Random Sequences (Discrete-time random processes)

- Further generalize the law random vectors
  or \( \{X_n\}_{n \in \mathbb{Z}} \)

- Define \( \{X_n\}_{n \in \mathbb{Z}} \) be a countable collection of random variables
  defined on prob. space \((\Omega, \mathcal{F}, P)\). This collection of
  r.v.'s is called a random sequence.

- Remark (i) The probabilistic behavior of \( \{X_n\} \) is well-defined
  for any \( A \in \mathcal{F} \) associated with the random seq. that can be obtained
  from a countable \# of set operations, from events involving
  the r.v.'s, i.e., \( A \in \mathcal{F} \) and \( P(A) \) is well-defined.

  For instance, for every \( \{i_1, i_2, \ldots, i_m\} \subset \mathbb{Z} \),

  \[
  F_{x_{i_1} x_{i_2} \cdots x_{i_m}}(x_{i_1}, x_{i_2}, \ldots, x_{i_m}) = P(X_{i_1} \leq x_{i_1}, \ldots, X_{i_m} \leq x_{i_m}).
  \]

  Further for \( \{i_1, i_2, \ldots, i_m\} \subset \mathbb{Z} \), by the continuity of prob. measures

  \[
  P\left( \bigcap_{n=1}^{\infty} \{X_n \leq x_{i_n}\} \right) = \lim_{k \to \infty} P\left( \bigcap_{i=1}^{k} \{X_{i_n} \leq x_{i_n}\} \right).
  \]

(ii) For each \( w \in \Omega \), \( \{X_n(w)\}_{n \in \mathbb{Z}} \) can be regarded as a
sequence of real numbers (or a discrete-time signal). Thus
we can view \( \{X_n\}_{n \in \mathbb{Z}} \) as a mapping from \( \Omega \) to the
set of sequences of real numbers (discrete-time signals).
This interpretation is convenient in many engineering
applications.
(iii) It is sometimes more convenient to replace the indexing set \( \mathbb{Z} \) by \( \mathbb{N} = \{0, 1, 2, \ldots\} \). This does not amount to any change, but a mere reindexing of the r.v.s in the random seq. (Why?)

Mean, autocorrelation and autocovariance functions of a random seq.

**Def.** The mean function \( \mu_{X} : \mathbb{Z} \to \mathbb{R} \) of the random seq. \( \{X_{n}\}_{n \in \mathbb{Z}} \) is defined as

\[
\mu_{X}(n) = E[X_{n}]
\]

(iii) The autocorrelation func \( R_{X}(m, n) : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R} \) of \( \{X_{n}\}_{n \in \mathbb{Z}} \) is defined as

\[
R_{X}(m, n) = E[X_{m}X_{n}]
\]

(iii) The autocovariance func \( K_{X}(m, n) : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R} \) of \( \{X_{n}\}_{n \in \mathbb{Z}} \) is defined as

\[
K_{X}(m, n) = E[(X_{m} - \mu_{X}(m))(X_{n} - \mu_{X}(n))]
\]

(When they exist!)

(iv) \( \sigma_{X}^{2}(n) = K_{X}(n, n) \) is called the variance func of \( \{X_{n}\}_{n \in \mathbb{Z}} \)

**Properties:**

(i) \( K_{X}(m, n) = R_{X}(m, n) - \mu_{X}(m)\mu_{X}(n) \).

(ii) (Schwarz) \( |R_{X}(m, n)|^{2} \leq R_{X}(m, m) R_{X}(n, n) \)

\[1K_{X}(m, n) \leq \sigma_{X}^{2}(m) \sigma_{X}^{2}(n) \]

(iii) \( K_{X}(m, n)[R_{X}(m, n)] \) is positive semi-definite func, i.e.

\[
\sum_{m} \sum_{n} Z_{m}K_{X}(m, n)Z_{n} \geq 0 \quad \text{for all } \{Z_{n}\}_{n \in \mathbb{Z}}.
\]
Special types of random seqs.

In general, a random seq. may not have enough structure for us to talk too much about. Here we focus on some specific subclasses of random seqs. that have some interesting structures:

(i) **Independent random sequence**

If \( Z \) is a sequence of independent random variables, then \( \{Z_i\}_{i=1}^{\infty} \) is an independent random sequence.

\( \{Z_i\}_{i=1}^{\infty} \) is called independent if for any \( \{\ell_1, \ell_2, ..., \ell_m\} \subset \mathbb{Z} \)

\( \{Z_{\ell_1}, Z_{\ell_2}, ..., Z_{\ell_m}\} \) are independent r.v.'s.

**Example:** Suppose \( \{Z_n\}_{n \in \mathbb{Z}} \) consists of the r.v.'s corresponding to the Bernoulli r.v.s obtained from a seq. of independent coin tosses, s.t.

\[ Z_n(\text{H}) = 1 \quad \text{and} \quad Z_n(\text{T}) = 0 \quad \text{and} \quad P(Z_n = 1) = p_n \quad \text{and} \quad P(Z_n = 0) = q_n \]

where \( p_n > 0 \) and \( q_n = 1 - p_n \).

Obviously \( \{Z_n\}_{n \in \mathbb{Z}} \) is then independent.

\[ E[Z_n] = P(Z_n = 1) = p_n \]

\[ E[Z_n^2] = E[Z_n] = p_n \]

\[ \text{Var}(Z_n) = p_n q_n \]

\[ \text{Cov}(Z_m, Z_n) = 0. \]

(In general, for any indep. \( \{Z_n\}_{n \in \mathbb{Z}} \), \( \text{K}_2(m,n) = 0 \) and \( \text{R}_2(m,n) = \text{K}_1(m) \text{K}_1(n) \))

For the special case where \( p_n = p \) for all \( n \in \mathbb{Z} \), the \( Z_n \)'s are identically distributed, we call this kind of random processes identically distributed (iid) random seqs.


