ when Θ is fixed, is called *a realization or sample path*

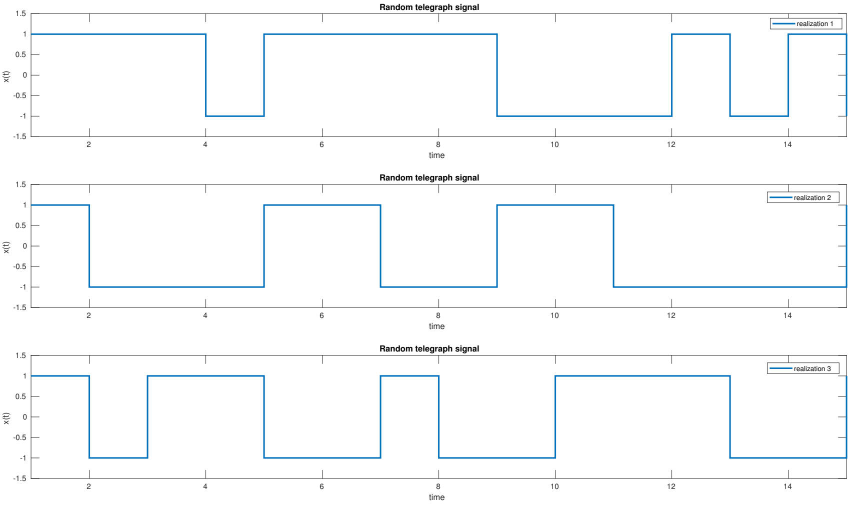
Example: Sinusoidal wave form

sinusoidal wave forms with random amplitude and phase





Example: Random telegraph signal



Specifying RPs

consider an RP $\{X(t,\Theta), t \in T\}$ when Θ is mapped from a sample space S

we often use the notation $X(t)$ to refer to an RP (just drop Θ)

- if T is a countable set then $X(t,\Theta)$ is called **discrete-time** RP
- if T is an uncountable set then $X(t,\Theta)$ is called **continuous-time** RP
- if S is a countable set then $X(t,\Theta)$ is called **discrete-valued** RP
- if S is an uncountable set then $X(t,\Theta)$ is called **continuous-valued** RP

another notation for discrete-time RP is $X[n]$ where n is the time index

From RV to RP

terms	RV	RP
cdf	$F_X(x)$	$F_{X(t)}(x)$
pdf (continuous-valued)	$f_X(x)$	$f_{X(t)}(x)$
pmf (discrete-valued)	$p(x)$	$p(x)$
mean	$m = \mathbf{E}[X]$	$m(t) = \mathbf{E}[X(t)]$
autocorrelation	$\mathbf{E}[X^2]$	$R(t_1, t_2) = \mathbf{E}[X(t_1)X(t_2)]$
variance	$\mathbf{var}[X]$	$\mathbf{var}[X(t)]$
autocovariance		$C(t_1, t_2) = \mathbf{cov}[X(t_1), X(t_2)]$
cross-correlation	$\mathbf{E}[XY]$	$R_{XY}(t_1, t_2) = \mathbf{E}[X(t_1)Y(t_2)]$
cross-covariance	$\mathbf{cov}(X, Y)$	$C_{XY}(t_1, t_2) = \mathbf{cov}[X(t_1), Y(t_2)]$

Distribution functions of RP (time sampled)

let sampling RP $X(t,\Theta)$ at times $t_1, t_2, \ldots, t_k$

$$X_1 = X(t_1,\Theta), \quad X_2 = X(t_2,\Theta), \ldots, X_k = X(t_k,\Theta)$$

this $(X_1, \ldots, X_k)$ is a *vector* RV

cdf of continuous-valued RV

$$F(x_1, x_2, \ldots, x_k) = P[X(t_1) \leq x_1, \ldots, X(t_k) \leq x_k]$$

pdf of continuous-valued RV

$$f(x_1, x_2, \ldots, x_k) dx_1 \cdots dx_k = P[x_1 < X(t_1) < x_1 + dx_1, \ldots, x_k < X(t_k) < x_k + dx_k]$$

PMF of discrete-valued RV

$$p(x_1, x_2, \ldots, x_k) = P[X(t_1) = x_1, \ldots, X(t_k) = x_k]$$

- we have specified distribution functions from any time samples of RV
- the distribution is specified by the collection of kth-order joint cdf/pdf/pmf
- we have droped notation $f_{X_1,\ldots,X_k}(x_1, \ldots, x_k)$ to simply $f(x_1, \ldots, x_k)$

Statistics

the **mean** function of an RP is defined by

$$m(t) = \mathbf{E}[X(t)] = \int_{-\infty}^{\infty} x f_{X(t)}(x) dx$$

the **variance** is defined by

$$\mathbf{var}[X(t)] = \mathbf{E}[(X(t) - m(t))^2] = \int_{-\infty}^{\infty} (x - m(t))^2 f_{X(t)}(x) dx$$

- both mean and variance functions are *deterministic* functions of time
- for discrete-time RV, another notation may be used: $m[n]$ where n is time index

Autocorrelation

the **autocorrelation** of $X(t)$ is the joint moment of RP at different times

$$R(t_1, t_2) = \mathbf{E}[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} xy f_{X(t_1)X(t_2)}(x, y) dx dy$$

the **autocovariance** of $X(t)$ is the covariance of $X(t_1)$ and $X(t_2)$

$$C(t_1, t_2) = \mathbf{E}[(X(t_1) - m(t_1))(X(t_2) - m(t_2))]$$

relations:

- $C(t_1, t_2) = R(t_1, t_2) - m(t_1)m(t_2)$
- $\mathbf{var}[X(t)] = C(t, t)$

another notation for discrete-time RV: $R(m, n)$ or $C(m, n)$ where m, n are (integer) time indices

Joint distribution of RPs

let $X(t)$ and $Y(t)$ be two RPs

let $(t_1, \ldots, t_k)$ and $(\tau_1, \ldots, \tau_k)$ be time samples of $X(t)$ and $Y(t)$, resp.

we specify joint distribution of $X(t)$ and $Y(t)$ from *all possible time choices* of time samples of two RPs

$$\begin{aligned} f_{XY}(x_1, \ldots, x_k, y_1, \ldots, y_k) dx_1 \cdots dx_k dy_1 \cdots dy_k = \\ P[x_1 < X(t_1) \leq x_1 + dx_1, \ldots, x_k < X(t_k) \leq x_k + dx_k, \\ y_1 < Y(\tau_1) \leq y_1 + dy_1, \ldots, y_k < Y(\tau_k) \leq y_k + dy_k] \end{aligned}$$

note that time indices of $X(t)$ and $Y(t)$ need not be the same

Statistics of multiple RPs

the **cross-correlation** of $X(t)$ and $Y(t)$ is defined by

$$R_{XY}(t_1, t_2) = \mathbf{E}[X(t_1)Y(t_2)]$$

(correlation of two RPs at different times)

the **cross-covariance** of $X(t)$ and $Y(t)$ is defined by

$$C_{XY}(t_1, t_2) = \mathbf{E}[(X(t_1) - m_X(t_1))(Y(t_2) - m_Y(t_2))]$$

relation: $C_{XY}(t_1, t_2) = R_{XY}(t_1, t_2) - m_X(t_1)m_Y(t_2)$

Independence, Uncorrelated, Orthogonal

more definitions:

two RPs $X(t)$ and $Y(t)$ are said to be

- **independent** if

  their joint cdf can be written as a product of two marginal cdf's

  mathematically,

$$F_{XY}(x_1, \ldots, x_k, y_1, \ldots, y_k) = F_X(x_1, \ldots, x_k) F_Y(y_1, \ldots, y_k)$$

- **uncorrelated** if

$$C_{XY}(t_1, t_2) = 0, \quad \text{for all } t_1 \text{ and } t_2$$

- **orthogonal** if

$$R_{XY}(t_1, t_2) = 0, \quad \text{for all } t_1 \text{ and } t_2$$

Stationary process

an RP is said to be **stationary** if the kth-order joint cdf's of

$$X(t_1), \dots, X(t_k), \quad \text{and} \quad X(t_1 + \tau), \dots, X(t_k + \tau)$$

are the *same*, for all time shifts τ and all k and all choices of $t_1, \dots, t_k$

in other words, randomness of RP does not change with time

results: a stationary process has the following properties

- the mean is constant and independent of time: $m(t) = m$ for all t
- the variance is constant and independent of time

more results on stationary process:

- the first-order cdf is independent of time

$$F_{X(t)}(x) = F_{X(t+\tau)}(x) = F_X(x), \quad \forall t, \tau$$

- the second-order cdf only depends on the time difference between samples

$$F_{X(t_1),X(t_2)}(x_1, x_2) = F_{X(0),X(t_2-t_1)}(x_1, x_2), \quad \forall t_1, t_2$$

- the autocovariance and autocorrelation can depend only on $t_2 - t_1$

$$R(t_1, t_2) = R(t_2 - t_1), \quad C(t_1, t_2) = C(t_2 - t_1), \quad \forall t_1, t_2$$

Wide-sense stationary process

if an RP $X(t)$ has the following two properties:

- the mean is constant: $m(t) = m$ for all t
- the autocovariance is a function of $t_2 - t_1$ only:

$$C(t_1, t_2) = C(t_1 - t_2), \quad \forall t_1, t_2$$

then $X(t)$ is said to be *wide-sense* stationary (WSS)

- all stationary RPs are wide-sense stationary (converse is not true)
- WSS is related to the concept of spectral density (later discussed)

Independent identically distributed processes

let $X[n]$ be a discrete-time RP and for any time instances $n_1, \ldots, n_k$

$$X_1 = X[n_1], X_2 = X[n_2], \quad X_k = X[n_k]$$

definition: iid RP $X[n]$ consists of a sequence of independent, identically distributed (iid) random variables

$$X_1, X_2, \ldots, X_k$$

with *common* cdf (in other words, same statistical properties)

this property is commonly assumed in applications for simplicity

IID processes

results: an iid process has the following properties

- the joint cdf of any time instances factors to the product of cdf's

$$F(x_1,\ldots,x_k) = P[X_1 \leq x_1,\ldots,X_k \leq x_k] = F(x_1)F(x_2)\cdots F(x_k)$$

- the mean is constant

$$m[n] = \mathbf{E}[X[n]] = m, \quad \forall n$$

- the autocovariance function is a delta function

$$C(n_1,n_2) = 0, \quad \text{for } n_1 \neq n_2, \quad C(n,n) = \sigma^2 \triangleq \mathbf{E}[(X[n]-m))^2]$$

- the autocorrelation function is given by

$$R(n_1,n_2) = C(n_1,n_2) + m^2$$

Independent and stationary increment property

let $X(t)$ be an RP and consider the interval $t_1 < t_2$

defitions:

- $X(t_2) - X(t_1)$ is called the **increment** of RP in the interval $t_1 < t < t_2$
- $X(t)$ is said to have **independent increments** if

$$X(t_2) - X(t_1), X(t_3) - X(t_2), \dots, X(t_k) - X(t_{k-1})$$

  are *independent* RV where $t_1 < t_2 < \cdots < t_k$ (non-overlapped times)
- $X(t)$ is said to have **stationary increments** if

$$P[X(t_2) - X(t_1) = y] = P[X(t_2 - t_1) = y]$$

  the increments in intervals of the same length have the same distribution regardless of when the interval begins results:

- the joint pdf of $X(t_1), \ldots, X(t_k)$ is given by the product of pdf of $X(t_1)$ and the marginals of individual *increments*

we will see this result in the properties of a sum process

Jointly stationary process

$X(t)$ and $Y(t)$ are said to be jointly stationary if the joint cdf's of

$$X(t_1, \ldots, t_k) \;\text{ and }\; Y(\tau_1, \ldots, \tau_k)$$

do not depend on the time origin for all k and all choices of $(t_1, \ldots, t_k)$ and $(\tau_1, \ldots, \tau_k)$

Periodic and Cyclostationary processes

$X(t)$ is called **wide-sense periodic** if there exists $T > 0$,

- $m(t) = m(t+T)$ for all t (mean is periodic)
- $C(t_1, t_2) = C(t_1 + T, t_2) = C(t_1, t_2 + T) = C(t_1 + T, t_2 + T)$,

  for all t_1, t_2, (covariance is periodic in each of *two arguments*)

$X(t)$ is called **wide-sense cyclostationary** if there exists $T > 0$,

- $m(t) = m(t+T)$ for all t (mean is periodic)
- $C(t_1, t_2) = C(t_1 + T, t_2 + T)$ for all t_1, t_2

  (covariance is periodic in *both of two arguments*)

Useful facts

- sample functions of a wide-sense periodic RP are periodic with probability 1

$$X(t) = X(t+T), \quad \text{for all } t$$

  except for a set of outcomes of probability zero
- sample functions of a wide-sense cyclostationary RP need NOT be periodic

examples:

- sinusoidal signal with random amplitude (page 88) is wide-sense cyclostationary and sample functions are periodic
- PAM signal (page 103) is wide-sense cyclostationary but sample functions are not periodic

Stochastic periodicity

definition: a continuous-time RP $X(t)$ is **mean-square periodic** with period T, if

$$\mathbf{E}[(X(t+T) - X(t))^2] = 0$$

let $X(t)$ be a wide-sense stationary RP

$X(t)$ is mean-square periodic if and only if

$$R(\tau) = R(\tau + T), \quad \text{for all } \tau$$

i.e., its autocorrelation function is periodic with period T

Ergodic random process

the **time average** of a realization of a WSS RP is defined by

$$\langle x(t) \rangle = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t) dt$$

the time-average autocorrelation function is defined by

$$\langle x(t)x(t+\tau) \rangle = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t+\tau) dt$$

- if the time average is equal to the ensemble average, we say the RP is **ergodic in mean**
- if the time-average autocorrelation is equal to ensemble autocorrelation then the RP is **ergodic in the autocorrelation**

Definition of ergodic RP

a WSS RP is **ergodic** if ensemble averages can be calculated using time averages of any realization of the process

- ergodic in mean: $\langle x(t) \rangle = \mathbf{E}[X(t)]$
- ergodic in autocorrelation: $\langle x(t)x(t+\tau) \rangle = \mathbf{E}[X(t)X(t+\tau)]$

calculus of random process (derivative, integrals) is discussed in mean-square sense

see Leon-Garcia, Section 9.7.2-9.7.3

References

1. Chapter 9-11 in A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd edition, Pearson Prentice Hall, 2009
2. Chapter 8-10 in H. Stark and J. W. Woods, *Probability, Statistics, and Random Processes for Engineers*, 4th edition, Pearson, 2012

Important random processes

Outlines

definitions, properties, and applications

- **Random walk:** genetic drifts, slowly-varying parameters, neuron firing
- **Gaussian:** popularly used by its tractability
- **Wiener/Brownian:** movement dynamics of particles
- **White noise:** widely used by its independence property
- **Markov:** population dynamics, market trends, page-rank algorithm
- **Poisson:** number of phone calls in a varying interval
- **ARMAX:** time series model in finance, engineering

Bernoulli random process

a (time) sequence of indepenent Bernoulli RV is an iid Bernoulli RP

example:

