# Discrete-Time Random Signals

Jose Krause Perin

Stanford University

June 27, 2018

# Last lecture

## Review of discrete-time signals and systems

- Systems can be linear, time-invariant, memoryless, causal, and stable
- Linear and time-invariant (LTI) systems are completely characterized by their impulse response
- We can use the convolution sum to calculate the output of an LTI system to any signal
- The complex exponential $e^{j\omega n}$, and more generally $z^n$, are eigenfunctions of LTI systems
- Frequency-domain representation of signals and systems
  - Discrete-time Fourier transform (DTFT)
  - $z$-transform and ROC. Without specifying the ROC the $z$-transform is ambiguous
- The DTFT is equivalent to the $z$-transform evaluated on the unit circle: $H(e^{j\omega}) = H(z = e^{j\omega})$, provided that the unit circle is in the ROC of the $z$-transform

# Today's lecture

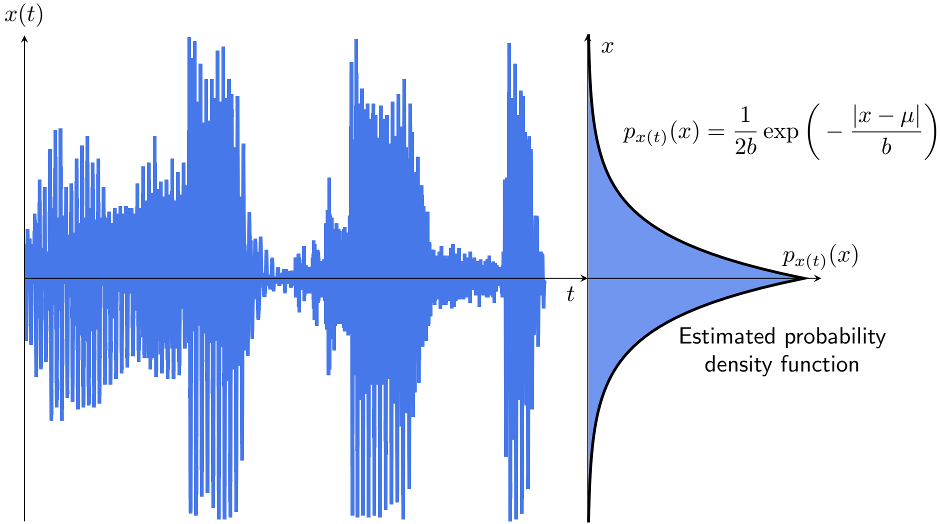
How to analyze linear and time-invariant (LTI) systems when the input is a random signal.

## Motivation

- Many signals vary in complicated patterns that cannot be easily described analytically
- It is often convenient and useful to consider such signals as being created by some sort of random mechanism

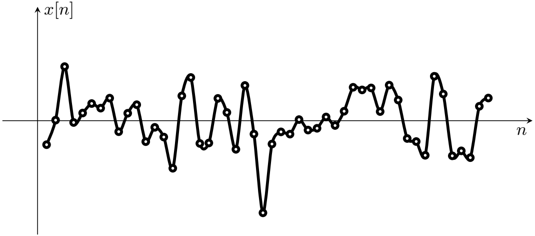
# Example: speech signals

Speech signals are well described by the **Laplacian distribution**



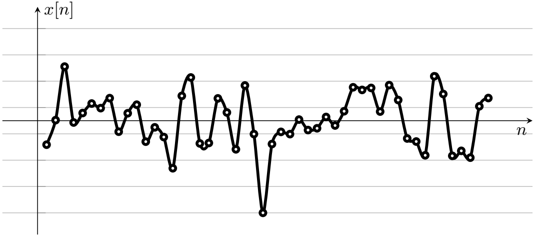
# Example: quantization

Quantization error is well described by an **uniform distribution**, even though quantization is a deterministic operation



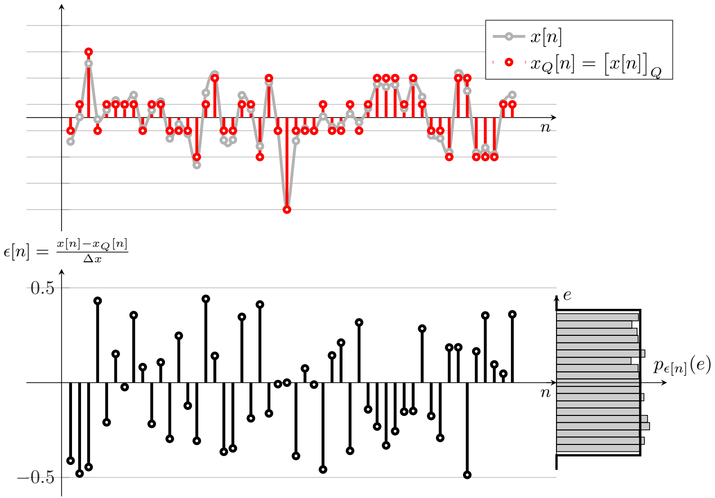
# Example: quantization

Quantization error is well described by an **uniform distribution**, even though quantization is a deterministic operation