To be more precise, we need to describe the prob. space underlying which such \( \{X_n\}_{n \in \mathbb{Z}} \) is defined. We will briefly do so for the iid Bernoulli random seq. mentioned above. For the other more complicated cases described hereafter, we will just believe that the underlying prob. space can be well-specified. In most of these cases, we don't really need a detailed description of the prob. space for us to handle the engineering problems of interest.

Going back to the iid Bernoulli random seq. \( \{X_1, X_2, \ldots\} \), \( \Omega \) = set of all infinite binary seqs. with elements \{H, T\}

Note that there is a 1-1 correspondence between \( \Omega \) and real numbers on the interval \([0, 1]\), via the dyadic numbers that we described in HW 2 for example.

Thus we can think events in \( \mathcal{B}[0, 1] \) and map them back to subsets of \( \Omega \). This collection of subsets will form our event space \( \mathcal{F} \).

For the prob. measure \( P \), we want it to satisfy, in addition to the 3 basic axioms, \( P(\Omega = 1) = p \) for all \( n \) and \( P(X_{i_1} = x_{i_1}, X_{i_2} = x_{i_2}, \ldots; X_{i_m} = x_{i_m}) = \prod_{k=1}^{m} P(X_{i_k} = x_{i_k}) \) for all \( i_1, \ldots, i_m \), \( m \) and \( x_{i_k} \in \{0, 1\} \). It turns out that there exists a unique prob. measure that satisfies these requirements.

Actually the latter requirement suggests us a way to construct \( \Omega \) from extension products of \( \{H, T\}, \{0, 1\}, P' \) with \( P'(H) = p \) and \( P'(T) = q \).
Gaussian random seqs.
- A seq. \( \{X_n\} \) is called a Gaussian random seq. if for any \( i_1, \ldots, i_m \in \mathbb{Z} \), \( \{X_{i_1}, \ldots, X_{i_m}\} \) is jointly Gaussian r.v.'s.

- A random seq. is called uncorrelated if
  \[
  K_{X}(m, n) = \begin{cases} 
  \sigma^2(m) & \text{if } m = n \\
  0 & \text{if } m \neq n.
  \end{cases}
  \]

- Obviously, independent random seq. is uncorrelated.
- For a Gaussian random seq., uncorrelatedness \( \leftrightarrow \) independence.

Independent Increment seqs.
- In most cases, we build more complicated random seqs. from simpler ones. One such example is independent increment seqs.

Consider an independent random seq. \( \{X_n\}_{n \in \mathbb{N}} \). Let
\[
Y_n = \sum_{i=1}^{n} X_i.
\]

Then for any \( i_1, \ldots, i_m \in \mathbb{N} \times \mathbb{N} \), \( Y_{i_1}, Y_{i_1} - Y_{i_1-1}, Y_{i_2} - Y_{i_2-1}, \ldots, Y_{i_m} - Y_{i_m-1} \) are independent r.v.'s.
Such a random seq. \( \{Y_n\}_{n \in \mathbb{N}} \) is said to have independent increments.
Let \( \{X_i, \ldots\} \) be a \( \overline{\text{iid Bernoulli}} \) seq. of parameter \( p > 0 \) 
\( \Pr(X_i = 1) = p \),

Consider \( Y_n = \begin{cases} \frac{1}{n} \sum_{i=1}^{n} X_i & n > 1 \\ 0 & n = 0 \end{cases} \).

By construction, \( \{Y_n\}_{n \in \mathbb{N}} \) is an \textit{indep.} incremenet seq.

\[
\mu_{Y}(n) = \mathbb{E}[Y_n] = \sum_{i=1}^{n} \mathbb{E}[X_i] = (2p-1)n
\]

\[
\nu_{Y}(m,n) = \text{cov}(Y_m, Y_n) = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{E}[(X_i - 2p+1)(X_j - 2p+1)] = \sum_{i=1}^{m} \text{var}(X_i)
\]

\[
\sigma_{Y}(n) = 2npq
\]

Note that

\[
\mathbb{E}[e^{j\omega X_i}] = p e^{j\omega} + (1-p)e^{-j\omega}
\]

Thus

\[
\mathbb{E}[Y_n] = (pe^{j\omega} + qe^{-j\omega})^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{j\omega(n-k)}
\]

\[
= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{j\omega(n-k)}
\]


Further, let \( m > n \),
\[
\Pr(Y_m = k | Y_n = l) = \Pr(k_m = k | Y_n = l) = \Pr(k_m = k-l | Y_n = l) = \Pr_m(k-l, Y_n = l)
\]

This approach can be further generalized to find the joint pdf of any finite subset from \( \{Y_n\}_{n \in \mathbb{N}} \).
Consider $Y_n = \lambda Y_{n-1} + Z_n$ where $0 < \lambda < 1$ and $Y_0 = 0$.

First of all, $X_0 = X_0$

$X_1 = \lambda X_0 + Z_1$

$X_2 = \lambda^2 X_0 + \lambda Z_1 + Z_2$

Thus, by induction, for any $\{X_0, X_1, \ldots, X_m\} \subseteq \mathbb{N}$, $\{Y_{i_0}, \ldots, Y_{i_m}\}$ are linear combinations of a set of jointly Gaussian r.v.'s.

Hence, $\{Y_{n}\}_{n=0}^{\infty}$ is a Gaussian random seq.

Obviously, $\mu_Y(n) = \lambda \mu_Y(n-1) + \mu$ for $n \geq 1$ and $\mu_Y(0) = \mu$.