- $I[n]$ is an indicator function of the event at time n where $I[n] = 1$ when success and $I[n] = 0$ when fail
- let $D[n] = 2I[n] - 1$ and it is called **random step** process

$$D[n] = 1 \text{ or } -1$$

$D[n]$ can represent the *deviation* of a particle movement along a line

Sum process

the sum of a sequence of iid random variables, $X_1, X_2, \ldots$

$$S[n] = X_1 + X_2 + \cdots + X_n, \quad n = 1, 2, \ldots$$

where $S[0] = 0$, is called the **sum process**

- we can write $S[n] = S[n-1] + X_n$ (recursively)
- the sum process has *independent increments* in nonoverlapping intervals

$$S[n] - S[n-1] = X_n, \quad S[n-1] - S[n-2] = X_{n-1}, \ldots, S[2] - S[1] = X_2$$

  (since X_k's are iid)
- the sum process has *stationary increments*

$$P(S[n] - S[k] = y) = P(S[n-k] = y), \quad n > k$$

Autocovariance of a sum process

- assume X_k's have mean m and variance σ^2
- $\mathbf{E}[S[n]] = nm$ (X_k's are iid)
- $\mathbf{var}[S[n]] = n\sigma^2$ (X_k's are iid)

we can show that

$$C(n,k) = \min(n,k)\sigma^2$$

the proof follows from letting $n \leq k$, and so $n = \min(n,k)$

$$\begin{aligned}
C(n,k) &= \mathbf{E}[(S[n]-nm)(S[k]-km)] \\
&= \mathbf{E}[(S[n]-nm)\{(S[n]-nm)+(S[k]-km)-(S[n]-nm)\}] \\
&= \mathbf{E}[(S[n]-nm)^2] + \mathbf{E}[(S[n]-nm)(S[k]-S[n]-(k-n)m)] \\
&= \mathbf{E}[(S[n]-nm)^2] + \mathbf{E}[(S[n]-nm)]\mathbf{E}[(S[k]-S[n]-(k-n)m)]
\end{aligned}$$

(apply that $S[n]$ has independent increments and $\mathbf{E}[S[n]-nm]=0$)

Properties of a sum process

- the joint pdf/pmf of $S(1), \ldots, S(n)$ is given by the product of pdf of $S(1)$ and the marginals of individual *increments*
  - X_k's are integer-valued
  - X_k's are continuous-valued
- the sum process is a Markov process (more on this)

Binomial counting process

let $I[n]$ be iid Bernoulli random process

the sum process $S[n]$ of $I[n]$ is then the **counting process**

- it gives the number successses in the first n Bernoulli trial
- the counting process is an increasing function
- $S[n]$ is binomial with parameter p (probability of success)

Random walk

let $D[n]$ be iid random step process where

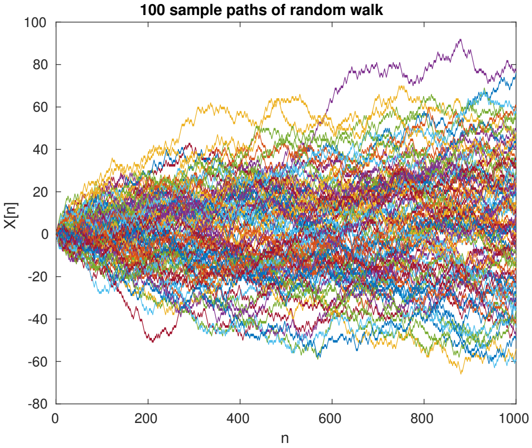
$$D[n] = \begin{cases} 1, & \text{with probability } p \\ -1, & \text{with probability } p \end{cases}$$

the random walk process $X[n]$ is defined by

$$X[0] = 0, \quad X[n] = \sum_{k=1}^{n} D[k], \quad k \geq 1$$

- the random walk is a sum process
- we can show that $\mathbf{E}[X[n]] = n(2p - 1)$
- the random walk has a tendency to either grow if $p > 1/2$ or to decrease if $p < 1/2$

a random walk example as the sum of Bernoulli sequences with $p = 1/2$



$\mathbf{E}[X(n)] = 0$ and $\mathbf{var}[X(n)] = n$ (variance grows over time)

Properties of a random walk

- $X[n]$ has **independent stationary increments** in nonoverlapping time intervals

$$P[X[m] - X[n] = y] = P[X[m-n] = y]$$

(increments in intervals of the same length have the same distribution)

- a random walk is related to an **autoregressive process** since

$$X[n+1] = X[n] + D[n+1]$$

(widely used to model financial time series, biological signals, etc)

$$\text{stock price:} \quad \log X[n+1] = \log X[n] + \beta D[n+1]$$

- extension: if $D[n]$ is a Gaussian process, we say $X[n]$ is a **Gaussian random walk**

Gaussian process

an RP $X(t)$ is a **Gaussian** process if the samples

$$X_1 = X(t_1), X_2 = X(t_2), \quad X_k = X(t_k)$$

are jointly Gaussian RV for all k and all choices of $t_1, \ldots, t_k$

that is the joint pdf of samples from time instants is given by

$$f_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-(1/2)(x-m)^T\Sigma^{-1}(x-m)}$$

$$m = \begin{bmatrix} m(t_1) \\ m(t_2) \\ \vdots \\ m(t_k) \end{bmatrix}, \Sigma = \begin{bmatrix} C(t_1,t_1) & C(t_1,t_2) & \cdots & C(t_1,t_k) \\ C(t_2,t_1) & C(t_2,t_2) & \cdots & C(t_2,t_k) \\ \vdots & \vdots & & \vdots \\ C(t_k,t_1) & \cdots & & C(t_k,t_k) \end{bmatrix}$$

Properties of Gaussian processes

- Gaussian RPs are specified completely by the mean and covariance functions
- Gaussian RPs can be both continuous-time and discrete-time
- linear operations on Gaussian RPs preserve Gaussian properties

Example of Gaussian process I

let $X(t)$ be a zero-mean Gaussian RP with

$$C(t_1, t_2) = 4e^{-3|t_1 - t_2|}$$

find the joint pdf of $X(t)$ and $X(t+s)$
we see that

$$C(t, t+s) = 4e^{-3s}, \quad \mathbf{var}[X(t)] = C(t,t) = 4$$

therefore, the joint of pdf of $X(t)$ and $X(t+s)$ is the Gaussian distribution parametrized by

$$f_{X(t),X(t+s)}(x_1, x_2) = \frac{1}{(2\pi)|\Sigma|^{1/2}} e^{-(1/2)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}, \quad \Sigma = \begin{bmatrix} 4 & 4e^{-3s} \\ 4e^{-3s} & 4 \end{bmatrix}$$

Example of Gaussian process II

let $X(t)$ be a Gaussian RP and let $Y(t) = X(t+d) - X(t)$

- mean of $Y(t)$ is

$$m_y(t) = \mathbf{E}[Y(t)] = m_x(t+d) - m_x(t)$$

- the autocorrelation of $Y(t)$ is

$$\begin{aligned} R_y(t_1, t_2) &= \mathbf{E}[(X(t_1+d) - X(t_1))(X(t_2+d) - X(t_2))] \\ &= R_x(t_1+d, t_2+d) - R_x(t_1+d, t_2) - R_x(t_1, t_2+d) + R_x(t_1, t_2) \end{aligned}$$

- the autocovariance of $Y(t)$ is then

$$\begin{aligned} C_y(t_1, t_2) &= \mathbf{E}[(X(t_1+d) - X(t_1) - m_y(t_1))(X(t_2+d) - X(t_2) - m_y(t_2))] \\ &= C_x(t_1+d, t_2+d) - C_x(t_1+d, t_2) - C_x(t_1, t_2+d) + C_x(t_1, t_2) \end{aligned}$$

since $Y(t)$ is the sum of two Gaussians then $Y(t)$ must be Gaussian

- any k-time samples of $Y(t)$

$$Y(t_1), Y(t_2), \ldots, Y(t_k)$$

  is linear transformation of jointly Gaussians, so $Y(t_1), \ldots, Y(t_k)$ have jointly Gaussian pdf
- for example, find joint pdf of $Y(t)$ and $Y(t+s)$: need only mean and covariance
  - $m_y(t)$ and $m_y(t+s)$
  - covariance is given by

$$\Sigma = \begin{bmatrix} C_y(t,t) & C_y(t,t+s) \\ C_y(t,t+s) & C_y(t+s,t+s) \end{bmatrix}$$

Wiener process

consider the random step on page 43

symmetric walk ($p = 1/2$), magnitude step of M, time step of h seconds

let $X_h(t)$ be the accumulated sum of random step up to time t

- $X_h(t) = M(D[1] + D[2] + \cdots + D[n]) = MS[n]$ where $n = [t/h]$
- $\mathbf{E}[X_h(t)] = 0$
- $\mathbf{var}[X_h(t)] = M^2 n$

Wiener process $X(t)$: obtained from $X_h(t)$ by shrinking the magnitude and time step to zero in a *precise way*

$$h \to 0, \quad M \to 0, \quad \text{with } M = \sqrt{\alpha h} \text{ where } \alpha > 0 \text{ is constant}$$

(meaning; if $v = M/h$ represents a particle speed then $v \to \infty$ as displacement M goes to 0)

Properties of Wiener (Wiener-Levy) process

- $\mathbf{E}[X(t)] = 0$ (zero mean of all time)
- $\mathbf{var}[X(t)] = (\sqrt{\alpha h})^2 \cdot (t/h) = \alpha t$ (stays finite and nonzero)
- $X(t) = \lim_{h \to 0} M(D[1] + \cdots + D[n]) = \lim_{n \to \infty} \sqrt{\alpha t} \frac{S[n]}{\sqrt{n}}$ approaching the sum of an *infinite* number of RV
- by CLT, pdf $X(t)$ approaches Gaussian with mean zero and variance αt

$$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-\frac{x^2}{2\alpha t}}$$

- $X(t)$ has independent stationary increments (from random walk form)
- Wiener process is a **Gaussian** random process ($X(t_k)$ is obtained as linear transformation of increments))

Properties of Wiener (Wiener-Levy) process

- used to model **Brownian motion** (movement of particles in fluid)
- the covariance function of Wiener process is

$$C(t_1, t_2) = \alpha \min(t_1, t_2), \quad \alpha > 0$$

to show this, let $t_1 \geq t_2$,

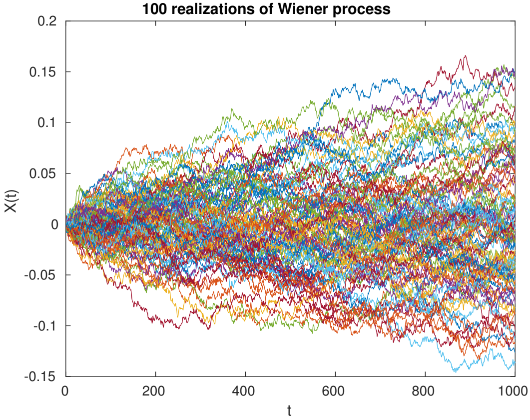
$$\begin{aligned} C(t_1, t_2) &= \mathbf{E}[X(t_1)X(t_2)] = \mathbf{E}[(X(t_1) - X(t_2) + X(t_2))X(t_2)] \\ &= \mathbf{E}[(X(t_1) - X(t_2))X(t_2)] + \mathbf{var}[X(t_2)] \\ &= 0 + \alpha t_2 \end{aligned}$$

using $X(t_1) - X(t_2)$ and $X(t_2)$ are independent (when $t_1 \geq t_2$)

if $t_2 < t_1$, we do the same and obtain $C(t_1, t_2) = \alpha t_1$

Sample paths of Wiener process

when $\alpha = 2$



$\mathbf{E}[X(t)] = 0$ and $\mathbf{var}[X(t)] = \alpha t$ (variance grows over time)

White noise process

definition: a random process $X(t)$ is white noise if

- $\mathbf{E}[X(t)] = 0$ (zero mean for all t)
- $\mathbf{E}[X(t)X(s)] = 0$ for $t \neq s$ (uncorrelated with another time sample)

in another word,

- the correlation function of a white noise is an **impulse function**

$$R(t_1, t_2) = \alpha\delta(t_1 - t_2), \quad \alpha > 0$$

- power spectral density is flat (more on this): $S(\omega) = \alpha, \quad \forall\omega$
- $X(t)$ has infinite power, varies extremely rapidly in time, and is most unpredictable

White noise and Wiener processes

those two properties of white noise are derived from the definition that

white Gaussian noise process is the *time derivative* of Wiener process

recall the correlation of Wiener process is $R_{\text{wiener}}(t_1, t_2) = \alpha \min(t_1, t_2)$

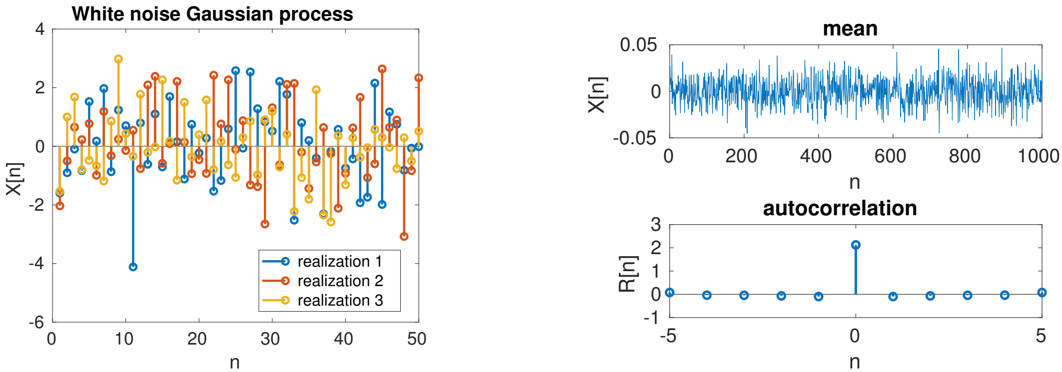
$$
\begin{aligned}
R(t_1, t_2) = \mathbf{E}[X(t_1)X(t_2)] &= \mathbf{E}\left[\frac{\partial}{\partial t_1} X_{\text{wiener}}(t_1) \cdot \frac{\partial}{\partial t_2} X_{\text{wiener}}(t_2)\right] \\
&= \frac{\partial}{\partial t_1}\frac{\partial}{\partial t_2} R_{\text{wiener}}(t_1, t_2) = \frac{\partial}{\partial t_1}\frac{\partial}{\partial t_2} \begin{cases} \alpha t_2, & t_2 < t_1 \\ \alpha t_1, & t_2 \geq t_1 \end{cases} \\
&= \frac{\partial}{\partial t_1} \alpha u(t_1 - t_2), \quad u \text{ is the step function}
\end{aligned}
$$

but u is not differentiable at $t_1 = t_2$, so the second derivative does not exist
instead, we generalize this notion using delta function

$$R(t_1, t_2) = \alpha \delta(t_1 - t_2)$$

Example of white noise

white noise Gaussian process with variance 2



- mean function (averaged over 10,000 realizations) is close to zero
- sample autocorrelation is close to a delta function where $R(0) \approx 2$

Poisson process

let $N(t)$ be the number of event occurrences in the interval $[0, t]$

properties:

- non-decreasing function (of time t)
- integer-valued and continuous-time RP

assumptions:

- events occur at an average rate of λ events per seconds
- the interval $[0, t]$ is divided into n subintervals and let $h = t/n$
- the probability of *more than one event* occurrences in a subinterval is negligible compared to the probability of observing one or zero events
- whether or not an event occurs in a subinterval is independent of the outcomes in other subintervals

