

Course material

Course:

Digital Signal Processing II

Video:

2.3.1.b Stochastic Processes

Concepts (extracted from automatically generated subtitles):

Realizations of stochastic processes. Random signals. Infinite number of random variables. Time index n . Set of samples. Second time. Random processes. Random process. Random variables. Zero mean. Sample of the realization. Joint probability density function. Simple example. Stochastic signal. Discrete time.



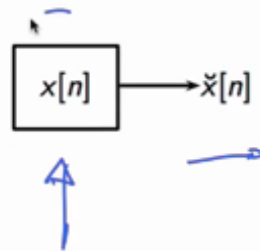
[to video sequence search](#)
(within Digital Signal Processing II.)



[to video](#)

Center for Digital Education. More educational support material here:

<https://exoset.epfl.ch/booc/>



In stochastic signal processing we are interested in describing and manipulating random signals and random signals are the realizations of stochastic processes or random processes. The best way to think of a random process is to think of a machine that contains an infinite number of random variables indexed by a time index n . So here inside you have an infinite number of random variables that could be all completely different and then this random variables are triggered in turn and produce the set of samples that we could actually measure if we ran this machine. As a simple example imagine a

notes

summary

0m 0s



For each new sample, toss a fair coin:

$$x[n] = \begin{cases} +1 & \text{if the outcome of the } n\text{-th toss is head} \\ -1 & \text{if the outcome of the } n\text{-th toss is tail} \end{cases}$$

17

process where the random variables are all the same and map the event of flipping a fair coin. So each sample of the realization will take value either plus 1 if the outcome of the n th coin toss is head and minus 1 if the outcome

notes

summary

0m 37s



For each new sample, toss a fair coin:

$$x[n] = \begin{cases} +1 & \text{if the outcome of the } n\text{-th toss is head} \\ -1 & \text{if the outcome of the } n\text{-th toss is tail} \end{cases}$$

17

of the n th toss is tail. So a realization of this process could look like so

notes

summary

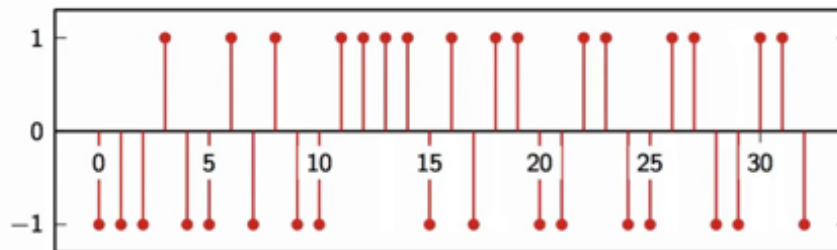
0m 56s



A simple discrete-time random signal generator



every time we turn on the generator we obtain a different *realization* of the signal



18

and if we run the machine a second time we would get a different realization that perhaps would look like so and a third time would give us yet another different realization and so on and so forth. So a discrete time random process

notes

summary

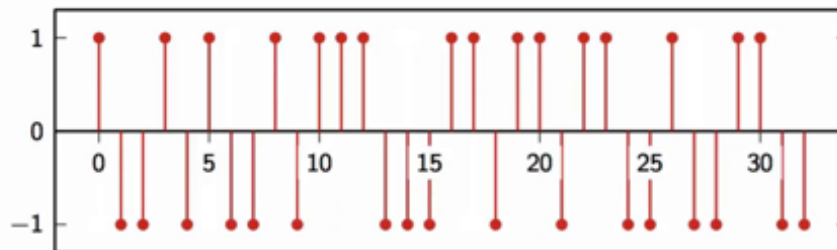
1m 1s



A simple discrete-time random signal generator



every time we turn on the generator we obtain a different *realization* of the signal



18

generates an infinite length sequence of random values. The questions are how do

notes

summary

1m 13s



DT random processes generate an infinite-length sequence of random values

- ▶ what is the distribution of each value ?
- ▶ what are the statistical relations between values?

19

we describe this process statistically? Well we need to know the statistical distribution of each value but also the interdependence between the infinite random variables that appear inside this process generation machine. Now you

notes

summary

1m 15s



DT random processes generate an infinite-length sequence of random values

- ▶ what is the distribution of each value ?
- ▶ what are the statistical relations between values?