# Example: quantization

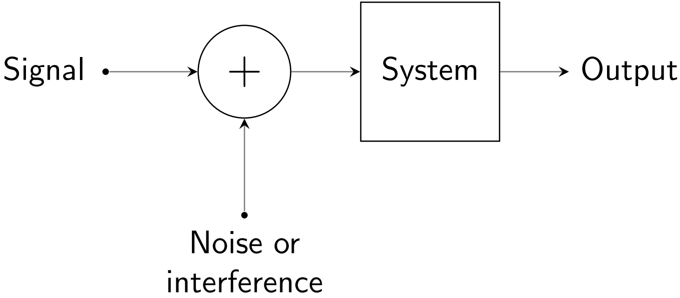
Quantization error is well described by an **uniform distribution**, even though quantization is a deterministic operation



# Example: noise and interference

Noise and interfering signals are typically modeled as **random processes**

1. What's the effect of the noise on the output?
2. How can we design the system to minimize the noise at the output?



# Outline

**Our goal:** analyze LTI systems when the input signal is random

1. Random processes
   - Averages of a random variable
   - Stationary random processes
   - Time averages and ergodic random processes
2. LTI systems with a random input
3. White noise
4. Examples

# Random processes

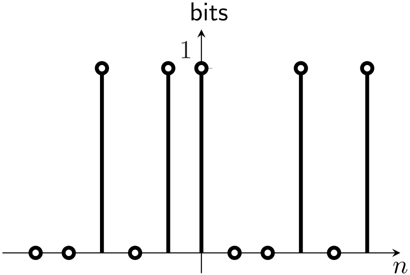
## Definition

A random process (or *stochastic process*) is an indexed set of random variables $x_n$, which are distributed according to some probability distribution $p_{x_n}(x)$

## Examples

Consecutive coin tosses

Random bit stream



# Random processes

We can think of a random process as a function $X(n, \chi)$ of two variables, time $n$ and the outcome of the underlying random experiment $\chi$.

- For fixed $n$, $X(n, \chi)$ is a random variable.
  In the example of the fair coin tossing,

$$X(n = 1, \chi) = \begin{cases} \mathrm{H}, & \text{with probability } 0.5 \\ \mathrm{T}, & \text{with probability } 0.5 \end{cases}$$

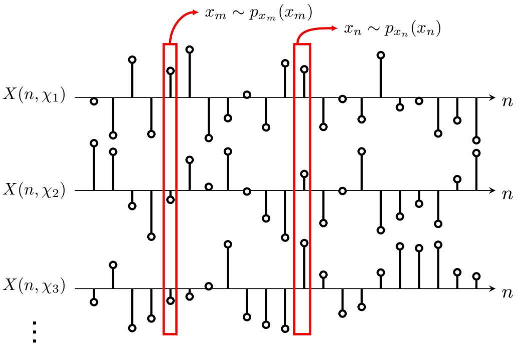
- For fixed $\chi$, $X(n, \chi)$ is a deterministic function of $n$ called a **sample function** or **sample sequence**


$\cdots$

$X(1, \chi) = \mathrm{H}$ $\cdots$ $X(4, \chi) = \mathrm{H}$

# Ensemble of sample sequences

The ensemble of sample sequences is a collection af all sequences generated by a random process.




**Recall:**

- For fixed $n$, we have a random variable (must use probability theory)
- For a fixed $\chi$, we have a deterministic signal

With this interpretation of random processes, we have two axes:

- The probability axis (fixed $n$), whereby the samples are random variables. Hence, we must use probability theory to analyze them.
- And we have the sample function axis (fixed $\chi$), whereby we have a deterministic signal generated by a realization of the random process.

We'll start by dealing with random variables. We'll see what we can learn about a random process by calculating probability averages such as mean, variance, and autocorrelation function.

Then, we'll move to dealing with sample functions, and we'll calculate their time averages.

Later we'll see how we can relate probability averages to time averages.

# Averages of a random variable

**Mean or expected value**

$$\mu_{x_n} = \mathbb{E}(x_n) = \int_{-\infty}^{\infty} x p_{x_n}(x) dx$$

**Average power or second moment**

$$\mathbb{E}(|x_n|^2) = \int_{-\infty}^{\infty} |x|^2 p_{x_n}(x) dx$$

**Variance**

$$\begin{aligned} \sigma_{x_n}^2 &= \mathbb{E}(|x_n - \mu_{x_n}|^2) = \int_{-\infty}^{\infty} |x - \mu_{x_n}|^2 p_{x_n}(x) dx \\ &= \mathbb{E}(|x_n|^2) - \mu_{x_n}^2 \end{aligned}$$

The integrals should be replaced by sums when the random variable is discrete.

# Joint averages of random variables

Expected value of a function of two random variables

$$\mathbb{E}(g(x_n, y_m)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) p_{x_n, y_n}(x, y) dx dy$$

Two random variables are **uncorrelated** if

$$\mathbb{E}(x_n y_m) = \mathbb{E}(x_n)\, \mathbb{E}(y_m)$$

Two random variables are **statistically independent** if

$$p_{x_n, y_n}(x, y) = p_{x_n}(x) p_{y_n}(y)$$

Independent random variables are also uncorrelated, but not all uncorrelated random variables are independent.
**Exception:** uncorrelated Gaussian random variables are always independent.