Meaning of Poisson process

- the outcome in each subinterval can be viewed as a Bernoulli trial
- these Bernoulli trials are independent
- $N(t)$ can be *approximated* by the binomial counting process

Binomial counting process:

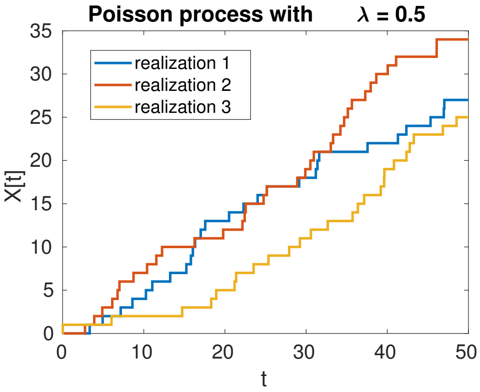
- let the probability of an event occurrence in subinterval is p
- average number of events in $[0, t]$ is $\lambda t = np$
- let $n \to \infty$ $(h = t/n \to 0)$ and $p \to 0$ while $np = \lambda t$ is fixed
- from the following approximation when n is large

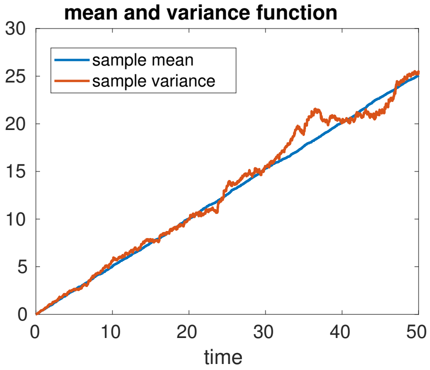
$$P(N(t) = k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k} \approx \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \ldots$$

$N(t)$ has a Poisson distribution and is called a **Poisson process**

Example of Poisson process ($\lambda = 0.5$)

- generated by taking cumulative sum of n-sequence Bernoulli with and $p = \lambda T/n$ where $n = 1000$ and $T = 50$





- the rate of Poisson process grows as λt for $t \in [0, T]$
- the mean and variance functions (approximate over 100 runs) have linear trend

Poisson process: joint pmf

joint pmf: for $t_1 < t_2$,

$$\begin{aligned}
P[N(t_1) = i, N(t_2) = j] &= P[N(t_1) = i]P[N(t_2) - N(t_1) = j - i] \\
&= P[N(t_1) = i]P[N(t_2 - t_1) = j - i] \\
&= \frac{(\lambda t_1)^i e^{-\lambda t_1}}{i!} \frac{(\lambda(t_2 - t_1))^j e^{-\lambda(t_2 - t_1)}}{(j - i)!}
\end{aligned}$$

autocovariance: $C(t_1, t_2) = \lambda \min(t_1, t_2)$

for $t_1 \leq t_2$,

$$\begin{aligned}
C(t_1, t_2) &= \mathbf{E}[(N(t_1) - \lambda t_1)(N(t_2) - \lambda t_2)] \\
&= \mathbf{E}[(N(t_1) - \lambda t_1)\{N(t_2) - N(t_1) - \lambda t_2 + \lambda t_1 + (N(t_1) - \lambda t_1)\}] \\
&= \mathbf{E}[(N(t_1) - \lambda t_1)]\mathbf{E}[(N(t_2) - N(t_1) - \lambda(t_2 - t_1)] + \mathbf{var}[N(t_1)] \\
&= \lambda t_1
\end{aligned}$$

we have used *independent and stationary increments* property

Applications of Poisson processes

examples:

- random telegraph signal
- the number of car accidents at a site or in an area
- the requests for individuals documents on a web server
- the number of customers arriving at a store

Time between events in Poisson process

let T be the time between event occurrences in a Poisson process

- the probability involving T follows

$$P[T > t] = P[\text{no events in } t \text{ seconds}] = (1-p)^n$$
$$= \left(1 - \frac{\lambda t}{n}\right)^n \to e^{-\lambda t}, \quad \text{as } n \to \infty$$

  T is an exponential RV with parameter λ
- the interarrival time in the underlying binomial proces are independent geometric RV
- the sequence of interarrival times $T[n]$ in a Poisson process form an iid sequence of *exponential* RVs with mean $1/\lambda$
- the sum $S[n] = T[1] + \cdots + T[n]$ has Erlang distribution

Markov process

for any time instants, $t_1 < t_2 < \cdots < t_k < t_{k+1}$, if

discrete-valued

$$P[X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k, \ldots, X(t_1) = x_1] = P[X(t_{k+1} = x_{k+1} \mid X(t_k) = x_k]$$

continuous-valued

$$f(x_{k+1} \mid X(t_k) = x_k, \ldots, X(t_1) = x_1) = f(x_{k+1} \mid X(t_k) = x_k)$$

then we say $X(t)$ is a **Markov** process

joint pdf conditioned on several time instants reduce to pdf conditioned on the **most recent** time instant

Properties of Markov process

- pmf and pdf of Markov processes are conditioned on several time instants can reduce to pmf/pdf that is only conditioned on the *most recent* time instant
- an integer-valued Markov process is called a **Markov chain** (more details on this)
- the sum of iid sequence where $S[0] = 0$ is a Markov process
- a Poisson process is a continuous-time Markov process
- a Wiener process is a continuous-valued Markov process
- in fact, any **independent-increment** process is also Markov

to apply the **independent-increment** property, consider a discrete-valued RP,

$$\begin{aligned}
&P[X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k, \ldots, X(t_1) = x_1] \\
&\quad = P[X(t_{k+1}) - X(t_k) = x_{k+1} - x_k \mid X(t_k) = x_k, \ldots, X(t_1) = x_1] \\
&\quad = P[X(t_{k+1}) - X(t_k) = x_{k+1} - x_k \mid X(t_k) = x_k] \quad \text{by independent increments} \\
&\quad = P[X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k]
\end{aligned}$$

more examples of Markov process

- birth-death Markov chains: transitions only between adjacent states are allowed

$$p(t+1) = Pp(t), \quad P \text{ is tri-diagonal}$$

- M/M/1 queue (a queuing model): continuous-time Markov chain

$$\dot{p}(t) = Qp(t)$$

Discrete-time Markov chain

a Markov chain is a random sequence that has n possible states:

$$X(t) \in \{1, 2, \dots, n\}$$

with the property that

$$\mathbf{prob}(\ X(t+1) = i \ | \ X(t) = j\ ) = p_{ij}$$

where $P = [p_{ij}] \in \mathbf{R}^{n \times n}$

- p_{ij} is the **transition probability** from state j to state i
- P is called the **transition matrix** of the Markov chain
- the state $X(t)$ still cannot be determined with *certainty*
- $\{1, 2, \dots, n\}$ is called *label* (simply mapped to integers)

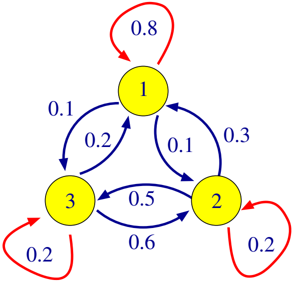
Example of DT Markov chain

a customer may rent a car from any of three locations and return to any of the three locations

Rented from location

1	2	3	
0.8	0.3	0.2	**1**
0.1	0.2	0.6	**2**
0.1	0.5	0.2	**3**

Returned to location



Properties of transition matrix

let P be the transition matrix of a Markov chain

- all entries of P are real *nonnegative* numbers
- the entries in any column are summed to 1 or $\mathbf{1}^T P = \mathbf{1}^T$:

$$p_{1j} + p_{2j} + \cdots + p_{nj} = 1$$

  (a property of a **stochastic matrix**)
- 1 is an eigenvalue of P
- if q is an eigenvector of P corresponding to eigenvalue 1, then

$$P^k q = q, \quad \text{for any } k = 0, 1, 2, \ldots$$

Probability vector

we can represent probability distribution of $x(t)$ as n-vector

$$p(t) = \begin{bmatrix} \mathbf{prob}(\ x(t) = 1\ ) \\ \vdots \\ \mathbf{prob}(\ x(t) = n\ ) \end{bmatrix}$$

- $p(t)$ is called a **state probability vector** at time t
- $\sum_{i=1}^{n} p_i(t) = 1$ or $\mathbf{1}^T p(t) = 1$
- the state probability propagates like a linear system:

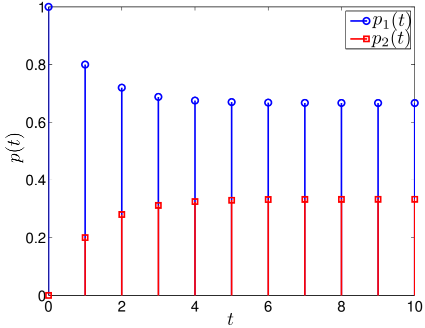
$$p(t+1) = Pp(t)$$

- the state PMF at time t is obtained by multiplying the initial PMF by P^t

$$p(t) = P^t p(0), \quad \text{for } t = 0, 1, \ldots$$

Example: a Markov model for packet speech

- two states of packet speech: contain 'silent activity' or 'speech activity'
- the transition matrix is $P = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix}$
- the initial state probability is $p(0) = (1, 0)$
- the packet in the first state is 'silent' with certainty



- eigenvalues of P are 1 and 0.4
- calculate P^t by using 'diagonalization' or 'Cayley-Hamilton theorem' or diagonalization approach

$$P^t = \begin{bmatrix} (5/3)(0.4 + 0.2 \cdot 0.4^t) & (2/3)(1 - 0.4^t) \\ (1/3)(1 - 0.4^t) & (5/3)(0.2 + 0.4^{t+1}) \end{bmatrix}$$

- $P^t \to \begin{bmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{bmatrix}$ as $t \to \infty$ (all columns are the same in limit!)
- $\lim_{t\to\infty} p(t) = \begin{bmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} p_1(0) \\ 1 - p_1(0) \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$

$p(t)$ does not depend on the *initial state probability* as $t \to \infty$

what if $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$?

- we can see that

$$P^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad P^3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \ldots$$

- P^t does not converge but oscillates between two values

under what condition $p(t)$ converges to a constant vector as $t \to \infty$?

definition: a transition matrix is **regular** if some integer power of it has all *positive* entries

fact: if P is regular and let w be *any* probability vector, then

$$\lim_{t\to\infty} P^t w = q$$

where q is a **fixed** probability vector, independent of t

Steady state probabilities

we are interested in the **steady state probability vector**

$$q = \lim_{t \to \infty} p(t) \qquad \text{(if converges)}$$

- the steady-state vector q of a regular transition matrix P satisfies

$$\lim_{t \to \infty} p(t+1) = P \lim_{t \to \infty} p(t) \qquad \Longrightarrow \qquad Pq = q$$

  (in other words, q is an eigenvector of P corresponding to eigenvalue 1)

- if we start with $p(0) = q$ then

$$p(t) = P^t p(0) = 1^t q = q, \quad \textbf{for all } t$$

  q is also called the **stationary state PMF** of the Markov chain

Example: weather model

probabilities of weather conditions given the weather on the preceding day:

$$P = \begin{bmatrix} 0.4 & 0.2 \\ 0.6 & 0.8 \end{bmatrix}$$

(probability that it will rain tomorrow given today is sunny, is 0.2)

given today is sunny with probability 1, calculate the probability of a rainy day in long term

Gauss-Markov process

let $W[n]$ be a white Gaussian noise process with $W[1] \sim \mathcal{N}(0, \sigma^2)$

definition: a Gauss-Markov process is a first-order autoregressive process

$$X[1] = W[1], \quad X[n] = aX[n-1] + W[n], \quad n \geq 1, \quad |a| < 1$$

- clearly, $X[n]$ is Markov since the state $X[n]$ only depends on $X[n-1]$
- $X[n]$ is Gaussian because if we let

$$X_k = X[k], \quad W_k = W[k], \quad k = 1, 2, \ldots, n \quad \text{(time instants)}$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n-1} \\ X_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ a & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a^{n-2} & a^{n-3} & \vdots & 1 & 0 \\ a^{n-1} & a^{n-2} & \cdots & a & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_{n-1} \\ W_n \end{bmatrix}$$

pdf of $(X_1, \ldots, X_n)$ is Gaussian for all n

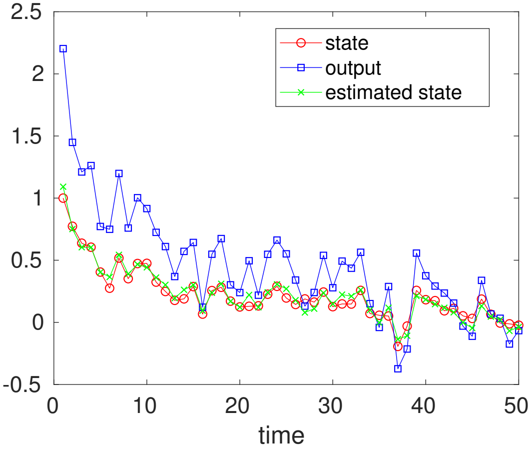
Questions involving a Gauss-Markov process

setting:

- we can observe $Y[n] = X[n] + V[n]$ where V represents a sensor noise
- only Y can be observed, but we do not know X

question: can we estimate $X[n]$ from information of $Y[n]$ and statistical properties of W and V?

solution: yes we can. one choice is to apply a Kalman filter example: $a = 0.8$, $Y[k] = 2X[k] + V[k]$



$X[k]$ is estimated by Kalman filter

References

1. Chapter 9 in A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd edition, Pearson Prentice Hall, 2009
2. Chapter 9 in H. Stark and J. W. Woods, *Probability, Statistics, and Random Processes for Engineers*, 4th edition, Pearson, 2012

Examples of random processes

Outlines

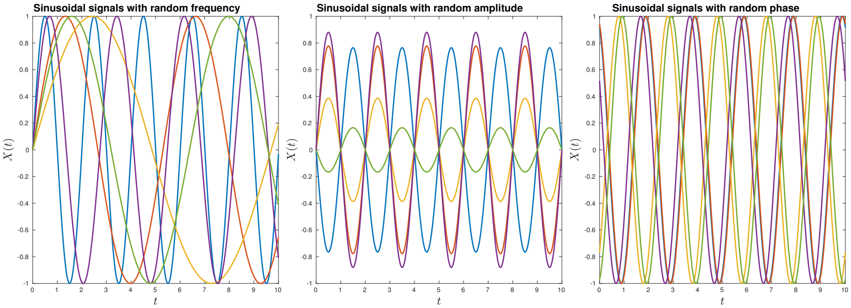
- sinusoidal signals
- random telegraph signals
- signal plus noise
- ARMA time series

Sinusoidal signals

consider a signal of the form

$$X(t) = A\sin(\omega t + \phi)$$

- randomness occurrs in in each of following settings: random frequency, random amplitude, random phase
- questions involving this example: find pdf, mean, variance, correlation function



Sinusoidal signal: random amplitude

$A \in \mathcal{U}[-1, 1]$ while $\omega = \pi$ and $\phi = 0$

$$X(t) = A\sin(\pi t)$$

(continuous-valued RP and sample function is periodic)

- find pdf of $X(t)$
  - when t is integer, we see $X(t) = 0$ for all A

$$P(X(t) = 0) = 1, \quad P(X(t) = \text{other values}) = 0$$

  - when t is not integer, $X(t)$ is just a scaled uniform RV

$$X(t) \in \mathcal{U}[-\sin(\pi t), \sin(\pi t)]$$

Sinusoidal signal: random amplitude

- find mean of $X(t)$

$$m(t) = \mathbf{E}[X(t)] = \mathbf{E}[A]\sin(\pi t)$$

(could have zero mean if $\sin(\pi t) = 0$)