19

remember from the previous discussion if we want to characterize statistically

notes

summary

1m 36s



- $$f_{x[n_0]x[n_1]\cdots x[n_{k-1}]}(x_0, x_1, \dots, x_{k-1})$$

J

- 20

notes

1m 37s



- ▶ infinite-length sequence of *interdependent* random variables
- ▶ a full characterization requires knowing

$$f_{x[n_0]x[n_1]\dots x[n_{k-1}]}(x_0, x_1, \dots, x_{k-1})$$



for *all* possible sets of k indices $\{n_0, n_1, \dots, n_{k-1}\}$ and for *all* $k \in \mathbb{N}$

- ▶ clearly too much to handle

20

set of random variables and so we would need to specify the PDF for any subset of random variables that take part in the generation of the process. In other words you should be able to pick any set of time indices, pick the random

notes

summary

1m 50s



- ▶ infinite-length sequence of *interdependent* random variables
- ▶ a full characterization requires knowing

$$f_{x[n_0]x[n_1]\dots x[n_{k-1}]}(x_0, x_1, \dots, x_{k-1})$$

for *all* possible sets of k indices $\{n_0, n_1, \dots, n_{k-1}\}$ and for *all* $k \in \mathbb{N}$

- ▶ clearly too much to handle

variables that are associated to the indices and then provide the PDF for this subset of random variables. This is clearly too much to handle so we need to simplify the problem. A first type of simplification involves considering

notes

summary

2m 2s



► first-order description:

- $f_{X[n]}(x[n]) \rightarrow$ time-varying mean $\bar{x}[n] = E[x[n]]$

► second-order description:

- $f_{X[n]}(x[n]) \rightarrow$ time-varying mean $\bar{x}[n] = E[x[n]]$
- $f_{X[n]X[m]}(x[n], x[m]) \rightarrow$ time-varying auto-correlation $r_x[n, m] = E[x[n]x[m]]$

► third-order description:

- time-varying mean
- time-varying auto-correlation
- $f_{X[n]X[m]X[p]}(x[n], x[m], x[p]) \rightarrow$ time-varying third moment

► ...

descriptions only up to a certain order. For instance a first order description

notes

summary

2m 18s



- ▶ first-order description:
 - $f_{X[n]}(x[n]) \rightarrow$ time-varying mean $\bar{x}[n] = E[x[n]]$
- ▶ second-order description:
 - $f_{X[n]}(x[n]) \rightarrow$ time-varying mean $\bar{x}[n] = E[x[n]]$
 - $f_{X[n]X[m]}(x[n], x[m]) \rightarrow$ time-varying auto-correlation $r_x[n, m] = E[x[n]x[m]]$
- ▶ third-order description:
 - time-varying mean
 - time-varying auto-correlation
 - $f_{X[n]X[m]X[p]}(x[n], x[m], x[p]) \rightarrow$ time-varying third moment
- ▶ ...

of the process would provide us with all the PDFs associated to all the random variables that are used in the process generation. So with this kind of description you could get for instance a time varying mean which is the expectation of the variable X of n which is the random variable associated to time index n . A second order description would provide all first and second order PDFs and the second order PDFs would involve any two time indices. With this kind of description we can compute the time varying mean and also the time varying autocorrelation which will depend on the chosen time indices. A third order description continues in this vein and adds a third order probability

notes

summary

2m 25s



- ▶ first-order description:
 - $f_{X[n]}(x[n]) \rightarrow$ time-varying mean $\bar{x}[n] = E[x[n]]$
- ▶ second-order description:
 - $f_{X[n]}(x[n]) \rightarrow$ time-varying mean $\bar{x}[n] = E[x[n]]$
 - $f_{X[n]X[m]}(x[n], x[m]) \rightarrow$ time-varying auto-correlation $r_x[n, m] = E[x[n]x[m]]$
- ▶ third-order description:
 - time-varying mean
 - time-varying auto-correlation
 - $f_{X[n]X[m]X[p]}(x[n], x[m], x[p]) \rightarrow$ time-varying third moment
- ▶ ...

density function for all possible triples of time indices. Now in all of these cases we need to still provide an infinite number of probability density functions so we're still dealing with a unmanageable amount of data. Things

notes

summary

3m 10s



for a stationary process, all partial-order descriptions are **time-invariant**:

$$f_{x[n_0]x[n_1]\dots x[n_{k-1}]}(\dots) = f_{x[n_0+M]x[n_1+M]\dots x[n_{k-1}+M]}(\dots)$$

