



Course material

Course:

Digital Signal Processing II

Video:

2.3.1.a Random Variables

Concepts (extracted from automatically generated subtitles):

Random variable. Random signals. Cumulative distribution function. Discrete random variable. Development of signal processing theory. Mean value. Probability theory. Random variable takes. Sample values. Measured temperature. Probability density function. Continuous random variable. Half of your measurements. Zero means. Possible values.



[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/>

- ▶ deterministic signals are known in advance: $x[n] = \sin(0.2n)$
- ▶ interesting signals are *not* known in advance: $s[n] = \text{what I'm going to say next}$
- ▶ we usually know something, though: $s[n]$ is a speech signal
- ▶ stochastic signals can be described probabilistically
- ▶ can we do signal processing with random signals? Yes!

So far in our development of signal processing theory we have assumed that all signals are deterministic, known in advance. We have written things like $x[n]$ is equal to $\sin(0.2n)$. So we describe the signal analytically and in so doing with a simple formula we cover all the sample values. But interesting signals in

notes

summary

0m 0s



- ▶ deterministic signals are known in advance: $x[n] = \sin(0.2n)$
- ▶ interesting signals are *not* known in advance: $s[n] = \text{what I'm going to say next}$
- ▶ we usually know something, though: $s[n]$ is a speech signal
- ▶ stochastic signals can be described probabilistically
- ▶ can we do signal processing with random signals? Yes!

the practice of signal processing are not known in advance. For instance here I am talking to you and the samples that represent my voice signal are not going to be known in advance otherwise it would be pretty boring to listen to what I'm saying. Maybe it is anyway. But even if we don't know the exact values of the samples we usually know something about the signals. For instance we know that

notes

summary

0m 21s



- ▶ deterministic signals are known in advance: $x[n] = \sin(0.2n)$
- ▶ interesting signals are *not* known in advance: $s[n] = \text{what I'm going to say next}$
- ▶ we usually know something, though: $s[n]$ is a speech signal
- ▶ stochastic signals can be described probabilistically
- ▶ can we do signal processing with random signals? Yes!

my voice is a speech signal and so this signal will have certain properties although the exact values of the samples are not known. So here we're going to

notes

summary

0m 40s



- ▶ deterministic signals are known in advance: $x[n] = \sin(0.2n)$
- ▶ interesting signals are *not* known in advance: $s[n] = \text{what I'm going to say next}$
- ▶ we usually know something, though: $s[n]$ is a speech signal
- ▶ stochastic signals can be described probabilistically
- ▶ can we do signal processing with random signals? Yes!

perform a paradigm shift in the way we think about signals. Instead of considering them deterministic objects we will look at them in a probabilistic

notes

summary

0m 54s



- 3

notes

1m 1s



- ▶ deterministic signals are known in advance: $x[n] = \sin(0.2n)$
- ▶ interesting signals are *not* known in advance: $s[n] = \text{what I'm going to say next}$
- ▶ we usually know something, though: $s[n]$ is a speech signal
- ▶ stochastic signals can be described probabilistically
- ▶ can we do signal processing with random signals? Yes!

developed so far both in terms of intuition and in terms of algorithmic procedures still applies to random signals and on top of it we will be able to introduce a new chapter in signal processing which is called adaptive signal processing. So we will use the statistical properties of the signals to change the way our filters process the information. But first let's start with a brief recap of the main concepts in probability theory. We will not be

notes

summary

1m 14s



- ▶ deterministic signals are known in advance: $x[n] = \sin(0.2n)$
- ▶ interesting signals are *not* known in advance: $s[n] = \text{what I'm going to say next}$
- ▶ we usually know something, though: $s[n]$ is a speech signal
- ▶ stochastic signals can be described probabilistically
- ▶ can we do signal processing with random signals? Yes!

exhaustive. We will just point out the fundamental ideas and concepts that will

notes

summary

1m 45s



4

notes

1m 51s



4

- ▶ tossing a coin: map heads to 0, tails to 1
- ▶ tossing a die: *discrete* r.v. is face value
- ▶ electric circuit: *continuous* r.v. is output voltage

notes

2m 10s



a *mapping* from a random event to a value $x \in \mathbb{R}$

Examples:

- ▶ tossing a coin: map heads to 0, tails to 1
- ▶ tossing a die: *discrete* r.v. is face value
- ▶ electric circuit: *continuous* r.v. is output voltage