- find correlation function

$$R(t_1, t_2) = \mathbf{E}[A\sin(\pi t_1)A\sin(\pi t_2)] = \mathbf{E}[A^2]\sin(\pi t_1)\sin(\pi t_2)$$

- find covariance function: $C(t_1, t_2) = R(t_1, t_2) - m(t_1)m(t_2)$

$$\begin{aligned} C(t_1, t_2) &= \mathbf{E}[A^2]\sin(\pi t_1)\sin(\pi t_2) - (\mathbf{E}[A])^2\sin(\pi t_1)\sin(\pi t_2) \\ &= \mathbf{var}[A]\sin(\pi t_1)\sin(\pi t_2) \end{aligned}$$

- $X(t)$ is wide-sense **cyclostationary**, *i.e.*, $m(t) = m(t+T)$ and $C(t_1, t_2) = C(t_1 + T, t_2 + T)$ for some T

Sinusoidal signal: random phase shift

$A = 1, \omega = 1$ and $\phi \sim \mathcal{U}[-\pi, \pi]$

$$X(t) = \sin(t + \phi) \qquad \text{(continuous-valued RP)}$$

- find pdf of $X(t)$: view $x = \sin(t + \phi)$ as a transformation of ϕ

$$x = \sin(t + \phi) \Leftrightarrow \phi_1 = \sin^{-1}(x) - t, \phi_2 = \pi - \sin^{-1}(x) - t$$

the pdf of $X(t)$ can be found from the formula

$$f_{X(t)}(x) = \sum_k f(\phi_k) \left| \frac{d\phi}{dx} \right|_{\phi = \phi_k} = \frac{1}{\pi\sqrt{1 - x^2}}, \quad -1 \leq x \leq 1$$

(pdf of $X(t)$ does not depend on t; hence, $X(t)$ is **strict-sense stationary**)

Sinusoidal signal: random phase shift

- find the mean function

$$\mathbf{E}[X(t)] = \mathbf{E}[\sin(t+\phi)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \sin(t+\phi)d\phi = 0$$

- find the covariance function

$$\begin{aligned} C(t_1,t_2) = R(t_1,t_2) &= \mathbf{E}[\sin(t_1+\phi)\sin(t_2+\phi)] \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{2}[\cos(t_1-t_2) - \cos(t_1+t_2+2\phi)]d\phi \\ &= (1/2)\cos(t_1-t_2) \end{aligned}$$

(depend only on $t_1 - t_2$)

- $X(t)$ is **wide-sense stationary** (also conclude from the fact that $X(t)$ is stationary)

Random telegraph signal

a signal $X(t)$ takes values in $\{1, -1\}$ randomly in the following setting:

- $X(0) = 1$ or $X(0) = -1$ with equal probability of $1/2$
- $X(t)$ changes the sign with each occurrence follows a Poisson process of rate α

questions involving this example:

- obviously, $X(t)$ is a discrete-valued RP, so let's find its pmf
- find the covariance function
- examine its stationary property

Random telegraph: PMF

based on the fact that

$$X(t) \text{ has the } \textit{same} \text{ sign as } X(0) \iff \text{number of sign changes in } [0,t] \text{ is } \textit{even}$$

$$X(t) \text{ and } X(0) \textit{ differ} \text{ in sign} \iff \text{number of sign changes in } [0,t] \text{ is } \textit{odd}$$

$$P(X(t)=1) = \underbrace{P(X(t)=1|X(0)=1)}_{\text{no. of sign change is even}} P(X(0)=1)$$

$$+ \underbrace{P(X(t)=1|X(0)=-1)}_{\text{no. of sign change is odd}} P(X(0)=-1)$$

let $N(t)$ be the number of sign changes in $[0,t]$ (which is Poisson)

$$P(N(t) = \text{ even integer}) = \sum_{k=0}^{\infty} \frac{(\alpha t)^{2k}}{(2k)!} e^{-\alpha t} = (1/2)(1+e^{-2\alpha t})$$

$$P(N(t) = \text{ odd integer}) = \sum_{k=0}^{\infty} \frac{(\alpha t)^{2k+1}}{(2k+1)!} e^{-\alpha t} = (1/2)(1-e^{-2\alpha t})$$

Random telegraph: PMF

pmf of $X(t)$ is then obtained as

$$\begin{aligned} P(X(t)=1) &= (1/2)(1+e^{-2\alpha t})(1/2) + (1/2)(1-e^{-2\alpha t})(1/2) \\ &= 1/2 \\ P(X(t)=-1) &= 1 - P(X(t)=1) = 1/2 \end{aligned}$$

- pmf of $X(t)$ does not depend on t
- $X(t)$ takes values in $\{-1, 1\}$ with equal probabilities

if $X(0) = 1$ with probability $p \neq 1/2$ then how the pmf of $X(t)$ would change ?

Random telegraph: mean and variance

- mean function:

$$\mathbf{E}[X(t)] = \sum_k x_k P(X(t) = x_k) = 1 \cdot (1/2) + (-1) \cdot (1/2) = 0$$

- variance:

$$\begin{aligned} \mathbf{var}[X(t)] = \mathbf{E}[X(t)^2] - (\mathbf{E}[X(t)])^2 &= \sum_k x_k^2 P(X(t) = x_k) \\ &= (1)^2 \cdot (1/2) + (-1)^2 \cdot (1/2) = 1 \end{aligned}$$

both mean and variance functions do not depend on time

Random telegraph: covariance function

since mean is zero and by definition

$$\begin{aligned} C(t_1,t_2) &= \mathbf{E}[X(t_1)X(t_2)] \\ &= (1)^2 P[X(t_1)=1, X(t_2)=1] + (-1)^2 P[X(t_1)=-1, X(t_2)=-1] \\ &+ (1)(-1)P[X(t_1)=1, X(t_2)=-1] + (-1)(1)P[X(t_1)=-1, X(t_2)=1] \end{aligned}$$

from above, we need to find joint pmf obtained via conditional pmf

$$P(X(t_1)=x_1, X(t_2)=x_2) = \underbrace{P(X(t_2)=x_2 \mid X(t_1)=x_1)}_{\text{depend on sign change}} \underbrace{P(X(t_1)=x_1)}_{\text{known}}$$

- $X(t_1)$ and $X(t_2)$ have the same sign

$$P(X(t_2)=x_1 \mid X(t_1)=x_1) = P(N(t_2-t_1) = \text{ even}) = (1/2)(1+e^{-2\alpha(t_2-t_1)})$$

- $X(t_1)$ and $X(t_2)$ have different signs

$$P(X(t_2)=-x_1 \mid X(t_1)=x_1) = P(N(t_2-t_1) = \text{ odd}) = (1/2)(1-e^{-2\alpha(t_2-t_1)})$$

Random telegraph: covariance function

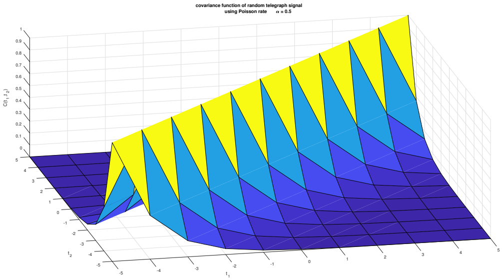
the covariance is obtained by

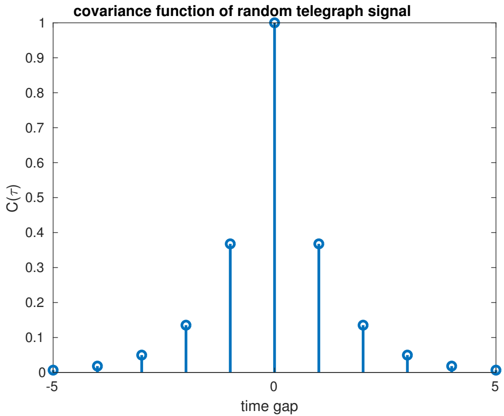
$$\begin{aligned} C(t_1, t_2) &= P(X(t_1) = X(t_2)) + P(X(t_1) \neq X(t_2)) \\ &= 2 \cdot (1/2)(1 + e^{-2\alpha(t_2 - t_1)})(1/2) - 2 \cdot (1/2)(1 - e^{-2\alpha(t_2 - t_1)}) \cdot (1/2) \\ &= e^{-2\alpha|t_2 - t_1|} \end{aligned}$$

- it depends only on the time gap $t_2 - t_1$, denoted as $\tau = t_2 - t_1$
- we can rewrite $C(\tau) = e^{-2\alpha|\tau|}$
- as $\tau \to \infty$, values of $X(t)$ at different times are less correlated
- $X(t)$ (based on the given setting) is wide-sense stationary

Random telegraph: covariance function

covariance function of random telegraph signal: set $\alpha = 0.5$





- left: $C(t_1, t_2) = e^{-2\alpha|t_2 - t_1|}$ as a function of (t_1, t_2)
- right: $C(t_1, t_2) = C(\tau)$ as a function of τ only

Random telegraph: revisit

revisit telegraph signal: when $X(0) = 1$ with probability $p \neq 1/2$

- how would pmf of $X(t)$ change ?
- examine stationary property under this setting

pmf of $X(t)$

$$\begin{aligned} P(X(t)=1) &= (1/2)(1+e^{-2\alpha t})(p) + (1/2)(1-e^{-2\alpha t})(1-p) \\ &= 1/2 + e^{-2\alpha t}(p-1/2) \\ P(X(t)=-1) &= 1 - P(X(t)=1) \\ &= 1/2 - e^{-2\alpha t}(p-1/2) \end{aligned}$$

- when $p \neq 1/2$, pmf of $X(t)$ varies over time but pmf converges to uniform as $t \to \infty$, regardless of the value of p
- if $p = 1$ ($X(0)$ is deterministic) then pmf still varies over time:

$$P(X(t)=1) = (1/2)(1+e^{-2\alpha t}), \quad P(X(t)=-1) = (1/2)(1-e^{-2\alpha t})$$

Random telegraph: stationary property

stationary property: $X(t)$ is stationary if

$$P(X(t_1) = x_1, \dots, X(t_k) = x_k) = P(X(t_1 + \tau) = x_k, \dots, X(t_k + \tau) = x_k)$$

for any $t_1 < t_2 < \cdots < t_k$ and any τ

examine by characterizing pmf as product of conditional pmf's

$$p(x_1, \dots, x_k) = p(x_k | x_{k-1}, \dots, x_1) p(x_{k-1} | x_{k-2}, \dots, x_1) \cdots p(x_2 | x_1) p(x_1)$$

which reduces to

$$P(X(t_1) = x_1, \dots, X(t_k) = x_k) =$$
$$P(X(t_k) = x_k | X(t_{k-1}) = x_{k-1}) \cdots P(X(t_2) = x_2 | X(t_1) = x_1) P(X(t_1) = x_1)$$

using *independent increments property* of Poisson process

Random telegraph: stationary property

because

- for example, if $X(t_k)$ don't change sign

$$P(X(t_k) = x_k | X(t_{k-1}) = x_{k-1}) = P(N(t_k - t_{k-1}) = \text{ even})$$

  if $X(t_{k-1})$ is given, values of $X(t_k)$ are determined solely by $N(t)$ in intervals (t_j, t_{j-1}) which is independent of the previous intervals

- only knowing x_{k-1} is enough to know conditional pmf:

$$P(x_k | x_{k-1}, x_{k-2}, \ldots, x_1) = P(x_k | x_{k-1})$$

then, we can find each of the conditional pmf's

$$P(X(t_k) = x_k | X(t_{k-1}) = x_{k-1}) = \begin{cases} (1/2)(1 + e^{-2\alpha(t_k - t_{k-1})}), \text{for } x_k = x_{k-1} \\ (1/2)(1 - e^{-2\alpha(t_k - t_{k-1})}), \text{for } x_k = -x_{k-1} \end{cases}$$

Random telegraph: stationary property

with the same reasoning, we can write the joint pmf (with time shift) as

$$\begin{aligned}P(X(t_1+\tau)=x_1,\ldots,X(t_k+\tau)=x_k) = \\ P(X(t_k+\tau)=x_k|X(t_{k-1}+\tau)=x_{k-1})\cdots \\ P(X(t_2+\tau)=x_2|X(t_1+\tau)=x_1)P(X(t_1+\tau)=x_1)\end{aligned}$$

where these are equal

$$P(X(t_k)=x_k|X(t_{k-1})=x_{k-1}) = P(X(t_k+\tau)=x_k|X(t_{k-1}+\tau)=x_{k-1})$$

because it depends only on the time gap (from page 101)

as a result, to examine stationary property, we only need to compare

$$P(X(t_1)=x_1) \quad \text{VS} \quad P(X(t_1+\tau)=x_1)$$

which only equal in **steady-state sense** (as $t_1 \to \infty$) from page 99

Pulse amplitude modulation (PAM)

setting: to send a sequence of binary data, transmit 1 or -1 for T seconds

$$X(t) = \sum_{n=-\infty}^{\infty} A_n p(t - nT)$$

where A_k is random amplitude (± 1) and $p(t)$ is a pulse of width T

- $m(t) = 0$ since $\mathbf{E}[A_n] = 0$
- $C(t_1, t_2)$ is given by

$$C(t_1, t_2) = \begin{cases} \mathbf{E}[X(t_1)^2] = 1, & \text{if } nT \leq t_1, t_2 < (n+1)T \\ \mathbf{E}[X(t_1)]\mathbf{E}[X(t_2)] = 0, & \text{otherwise} \end{cases}$$

- $X(t)$ is wide-sense cyclostationary but clearly sample function of $X(t)$ is not periodic

Signal with additive noise

most applications encounter a random process of the form

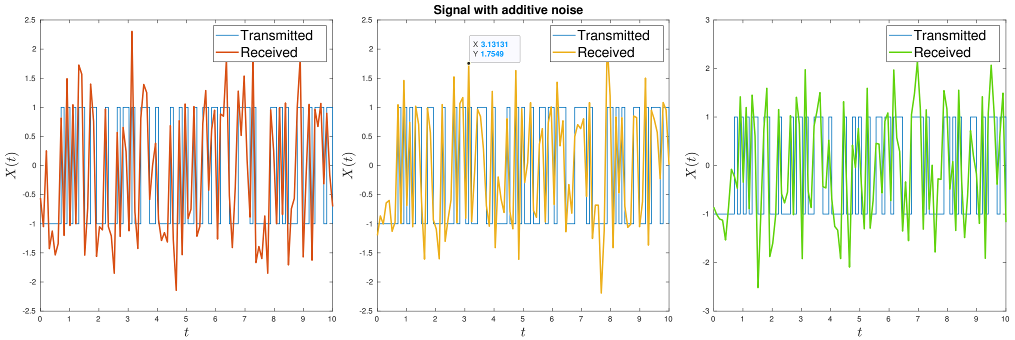
$$Y(t) = X(t) + W(t)$$

- $X(t)$ is transmitted signal (could be deterministic or random) but unknown
- $Y(t)$ is the measurement (observable to users)
- $W(t)$ is noise that corrupts the transmitted signal

common questions regarding this model:

- if only $Y(t)$ is measurable can we reconstruct/estimate what $X(t)$ is ?
- if we can, what kind of statistical information about $W(t)$ do we need ?
- if $X(t)$ is deterministic, how much W affect to Y in terms of fluctuation?