22

start to be a little bit more reasonable in the case of stationary processes. For a stationary process all partial order descriptions are time invariant. In other words if you have a joint probability density function for any set of time

notes

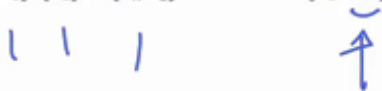
summary

3m 25s



for a stationary process, all partial-order descriptions are **time-invariant**:

$$f_{x[n_0]x[n_1]\dots x[n_{k-1}]}(\dots) = f_{x[n_0+M]x[n_1+M]\dots x[n_{k-1}+M]}(\dots)$$



indices and you shift this time indices by capital M samples the probability density function does not change. Physically this is really the same thing as time invariance in the case of filters and it means that if you have a random signal generating machine it doesn't really matter when you switch this machine on. Its internal properties will translate uniformly in time so it's a

notes

summary

3m 40s



for a stationary process, all partial-order descriptions are **time-invariant**:

$$f_{x[n_0]x[n_1]\dots x[n_{k-1}]}(\dots) = f_{x[n_0+M]x[n_1+M]\dots x[n_{k-1}+M]}(\dots)$$

| | |
↑

22

reasonable requirement to impose on a random process. For a stationary random

notes

summary

4m 1s



For stationary random processes:

- ▶ mean is time-invariant: $E[x[n]] = m_x$
- ▶ autocorrelation depends only on time lag: $E[\underbrace{x[n]} \underbrace{x[m]}] = \underbrace{r_x[n - m]}$
- ▶ (higher-order moments depend only on relative time differences, etc...)

23

process the mean becomes time invariant because it no longer depends on when you switch on the machine. And the autocorrelation only depends on the time lag namely on the difference between the two indices that you choose. So the expectation of the product between the random variable at time n and the random variable at time m only depends on the time difference n minus m . And if you go to higher order descriptions once again they will only depend on the relative time differences between the chosen time indices. Okay so in a fully stationary random process all order descriptions are time invariant. Things become even simpler if we just care about the first and second order descriptions. So a wide

notes

summary

4m 11s



For WSS random processes we only care about the first two moments:

- ▶ $E[x[n]] = m_x$
- ▶ $E[x[n]x[m]] = r[n - m]$

24

sense stationary process or WSS for short is characterized by having a time invariant mean and an autocorrelation that only depends on the difference

notes

summary

4m 56s



For WSS random processes we only care about the first two moments:

▶ $E[x[n]] = m_x$

▶ $E[x[n]x[m]] = r[n - m]$

24

between the time indices. Why is this enough for us? We will see that in

notes

summary

5m 10s



Why WSS?

- ▶ most stochastic SP techniques use quadratic “cost” functions
- ▶ algorithms require only the first and second moments
- ▶ quadratic optimization (Mean Square Error) mathematically well-behaved

25

stochastic signal processing most of the techniques use a quadratic cost function and this only requires the first and second moments. Using a quadratic cost

notes

summary

5m 13s



Why WSS?

- ▶ most stochastic SP techniques use quadratic “cost” functions
- ▶ algorithms require only the first and second moments
- ▶ quadratic optimization (Mean Square Error) mathematically well-behaved

25

function is also very well behaved mathematically and it makes sense in terms of minimizing an error measure. We will see this in the next module. A very

notes

summary

5m 25s



White noise process:

- ▶ zero-mean: $E[x[n]] = 0$
- ▶ uncorrelated: $E[x[n]x[m]] = E[x[n]]E[x[m]]$ for $m \neq n$
- ▶ autocorrelation $r_x[n] = \sigma_x^2 \delta[n]$

According to underlying distribution:

- ▶ Gaussian white noise
- ▶ uniform white noise
- ▶ ...

26

common type of wide sense stationary process is white noise. And a white noise process is characterized by having zero mean and by being uncorrelated namely the expectation of the product of any two random variables in the process is equal to the product of the expectations. But now since the process is zero mean the autocorrelation sequence will be equal to zero everywhere except in zero where it is equal to the variance of the process. Now the variance will depend on the underlying distribution for the random variables that describe the process. So if we have a Gaussian random variable at each sample we will have a Gaussian white noise or we could have a uniform white noise. The coin toss process that we saw before is also a white process. Each new sample is

notes

summary

5m 40s



For each new sample, toss a fair coin:

$$x[n] = \begin{cases} +1 & \text{if the outcome of the } n\text{-th toss is head} \\ -1 & \text{if the outcome of the } n\text{-th toss is tail} \end{cases}$$

- ▶ each sample is independent from all others
- ▶ each sample value has a 50% probability: $f_x(x) = \delta(x \pm 1)/2$

white noise process with $r_x[n] = \delta[n]$



27

independent because we assume that each sample is generated by tossing an independent coin and each sample has 50% probability of being plus 1 or minus 1 so that its PDF is equal to delta of x plus 1 plus delta of x minus 1 divided by 2. These deltas are the Dirac deltas. And the autocorrelation will be equal to

notes

summary

6m 22s



For each new sample, toss a fair coin:

$$x[n] = \begin{cases} +1 & \text{if the outcome of the } n\text{-th toss is head} \\ -1 & \text{if the outcome of the } n\text{-th toss is tail} \end{cases}$$