4

you have a cubic die with six faces you have another discrete random variable that takes values from 1 to 6. You can have a continuous valued random variable. You can imagine for instance measuring the voltage across a circuit and the continuous random variable encodes this unknown voltage that you're measuring. So a random variable encodes an event but we're interested in measuring the

notes

summary

2m 25s



- ▶ cumulative distribution function (cdf):

$$F_x(\alpha) = P[x \leq \alpha], \quad \alpha \in \mathbb{R}$$

$$\lim_{\alpha \rightarrow \infty} F_x(\alpha) = 1$$

- ▶ probability density function (pdf):

$$f_x(\alpha) = \frac{dF_x(\alpha)}{d\alpha}, \quad \alpha \in \mathbb{R}$$

$$F_x(\alpha) = \int_{-\infty}^{\alpha} f_x(x) dx$$

probability that an event occurs. To do so we define the cumulative distribution function or CDF for short called F_x of α and this measures the probability

notes

summary

2m 46s



- ▶ cumulative distribution function (cdf):

$$F_x(\alpha) = P[x \leq \alpha], \quad \alpha \in \mathbb{R}$$

$$\lim_{\alpha \rightarrow \infty} F_x(\alpha) = 1$$

- ▶ probability density function (pdf):

$$f_x(\alpha) = \frac{dF_x(\alpha)}{d\alpha}, \quad \alpha \in \mathbb{R}$$

$$F_x(\alpha) = \int_{-\infty}^{\alpha} f_x(x) dx$$

that the random variable takes values less than or equal to alpha and alpha is

notes

summary

3m 0s



- cumulative distribution function (cdf):

$$F_x(\alpha) = P[x \leq \alpha], \quad \alpha \in \mathbb{R}$$

$$\lim_{\alpha \rightarrow \infty} F_x(\alpha) = 1$$

- probability density function (pdf):

$$f_x(\alpha) = \frac{dF_x(\alpha)}{d\alpha}, \quad \alpha \in \mathbb{R}$$

$$F_x(\alpha) = \int_{-\infty}^{\alpha} f_x(x) dx$$

a real number again. The key property of the CDF is that if you take the limit of this function for alpha that goes to infinity this limit is equal to 1. That

notes

summary

3m 3s



- cumulative distribution function (cdf):

$$F_x(\alpha) = P[x \leq \alpha], \quad \alpha \in \mathbb{R}$$

$$\lim_{\alpha \rightarrow \infty} F_x(\alpha) = 1$$

- probability density function (pdf):

$$f_x(\alpha) = \frac{dF_x(\alpha)}{d\alpha}, \quad \alpha \in \mathbb{R}$$

$$F_x(\alpha) = \int_{-\infty}^{\alpha} f_x(x) dx$$

means that as you span all the possible values that the random variable can take the total probability converges to 1 which is the full 100% probability. You can differentiate the CDF and obtain the probability density function. This is a

notes

summary

3m 16s



- cumulative distribution function (cdf):

$$F_x(\alpha) = P[x \leq \alpha], \quad \alpha \in \mathbb{R}$$

$$\lim_{\alpha \rightarrow \infty} F_x(\alpha) = 1$$

- probability density function (pdf):

$$f_x(\alpha) = \frac{dF_x(\alpha)}{d\alpha}, \quad \alpha \in \mathbb{R}$$

$$F_x(\alpha) = \int_{-\infty}^{\alpha} f_x(x) dx$$

function that for a value alpha of the argument gives you the point-wise probability that the random variable takes on the value alpha. If you integrate the probability density function you go back to the cumulative

notes

summary

3m 28s



- cumulative distribution function (cdf):

$$F_x(\alpha) = P[x \leq \alpha], \quad \alpha \in \mathbb{R}$$

$$\lim_{\alpha \rightarrow \infty} F_x(\alpha) = 1$$

- probability density function (pdf):

$$f_x(\alpha) = \frac{dF_x(\alpha)}{d\alpha}, \quad \alpha \in \mathbb{R}$$

$$F_x(\alpha) = \int_{-\infty}^{\alpha} f_x(x) dx$$



distribution function. So let's make an example. Suppose you want to measure the

notes

summary

3m 44s



Example



Measure repeatedly the temperature of melting ice:

- ▶ continuous random variable is the measured temperature
- ▶ should be zero Celsius
- ▶ changes in barometric pressure
- ▶ different mineral content in water
- ▶ inaccurate thermometer...

6

temperature of melting ice. So you take a beaker full of ice and you put a thermometer in it and you repeat this experiment several times. So the continuous random variable that models this event is the measured temperature

notes

summary

3m 49s



Example



Measure repeatedly the temperature of melting ice:

- ▶ continuous random variable is the measured temperature
- ▶ should be zero Celsius
- ▶ changes in barometric pressure
- ▶ different mineral content in water
- ▶ inaccurate thermometer...