Thus, $\mu_Y(n) = \mu \sum_{i=0}^{n} \lambda^i = \frac{\mu(1 - \lambda^{n+1})}{1 - \lambda}$.

$K_Y(m,n) = \text{cov}(Y_m, Y_n) = \sum_{k=0}^{m} \sum_{l=0}^{n} \lambda^{m+k} \text{cov}(Z_k, Z_l)$

$= \sum_{k=0}^{m} \lambda^{m+k} \text{var}(Z_k)$

$= \sigma^2 \sum_{k=0}^{m} \lambda^{2k}$

$= \sigma^2 \frac{\lambda^{2(m+1)} - \lambda^2}{\lambda^2 - 1}$

$\sigma_Y(n) = \sigma \sqrt{\sum_{k=0}^{n} \lambda^{2k}} = \sigma \sqrt{\frac{1 - \lambda^{2(n+1)}}{1 - \lambda^2}}$

Here, $f_Y(y) = \frac{1}{\sqrt{2\pi} \sigma_Y(n)} \exp \left\{ \frac{-(y - \mu_Y(n))^2}{2\sigma_Y(n)^2} \right\}$

How about $f_{Y_1, \ldots, Y_m}(y_1, \ldots, y_m)$?

How about general $f_{Y_1, \ldots, Y_m}$?
A random seq. \( \{x_n\}_{n \in \mathbb{Z}} \) is called stationary if for any \( \{i_1, i_2, \ldots, i_m\} \subset \mathbb{Z} \) and \( k \in \mathbb{Z} \) it holds:

\[ F_{x_{i_1}, x_{i_2}, \ldots, x_{i_m}}(x_{i_1}, x_{i_2}, \ldots, x_{i_m}) = F_{x_{i_1+k}, x_{i_2+k}, \ldots, x_{i_m+k}}(x_{i_1}, x_{i_2}, \ldots, x_{i_m}). \]

i.e., all statistical properties of \( \{x_n\}_{n \in \mathbb{Z}} \) is shift (time) invariant.

Wide-sense Stationary Seq.

A random seq. \( \{x_n\}_{n \in \mathbb{Z}} \) is called wide-sense stationary (WSS) if:

1. \( \mu_x(n) = \mu \) for all \( n \in \mathbb{Z} \)
2. \( R_x(m,n) = R_x(m+k, n+k) \) for all \( k \in \mathbb{Z} \)

(i.e., \( R_x(m,n) \) depends only on the difference \( m-n \).

(Usually we write \( R_x(m,n) = R_x(m-n) \).

i.e., the 1st and 2nd order statistics of \( \{x_n\}_{n \in \mathbb{Z}} \) is shift invariant.

Do we have \( R_x(m,n) = R_x(m+k, n+k) \)? why?

Which of the previously considered random seqs are stationary or WSS?

**Fact:**

(i) Stationary \( \Rightarrow \) WSS  
(But WSS \( \not\Rightarrow \) Stationary)

(ii) For Gaussian random seqs., Stationary \( \Rightarrow \) WSS (Why?)
Cyclostationary random seq.

A random seq \( \{ x_n \}_{n \in \mathbb{Z}} \) is called cyclostationary if

(i) \( p_x(n) = p_x(n+k) \)

(ii) \( R_x(m, n) = R_x(m+k, n+k) \) (similar for \( K_x(m, n) \))

for all \( m, n \in \mathbb{Z} \).

Example: Let \( \{ x_n \}_{n \in \mathbb{Z}} \) be an i.i.d. Bernoulli seq with \( p_x(0) = p \) and \( p_x(1) = 1-p = q \).

Consider \( Y_n = \sin \frac{2\pi n x_n}{N} \) and the random seq. \( \{ Y_n \} \).

\[
\mu_Y(n) = E \left[ \sin \frac{2\pi n x_n}{N} \right] = q \sin \frac{2\pi n \cdot 0}{N} + p \sin \frac{2\pi n \cdot 1}{N} = p \sin \frac{2\pi n}{N}.
\]

\[
K_Y(m, n) = E \left[ (\sin \frac{2\pi m x_n}{N} - p \sin \frac{2\pi m \cdot 1}{N})(\sin \frac{2\pi n x_n}{N} - p \sin \frac{2\pi n \cdot 1}{N}) \right]
\]

\[
= \begin{cases} 
   pq \sin \frac{2\pi m}{N} & \text{if } m=n \\
   0 & \text{if } m \neq n
\end{cases}
\]

Thus \( \{ Y_n \} \) is cyclostationary with period \( N \).

Now consider \( Z_n = Y_n - T \) where \( T \) is a uniform random variable taking values from \( 0, 1, \ldots, N-1 \) with equal probs. and is
adapt. of \( \{ x_n \} \). Consider the random seq \( \{ Z_n \} \).

\[
\mu_Z(n) = E[Z_n] = E[E[Z_n|T]] = E[p \sin \frac{2\pi (n-T)}{N}]
\]

\[
= \frac{p}{N} \sum_{m=0}^{N-1} \sin \frac{2\pi m}{N} = 0
\]

\[
K_Z(m, n) = E[(Z_m - Z_n)(Z_m - Z_n)|T]] = \begin{cases} 
   pq \sin \frac{2\pi (m-T)}{N} & \text{if } m=n \\
   0 & \text{if } m \neq n
\end{cases}
\]

Thus \( \{ Z_n \} \) is WSS! Is \( \{ Z_n \} \) stationary?
Power Spectral Density (PSD)

Let \( \{Z_n\}_{n \in \mathbb{Z}} \) be a WSS random seq.
Recall that the auto-correlation \( R_z(m,n) = \mathbb{E}[Z_m Z_n] \) depends only on a lag between the two arguments. Thus it can be written as \( R_z(n) \).