- ▶ each sample is independent from all others
- ▶ each sample value has a 50% probability: $f_x(x) = \delta(x \pm 1)/2$

white noise process with $r_x[n] = \delta[n]$

27

the discrete time delta because it will be zero everywhere except for n equal to zero where it will be equal to 1 which is the variance of this random variable. So how do we compute the moments of a process, of a wide sense stationary

notes

summary

6m 49s



With access to the theoretical univariate and bivariate pdfs of the WSS process:

$$m_x = E[x[n]] = \int_{-\infty}^{\infty} a f_{x[0]}(a) da$$

$$r_x[k] = E[x[0]x[k]] = \int \int_{-\infty}^{\infty} ab f_{x[0]x[k]}(a, b) da db$$

28

process? Well if we have access to the theoretical first and second order probability density functions for the process we can just apply the definition and the mean will be the expectation of the generic random variable. Remember now this is time invariant because of the wide sense stationarity. And for the autocorrelation we will pick an index k which represents the time lag and then we compute the expectations of the product of the random variables at say time 0 and time k . These actual values are immaterial. All that matters is the time difference between the two. Since we have access to the bivariate PDF we can

notes

summary

7m 0s



With access to the theoretical univariate and bivariate pdfs of the WSS process:

$$m_x = E[x[n]] = \int_{-\infty}^{\infty} a f_{x[0]}(a) da$$

$$r_x[k] = E[x[0]x[k]] = \int \int_{-\infty}^{\infty} ab f_{x[0]x[k]}(a, b) da db$$

28

compute this double integral and obtain the autocorrelation sequence. Of course this is just in theory because in practice we don't have access to the theoretical probability density functions. What we have access to is if we're lucky a set of realizations of the process. So suppose we have access to the

notes

summary

7m 41s



With access to M realizations of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{i=0}^{M-1} \check{x}_i[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{i=0}^{M-1} \check{x}_i[n] \check{x}_i[n+k]$$

machine that generates the stochastic signal we can run this machine several

notes

summary

7m 55s



With access to M realizations of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{i=0}^{M-1} \check{x}_i[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{i=0}^{M-1} \check{x}_i[n] \check{x}_i[n+k]$$

29

times and we collect say capital M realizations of the wide sense stationary process. At that point we can compute the mean as an ensemble average and the

notes

summary

8m 1s



With access to M realizations of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{i=0}^{M-1} \check{x}_i[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{i=0}^{M-1} \check{x}_i[n] \check{x}_i[n+k]$$

29

autocorrelation as an ensemble autocorrelation. We have several realizations.

notes

summary

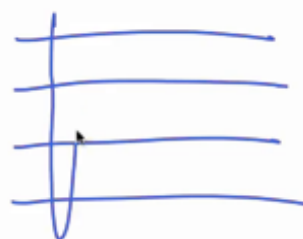
8m 13s



With access to M realizations of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{i=0}^{M-1} \check{x}_i[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{i=0}^{M-1} \check{x}_i[n] \check{x}_i[n+k]$$



29

To compute the mean we average across realizations. And of course it

notes

summary

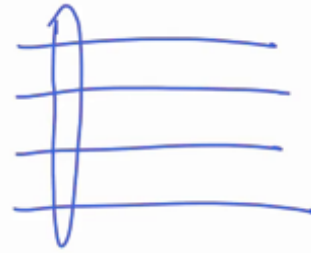
8m 20s



With access to M realizations of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{i=0}^{M-1} \tilde{x}_i[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{i=0}^{M-1} \tilde{x}_i[n] \tilde{x}_i[n+k]$$



29

doesn't matter where we do this because the mean is time invariant. And to compute the autocorrelation we take an index k and then we compute the sum of the product of two samples of the process. There are k samples apart and

notes

summary

8m 25s

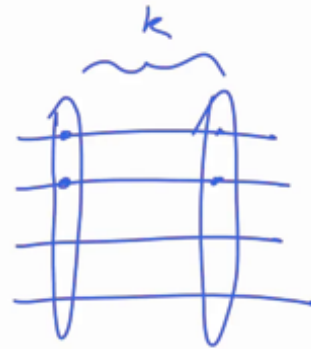


With access to M realizations of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{i=0}^{M-1} \tilde{x}_i[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{i=0}^{M-1} \tilde{x}_i[n] \tilde{x}_i[n+k]$$