# Correlation functions

**Autocorrelation**

$$\phi_{xx}[n,m] = \mathbb{E}(x_n x_m^*)$$

**Autocovariance**

$$\gamma_{xx}[n,m] = \mathbb{E}((x_n - \mu_{x_n})(x_m - \mu_{x_m})^*)$$

**Cross-correlation**

$$\phi_{xy}[n,m] = \mathbb{E}(x_n y_m^*)$$

**Cross-covariance**

$$\gamma_{xy}[n,m] = \mathbb{E}((x_n - \mu_{x_n})(y_m - \mu_{y_m})^*)$$

**Note:** By knowing the mean and the autocorrelation function, we can determine average power, variance, and autocovariance.

# Example: Bernoulli random process

- A Bernoulli random process is a sequence of binary random variables $\{x_n \sim \mathcal{B}(\rho)\}$. Canonically,

$$x_n = \begin{cases} 1, & \text{with probability } \rho, \\ 0, & \text{with probability } 1-\rho, \end{cases} \qquad p_{x_n}(x) = \begin{cases} \rho, & x = 1 \\ 1-\rho, & x = 0 \\ 0, & \text{otherwise} \end{cases}$$

- A Bernoulli process is **independent and identically distributed (IID)**. That is, each $x_n$ is picked independently from the same distribution $\mathcal{B}(\rho)$.

From this we can conclude:

$$\begin{aligned} \mu &= 1 \cdot \rho + 0 \cdot (1-\rho) = \rho \\ \mathbb{E}(x_n^2) &= 1^2 \cdot \rho + 0^2 \cdot (1-\rho) = \rho \\ \sigma^2 &= \mathbb{E}(x_n^2) - \mu^2 = \rho(1-\rho) \\ \phi_{xx}[m] &= \mathbb{E}(x_{n+m}x_n) = \rho\delta[m] && \text{(since it is IID)} \end{aligned}$$

# Example: Uniform random process

- An uniform random process is a sequence of uniform random variables $\{x_n \sim \mathcal{U}[a,b]\}$.

$$p_{x_n}(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

From this we can conclude:

$$\mu = \int_a^b \frac{x}{b-a} dx = \frac{b+a}{2}$$

$$\mathbb{E}(x_n^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{b^2+ab+a^2}{3}$$

$$\sigma^2 = \mathbb{E}(x_n^2) - \mu^2 = \frac{(b-a)^2}{12}$$

$$\phi_{xx}[m] = \mathbb{E}(x_{n+m}x_n) = \mathbb{E}(x_n^2)\delta[m] \qquad \text{(Assuming IID)}$$

# Example: Gaussian random process

- A Gaussian random process is a sequence of Gaussian random variables $\{x_n \sim \mathcal{N}(\mu, \sigma^2)\}$.

$$p_{x_n}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

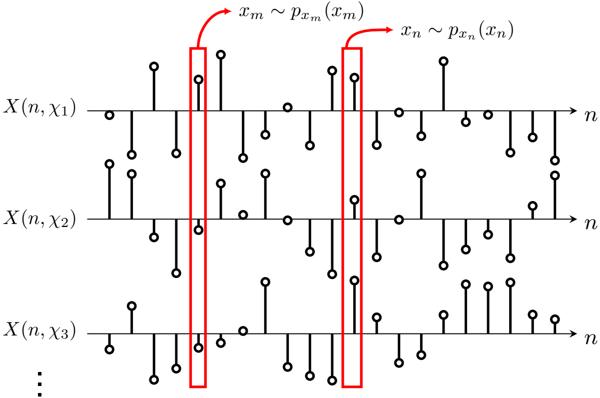
From this we can conclude:

$$\mathbb{E}(x_n^2) = \sigma^2 + \mu^2$$

$$\phi_{xx}[m] = \mathbb{E}(x_{n+m}x_n) = \mathbb{E}(x_n^2)\delta[m] \qquad \text{(Assuming IID)}$$

# Stationary random processes

- Stationarity refers to **time invariance** of some, or all, of the statistics of a random process, such as mean, autocorrelation, joint distributions, etc
- A random process is **strict-sense stationary (SSS)**, if its finite-order distributions do not change over time. For the first-order distributions, this means $p_{x_m}(x_m) = p_{x_n}(x_n),\ \forall n, m$.



# Stationary random processes

All statistics of a SSS random process are time invariant.
As a result, the mean, average power, and variance are constant with $n$:

$$\mu = \mathbb{E}(x_n)$$
$$\sigma^2 = \mathbb{E}(|x_n|^2) - \mu^2$$

And the autocorrelation function only depends on the time difference $m$:

$$\phi_{xx}[m] = \phi_{xx}[n+m, n] = \mathbb{E}(x_{n+m}x_n^*)$$

**Question:** What is an example of SSS random process?

# Wide-sense stationary (WSS) random processes

- Strict sense stationarity is a strong condition that is hard to verify in practice.
- A weaker (and more useful) condition is **wide-sense stationarity**
- A random process is **wide-sense stationary (WSS)** if its mean and autocorrelation function are **time invariant**.

$$\mu = \mathbb{E}(x_n)$$
$$\phi_{xx}[m] = \phi_{xx}[n+m, n] = \mathbb{E}(x_{n+m}x_n^*)$$

  The mean is constant, and the autocorrelation function only depends on the time difference $m$.
- SSS implies WSS, but WSS does not mean SSS.
  **Exception:** WSS Gaussian random processes are also SSS.