Taking (discrete-time) Fourier Transform on \( R_z(n) \), we have
\[
S_z(w) = \sum_{n=-\infty}^{\infty} R_z(n) e^{-jwn}
\]
for \(-\pi \leq w \leq \pi\).

\( S_z(w) \) is called the **power spectral density** of \( \{Z_n\}_n \).

Obviously, we can obtain \( R_z(n) \) back from \( S_z(w) \) by taking inverse Fourier transform,
\[
R_z(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_z(w) e^{jwn} \, dw.
\]

In particular,
\[
E[Z_n^2] = R_z(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_z(w) \, dw.
\]

Since \( E[Z_n^2] \) can be interpreted as the average power of the random seq \( \{Z_n\}_n \), \( S_z(w) \) has the meaning of the density of power of \( \{Z_n\}_n \) at the freq. \( w \).

**Properties of \( S_z(w) \):**

(i) \( S_z(w) \geq 0 \) for all \( w \) (since \( R_z(m,n) \) is positive semi-definite.

(ii) \( S_z(-w) = S_z(w) \)
Remark: A similar definition can be made for a cyclostationary random seq. with period \( k \).

To do so, first look at \( \overline{R}_x(m,n) = \frac{1}{k} \sum_{i=1}^{k} R((m+i)n) \). A careful inspection of \( \overline{R}_x(m,n) \) reveals that it depends only on \( m-n \). Thus we can write it as \( \overline{R}_x(n) \).

Then the PSD of a cyclostationary random seq. can be defined as \( \overline{S}_x(w) = \sum_{m=-\infty}^{\infty} \overline{R}_x(n) e^{-jwn} \).

Note that this definition has the exact same meaning as defining \( \overline{X}_n = X_{n-N} \) where \( N \) is a r.v. taking values from \( \{0,1, \ldots, k-1\} \) with equal probs and independent of \( \{X_n\} \) and considering the PSD of \( \overline{X}_n \).

Linear Time Invariant Systems and Random Seqs.

Previously we discuss that complicated random seqs. can be generated from simple random seqs. One such way is to pass a random seq. into a linear time invariant system. Recall that any LTI system can be represented by an impulse response \( h_n \). Thus we generate the new random seq. \( \{Y_n\} \) (if it is well-defined) by

\[
Y_n = \sum_{k=-\infty}^{\infty} h_k \overline{X}(n-k) = h_n * \overline{X}_n
\]

or write as \( \{Y_n\} = L \{\overline{X}_n\} \).
Another way to interpret this process is to view \( \{Y_n\} \) as the output of \( L \) when \( \{Z_n\} \) is input to the system:

Recall that for each \( \omega \in \Omega \), we can interpret \( \{Z_n(\omega)\} \) as a discrete-time signal. Then \( \{Y_n(\omega)\} \) is simply the output of \( L \) with \( \{Z_n(\omega)\} \) as input to the system. In this way, we obtain the interpretation of putting a random signal into a LTI system.

Our goal is to describe the statistical properties of \( \{Y_n\} \). This is generally difficult, so we focus mainly on the 1st and 2nd order statistics.

To do so, we need to introduce the idea of cross-covariance (cross-covariance) function of \( \{Z_n\} \) and \( \{Y_n\} \),

\[
R_{YZ}(m,n) \equiv E[Z_m Y_n]
\]

\[
(\text{K}2_{xy}(m,n) \equiv \text{cov}(Z_m, Y_n))
\]

When \( \{Y_n\} = L(\{Z_n\}) \),

\[
R_{YZ}(m,n) \equiv E[Z_m Y_n] = E[Z_m \cdot \sum_{k=-\infty}^{\infty} R_Z Z_{n-k}]
\]

\[
= E\left[\sum_{k=-\infty}^{\infty} R_Z Z_m Z_{n-k}\right]
\]

\[
= \sum_{k=-\infty}^{\infty} R_Z R_Z(m,n-k)
\]

* A sufficient condition for this is \( \sum_{k=-\infty}^{\infty} |R_Z(k)| < \infty \) and \( |R_Z(m,n)| < \infty \) for all \( m,n \).
- Fixing $m$ and treating $R_{X \cdot R}(m,n)$ and $R_{X}(m,n)$ as does indexed by $n$, we have

$$R_{X \cdot R}(m,n) = R_{X}(m,n) \times R_{R}.$$ 

or $$\{R_{X \cdot R}(m,n)\} = L_{m}\{R_{X}(m,n)\}.$$ 

- Further $R_{Y}(m,n) = E[X_{m} \cdot Y_{n}] = E[\sum_{k=-\infty}^{\infty} R_{X \cdot R}(m-k,n) \times X_{m-k} \cdot Y_{n}]$ 

$$= \sum_{k=-\infty}^{\infty} R_{X \cdot R}(m-k,n) \times X_{m-k} \cdot Y_{n}$$ 

or $$\{R_{Y}(m,n)\} = L_{m}\{R_{X \cdot R}(m,n)\} = L_{m}\{L_{n}\{R_{X \cdot R}(m,n)\}\}.$$ 

$$R_{Y}(m,n) = R_{X \cdot R}(m,n) \times R_{R} = (R_{X}(m,n) \times R_{R}) \times R_{R}.$$ 

- $R_{y} \mu_{r}(n) = E[Y_{n}] = E[\sum_{k=-\infty}^{\infty} R_{X \cdot R}(m-k) \times X_{m-k}] = \sum_{k=-\infty}^{\infty} R_{X \cdot R}(m-k) \times \mu_{X}(m-k)$ 

$$\mu_{y}(n) = R_{R} \mu_{r}(n).$$ 

Remark: The development above can be extended to the case of linear time-varying system, whose output can be described by the time-varying impulse response $h(n;k)$. All the results above are still valid if we replace $R_{X \cdot R}$ by $h(n;k)$ in the corresponding expressions.
For \( \{Z_n\} \),

(i) \( \mu_X(n) = \mu_X \sum_{k=\infty}^{\infty} \hat{R}_k = \mu_X \sum_{k=\infty}^{\infty} R_k = \text{constant} \).