Example: signal with additive noise



- $X(t)$ is a pulse (deterministic)
- $W(t)$ is white Gaussian noise with variance 0.5

Signal with additive noise

simple setting: let us make X and W independent

cross-covariance: let $\tilde{X}$ and $\tilde{W}$ be the mean removed versions

$$\begin{aligned} C_{xy}(t_1,t_2) &= \mathbf{E}[(X(t_1)-m_x(t_1))(Y(t_2)-m_y(t_2))] \\ &= \mathbf{E}[(X(t_1)-m_x(t_1))((X(t_2)+W(t_2)-m_x(t_2)-m_w(t_2))] \\ &= \mathbf{E}[\tilde{X}(t_1)(\tilde{X}(t_2)+\tilde{W}(t_2))] \\ &= C_x(t_1,t_2)+0 \end{aligned}$$

cross-covariance can never be zero as Y is a function of X

autocovariance:

$$\begin{aligned} C_y(t_1,t_2) &= \mathbf{E}[(Y(t_1)-m_y(t_1))(Y(t_2)-m_y(t_2))] \\ &= \mathbf{E}[(\tilde{X}(t_1)+\tilde{W}(t_1))(\tilde{X}(t_2)+\tilde{W}(t_2))] \\ &= C_x(t_1,t_2)+C_w(t_1,t_2)+0 \end{aligned}$$

the variance in Y is always higher than X; the increase is from the noise

Signal with additive noise

simple setting: let us make X and W independent

cross-covariance:

$$
\begin{aligned}
R_{xy}(t_1,t_2) &= \mathbf{E}[X(t_1)Y(t_2)] = \mathbf{E}[X(t_1)(X(t_2)+W(t_2)] \\
&= \mathbf{E}[X(t_1)X(t_2)] + \mathbf{E}[X(t_1)W(t_2)] \\
&= R_x(t_1,t_2) + m_x(t_1)m_w(t_2)
\end{aligned}
$$

cross-covariance can never be zero as Y is a function of X

autocovariance:

$$
\begin{aligned}
R_y(t_1,t_2) &= \mathbf{E}[Y(t_1)Y(t_2)] = \mathbf{E}[(X(t_1)+W(t_1))(X(t_2)+W(t_2)] \\
&= \mathbf{E}[X(t_1)X(t_2)] + \mathbf{E}[W(t_1)W(t_2)] + \mathbf{E}[X(t_1)W(t_2)+W(t_1)X(t_2)] \\
&= R_x(t_1,t_2) + R_w(t_1,t_2) + m_x(t_1)m_w(t_2) + m_x(t_2)m_w(t_1)
\end{aligned}
$$

the variance in Y is always higher than X; the increase is from the noise

Autoregressive Moving Average

let $e(t)$ be a white noise process, an ARMA process is described by

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + \cdots + a_p y(t-p) \\ + e(t) + c_1 e(t-1) + \cdots + c_q e(t-q)$$

$y(t)$ depends on its own history (autoregressive) and noise history (moving average)

define the **lag operator**, $Ly(t) = y(t-1)$

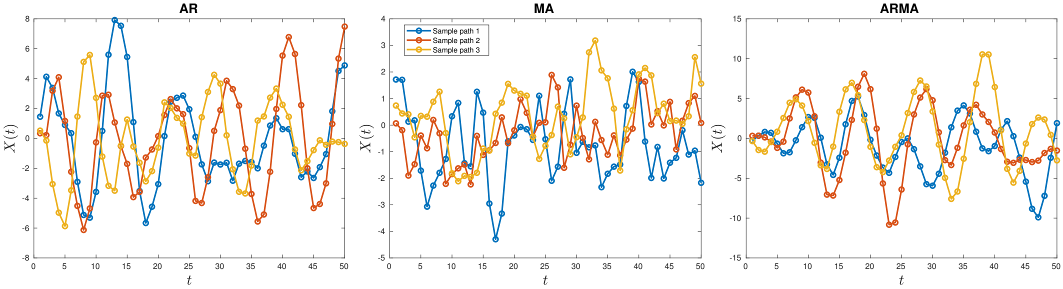
recursive equation of ARMA can be expressed as

$$[1 - (a_1 L + \cdots + a_p L^p)]y(t) = [1 + c_1 L + \cdots + c_q L^q]e(t) \Leftrightarrow A(L)y(t) = C(L)e(t)$$

- $A(L) = 1 - (a_1 L + \cdots + a_p L^p)$: autoregressive (AR) polynomial of order p
- $C(L) = 1 + c_1 L + \cdots + c_q L^q$: moving average (MA) polynomial of order q

coefficients of AR and MA polynomials affect several properties of ARMA processes

Sample paths of ARMA



- $A(L) = 1 - (1.4L - 0.8L^2)$ for AR and $C(L) = 1 + 0.7L + 0.2L$ for MA
- each sample path is driven by different realizations of white noise

Stationary ARMA processes

- an ARMA process is **wide-sense stationary (WSS)** if the roots of

  AR polynomial: $A(L) = 1 - (a_1 L + a_2 L + \cdots + a_p L^p)$ lie outside the unit circle

- the ARMA process is **invertible** if the roots of

  MA polynomial: $C(L) = 1 + c_1 L + c_2 L + \cdots + c_q L^q)$ lie outside the unit circle

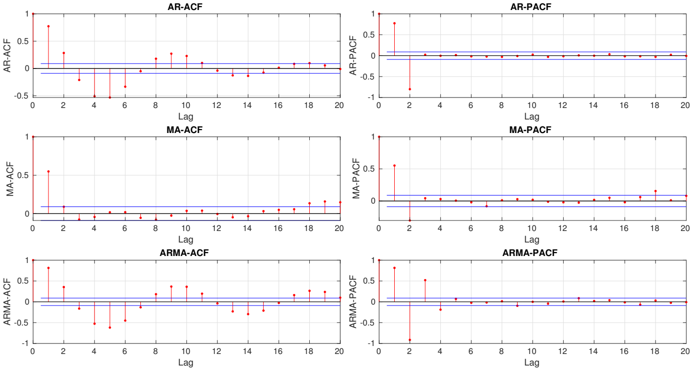
the transfer function from e to y is

$$H(z) = \frac{N(z)}{D(z)} = \frac{1 + c_1 z^{-1} + \cdots + c_q z^{-q}}{1 - (a_1 z^{-1} + a_2 z^{-2} + \cdots + a_p z^{-p})}$$

refer to page 154, X is WSS if $H(z)$ is **stable**, *i.e.*, poles of $H(z)$ or roots of $D(z)$ lie **inside** the unit circle – equivalent to condition on $A(L)$

ACF and PACF of ARMA processes

MATLAB shows ACF (autocovariance) and PACF (partial autocovariance)



- PACF of AR(p) cuts off after lag p
- ACF of MA(q) cuts off after lag q

AR process: autocorrelation

the autocorrelation of AR(p) process:

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + \cdots + a_p y(t-p) + e(t)$$

also progresses as another autoregressive process known as Yule-Walker equation

$$R(\tau) = a_1 R(\tau - 1) + a_2 R(\tau - 2) + \cdots + a_p R(t-p)$$

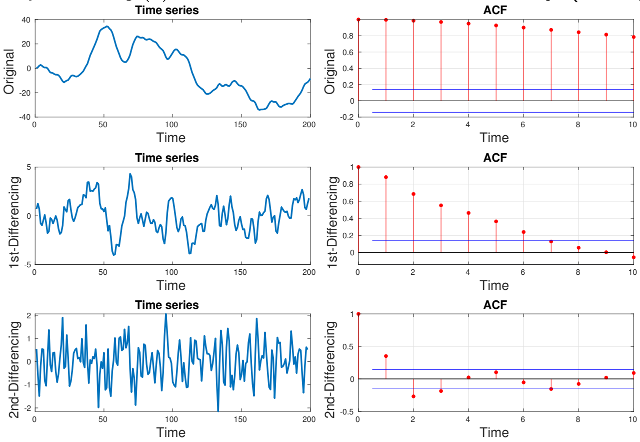
YW equation can be expressed as a **Toeplitz** system, *e.g.*, AR(3)

$$\begin{bmatrix} R(1) \\ R(2) \\ R(3) \end{bmatrix} = \begin{bmatrix} R(0) & R(-1) & R(-2) \\ R(1) & R(0) & R(-1) \\ R(2) & R(1) & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

we can use Toeplitz structure in Yule-Walker equation to solve AR coefficients

Stationarity via differencing

a process $y(t)$ does not seem to be stationary (slowly decaying ACF)



but after differencing twice:

- $y_1(t) = y(t) - y(t-1)$
- $y_2(t) = y_1(t) - y_2(t-1)$

y_2 fluctuates around a constant and ACF decays to zero

Integrated process

denote L a lag operator; a process $y(t)$ is **integrated** of order d if

$$(I - L)^d y(t)$$

is WSS (after d^{th} differencing)

- we use $I(d)$ to denote the integrated model of order d
- random walk is the first-order integrated model
- the lag of differencing is used to reduce a series with a trend

ARIMA process

$y(t)$ is an ARIMA(p, d, q) process if the dth differences of $y(t)$ is an ARMA(p,q)

$$A(L)(I - L)^d y(t) = C(L)e(t)$$

examples of scalar ARIMA models

- $y(t) = y(t-1) + e(t) + ce(t-1)$ can be arranged as

$$(1 - L)y(t) = (1 + cL)e(t)$$

  which is ARIMA(0,1,1) or sometimes called **integrated moving average**
- $y(t) = ay(t-1) + y(t-1) - ay(t-2) + e(t)$ can be arranged as

$$(1 - aL)(1 - L)y(t) = e(t)$$

  which is ARIMA(1,1,0)

Pure seasonal ARMA

an ARMA$(P, Q)_s$ process takes the form

$$\tilde{A}(L^s)y(t) = \tilde{C}(L^s)$$

where s is the **seasonal period** (positive integer)

- $\tilde{A}(L^s) = 1 - (a_1 L^s + a_2 L^{2s} + \cdots + a_P L^{Ps})$ is called **seasonal AR polynomial**
- $\tilde{C}(L^s) = 1 + c_1 L^s + c_2 L^{2s} + \cdots + c_Q L^{Qs})$ is called **seasonal MA polynomial**

example: $y(t) = a_1 y(t-12) + a_2 y(t-24) + e(t) + c_1 e(t-12)$

$$[1 - (a_1 L^s + a_2 L^{2s})]y(t) = [1 + c_1 L^s]e(t)$$

and $s = 12, P = 2, Q = 1$

Behavior of ACF and PACF

stationary ARMA processes

	AR(p)	MA(q)	ARMA(p,q)
ACF	tails off	cuts off after lag q	tails off
PACF	cuts off after lag p	tails off	tails off

pure SARMA processes

	AR$(P)_s$	MA$(Q)_s$	ARMA$(P,Q)_s$
ACF	tails off at lags ks, $k=1,2,\dots,$	cuts off after lag Qs	tails off at lags ks
PACF	cuts off after lag Ps	tails off at lags ks, $k=1,2,\dots,$	tails off at lags ks

note: the values at nonseason lags $\tau \neq ks$, for $k=1,2,\dots,$ are zero

Wide-sense stationary processes

Outlines

- definition
- properties of correlation function
- power spectral density (Wiener – Khinchin theorem)
- cross-correlation
- cross spectrum
- linear system with random inputs
- non-stationary processes

Definition

the second-order joint cdf of an RP $X(t)$ is

$$F_{X(t_1),X(t_2)}(x_1, x_2)$$

(joint cdf of two different times)

we say $X(t)$ is wide-sense (or second-order) stationary if

$$F_{X(t_1),X(t_2)}(x_1, x_2) = F_{X(t_1+\tau),X(t_2+\tau)}(x_1, x_2)$$

the second-order joint cdf do not change for all t_1, t_2 and for all τ

results:

- $\mathbf{E}[X(t)] = m$ (mean is constant)
- $R(t_1, t_2) = R(t_2 - t_1)$ (correlation depends only on the time gap)

Properties of correlation function

let $X(t)$ be a wide-sense scalar real-valued RP with correlation function $R(t_1, t_2)$

- since $R(t_1, t_2)$ depends only on $t_1 - t_2$, we usually write $R(\tau)$ with $\tau = t_1 - t_2$
- $R(0) = \mathbf{E}[X(t)^2]$ for all t
- $R(\tau)$ is an even function of τ

$$R(\tau) \triangleq \mathbf{E}[X(t+\tau)X(t)] = \mathbf{E}[X(t)X(t+\tau)] \triangleq R(-\tau)$$

- $|R(\tau)| \leq R(0)$ (correlation is maximum at lag zero)

$$\mathbf{E}[(X(t+\tau) - X(t))^2] \geq 0 \Longrightarrow 2\mathbf{E}[X(t+\tau)X(t)] \leq \mathbf{E}[X(t+\tau)^2] + \mathbf{E}[X(t)^2]$$

- the autocorrelation is a measure of **rate of change** of a WSS

$$P(|X(t+\tau) - X(t)| > \epsilon) = P(|X(t+\tau) - X(t)|^2 > \epsilon^2)$$
$$\leq \frac{\mathbf{E}[|X(t+\tau) - X(t)|^2]}{\epsilon^2} = \frac{2(R(0) - R(\tau))}{\epsilon^2}$$

- for complex-valued RP, $R(\tau) = R^*(-\tau)$

$$R(\tau) \triangleq \mathbf{E}[X(t+\tau)X^*(t)] = \mathbf{E}[X(t)X^*(t-\tau)] = \overline{\mathbf{E}[X(t-\tau)X^*(t)]} \triangleq R^*(-\tau)$$

- if $R(0) = R(T)$ for some T then $R(\tau)$ is **periodic** with period T and $X(t)$ is **mean square periodic**, *i.e.*,

$$\mathbf{E}[(X(t+T) - X(t))^2] = 0$$

$R(\tau)$ is periodic because

$$\begin{aligned}(R(\tau+T) - R(\tau))^2 &= \{\mathbf{E}[(X(t+\tau+T) - X(t+\tau))X(t)]\}^2 \\ &\leq \mathbf{E}[(X(t+\tau+T) - X(t+\tau))^2]\mathbf{E}[X^2(t)] \quad \text{(Cauchy-Schwarz ineq)} \\ &= 2[R(0) - R(T)]R(0) = 0\end{aligned}$$

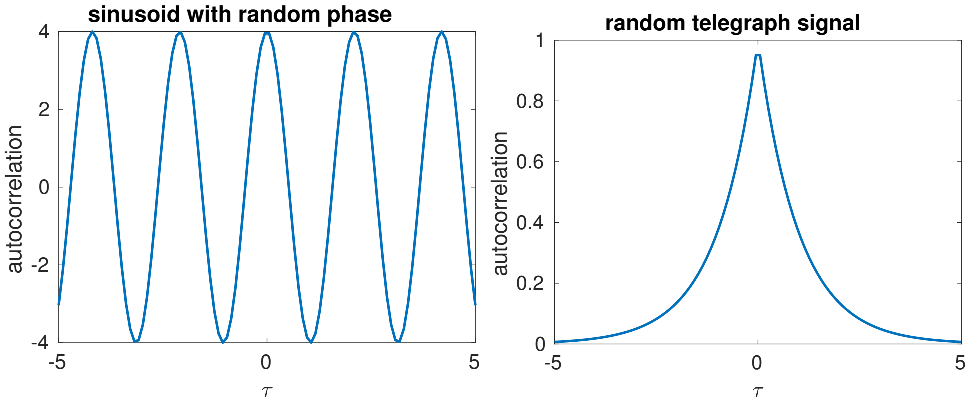
$X(t)$ is mean square periodic because

$$\mathbf{E}[(X(t+T) - X(t))^2] = 2(R(0) - R(T)) = 0$$

- let $X(t) = m + Y(t)$ where $Y(t)$ is a zero-mean process

$$R_x(\tau) = m^2 + R_y(\tau)$$

Example of WSS processes



- sinusoids with random phase: $R(\tau) = \frac{A^2}{2}\cos(\omega\tau)$
- random telegraph signal: $R(\tau) = e^{-2\alpha|\tau|}$