↑ —



29

then we sum this across realization. So imagine this is a time difference of k and so we take these two samples we multiply them together then we take these two samples we multiply them together we sum them all together and then we normalize by the number of realizations that we have. If we're even less lucky though we will not have access to the generation machine we will

notes

summary

8m 37s

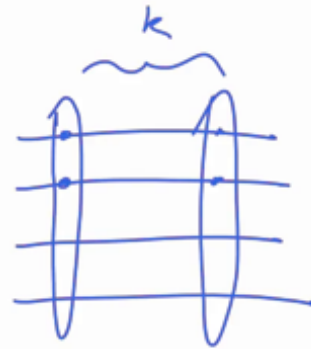


With access to M realizations of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{i=0}^{M-1} \tilde{x}_i[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{i=0}^{M-1} \tilde{x}_i[n] \tilde{x}_i[n+k]$$

↑ —



29

have access to a single realization of the stochastic process as in the case

notes

summary

9m 0s

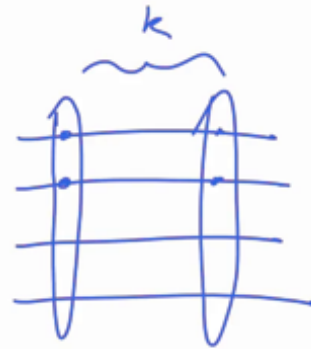


With access to M realizations of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{i=0}^{M-1} \tilde{x}_i[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{i=0}^{M-1} \tilde{x}_i[n] \tilde{x}_i[n+k]$$

↑ —



29

for instance of someone uttering a sentence. This is a speech signal of which you only have one realization. At that point what you try to do is to take

notes

summary

9m 1s



With access to M samples of a single realization of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{n=0}^{M-1} \tilde{x}[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{n=0}^{M-|k|-1} \tilde{x}[n] \tilde{x}[n+|k|]$$



- ▶ processes for which this works are called *ergodic*
- ▶ at least $M > 4k_{\max}$

30

time averages. So you have a single realization and to compute the mean you will just get all the samples and sum them together and divide by the length

notes

summary

9m 14s



With access to M samples of a single realization of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{n=0}^{M-1} \check{x}[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{n=0}^{M-|k|-1} \check{x}[n] \check{x}[n+|k|]$$



- ▶ processes for which this works are called *ergodic*
- ▶ at least $M > 4k_{\max}$

30

of the samples. And to compute the autocorrelation you will choose a time lag k and then take a sample and the sample k samples apart, compute the

notes

summary

9m 25s



With access to M samples of a single realization of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{n=0}^{M-1} \tilde{x}[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{n=0}^{M-|k|-1} \tilde{x}[n] \tilde{x}[n+|k|]$$



- ▶ processes for which this works are called *ergodic*
- ▶ at least $M > 4k_{\max}$

30

product then move by one and take this sample and go k samples apart and compute the product and sum them all together. In reality we should be performing ensemble averages and we're performing time averages. For some processes this will work and the class of processes for which we can replace ensemble averages with time averages is called ergodic. In practice we should use a number of samples in our realization that is at least four times bigger than

notes

summary

9m 37s



With access to M samples of a single realization of the WSS process:

$$m_x \approx \frac{1}{M} \sum_{n=0}^{M-1} \check{x}[n]$$

$$r_x[k] \approx \frac{1}{M} \sum_{n=0}^{M-|k|-1} \check{x}[n] \check{x}[n+|k|]$$



- ▶ processes for which this works are called ergodic
- ▶ at least $M > 4k_{\max}$

30

the maximum index in the autocorrelation sequence that we will need in our application. As a last observation consider the formula for the correlation

notes

summary

10m 6s



Correlation (via ensemble average):

$$E[xy] = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=0}^{M-1} \check{x}_i \check{y}_i.$$

Inner product in $\ell_2(\mathbb{Z})$:

$$\langle x, y \rangle = \sum_{n=-\infty}^{\infty} x_n y_n.$$

31

between two random variables computed as an ensemble average. So apart from the normalization factor this is very very close to the formula for the inner

notes

summary

10m 17s



$$E[xy] = 0 \implies \text{orthogonal random variables}$$

- ▶ if x, y zero mean: variables are uncorrelated
- ▶ no linear relationship between variables
- ▶ variables are maximally different

product between two sequences in L_2 of z . So by analogy we say that if the correlation between two variables is zero the two variables are orthogonal and the interpretation of this orthogonality is the same as in L_2 of z . Orthogonality means that there is no linear relationship between the variables and that the variables are maximally different. This is a concept that would be very useful in adaptive signal processing.

notes

summary

10m 28s