(ii) \( -R_{XY}(m,n) = \sum_{k=\infty}^{\infty} \hat{R}_k R_Z(m-n-k) = \sum_{k=\infty}^{\infty} R_k \tilde{R}_Z(m-n-k) \)

where \( \hat{R}_k \) is the mirror image of \( R_k \) about \( k=0 \).

\[ \beta_{Z}\tilde{Z}(m-n) = \]

Here \( R_{XY}(m,n) = R_Z \ast \tilde{R}_Z(m-n) \), which depends only on \((m-n)\)!

Max so, \( R_Y(m,n) = \sum_{k=\infty}^{\infty} \hat{R}_k R_Z(m-k,n) \)

\[ = \sum_{k=\infty}^{\infty} R_k R_{ZX}(m-n-k) \]

\[ = R_{ZX} \ast \tilde{R}_Z \ast R(m-n) \]

Here \( R_Y(m,n) = R_Y(m-n) = R_{ZX} \ast \tilde{R}_Z \ast R(m-n) \) depends only on \((m-n)\)!

**Summarizing (with better notation):**

If \( \{X_n\} \) is WSS and \( \{Y_n\} \) is the output of \( \{X_n\} \) through a LTI filter with impulse response \( R_X \), then

\( \{X_n\} \) is WSS where \( \mu_X = \mu_\tilde{X} \sum_{k=\infty}^{\infty} \hat{R}_k \)

\( R_Y(n) = R_Z \ast \tilde{R}_Z \ast R(n) \).

Also \( \{X_n\}, \{Y_n\} \) are jointly WSS and \( R_{XY}(n) = R_Z \ast \tilde{R}(n) \).

\( \ast \{X_n\}, \{Y_n\} \) are jointly WSS if

(i) \( \{Z_n\} \) and \( \{Y_n\} \) are both WSS

(ii) \( R_{XY}(m,n) \) depends only on \((m-n)\)
The same result for WSS input to LTI filter can be expressed in terms of PSDs. in the freq domain.

* If \( \{X_n\} \) is WSS and \( \{Y_n\} \) is the output of \( \{X_n\} \) through LTI \( h_k \), then \( \{Y_n\} \) is WSS. with \( H(w) = \sum_{k=-\infty}^{\infty} h_k e^{-jwk} \)

\[
S_{xy}(w) = \mathcal{F}\{R_{xy}(n)\} = S_z(w) \cdot H^*(w),
\]

\[
S_y(w) = S_z(w) \cdot H^*(w) \cdot H(w) = S_z(w) \cdot |H(w)|^2,
\]

where \( H(w) \) is the transfer function of LTI filters.

**Example:**

\[
Y_n = \sum_{k=1}^{M} a_k X_{n-k} + \sum_{k=0}^{N} b_k Z_{n-k}
\]

where \( \{Z_n\} \) is a zero-mean Gaussian rand. process

with PSD \( S_z(w) \). This type of rand. seq. is called an ARMA process.

First find the LTI \( h_k \) s.t. \( \{X_n\} \rightarrow [h_k] \rightarrow \{Y_n\} \)

Then use result above to find \( R_Y(n) \) and \( \mu_Y \), since \( \{Y_n\} \) is WSS as \( \{Z_n\} \) is WSS. Also \( \{X_n\} \) is obviously Gaussian.

It is easier to work in the freq domain for this problem.

Consider \( Y_n = \sum_{k=1}^{M} a_k Y_{n-k} + \sum_{k=0}^{N} b_k X_{n-k} \)

Taking Fourier transform,

\[
Y(w) = \sum_{k=1}^{M} a_k e^{-jwk} Y(w) + \sum_{k=0}^{N} b_k e^{-jwk} X(w)
\]

Thus \( H(w) = \frac{Y(w)}{X(w)} = \frac{\sum_{k=0}^{N} b_k e^{-jwk}}{1 - \sum_{k=1}^{M} a_k e^{-jwk}} \)

From above \( S_Y(w) = S_X(w) \cdot |H(w)|^2 \). To get \( R_Y(n) \) take inverse FT.
Convergence of Random Sequences

So far most of the interesting discussions we have had about random sequences involve some form of stationarity. "Stationarity" is a somewhat restrictive requirement. There are many interesting random sequences which are not stationary (WSS) in some transient period but "eventually" behave pretty much like stationary random sequences in the "steady state".

We would next to formalize the vague notion of "eventually" & "steady state" by considering limits of random sequences.

Recall that a random variable is a function that maps the sample space to the real line. Thus the notion of limits of a seq. of random variables is not as simple as the limit of a seq. of real numbers.

In fact, there are many ways we can talk about the "limit" of a random seq. The common ways are described below.
Let \( \{ Z_1, Z_2, \ldots \} \) be a random seq and \( X \) be a r.v. defined on the same prob space \((\Omega, \mathcal{F}, P)\).

(i) **Surely (pointwise) convergence**

\( \{ Z_n \} \) is said to converge surely to \( X \) if for every \( w \in \Omega \), \( \lim_{n \to \infty} Z_n(w) = X(w) \).

This is usually too restrictive and we are not interested in this type of convergence.