6

that you read on your thermometer and you know that this should be 0 degrees Celsius. But many things can affect the experiment. Changes in barometric pressure, whether the water is purified or not, whether the thermometer has some inaccuracies and so on so forth. So if you repeat this experiment many times the value for the random variable that measures the temperature of melting ice will never be exactly zero but will vary and sometimes take positive values sometimes take negative values. If you plot the cumulative distribution function for this temperature variable you would get probably a curve like this. What this

notes

summary

4m 1s



$$E[x] = \int_{-\infty}^{\infty} x f_x(x) dx$$

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_x(x) dx$$

means is that it's very unlikely that your thermometer will measure something like minus 5 degrees Celsius and as you move up the scale you see the zero corresponds to 50% probability. Half of your measurements will be slightly below zero and half of your measurements will be slightly above zero but as you see as you go up in temperature the curve flattens out which means that these temperatures are again very unlikely. This is easier to see if you differentiate the CDF and turn it into a PDF, a probability density function, and you get the standard Gaussian distribution that is the norm in repeated physical experiments. The most likely temperature that we can measure is zero and as we move towards higher or lower temperatures their probability decreases very fast. If we move on to a discrete random variable like a die-toss the cumulative distribution function is always piecewise constant because there are no values that the random variable can take except in the set of discrete possible values. So here for a die you have six possible outcomes and so you have six steps in the CDF. You can still differentiate such a function if you use the Dirac Delta formalism. You will get a Dirac Delta at each discontinuity point and the amplitude of the Dirac Delta here is one over six. Now remember if you integrate the PDF over an interval you get the probability that the underlying random variable takes values over that interval. So if I integrate between 2.5 and 4.5 for instance I'm actually just considering the discrete events of the die showing the face number three or number four

notes

summary

4m 39s



$$E[x] = \int_{-\infty}^{\infty} x f_x(x) dx$$

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_x(x) dx$$

and of course the result would be one third. The expectation of a random variable is a way to express concisely some information about the random variable itself and what we do by taking the expectation of the random variable is to compute its mean value. Now there's a very important theorem in probability theory, the expectation

notes

summary

$$E[x] = \int_{-\infty}^{\infty} x f_x(x) dx$$

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_x(x) dx$$



theorem, that says that if we take the expected value of a function of a random variable we can compute that just by knowing the probability density function and computing the integral from minus infinity to plus infinity of the

notes

summary

6m 47s



$$E[x] = \int_{-\infty}^{\infty} x f_x(x) dx$$

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_x(x) dx$$

↑
⏟

function of a dummy variable x that we use for the integration times the PDF of the random variable itself. This will be very useful when we compute for instance the moments of a random variable. These are concise scalar

notes

summary

6m 56s



- ▶ raw moments: $E[x^n] = \int_{-\infty}^{\infty} x^n f_x(x) dx$
- ▶ special case: mean: $m_x = E[x] = \int_{-\infty}^{\infty} x f_x(x) dx$
- ▶ central moments: $E[(x - m_x)^n] = \int_{-\infty}^{\infty} (x - m_x)^n f_x(x) dx$
- ▶ special case: variance: $\sigma_x^2 = E[(x - m_x)^2]$

descriptors of the random variable. Order number one you have the mean, order number two you measure the spread of the random variable and higher order have different applications that we will not really be concerned with. So the raw moments are the expected values of increasing powers of the random variable itself. For n equal to 1 you have the mean that we have seen before. You also have the central moments which are moments of the random variable centered around its mean and by the expectation theorem you can compute this by taking the integral of the dummy variable minus the mean of the random

notes

summary

7m 11s



- ▶ raw moments: $E[x^n] = \int_{-\infty}^{\infty} x^n f_x(x) dx$
- ▶ special case: mean: $m_x = E[x] = \int_{-\infty}^{\infty} x f_x(x) dx$
- ▶ central moments: $E[(x - m_x)^n] = \int_{-\infty}^{\infty} (x - m_x)^n f_x(x) dx$
- ▶ special case: variance: $\sigma_x^2 = E[(x - m_x)^2]$

variable to the power that corresponds to the order of the moment you're computing times the probability density function. A very special case is the variance of a random variable that measures how spread the PDF is around

notes

summary

7m 49s



- ▶ raw moments: $E[x^n] = \int_{-\infty}^{\infty} x^n f_x(x) dx$
- ▶ special case: mean: $m_x = E[x] = \int_{-\infty}^{\infty} x f_x(x) dx$
- ▶ central moments: $E[(x - m_x)^n] = \int_{-\infty}^{\infty} (x - m_x)^n f_x(x) dx$
- ▶ special case: variance: $\sigma_x^2 = E[(x - m_x)^2]$