# Autocorrelation function of WSS processes

The autocorrelation function $\phi_{xx}[m]$ of a WSS process $x[n]$ has the following properties

1. $\phi_{xx}[m]$ is **real valued**
2. $\phi_{xx}[m]$ is **even symmetric**, i.e., $\phi_{xx}[m] = \phi_{xx}[-m]$
3. The DTFT of $\phi_{xx}[m]$ must be **non-negative at all frequencies**

$$\mathcal{F}\{\phi_{xx}[m]\} \geq 0, \forall\, \omega \in [-\pi, \pi] \tag{1}$$

   $\mathcal{F}\{\cdot\}$ denotes the DTFT.

Properties 1 to 3 are **necessary and sufficient** for a function to be an autocorrelation function of a WSS process.

More properties

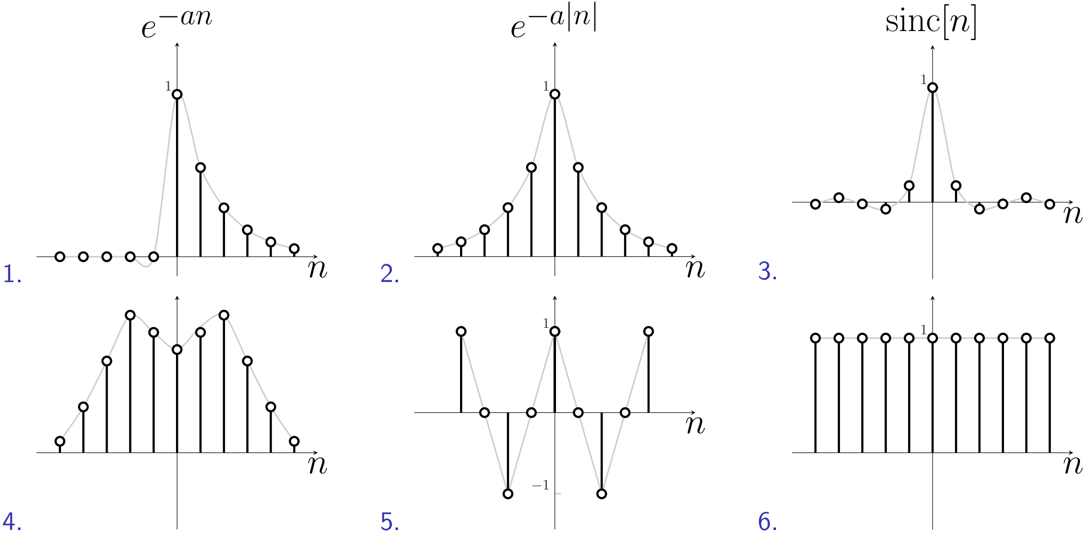
4. $|\phi_{xx}[m]| \leq \phi_{xx}[0] = \mathbb{E}(|x[n]|^2) =$ **average power** of $x[n]$

   *Proof:*

$$\begin{aligned}
\phi_{xx}^2[m] &= [\mathbb{E}(x[m+n]x[n])]^2 \\
&\leq \mathbb{E}(|x[m+n]|^2)\,\mathbb{E}(|x[n]|^2) && \text{(by Schwarz inequality)} \\
&= \phi_{xx}^2[0] && \text{(by stationarity)}
\end{aligned}$$

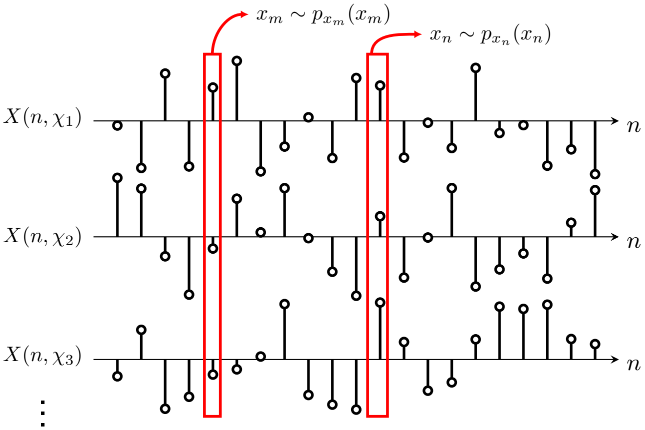
5. If $\phi_{xx}[T] = \phi_{xx}[0]$ for some $T \neq 0$, then $\phi_{xx}[m]$ is periodic with period $T$.

# Which functions can be $\phi_{xx}[m]$ of a WSS process?



# Ensemble of sample functions

- So far we have focused on random variables
- What can we learn from a sample function?



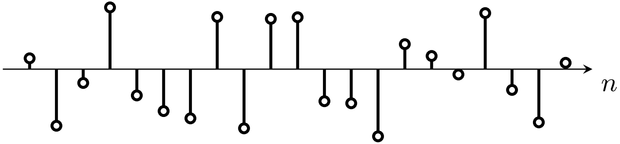
# Time averages

- Until now we have focused on probability averages $\mathbb{E}(\cdot)$
- We can also define time averages $\langle\cdot\rangle$

$$\langle x[n]\rangle = \lim_{L\to\infty} \frac{1}{2L+1} \sum_{n=-L}^{L} x[n]$$

$$\langle x[n+m]x^*[n]\rangle = \lim_{L\to\infty} \frac{1}{2L+1} \sum_{n=-L}^{L} x[n+m]x^*[n]$$

We can calculate time averages of any deterministic signal




**Careful with notation:** We use $x_n$ to refer to the *random variables* of some random process, whereas $x[n]$ denotes a *sample function* of some random process.