(ii) **Almost surely convergence**

Consider the event \( A = \{ w \in \Omega : \lim_{n \to \infty} Z_n(w) = X(w) \} \) written as \( \{ \lim_{n \to \infty} Z_n = X \} \).

(Why this is an event, i.e., why \( A \in \mathcal{F} \)?)

\( \{ Z_n \} \) is said to converge almost surely (a.s.) or with prob. 1 to \( X \) if \( P(A) = 1 \).

Notation: will write \( \lim_{n \to \infty} Z_n = X \) a.s. or \( Z_n \overset{a.s.}{\to} X \).

Obviously, a.s. convergence is weaker than (less restrictive than) surely convergence. But it is still quite strong since only on a set of zero prob. \( \{ Z_n \} \) doesn't converge to \( X \). Most of the time this zero prob set is of no significance!
(iii) Mean square convergence

Suppose that $\mathbb{E}[|X|^2] < \infty$ for all $n \geq 1$.

$\{X_n\}$ is said to converge in mean square to $X$ if
\[
\lim_{n \to \infty} \mathbb{E}[(X_n - X)^2] = 0.
\]

Notation: $\lim_{n \to \infty} X_n = X$ m.s. or $X_n \xrightarrow{m.s.} X$

This is not the same as a.s. convergence, and may not be weaker than surely convergence (why?).

(iv) Convergence in Probability

$\{X_n\}$ is said to converge in prob. to $X$ if for any $\varepsilon > 0$,
\[
\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.
\]

Notation: $\lim_{n \to \infty} X_n = X$ p. or $X_n \xrightarrow{P} X$

Notice the subtle difference between conv. in prob. and a.s. convergence. It turns out that conv. in prob. is
less restrictive (weaker) than both a.s. and m.s. convergence.

(v) Convergence in Distribution

$\{X_n\}$ is said to converge in distribution to $X$ if
\[
\lim_{n \to \infty} F_{X_n}(x) = F_X(x) \quad \text{for every continuity pt } x \text{ of } F_X(x).
\]

or equivalently
\[
\lim_{n \to \infty} \mathbb{P}_{X_n}(w) = \mathbb{P}_X(w) \quad \text{for all } w.
\]

Notation: $\lim_{n \to \infty} X_n = X$ d. or $X_n \xrightarrow{d} X$

It turns out that conv. in distr. is the weakest among conv
Example: Consider a Cauchy r.v. \( X \) with pdf \( f_X(x) = \frac{1}{\pi(1+x^2)} \).

Consider the prob space induced by \( X : (\mathbb{R}, \mathcal{B}, P_X) \).

We will treat this as our underlying prob. space on \((\mathbb{R}, \mathcal{B}, P)\) set.

Now consider the r.v. \( Y_n \) \((n \geq 1)\) defined \( Y_n(w) = \frac{X}{n} \) for \( w \in \mathbb{R} \). One can write \( Y_n = \frac{X}{n} \).

It is easy to see that \( Y_n \) is Cauchy with pdf \( f_{Y_n}(y) = \frac{1}{\pi(\frac{1}{n}+y^2)} \).

- Obviously, for every \( w \in \mathbb{R} \), \( \lim_{n \to \infty} Y_n(w) = \lim_{n \to \infty} \frac{X}{n} = 0 \).

Then \( \{Y_n\} \) converges surely (and a.s.) to the r.v. 0.

It is also easy to see that for every \( \varepsilon > 0 \),

\[
\lim_{n \to \infty} P(|Y_n - 0| > \varepsilon) = \lim_{n \to \infty} \left( 1 - \int_{-\varepsilon}^{\varepsilon} \frac{1}{\pi(\frac{1}{n}+y^2)} \, dy \right) = 0.
\]

Thus \( Y_n \to 0 \).

- Not hard to see that \( \lim_{n \to \infty} F_{Y_n}(y) = \lim_{n \to \infty} \int_{-\infty}^{y} \frac{1}{\pi(\frac{1}{n}+u^2)} \, du = \begin{cases} 
0 & \text{if } y < 0, \\
1 & \text{if } y \geq 0.
\end{cases} \)

Thus \( Y_n \to 0 \).

But since \( E[Y_n^2] = \infty \) (or undefined), hence \( Y_n \) cannot converge in mean square!
Example (from Hajek). Consider $(\Omega, \mathcal{F}, \mathbb{P})$ let $\Omega = [0, 1]$ and $\mathcal{F} = \mathcal{B}[0, 1]$ and $\mathbb{P}$ be such that $\mathbb{P}([a, b]) = b - a$ for any $a < b \in [0, 1]$. This corresponds to the uniform distribution on $[0, 1]$.

Now consider $\{Z_n\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ as shown.

In general, write $n = 2^k + j$ where $k = \lfloor \log_2 n \rfloor$ and $0 \leq j < 2^k$.

Then $\mathbb{I}_{\lfloor n/2^k \rfloor}(w) = \begin{cases} 1, & w \in \left(\frac{j}{2^k}, \frac{j+1}{2^k}\right) \\ 1, & \text{otherwise} \end{cases}$

Length for $\mathbb{I}_{\lfloor n/2^k \rfloor}(w) = 1$ is $2^{-k}$.

Notice that for each $k > 1$, there is $2^k \leq n < 2^{k+1}$ s.t. $\mathbb{I}_{\lfloor n/2^k \rfloor}(w) = 0$ and $\mathbb{I}_{\lfloor n/2^k \rfloor}(w) = 1$.

Thus $\lim_{n \to \infty} \mathbb{I}_{\lfloor n/2^k \rfloor}(w)$ does not exist for any $w \in [0, 1]$.

Here $\{Z_n\}$ does not converge a.s. and does not converge weakly.