the mean of the random variable. For example a Gaussian random variable like

notes

summary

8m 1s



- ▶ m is the mean
- ▶ σ^2 is the variance

11

notes

8m 9s



$$f(x) = \mathcal{N}(m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

- ▶ m is the mean
- ▶ σ^2 is the variance



11

the real axis and the variance that measures the spread of the bell curve. And these two parameters enter explicitly in the definition of the PDF

notes

summary

8m 20s



$$f(x) = \mathcal{N}(m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

- ▶ m is the mean
- ▶ σ^2 is the variance



11

of the random variable as you can see here. If you have a uniform random

notes

summary

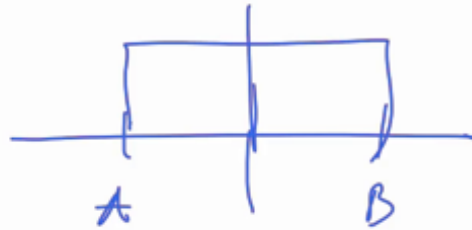
8m 31s



$$f(x) = \mathcal{U}(A, B) = \frac{1}{B - A}$$

► $m = \frac{A + B}{2}$

► $\sigma^2 = \frac{(B - A)^2}{12}$



12

variable, so a variable that is continuous valued but takes values over a certain interval from a to b with uniform probability for all possible values on the interval then you have that the mean is of course the midpoint of this interval here which is a plus b divided by 2 and the variance is computed as b

notes

summary

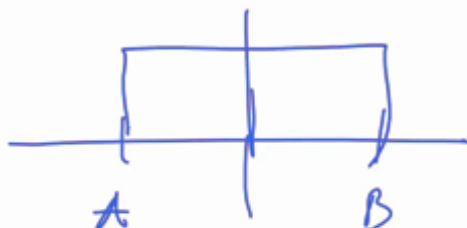
8m 37s



$$f(x) = \mathcal{U}(A, B) = \frac{1}{B - A}$$

► $m = \frac{A + B}{2}$

► $\sigma^2 = \frac{(B - A)^2}{12}$



12

minus a squared over 12. This is an important result that we will see again when we talk about quantization in a future lecture. Of course having a single

notes

summary

8m 55s



- ▶ cross-correlation: $R_{xy} = E[x y]$.
- ▶ covariance: $C_{xy} = E[(x - m_x)(y - m_y)]$.
- ▶ if zero-mean: $C_{xy} = R_{xy}$

to compute the covariance we need to know the *joint* pdf $f_{xy}(x, y)$:

$$E[g(x, y)] = \int \int_{-\infty}^{\infty} g(x, y) f_{xy}(x, y) dx dy$$

14

random variable is hardly any good so we probably will have several to play with and once we have several random variables we're interested in the relations between them. Of course we would like to keep things as simple as possible so it would be nice as in the case of the moments to have some simple scalar descriptor of the interplay between two random variables. So the

notes

summary

9m 4s



- ▶ cross-correlation: $R_{xy} = E[x y]$.
- ▶ covariance: $C_{xy} = E[(x - m_x)(y - m_y)]$.
- ▶ if zero-mean: $C_{xy} = R_{xy}$

to compute the covariance we need to know the *joint* pdf $f_{xy}(x, y)$:

$$E[g(x, y)] = \int \int_{-\infty}^{\infty} g(x, y) \underbrace{f_{xy}(x, y)}_{\text{joint pdf}} dx dy$$

first thing we want to compute is what is called the cross correlation, a measure of how intertwined two random variables are. This is computed as the expectation of the product of the two random variables. The covariance is very similar to the cross correlation but each random variable is re-centered around its mean. If both variables are zero means then the cross correlation and the covariance coincide. Now to compute either cross correlation or covariance we need to know the joint probability density function of the two variables. Now this could be very complicated to obtain or estimate if the interplay between the variables is complicated so we always try to shoot for the simpler cases and we have two such cases. If the variables are uncorrelated then their cross

notes

summary

9m 25s



- ▶ uncorrelated elements: $E[xy] = E[x] E[y] = m_x m_y$
(no linear relationship)
- ▶ independent elements: $f_{xy}(x, y) = f_X(x)f_Y(y)$
(no relationship)

correlation is simply equal to the product of their means. This implies that the two variables have no linear relationship between them. A stronger condition is the independence of the variables. That means that the joint

notes

summary

10m 15s



- ▶ uncorrelated elements: $E[xy] = E[x] E[y] = m_x m_y$
(no linear relationship)
- ▶ independent elements: $f_{xy}(x, y) = f_x(x) f_y(y)$
(no relationship)

probability density function is actually given by the product of the two independent probability density functions for the two variables and this means that there is no relationship at all between the variables.

notes

summary

10m 30s