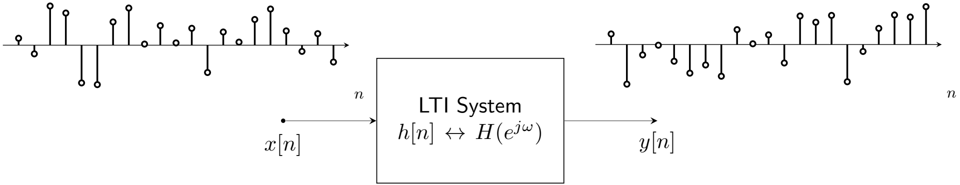
# Ergodic random processes

A random process is **ergodic** if its time averages are equal to its probability averages:

$$\begin{aligned} \text{Time averages} &= \text{Probability averages} \\ \langle x[n] \rangle &= \mathbb{E}(x_n) = \mu && \text{(expected value)} \\ \langle x[n+m]x^*[n] \rangle &= \mathbb{E}(x_{n+m}x_n^*) = \phi_{xx}[m] && \text{(autocorrelation function)} \end{aligned}$$

- In practice, we don't have an ensemble of sample functions that we can use to estimate the mean and autocorrelation function.
- We generally have only one sample function.
- With the **ergodic assumption**, we can estimate probability averages from a single sample function

# LTI system with a random input




For a given **sample function** $x[n]$, we can apply the convolution sum as usual:

$$y[n] = \sum_{m=-\infty}^{\infty} x[n-m]h[m]$$

**More importantly:** what is the effect of the system on the statistics (e.g., mean and autocorrelation function) of the random process?

# LTI system with a random input

## Mean or expected value

$$\begin{aligned}
\mu_y = \mathbb{E}(y[n]) &= \mathbb{E}\left(\sum_{n=-\infty}^{\infty} x[m-n]h[m]\right) \\
&= \sum_{n=-\infty}^{\infty} \mathbb{E}(x[m-n])h[m] \\
&\qquad \text{(Expectation is a linear operator and } h[n] \text{ is not random)} \\
&= \mu_x \sum_{n=-\infty}^{\infty} h[m] \qquad \text{(Assuming } x[n] \text{ is WSS)} \\
&= \mu_x H(e^{j0})
\end{aligned}$$

The mean is scaled by the gain of the LTI system at zero frequency.

# LTI system with a random input

## Autocorrelation function

$$
\begin{aligned}
\phi_{yy}[m] &= \mathbb{E}(y[n+m]y^*[n]) && \text{(by definition)} \\
&= \mathbb{E}\left\{\left(\sum_{r=-\infty}^{\infty} x[n+m-r]h[r]\right) \cdot \left(\sum_{k=-\infty}^{\infty} x^*[n-r]h^*[k]\right)\right\} \\
&= \sum_{r=-\infty}^{\infty} h[r] \sum_{k=-\infty}^{\infty} h^*[k]\, \mathbb{E}(x[n+m-r]x^*[n-k]) \\
&= \sum_{l=-\infty}^{\infty} \left(\sum_{k=-\infty}^{\infty} h[l+k]h^*[k]\right)\phi_{xx}[m-l] && \text{(change variables } r = l+k\text{)}
\end{aligned}
$$

Let's define the **autocorrelation function of deterministic signals**

$$c_{hh}[l] \equiv \sum_{k=-\infty}^{\infty} h[l+k]h^*[k]$$

Note that the autocorrelation function of deterministic signals and convolution are closely related:

$$c_{hh}[l] = h[l] * h^*[-l]$$

Now we can rewrite the autocorrelation function of the output of a LTI system to a random process more compactly:

$$\begin{aligned}\phi_{yy}[m] &= \sum_{l=-\infty}^{\infty} c_{hh}[l]\phi_{xx}[m-l] \\ &= c_{hh}[m] * \phi_{xx}[m] = h[l] * h^*[-l] * \phi_{xx}[m]\end{aligned}$$

The autocorrelation function of the input random process is *filtered* by the deterministic autocorrelation function of the impulse response.

# In the frequency domain

From the previous derivation in the time domain:

$$\begin{aligned}\phi_{yy}[m] &= c_{hh}[m] * \phi_{xx}[m] \\ &= h[l] * h^*[-l] * \phi_{xx}[m]\end{aligned}$$

In the frequency domain:

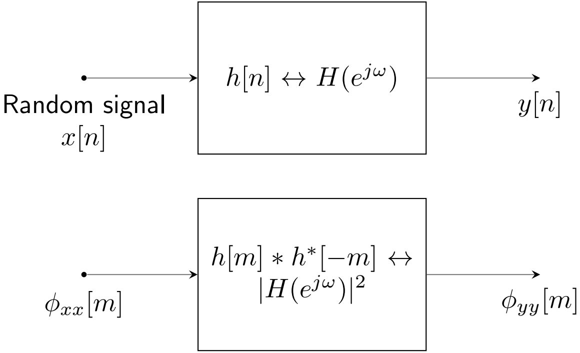
$$\begin{aligned}\mathcal{F}\{\phi_{yy}[m]\} &= H(e^{j\omega}) \cdot H^*(e^{j\omega}) \cdot \mathcal{F}\{\phi_{xx}[m]\} \\ &= |H(e^{j\omega})|^2 \cdot \mathcal{F}\{\phi_{xx}[m]\}\end{aligned}$$