Nonnegativity of correlation function

let $X(t)$ be a real-valued WSS and let $Z = (X(t_1), X(t_2), \ldots, X(t_N))$

the correlation matrix of Z, which is always nonnegative, takes the form

$$\mathbf{R} = \begin{bmatrix} R(0) & R(t_1 - t_2) & \cdots & R(t_1 - t_N) \\ R(t_2 - t_1) & R(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & R(t_{N-1} - t_N) \\ R(t_N - t_1) & \cdots & R(t_N - t_{N-1}) & R(0) \end{bmatrix} \quad \text{(symmetric)}$$

since by assumption,

- $X(t)$ can be either CT or DT random process
- N (the number of time samples) can be any number
- the choice of t_k's are arbitrary

we then conclude that $\mathbf{R} \succeq 0$ holds for all sizes of $\mathbf{R}$ $(N = 1, 2, \ldots)$

Nonnegativity of correlation matrix

the nonnegativity of $\mathbf{R}$ can also be checked from the definition:

$$a^T \mathbf{R} a \geq 0, \quad \text{for all} \quad a = (a_1, a_2, \ldots, a_N)$$

which follows from

$$\begin{aligned} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i^T R(t_i - t_j) a_j &= \sum_{i}^{N} \sum_{j}^{N} \mathbf{E}[a_i^T X(t_i) X(t_j)^T a_j] \\ &= \mathbf{E}\left[\left(\sum_{i=1}^{N} a_i^T X(t_i) \right)^2 \right] \geq 0 \end{aligned}$$

important note: the value of $R(t)$ at some fixed t can be negative !

Example

example: $R(\tau) = e^{-|\tau|/2}$ and let $t = (t_1, t_2, \ldots, t_5)$

```
k=4 ; rng('default'); t = abs(randn(k,1)); t = sort(t); % t = (t1,...,tk)
R1 = exp(-0.5*abs(t-t'));  % broadcast t-t' as all possible subtractions
R = zeros(k);
for i=1:k
    for j=1:k
        R(i,j) = exp(-0.5*abs(t(i)-t(j))); % Slower in loop
    end
end
R1 =
    1.0000    0.8502    0.5230    0.4229
    0.8502    1.0000    0.6152    0.4974
    0.5230    0.6152    1.0000    0.8086
    0.4229    0.4974    0.8086    1.0000

eig(R) =
0.1385    0.1847    0.8144    2.8624
```

showing that $\mathbf{R} \succeq 0$ (try with any k)

Block Toeplitz structure of correlation matrix

CT process: if $X(t)$ are sampled as $Z = (X(t_1), X(t_2), \ldots, X(t_N))$ where

$$t_{i+1} - t_i = \text{constant} = s \quad , i = 1, \ldots, N-1$$

(times have **constant spacing**, $s > 0$ and no need to be an integer)

we see that $\mathbf{R} = \mathbf{E}[ZZ^T]$ has a symmetric **block Toeplitz structure**

$$\mathbf{R} = \begin{bmatrix} R(0) & R(-s) & \cdots & R(-(N-1)s) \\ R(s) & R(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & R(-s) \\ R((N-1)s) & \cdots & R(s) & R(0) \end{bmatrix} \quad \text{(symmetric)}$$

if $X(t)$ is WSS then $\mathbf{R} \succeq 0$ for any integer N and any $s > 0$

Example

example: $R(\tau) = e^{-|\tau|/2}$

```
>> t=0:0.5:2; R = exp(-0.5*abs(t)); T = Toeplitz(R)
R =

    1.0000    0.7788    0.6065    0.4724    0.3679


T =

    1.0000    0.7788    0.6065    0.4724    0.3679
    0.7788    1.0000    0.7788    0.6065    0.4724
    0.6065    0.7788    1.0000    0.7788    0.6065
    0.4724    0.6065    0.7788    1.0000    0.7788
    0.3679    0.4724    0.6065    0.7788    1.0000

eig(T) =

    0.1366    0.1839    0.3225    0.8416    3.5154
```

Covariance matrix of DT process

DT process: time indices are integers, so $Z = (X(1), X(2), \ldots, X(N))$

times also have **constant spacing**

$\mathbf{R} = \mathbf{E}[ZZ^T]$ also has a symmetric **block Toeplitz structure**

$$\begin{bmatrix} R(0) & R(-1) & \cdots & R(1-N) \\ R(1) & R(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & R(-1) \\ R(N-1) & \cdots & R(1) & R(0) \end{bmatrix}$$

if $X(t)$ is WSS then $\mathbf{R} \succeq 0$ for any positive integer N

Example

example: $R(\tau) = \cos(\tau)$

```
>> t=0:2; R = cos(t); T = Toeplitz(R)

R =
    1.0000    0.5403   -0.4161

T =
    1.0000    0.5403   -0.4161
    0.5403    1.0000    0.5403
   -0.4161    0.5403    1.0000

eig(T) =

    0.0000
    1.4161
    1.5839
```

$R(\tau)$ at some τ can be negative !

Power spectral density

Wiener-Khinchin Theorem: if a process is wide-sense stationary, the autocorrelation function and the power spectral density form a Fourier transform pair:

$$S(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau) d\tau \qquad \text{continuous-time FT}$$

$$S(\omega) = \sum_{k=-\infty}^{\infty} R(k) e^{-i\omega k} \qquad \text{discrete-time FT}$$

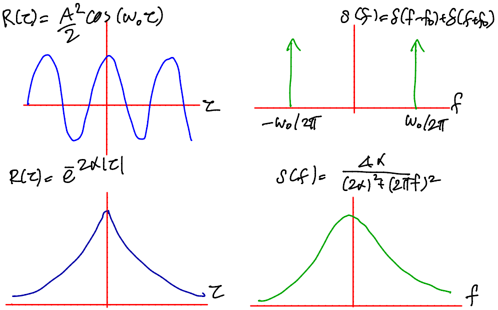
$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} S(\omega) d\omega \qquad \text{continuous-time IFT}$$

$$R(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega\tau} S(\omega) d\omega \qquad \text{discrete-time IFT}$$

$S(\omega)$ indicates a density function for average power versus frequency

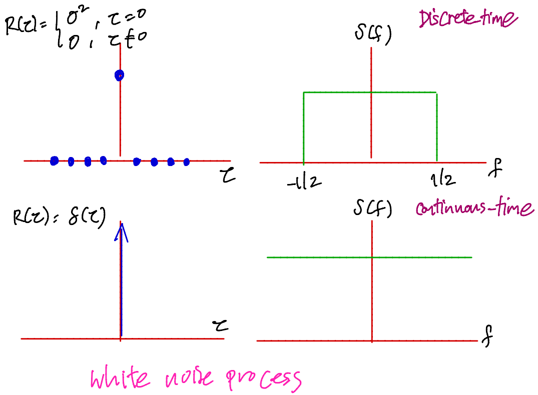
Example: PSD

examples: sinusoid with random phase and random telegraph



- (left) $X(t) = A\sin(\omega_0 t + \phi)$ and $\phi \sim \mathcal{U}(-\pi, \pi)$
- (right) $X(t)$ is random telegraph signal

Example: PSD of white noise



- (left) DT white noise process has a spectrum as a rectangular window
- (right) CT white noise process has a flat spectrum

Spectrum of MA

let $X(n)$ be a DT white noise process with variance σ^2

$$Y(n) = X(n) + \alpha X(n-1), \quad \alpha \in \mathbf{R}$$

then $Y(n)$ is an RP with autocorrelation function

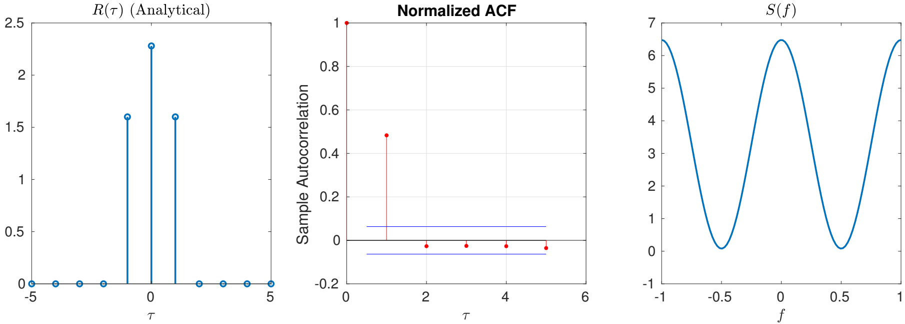
$$R_Y(\tau) = \begin{cases} (1+\alpha^2\sigma^2), & \tau = 0, \\ \alpha\sigma^2, & |\tau| = 1, \\ 0, & \text{otherwise} \end{cases}$$

the spectrum of DT process (is periodic in $f \in [-1/2, 1/2]$) is given by

$$\begin{aligned} S(f) &= \sum_{k=-\infty}^{\infty} R_Y(k) e^{-i2\pi f k} \\ &= (1+\alpha^2\sigma^2) + \alpha\sigma^2(e^{i2\pi f} + e^{-i2\pi f}) \\ &= \sigma^2(1 + \alpha^2 + 2\alpha\cos(2\pi f)) \end{aligned}$$

Spectrum of MA

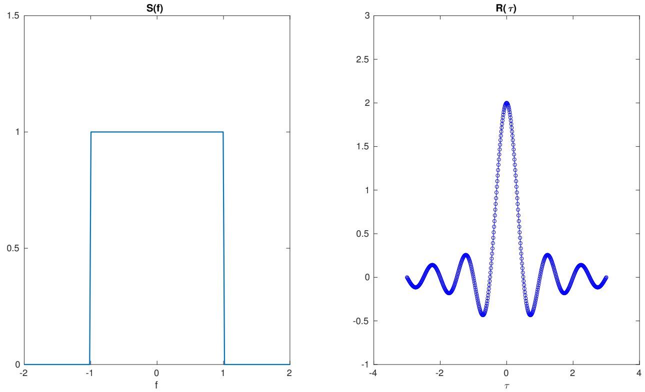
examples: moving average process with $\sigma^2 = 2$ and $\alpha = 0.8$



- $R(\tau)$ cuts off at lag 2
- normalized ACF is calculated based on sample auto-correlation (tails at lag > 2)
- spectrum is periodic in $f \in [-1/2, 1/2]$

Band-limited white noise

given a (white) process whose spectrum is *flat* in the range $-B \leq f \leq B$



the magnitude of the spectrum is $N/2$

what will the (continuous-valued) process look like ?

Autocorrelation via IFT

autocorrelation function is obtained from IFT

$$\begin{aligned} R(\tau) &= (N/2) \int_{-B}^{B} e^{i2\pi f\tau} df \\ &= \frac{N}{2} \cdot \frac{e^{i2\pi B\tau} - e^{-i2\pi B\tau}}{i2\pi\tau} \\ &= \frac{N \sin(2\pi B\tau)}{2\pi\tau} = NB\mathrm{sinc}(2\pi B\tau) \end{aligned}$$

- $X(t)$ and $X(t+\tau)$ are uncorrelated at $\tau = \pm k/2B$ for $k = 1, 2, \ldots$
- if $B \to \infty$, the band-limited white noise becomes a white noise

$$S(f) = \frac{N}{2}, \quad \forall f, \quad R(\tau) = \frac{N}{2}\delta(\tau)$$

Properties of power spectral density

consider real-valued RPs, so $R(\tau)$ is real-valued

- $S(\omega)$ is real-valued and even function ($\because R(\tau)$ is real and even)
- $R(0)$ indicates the **average power**

$$R(0) = \mathbf{E}[X(t)^2] = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)d\omega$$

- $S(\omega) \geq 0$ for all ω and for all $\omega_2 \geq \omega_1$

$$\frac{1}{2\pi}\int_{\omega_1}^{\omega_2} S(\omega)d\omega$$

  is the average power in the frequency band (ω_2, ω_1)

(see proof in Chapter 9 of H. Stark)

Power spectral density as a time average

let $X[0], X[1], \ldots, X[N-1]$ be N observations from DT WSS process

discrete Fourier transform of the time-domain sequence is

$$\tilde{X}[k] = \sum_{n=0}^{N-1} X[n] e^{-\frac{i2\pi}{N}kn}, \quad k = 0, 1, \ldots, N-1$$

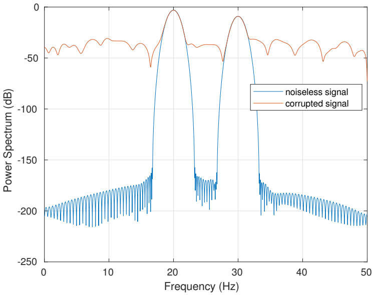
- $\tilde{X}[k]$ is a complex-valued sequence describing DT Fourier transform with only *discrete frequency points*
- $\tilde{X}[k]$ is a measure of *energy* at frequency $2\pi k/N$
- an estimate of *power* at a frequency is then

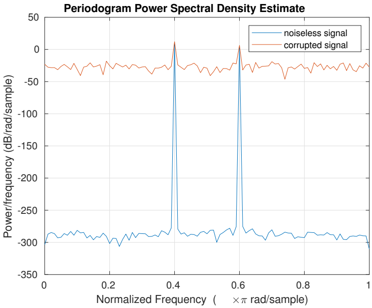
$$\tilde{S}(k) = \frac{1}{N} |\tilde{X}[k]|^2$$

and is called **periodogram estimate** for the power spectral density

Example of PSD

example: $X(t) = \sin(40\pi t) + 0.5\sin(60\pi t)$

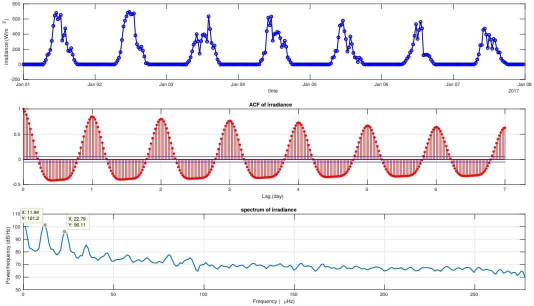




- signal has frequency components at 20 and 30 Hz
- peaks at 20 and 30 Hz are clearly seen
- when signal is corrupted by noise, spectrum peaks can be less distinct
- the plots are done using `pspectrum` and `periodogram` in MATLAB

Frequency analysis of solar irradiance

data are irradiance with sampling period of $T = 30$ min



- ACF is a normalized autocorrelation function (by $R(0)$) and appears to be periodic
- spectral density appears to have three peaks corresponding to $0, 12, 24$ μHz
- the frequencies of $12, 24$ μHz correspond to the periods of one day and half day respectively
- ACF and spectral density are computed by `autocorr` and `pwelch`