On the other hand, $\mathbb{E}[\|Z_n - 0\|^2] = 2^{-k} \to 0$ as $n \to \infty$. Thus $\mathbb{E}[Z_n] \to 0$. 
Also for any \( \lambda > 0 \), \( P(12z - 01 \geq \lambda) = 2^\lambda \).

Thus \( \lim_{n \to \infty} P(12z - 01 \geq \lambda) = 0 \), i.e. \( Z_n \to \mu^0 \).

Finally, \( F_n(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - 2^{-k} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \Rightarrow F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - 2^{-k} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \)

Thus \( Z_n \to \mu^0 \).

Example: Recall our previous Gaussian seq example:
Let \( Z_{n}\to D \) be iid Gaussian with mean zero and unit variance.
\( Y_n = \lambda Y_{n-1} + Z_n \) where \( |\lambda| < 1 \) and \( Z_1 = 0 \).

\[
P(1|Y_n - Y_{n-1}| > 2) \geq P(Y_n > 0, Z_n \leq -2) + P(Y_n < 0, Z_n > 2)
\]

\[
(Y_n + Z_n, n \geq 1) \Rightarrow P(Y_n > 0)P(Z_n \leq -2) + P(Y_n < 0)P(Z_n > 2)
\]

\[
(\lambda < 2) \Rightarrow P(Z_n \leq -2) > 0.02
\]

Now for any r.v. \( X \), \( P(1|Y_n - X| > 1) + P(1|Y_n + X| > 1) \geq P(1|Y_n - X| > 1) + P(1|Y_n + X| > 1) \)

\[
\geq P(1|Y_n - X| > 1) U (1|Y_n + X| > 1)
\]

\[
\geq P(1|Y_n - Y_{n-1}| > 2) \geq 0.02.
\]

Hence \( Y_n \) cannot converge to \( X \) in prob.

But from before, we know that \( Y_n \sim N(0, \sigma^2_{Y_n}) \)
where \( \sigma^2_{Y_n} = \frac{\lambda^2}{1 - \lambda^2} \xrightarrow{n \to \infty} \frac{1}{1 - \lambda^2} \).

Thus \( Y_n \to N(0, \frac{1}{1 - \lambda^2}) \).
From the examples above, we observe certain relations between different types of convergence. The following list of properties state the main ones:

(i) If \( X_n \overset{a.s.}{\to} X \), then \( X_n \overset{P}{\to} X \).

(ii) If \( X_n \overset{w.s.}{\to} X \), then \( X_n \overset{P}{\to} X \).

(iii) If \( P[|X_n| < L] = 1 \) for all \( n \) for some \( L < \infty \), and

\[ X_n \overset{P}{\to} X, \] then \( X_n \overset{w.s.}{\to} X \)

(iv) If \( X_n \overset{P}{\to} X \), then \( X_n \overset{a.s.}{\to} X \).

Proof: (ii) Suppose \( X_n \overset{w.s.}{\to} X \) and let \( \varepsilon > 0 \). From Chebyshev,

\[ P(\|X - X_n\| > \varepsilon) \leq \frac{E[\|X - X_n\|^2]}{\varepsilon^2} \to 0 \quad \text{as} \quad n \to \infty. \]

Thus \( X_n \overset{P}{\to} X \).

(iii) Suppose \( X_n \overset{a.s.}{\to} X \) and let \( \varepsilon > 0 \). Define \( A_n = \{ l > 0 : \|X_n - X_l\| < \varepsilon \} \) and \( B_n = \{ l > 0 : \|X_n - X_l\| < \varepsilon \} \) for all \( l > n \).

Obviously, \( B_n \subseteq A_n \) and \( B_1 \subseteq B_2 \subseteq \cdots \). Thus by continuity of probability measure, \( \lim_{n \to \infty} P(B_n) = P(\cap_{n=1}^{\infty} B_n) \).

But \( \bigcap_{n=1}^{\infty} \overline{B_n} = \{ l > 0 : \lim_{n \to \infty} \|X_n - X_l\| = \varepsilon \} \implies P(\bigcup_{n=1}^{\infty} B_n) = \lim_{n \to \infty} P(B_n) = 1 \).

Now \( P(B_n) = P(A_n) \leq 1 \). Taking limit, we have \( \lim_{n \to \infty} P(A_n) = 1 \).

Thus \( X_n \overset{P}{\to} X \).

(iv) Suppose that \( X_n \overset{a.s.}{\to} X \) and \( P(|X_n| \leq L) = 1 \). For any \( \varepsilon > 0 \),

\[ P(|X| \geq L + \varepsilon) \leq P(|X - X_n| \geq \varepsilon) \to 0 \quad \Rightarrow \quad P(|X| \leq L) = 1. \]

\[ \Rightarrow \quad P(\|X - X_n\|^2 \leq 4L^2) = 1. \]

Finally, \( E[\|X - X_n\|^2] \leq 4L^2 \frac{P(\|X - X_n\|^2 \geq \varepsilon)}{\varepsilon} + \varepsilon^2 \frac{P(\|X - X_n\|^2 \leq \varepsilon)}{\varepsilon} \to 0 \).
(iv) Suppose that \( X_n \to X \). For any cdf \( F \) of \( F \), we want to show that \( \lim_{n \to \infty} F_{X_n}(x) = F(x) \).

Let \( \varepsilon > 0 \). Since \( F \) is cdf at \( x \), then there exists \( S > 0 \), s.t. \( F_x(x - S) = F_x(x) - \frac{\varepsilon}{2} \).

Since \( \{X_n \leq x\} \cup \{X_n - X \geq S\} = \{X \leq x - S\} \),

\[
F_{X_n}(x) + P(\{X_n - X \geq S\}) \geq F_x(x - S) > F_x(x) - \frac{\varepsilon}{2}
\]

\[
\leq \frac{\varepsilon}{2}
\]

for large \( n \).

\[
F_{X_n}(x) \geq F_x(x) - \varepsilon.
\]

for suff. large \( n \).