- The DTFT of the autocorrelation function of a random process is called **power spectrum density (PSD)**
- The PSD has units of W/Hz or dBm/Hz
- **Notation:** $\Phi_{xx}(e^{j\omega}) \equiv \mathcal{F}\{\phi_{xx}[m]\}$

# Properties of the power spectrum density

1. The PSD is **real valued**
$$\Phi_{xx}(e^{j\omega}) = \Phi_{xx}^*(e^{j\omega}),$$
since the autocorrelation function has **even symmetry**: $\phi_{xx}[m] = \phi_{xx}[-m]$.
2. The PSD is **even symmetric**
$$\Phi_{xx}(e^{j\omega}) = \Phi_{xx}(e^{-j\omega}),$$
since the autocorrelation function is always real.
3. The PSD is **non negative**
$$\Phi_{xx}(e^{j\omega}) \geq 0,$$
This is the same condition required by an autocorrelation function of a WSS random process.
4. The area under $\Phi_{xx}(e^{j\omega})$ is the **average power**
$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{xx}(e^{j\omega})d\omega = \phi_{xx}[0] = E(|x[n]|^2)$$

# Effect of LTI system on the autocorrelation function

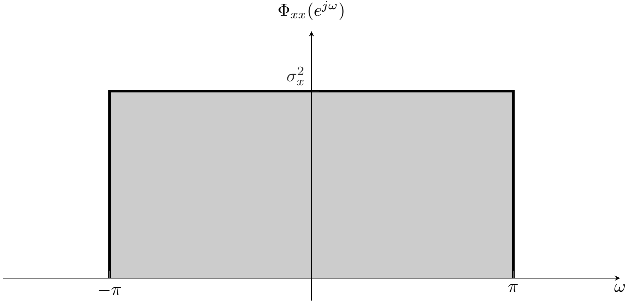



- The input autocorrelation function is *filtered* by the LTI system defined by $h[n] * h^*[-n] \leftrightarrow |H(e^{j\omega})|^2$
- The random process remains WSS after an LTI system

# White noise

- White noise is a particularly important class of random process
- A **zero-mean** random process whose **PSD is constant over all frequencies** is commonly referred to as **white noise**. Constant PSD over all frequencies implies that the samples are **uncorrelated** in the time domain

$$\Phi_{xx}(e^{j\omega}) = \sigma_x^2, |\omega| \leq \pi \iff \phi_{xx}[m] = \sigma_x^2 \delta[m]$$



# White noise into LTI system

For a white noise input,

$$\phi_{yy}[m] = c_{hh}[m] * \phi_{xx}[m] = \sigma_x^2 c_{hh}[m] \qquad \text{(output autocorrelation function)}$$

$$\begin{aligned}\Phi_{yy}(e^{j\omega}) &= |H(e^{j\omega})|^2 \Phi_{xx}(e^{j\omega}) \\ &= \sigma_x^2 |H(e^{j\omega})|^2 \qquad \text{(output PSD)}\end{aligned}$$

- Note that the output noise PSD may not be white. Hence, we say that the filter $H(e^{j\omega})$ **colored** the noise or **shaped** the noise.
- It is typically easier to analyze systems with white noise. As a result, it is common to employ a **noise whitening filters** to make the noise white.

# Example: moving average system

Recall the 4-point moving average system defined by the difference equation:

$$y[n] = \frac{1}{4}(x[n] + x[n-1] + x[n-2] + x[n-3])$$

This system has impulse response:

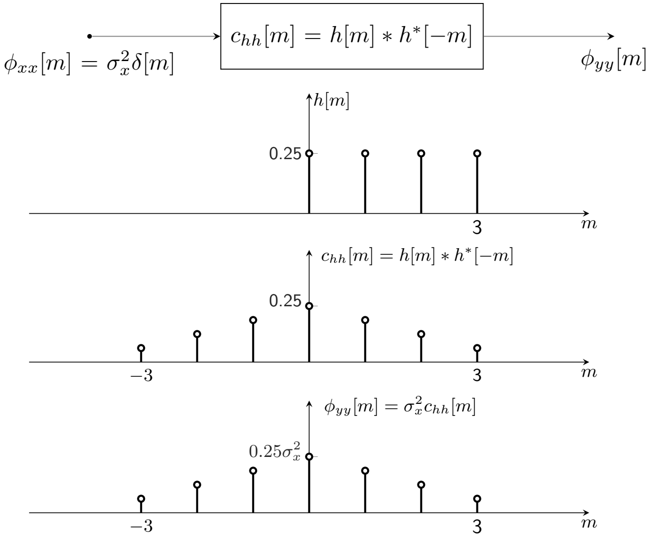
$$h[n] = \frac{1}{4}(\delta[n] + \delta[n-1] + \delta[n-2] + \delta[n-3])$$