Cross correlation and cross spectrum

cross correlation between processes $X(t)$ and $Y(t)$ is defined as

$$R_{XY}(\tau) = \mathbf{E}[X(t+\tau)Y(t)]$$

cross-power spectral density between $X(t)$ and $Y(t)$ is defined as

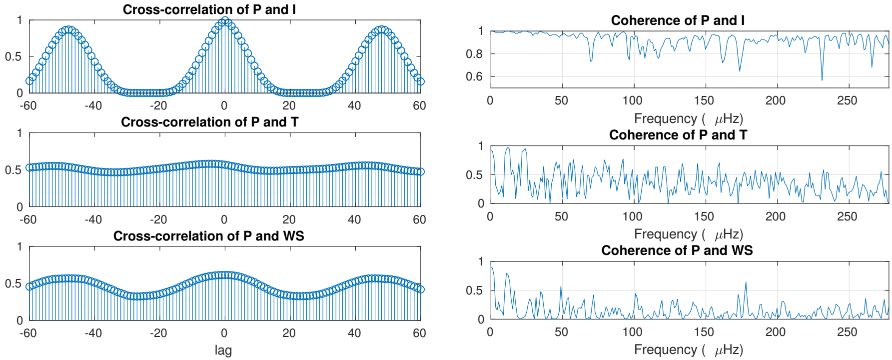
$$S_{XY}(\omega) = \int_{-\infty}^{\infty} e^{-\mathrm{i}\omega\tau} R_{XY}(\tau) d\tau$$

properties:

- $S_{XY}(\omega)$ is complex-valued in general, even $X(t)$ and $Y(t)$ are real
- $R_{YX}(\tau) = R_{XY}(-\tau)$
- $S_{YX}(\omega) = S_{XY}(-\omega)$

Example: Solar time series

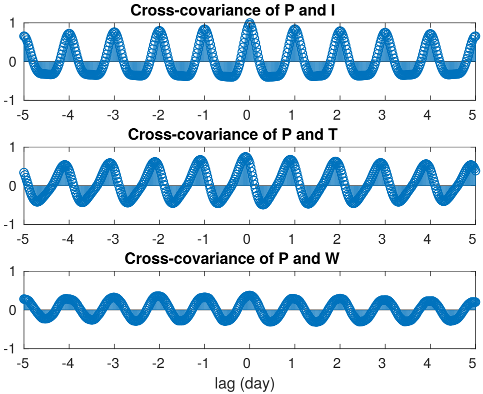
solar power (P), solar irradiance (I), temperature (T), wind speed (WS)



- (normalized) cross correlations are computed by `xcorr` in MATLAB
- (normalized) coherence functions are computed by `mscohere`:

$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_x(f)S_y(f)}$$

Example: cross covariance function



- P and I are highly correlated while P and WS are least correlated
- cross covariance functions are almost periodic (daily cycle) with slightly decaying envelopes

Extended definitions

extension: let $X(t)$ be a *complex-valued vector* random process

- denote $*$ Hermittian transpose, *i.e.*, $X^* = \overline{X}^T$
- correlation function: $R(\tau) = \mathbf{E}[X(t+\tau)X(t)^*]$
- covariance function: $C(\tau) = R(\tau) - \mu\mu^*$
- $R_{YX}(\tau) = R_{XY}^*(-\tau)$
- $S_{YX}(\omega) = S_{XY}^*(-\omega)$
- $S(\omega)$ is self-adjoint, *i.e.*, $S(\omega) = S^*(\omega)$ and $S(\omega) \succeq 0$

(cross) correlation and (cross) spectral density functions are *matrices*

Theorems on correlation function and spectrum

Theorem 1: a necessary and sufficient condition for $R(\tau)$ to be a correlation function of a WSS is that it is positive semidefinite

- proof of sufficiency part: if $R(\tau)$ is positive semidefinite then there exists a WSS whose correlaction function is $R(\tau)$
  - if $R(\tau)$ is psdf then its Fourier transform is positive semidefinite (a proof is not obvious)
  - let us call $S(\omega) = \mathcal{F}(R(\tau)) \succeq 0$
  - by spectral factorization theorem, there exists a stable filter $H(\omega)$ such that $S(\omega) = H(\omega)H^*(\omega)$ – more advanced topic
  - the existence of a WSS is given by applying a white noise to the filter $H(\omega)$ – the topic we will learn next on page 154
- proof of necessity part: if a process is WSS then $R(\tau)$ is positive semidefinite – shown on page 124

Theorem: Fourier pair

Theorem 2: let $S(\omega)$ be a self-adjoint and nonnegative matrix and

$$\int_{-\infty}^{\infty} \mathbf{tr}(S(\omega))d\omega < \infty$$

then its inverse Fourier transform:

$$R(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega t} S(\omega) d\omega$$

is **nonnegative**, *i.e.*, $\sum_{j=1}^{N}\sum_{k=1}^{N} a_j^* R(t_j - t_k) a_k \geq 0$

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}^* \begin{bmatrix} R(0) & R(t_1 - t_2) & \cdots & R(t_1 - t_N) \\ R(t_2 - t_1) & R(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & R(t_{N-1} - t_N) \\ R(t_N - t_1) & \cdots & R(t_N - t_{N-1}) & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} \succeq 0$$

Proof: non-negativity of $R(\tau)$

consider $N = 3$ case (can be extended easily)

$$A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}^* \begin{bmatrix} R(0) & R(t_1 - t_2) & R(t_1 - t_3) \\ R(t_2 - t_1) & R(0) & R(t_2 - t_3) \\ R(t_3 - t_1) & R(t_3 - t_2) & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

$$= \int_{-\infty}^{\infty} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}^* \begin{bmatrix} e^{i\omega(t_1-t_1)}S(\omega) & e^{i\omega(t_1-t_2)}S(\omega) & e^{i\omega(t_1-t_3)}S(\omega) \\ e^{i\omega(t_2-t_1)}S(\omega) & e^{i\omega(t_2-t_2)}S(\omega) & e^{i\omega(t_2-t_3)}S(\omega) \\ e^{i\omega(t_3-t_1)}S(\omega) & e^{i\omega(t_3-t_2)}S(\omega) & e^{i\omega(t_3-t_3)}S(\omega) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} d\omega$$

$$= \int_{-\infty}^{\infty} \begin{bmatrix} e^{-i\omega t_1}a_1 \\ e^{-i\omega t_2}a_2 \\ e^{-i\omega t_3}a_3 \end{bmatrix}^* \begin{bmatrix} S^{1/2}(\omega) \\ S^{1/2}(\omega) \\ S^{1/2}(\omega) \end{bmatrix} \begin{bmatrix} S^{1/2}(\omega) & S^{1/2}(\omega) & S^{1/2}(\omega) \end{bmatrix} \begin{bmatrix} e^{-i\omega t_1}a_1 \\ e^{-i\omega t_2}a_2 \\ e^{-i\omega t_3}a_3 \end{bmatrix} d\omega$$

$$\triangleq \int_{-\infty}^{\infty} Y^*(\omega)Y(\omega)d\omega \succeq 0$$

because the integrand is nonnegative definite for all ω

(we have used the fact that $S(\omega) \succeq 0$ and has a square root)

Theorem: non-negativity of PSD

Theorem 3: let $R(t)$ be a continuous correlation matrix function such that

$$\int_{-\infty}^{\infty} |R_{ij}(t)| dt < \infty, \quad \forall i, j$$

then the spectral density matrix

$$S(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} R(t) dt$$

is **self-adjoint** and **positive semidefinite**

- matrix case: proof by Balakrishnan, Introduction to Random Process in Engineering, page 79
- scalar case: proof by Starks and Woods, page 607 (need to learn the topic on page 154 first)

simple proof (from Starks): let $\omega_2 > \omega_1$, define a filter transfer function

$$H(\omega) = 1, \quad \omega \in (\omega_1, \omega_2), \qquad H(\omega) = 0, \quad \text{otherwise}$$

let $X(t)$ and $Y(t)$ be input/output to this filter, then

$$S_{YY}(\omega) = S_{XX}(\omega), \quad \omega \in (\omega_1, \omega_2), \qquad S_{YY}(\omega) = 0, \quad \text{elsewhere}$$

since $\mathbf{E}[Y(t)^2] = R_y(0)$ and it is nonnegative, it follows that

$$R_y(0) = \frac{1}{2\pi} \int_{\omega_1}^{\omega_2} S_x(\omega) d\omega \geq 0$$

this must holds for any $\omega_2 > \omega_1$

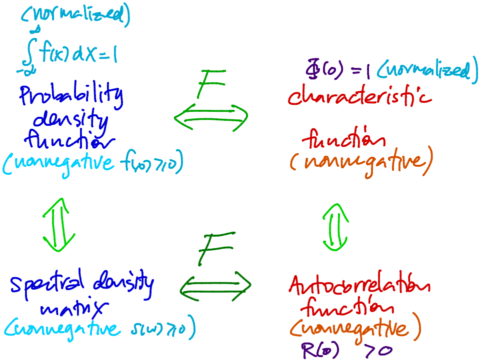
hence, choosing $\omega_2 \approx \omega_1$ we must have $S_x(\omega) \geq 0$ — the power spectral density must be nonnegative

Conclusion

a function $R(\tau)$ is nonnegative if and only if

it has a nonnegative Fourier transform

- a valid spectral density function therefore can be checked by its nonnegativity and it is easier than checking the nonnegativity condition of $R(\tau)$
- analogy for probability density function



Linear system with random inputs

consider a linear system with input and output relationship through

$$y = Hx$$

which represents many applications (filter, transformation of signals, etc.)

questions regarding this setting:

- if x is a random signal, how can we explain about randomness of y?
- if x is wide-sense stationary, how about y? under what condition on H?
- if y is also wide-sense, how about relations between correlation/power spectral density of x and y?

Recap on linear systems

recall the definitions

- **linear system:**

$$H(x_1 + \alpha x_2) = Hx_1 + \alpha H x_2$$

- **time-invariant system:** it commutes with shift operator

$$Hx(t - T) = y(t - T)$$

(time shift in the input causes the same time shift in the output)

- response of linear time-invariant system: denote h the impulse response

$$y(t) = h(t) * x(t) = \begin{cases} \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau & \text{continous-time} \\ = \sum_{k=-\infty}^{\infty} h(t-k)x(k) & \text{discrete-time} \end{cases}$$

- **stable:** poles of H are in stability region (LHP or inside unit circle)
- **causal system:** response of y at t depends only on *past* values of x

$$\text{impulse response} \;\; h(t) = 0, \quad \text{for } t < 0$$

Properties of output from LTI system

let $Y = HX$ where H is linear time-invariant system and **stable**

if $X(t)$ is wide-sense stationary then

- $m_Y(t) = H(0)m_X(t)$
- $Y(t)$ is also wide-sense stationary
  (in steady-state sense if $X(t)$ is applied when $t \geq 0$)
- correlations and spectra are given by

time-domain	frequency-domain
$R_{YX}(\tau) = h(\tau) * R_X(\tau)$	$S_{YX}(\omega) = H(\omega)S_X(\omega)$
$R_{XY}(\tau) = R_X(\tau) * h^*(-\tau)$	$S_{XY}(\omega) = S_X(\omega)H^*(\omega)$
$R_Y(\tau) = R_{YX}(\tau) * h^*(-\tau)$	$S_Y(\omega) = S_{YX}(\omega)H^*(\omega)$
$R_Y(\tau) = h(\tau) * R_X(\tau) * h^*(-\tau)$	$S_Y(\omega) = H(\omega)S_X(\omega)H^*(\omega)$

using $\mathcal{F}(f(t) * g(t)) = F(\omega)G(\omega)$ and $\mathcal{F}(f^*(-t)) = F^*(\omega)$

Proof: mean of output

show: $m_Y(t) = H(0)m_X(t)$

$$\begin{aligned} Y(t) &= \int_{-\infty}^{\infty} h(s)X(t-s)ds \\ \mathbf{E}[Y(t)] &= \int_{-\infty}^{\infty} h(s)\mathbf{E}[X(t-s)]ds \\ &= \int_{-\infty}^{\infty} h(s)ds \cdot m_x \qquad \text{(since } X(t) \text{ is WSS)} \\ &= H(0)m_x \end{aligned}$$

mean of Y is transformed by the DC gain of the system

Proof: WSS of Y

$$
\begin{aligned}
R_y(t+\tau, t) &= \mathbf{E}[Y(t+\tau)Y(t)^T] \\
&= \mathbf{E}\left[\left(\int_{-\infty}^{\infty} h(\sigma)X(t+\tau-\sigma)ds\right)\left(\int_{-\infty}^{\infty} h(s)X(t-s)ds\right)^T\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\sigma)\mathbf{E}[X(t+\tau-\sigma)X(t-s)^T]h(s)^T d\sigma ds \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\sigma)R_x(\tau+s-\sigma)h(s)^T d\sigma ds \quad (X \text{ is WSS})
\end{aligned}
$$

we see that $R_y(t+\tau, t)$ does not depend on t anymore but only on τ

- we have shown that $Y(t)$ has a constant mean and the autocorrelation function depends only on the time gap τ
- hence, $Y(t)$ is also a wide-sense stationary process

Proof: Cross-correlation of input and output

using $Y(t) = \int_{-\infty}^{\infty} h(\alpha)X(t-\alpha)d\alpha$

- $R_{YX}(\tau) = h(\tau) * R_X(\tau)$

$$R_{YX}(\tau) = \mathbf{E}[Y(t)X^*(t-\tau)] = \int_{-\infty}^{\infty} h(\alpha)\mathbf{E}[X(t-\alpha)X^*(t-\tau)]d\alpha$$
$$= \int_{-\infty}^{\infty} h(\alpha)R_X(\tau-\alpha)d\alpha$$

- $R_Y(\tau) = R_{YX}(\tau) * H^*(-\tau)$

$$R_Y(\tau) = \mathbf{E}[Y(t)Y^*(t-\tau)] = \int_{-\infty}^{\infty} \mathbf{E}[Y(t)X^*(t-(\tau+\alpha))]h^*(\alpha)d\alpha$$
$$= \int_{-\infty}^{\infty} R_{YX}(\tau+\alpha)h^*(\alpha)d\alpha = \int_{-\infty}^{\infty} R_{YX}(\tau-\sigma)h^*(-\sigma)d\sigma$$

Power spectrum of output process

the relation $S_Y(\omega) = H(\omega)S_X(\omega)H^*(\omega)$ reduces to

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega)$$

for *scalar* processes $X(t)$ and $Y(t)$

- average power of the output depends on the input power at that frequency multiplied by power gain at the same frequency
- we call $|H(\omega)|^2$ the **power spectral density (PSD) transfer function**

this relation gives a procedure to estimate $H(\omega)$ when signals $X(t)$ and $Y(t)$ can be observed