And frequency response:

$$H(e^{j\omega}) = \frac{1}{4}(1 + e^{-j\omega} + z^{-j2\omega} + z^{-j3\omega}) = \frac{\sin(2\omega)}{4\sin(\omega/2)}e^{-j2\omega}$$

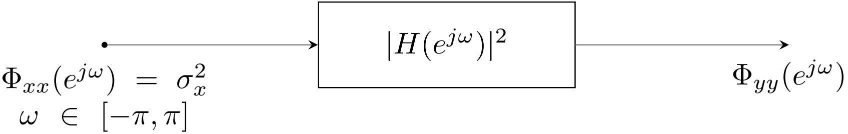
# White noise into moving average filter

In **time domain**:



# White noise into moving average filter

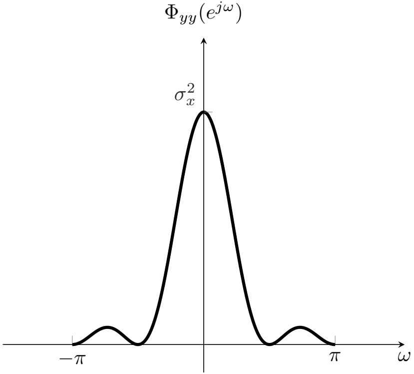
In the **frequency domain**:




$$H(e^{j\omega}) = \frac{\sin(2\omega)}{4\sin(\omega/2)} e^{-j2\omega}$$

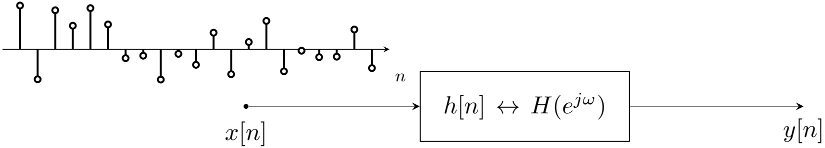
The output PSD is therefore

$$\begin{aligned}\Phi_{yy}(e^{j\omega}) &= \sigma_x^2 |H(e^{j\omega})|^2 \\ &= \sigma_x^2 \left( \frac{\sin(2\omega)}{4\sin(\omega/2)} \right)^2\end{aligned}$$



# Simulation example

We'll generate an uniform random process and filter it by a 4-point moving average filter:
$h[n] = \frac{1}{4}(\delta[n] + \delta[n-1] + \delta[n-2] + \delta[n-3])$




In Matlab:
Generate $1 \times 1000$ random vector uniformly distributed in $[-1, 1]$

```
>> x = 2*rand(1, 1000) - 1
```

Calculate output:

```
>> y = filter([1, 1, 1, 1]/4, 1, x)
```

# Estimating the autocorrelation function

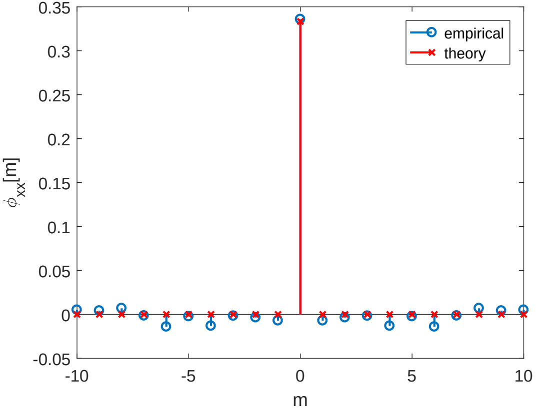
## Theoretical autocorrelation

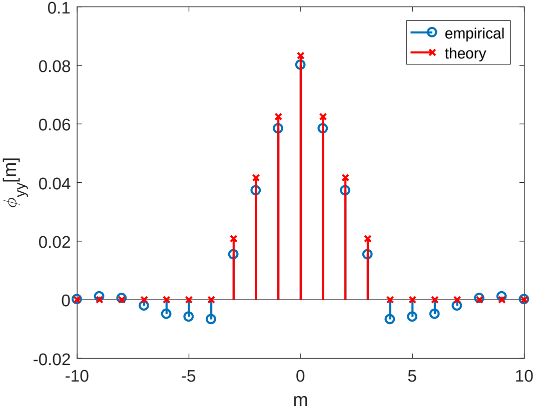
```
>> maxLag = 10 % maximum lag of our autocorrelation function
>> h = [1, 1, 1, 1]/4 % impulse response
>> chh = conv(h, conj(fliplr(h))) % deterministic autocorrelation
>> phi_xx_theory = zeros(1, 2*maxLag+1)
>> phi_xx_theory(maxLag+1) = 1/3 % theoretical ϕxx[0] for xn ∼ U[−1, 1]
>> phi_yy_theory = conv(phi_xx_theory, chh, 'same')
>> stem(-maxLag:maxLag, phi_xx_theory) % Plot
>> stem(-maxLag:maxLag, phi_yy_theory)
```

# Estimating the autocorrelation function

## Empirical autocorrelation

```
>> phi_yy = xcorr(y, y, maxLag, 'unbiased') % estimate autocorrelation
>> stem(-maxLag:maxLag, phi_yy)
```

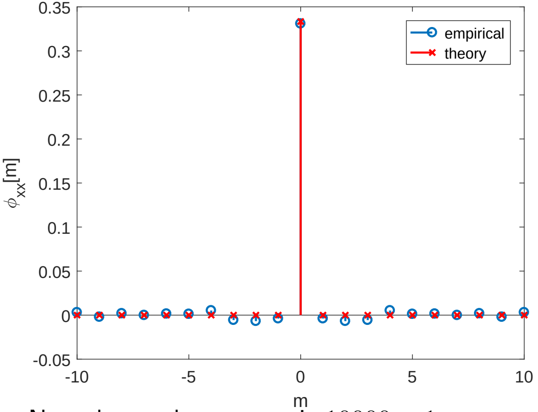


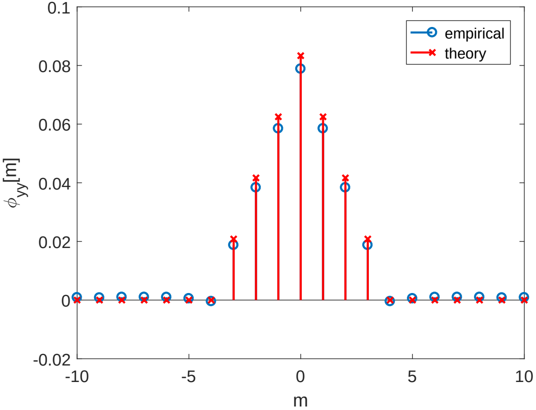



# Estimating the autocorrelation function

## Empirical autocorrelation

```
>> phi_yy = xcorr(y, y, maxLag, 'unbiased') % estimate autocorrelation
>> stem(-maxLag:maxLag, phi_yy)
```

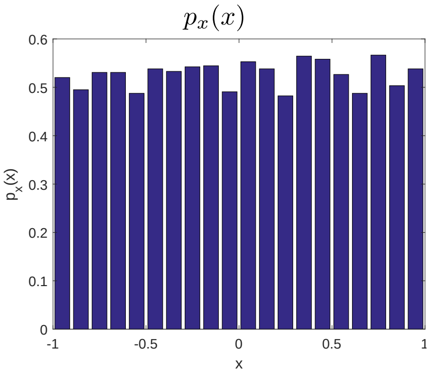


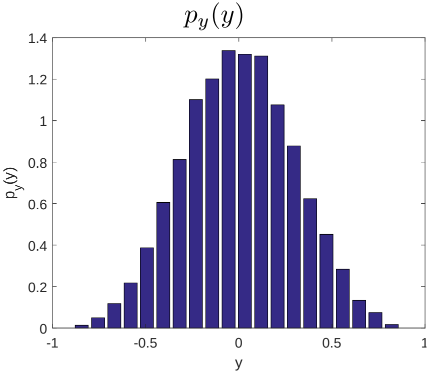




Now the random vector is $10000 \times 1$

# Estimating the probability density function

```
>> Nbins = 20
>> [counts, centers] = hist(x, Nbins)
% normalize to make area under pdf equal to 1
>> x_pdf = Nbins/(centers(end) - centers(1))*counts/sum(counts)
>> bar(centers, x_pdf) % plot
```



# Central limit theorem

The **central limit theorem** states that the probability density function of the sum of a large number of independent random variables approaches a Gaussian distribution.

$$Z = X_1 + X_2 + \ldots + X_N \implies p_Z(z) \xrightarrow{N\to\infty} \mathcal{N}(\mu, \sigma^2)$$

- In digital filters, we're essentially performing a weighted sum of samples of the input. As a result, the output of a digital filter for a random input is approximately Gaussian distributed.
- In the example of the 4-point moving average system, we were only adding 4 random variables and the resulting output pdf was close to Gaussian.
- This effect is more visible in filters with larger memory, and in particular in filters with **infinite impulse response (IIR)**

# Another probability theorem

If we add two independent random variables $Z = X + Y$, the pdf of $Z$ is given by the convolution of the pdfs of $X$ and $Y$

$$p_Z = p_X * p_Y$$

We can extend that to a sum of several independent random variables:

$$Z = \sum_{k=1}^{N} X_k \implies p_Z = p_{X_1} * \ldots * p_{X_N}$$

From the **central limit theorem**, we know that as $N \to \infty$, $p_Z \to \mathcal{N}(\mu, \sigma^2)$

- This shows that the convolution of many signals $h_1(t) * h_2(t) * \ldots * h_N(t) \approx g(t)$, where $g(t)$ is the Gaussian function.
- Hence, if we cascade many LTI systems, the impulse response of the equivalent LTI system is approximately the Gaussian function.

# Summary

- A random process is an indexed collection of random variables
- A random process is strict-sense stationary (SSS) if all its finite-order statistics are time invariant. That's hard to verify in practice.
- A random process is wide-sense stationary (WSS) if its mean is constant and if its autocorrelation function is only a function of the time difference.
- A random process is ergodic if its time averages are equal to its probability averages
- The Fourier transform of the autocorrelation function is called the power spectrum density (PSD). The PSD has units of W/Hz or dBm/Hz.
- When a random signal is filtered by an LTI system defined by $h[n] \leftrightarrow H(e^{j\omega})$, its autocorrelation function is filtered by an LTI system defined by $h[n] * h^*[-n]$, and its PSD is shaped by $|H(e^{j\omega})|^2$
- Random processes that have PSD constant over all frequencies are called white noise
- By the central limit theorem, the output of an LTI system to a random input is approximately Gaussian distributed