Example: random telegraph signal

a random telegraph signal with transition rate α is passed thru an RC filter with

$$H(s) = \frac{\tau}{s+\tau}, \quad \tau = 1/RC$$

question: find psd and autocorrelation of the output

random telegraph signal has the spectrum: $\quad S_x(f) = \dfrac{4\alpha}{4\alpha + 4\pi^2 f^2}$

from $S_y(f) = |H(f)|^2 S_x(f)$ and $R_y(t) = \mathcal{F}^{-1}[S_y(f)]$

$$S_y(f) = \left(\frac{\tau^2}{\tau^2 + 4\pi^2 f^2}\right) \frac{4\alpha}{4\alpha + 4\pi^2 f^2} = \frac{4\alpha\tau^2}{\tau^2 - 4\alpha^2} \left\{ \frac{1}{4\alpha^2 + 4\pi^2 f^2} - \frac{1}{\tau^2 + 4\pi^2 f^2} \right\}$$

$$R_y(t) = \frac{1}{\tau^2 - 4\alpha^2} \left(\tau^2 e^{-2\alpha|t|} - 2\alpha\tau e^{-\tau|t|}\right)$$

(we have used $\mathcal{F}[e^{-at}] = 2a/(a^2 + \omega^2)$) and $\omega = 2\pi f$

Example: PSD of AR process

first-order AR process

$$Y(n) = aY(n-1) + X(n)$$

$X(n)$ is i.i.d white noise with variance of σ^2

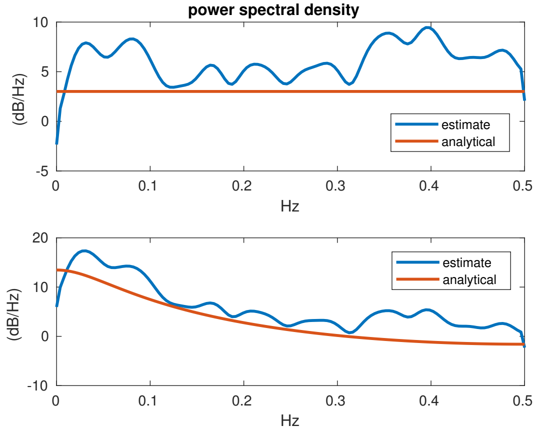
- $H(z) = \frac{1}{1-az^{-1}}$ or $H(e^{i\omega}) = \frac{1}{1-ae^{-i\omega}}$
- spectral density is obtained by

$$S_y(\omega) = |H(\omega)|^2 S_x(\omega) = \frac{\sigma^2}{(1-ae^{-i\omega})(1-ae^{i\omega})}$$

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

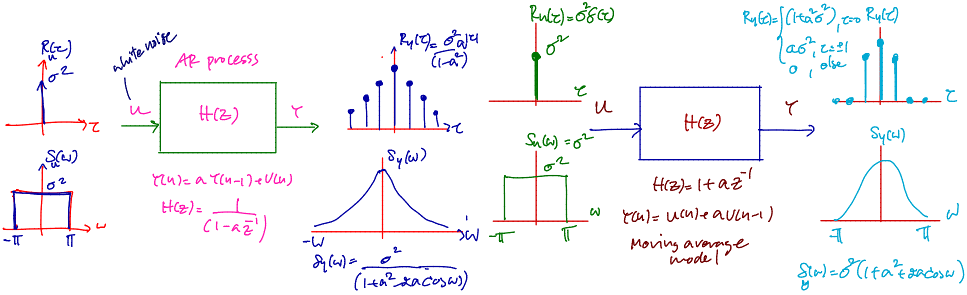
Example: PSD of AR

spectral density of AR process: $a = 0.7$ and $\sigma^2 = 2$



Input and output spectra

in conclusion, when input is white noise, the spectrum is flat



when white noise is passed through a filter, the output spectrum is no longer flat

Response to linear system: state-space models

consider a discrete-time linear system via a state-space model

$$X(k+1) = AX(k) + BU(k), \quad Y(k) = HX(k)$$

where $X \in \mathbf{R}^n, Y \in \mathbf{R}^p, U \in \mathbf{R}^m$

known results:

- two forms of solutions of state and output variables are

$$\begin{aligned} X(t) &= A^t X(0) + \sum_{\tau=0}^{t-1} A^\tau BU(t-1-\tau), \quad Y(t) = HX(t) \\ &= A^{t-s} X(s) + \sum_{\tau=s}^{t-1} A^{t-1-s} BU(\tau), \quad Y(t) = HX(t) \end{aligned}$$

- the autonomous system (when $U = 0$) is stable if $|\lambda(A)| < 1$

State-space models: autocovariance function

Theorem: let U be a i.i.d white noise sequence with covariance Σ_u and if i) A is **stable** and ii) $X(0)$ is uncorrelated with $U(k)$ for all $k \geq 0$ then

- $\lim_{n\to\infty} \mathbf{E}[X(n)] = 0$
- $C(n,n) \to \Sigma$ as $n \to \infty$ where

$$\Sigma = A\Sigma A^T + B\Sigma_u B^T$$

  (Σ is a **unique** solution to the **Lyapunov equation**)

- $X(t)$ is wide-sense stationary in **steady-state** sense, *i.e.*,

$$\lim_{n\to\infty} C(n+k,n) = C(k) = \begin{cases} A^k\Sigma, & k \geq 0 \\ \Sigma(A^T)^{|k|}, & k < 0 \end{cases}$$

Proof: mean of state

the mean of $X(t)$ converges to zero

let $m(n) = \mathbf{E}[X(n)]$ and it's easy to see

$$m(n) = \mathbf{E}[X(n)] = A\mathbf{E}[X(n-1)] + B\mathbf{E}[U(n-1)] = Am(n-1)$$

hence, $m(n)$ propagates like a linear system:

$$m(n) = A^n m(0)$$

and goes to zero as $n \to \infty$ since A is stable ■

zero-mean system: $\tilde{X}(n) = X(n) - m(n)$

$$\tilde{X}(n) = A\tilde{X}(n-1) + BU(n-1)$$

mean-removed process also follow the same state-space equation

Proof: covariance function of state

show: $\lim_{n\to\infty} C(n,n) = \Sigma$ and satisfies the Lyapunov equation

- $\tilde{X}(n)$ is uncorrelated with $U(k)$ for all $k \geq n$

$$\tilde{X}(t) = A^t \tilde{X}(0) + \sum_{\tau=0}^{t-1} A^\tau BU(t-1-\tau)$$

  because $\tilde{X}(0)$ is uncorrelated with $U(t)$ for all t and $\tilde{X}(t)$ is only a function of $U(t-1), U(t-2), \ldots, U(0)$

- since $\tilde{X}(n-1)$ is uncorrelated with $U(n-1)$, we obtain

$$C(n,n) = AC(n-1,n-1)A^T + B\Sigma_u B^T$$

  from the state equation: $\tilde{X}(n) = A\tilde{X}(n-1) + BU(n-1)$

- then we can write $C(n,n)$ recursively

$$C(n,n) = \underbrace{A^n C(0,0)(A^T)^n}_{\text{go to zero}} + \underbrace{\sum_{k=0}^{n-1} A^k B\Sigma_u B^T (A^T)^k}_{\text{converges}}$$

  and observe its asymptotic behaviour when $n \to \infty$

Proof: covariance function of state

Theorem: let $A \in \mathbf{R}^{n\times n}$ with spectral radius $\rho(A)$. We have $\rho(A) < 1$ if and only if $\lim_{k\to\infty} A^k = 0$ (proved by Jordan canonical form of A)

- if A is stable, the spectral radius of A is less than one, hence $A^n \to 0$ as $n \to \infty$
- let $\Sigma = \sum_{k=0}^{\infty} A^k B \Sigma_u B^T (A^T)^k$, we can check that

$$\Sigma = A\Sigma A^T + B\Sigma_u B^T$$

- Σ is unique, otherwise, by contradiction

$$\Sigma_1 = A\Sigma_1 A^T + B\Sigma_u B^T, \quad \Sigma_2 = A\Sigma_2 A^T + B\Sigma_u B^T$$

  we can subtract one from another and see that

$$\Sigma_1 - \Sigma_2 = A(\Sigma_1 - \Sigma_2)A^T = A^2(\Sigma_1 - \Sigma_2)(A^T)^2 = \cdots = A^n(\Sigma_1 - \Sigma_2)(A^T)^n$$

  this goes to zero since A is stable ($\|A^k\| \to 0$)

$$\|\Sigma_1 - \Sigma_2\| = \|A^n(\Sigma_1 - \Sigma_2)(A^T)^n\| \leq \|A\|^{2n}\|\Sigma_1 - \Sigma_2\| \to 0$$

this completes the proof

Proof: WSS in steady-state

show that $\tilde{X}(n)$ is wide-sense stationary in steady-state

- $\tilde{X}(k)$ is uncorrelated with $\{U(k), U(k+1), \ldots, U(n-1)\}$
- from the solution of $\tilde{X}(n)$

$$\tilde{X}(n) = A^{n-k}\tilde{X}(k) + \sum_{\tau=k}^{n-1} A^{n-1-\tau} BU(\tau), \quad k < n$$

  the two terms on RHS are uncorrelated

- the autocovariance function is obtained by (for $n > k$)

$$\begin{aligned} C(n,k) &= \mathbf{E}[\tilde{X}(n)\tilde{X}(k)^T] \\ &= A^{n-k}\mathbf{E}[\tilde{X}(k)\tilde{X}(k)^T] + \sum_{\tau=k}^{n-1} A^{n-1-\tau} B\mathbf{E}[U(\tau)\tilde{X}(k)^T] \\ &= A^{n-k}C(k,k) + 0 \end{aligned}$$

  which converges to $A^{n-k}\Sigma$ as $n, k \to \infty$ if A is stable

State-space models: autocovariance of output

output equation:

$$Y(n) = HX(n), \quad \tilde{Y}(n) = H\tilde{X}(n)$$

when $X(n)$ is wide-sense stationary (in steady-state) then

when $n, k \to \infty$, we have

$$C_y(n,k) = HC_x(n,k)H^T = HA^{n-k}C_x(k,k)H^T, \quad n \geq k$$

and

$$\lim_{n\to\infty} C_y(n,n) = \lim_{n\to\infty} HC_x(n,n)H^T = H\Sigma H^T$$

where Σ is the solution to the Lyapunov equation: $\Sigma = A\Sigma A^T + B\Sigma_u B^T$

Example: AR process

AR process with $a = 0.7$ and U is i.i.d. white noise with $\sigma^2 = 2$

$$Y(n) = aY(n-1) + U(n-1)$$

1st-order AR process is already in state-space equation

- in steady-state, the covariance function at lag 0 converges to α where

$$\alpha = a\alpha^2 + \sigma^2 \quad \Longrightarrow \quad \alpha = \frac{\sigma^2}{1-a^2}$$

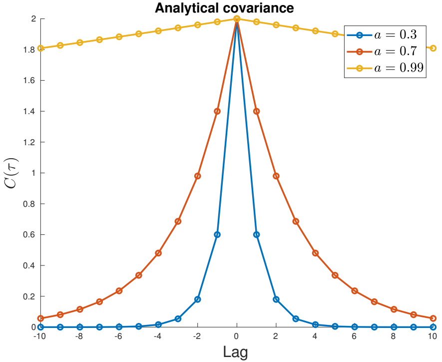
  (we have solved the Lyapunov equation)

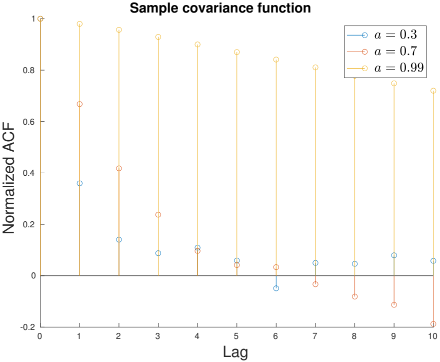
- in steady-state, the covariance function is given by

$$C(\tau) = \frac{\sigma^2 a^{|\tau|}}{1-a^2}$$

Example: Covariance function of AR

vary $a = 0.3, 0.7, 0.99$



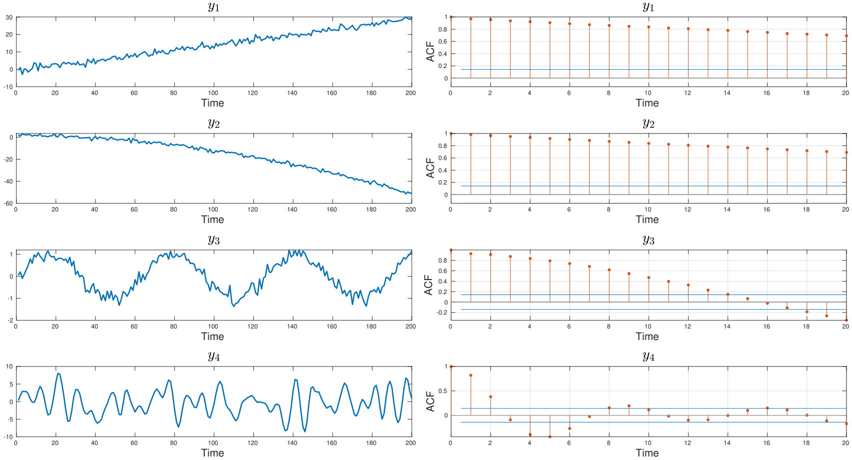


- $C(\tau)$ decays with rate a
- normalized ACF plots $C(\tau)/C(0)$ (maximum peak is always unit)

Common causes of non-stationarity

- time-varying mean: processes with a static trend, drift
- time-depending covariance: $C(t_1, t_2)$ is not a function of $|t_2 - t_1|$

Which process seems to be non-stationary?



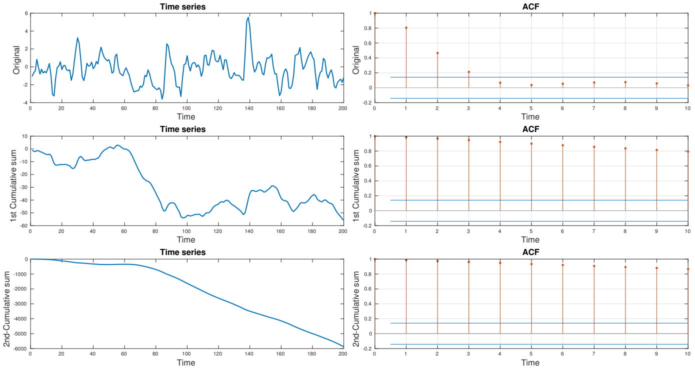
- y_1, y_2, y_3 are clearly not non-stationary because they have static trends; their sample ACFs seem to decay slowly
- y_4 fluctuates around a constant and its sample ACF decays to zero (as if y_4 was generated from a stable system)
- in fact, checking stationarity cannot merely be done just by looking at time series
- further reading: several stationary tests are available, *e.g.*, Augmented Dickey–Fuller test

Cumulative sum of WSS process

as an illustrative example, suppose $y(t)$ is WSS (*e.g.*, stationary ARMA process)

$$s(t) = \sum_{\tau=0}^{t} y(\tau) = y(0) + y(1) + \cdots + y(t)$$

question: is $s(t)$ WSS ? ✎: sketch the mean and autocorrelation



- row 1: $y(t)$, row 2: $s(t)$ (as 1st cum sum), row 3: cum sum of $s(t)$
- how do the profile of time series and ACF suggest stationarity of cum sum process ?

Common forms of non-stationary signals

- $y(t) = s(t) + u(t)$, $s(t)$ is deterministic, and $u(t)$ is WSS
- $y(t)$ is **intregrated process** of some WSS process, *e.g.*, ARIMA process

References

1. Chapter 9-10 in A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd edition, Pearson Prentice Hall, 2009
2. Chapter 9 in H. Stark and J. W. Woods, *Probability, Statistics, and Random Processes for Engineers*, 4th edition, Pearson, 2012