A similar argument can be used to show \( F_{X_n}(x) \leq F_x(x) + \varepsilon \) for suff. large \( n \).

Thus \( |F_{X_n}(x) - F(x)| < \varepsilon \) for suff. large \( n \).

**Laws of Large Numbers.**

- Sample average converges to mean!

**Theorem:** Let \( \{X_n\}_{n \geq 1} \) be a seq of r.v.s with same finite mean \( \mu \). Let \( S_n = \sum_{i=1}^{n} X_i \). Then

- (i) \( \frac{S_n}{n} \xrightarrow{m.s.} \mu \) if \( \text{Var}(X_i) \leq C \) for some \( C \) and all \( i \geq 1 \), and \( \text{Cov}(X_i, X_j) = 0 \) for \( i \neq j \).

- (ii) \( \frac{S_n}{n} \xrightarrow{p} \mu \) if \( \{X_i\} \) are iid. (Weak law)

- (iii) \( \frac{S_n}{n} \xrightarrow{a.s.} \mu \) if \( \{X_i\} \) are iid. (Strong law).

**Proof:** (i) \( \mathbb{E}[(\frac{S_n}{n} - \mu)^2] = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i, X_j) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) \leq \frac{C}{n} \to 0 \)

Thus \( \frac{S_n}{n} \xrightarrow{m.s.} \mu \).
Proof. (ii) First consider the char. of $\frac{Z_n}{n}$:

$$E[e^{jw\frac{Z_n}{n}}] = \overline{F}_{Z_n}(w),$$

where $\overline{F}_Z(w)$ is char. of $Z_t$.

Thus then $E[e^{jw\frac{Z_n}{n}}] = [\overline{F}_Z(\frac{w}{n})]^n$.

Since $E[Z_t] = \mu$, then $\overline{F}_Z(w)$ is differentiable with $\overline{F}_Z(0) = 1$ and $\overline{F}_Z'(0) = j\mu$ and $\overline{F}_Z$ is continuous. By Taylor's theorem, for any $w$

$$\overline{F}_Z(\frac{w}{n}) \approx 1 + \frac{w}{n} \overline{F}_Z'(\frac{w}{n})$$

where $w \in (0, \frac{w}{n})$ for all $n$.

Thus $\overline{F}_Z(\frac{w}{n}) \to j\mu$ as $n \to \infty$. \Rightarrow $E[e^{jw\frac{Z_n}{n}}] = [1 + \frac{w}{n} \overline{F}_Z'(\frac{w}{n})]^n \to e^{j\mu}$

Hence $\frac{S_n}{n} \Rightarrow \mu$. Now fix $\varepsilon > 0$.

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = \frac{F_{S_n}(\mu - \varepsilon) + (1 - F_{S_n}(\mu + \varepsilon))}{\varepsilon} \to 0$$

as $n \to \infty$.

Hence $\frac{S_n}{n} \Rightarrow \mu$.

Central Limit Theorem

- Correctly scaled: sum of iid r.v.s. converges to Gaussian.

Then: Let $\{X_n\}$ be iid r.v.s. with mean $\mu$ and variance $\sigma^2$. Let $S_n = \sum_{i=1}^{n} X_i$, then $\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2)$.

Proof. Let $Y_n = \frac{S_n - n\mu}{\sqrt{n}}$, then $\overline{F}_{Y_n}(w) = \overline{F}_{Z_n}(\frac{w}{n}) e^{j\mu} \overline{F}_{\frac{w}{n}}$, where $\overline{F}_Z(w)$ is the char. of $Z_i$.

But mean and variance of $Z_i$ exist implies $\overline{F}_Z(w)$ is twice-diff. with $\overline{F}_Z(0) = 1$, $\overline{F}'_Z(0) = j\mu$ and $\overline{F}''_Z(0) = -\mu^2 - \sigma^2$, $\overline{F}$ is continuous.

By Taylor's theorem, $\overline{F}_Z(\frac{w}{n}) e^{j\mu} \overline{F}_{\frac{w}{n}} \approx 1 + \frac{w^2}{2n} \overline{F}''_Z(\frac{w}{n})$ for some $w \in (0, \frac{w}{n})$.

Thus $\overline{F}_{Y_n}(w) = (1 + \frac{w^2}{2n} \overline{F}''_Z(\frac{w}{n}))^n \to e^{j\mu w^2/2} \Rightarrow Y_n \xrightarrow{d} N(\mu, \sigma^2)$.
Markov Random Sequences

- Describe scenarios in which "the current depends on the past only through the immediate past".

- For simplicity, we will separate our discussion into 2 cases: In general, a unified treatment is possible but beyond us here.

"Continuous-valued" Markov rand. seqs.

- Let \( \{Z_n\}_{n=0}^{\infty} \) be a seq. of jointly continuous rv's defined on the same prob. space. It is called a Markov rand. seq. if

\[ f_{Z_{n+1}, \ldots, Z_0} (X_{n+1}, \ldots, X_0) = f_{Z_{n+1}} (X_{n+1} | X_n, \ldots, X_0) \]

for all \( n \geq 1 \).

- Immediately we have

\[ f_{Z_0, Z_1, \ldots, Z_n} (x_0, x_1, \ldots, x_n) = f_{Z_0} (x_0) \prod_{\ell=1}^{n} f_{Z_{\ell+1}} (x_{\ell+1} | x_{\ell}) \]

- When \( f_{Z_{\ell+1}} (x_{\ell+1} | x_{\ell}) \) is shift-invariant, i.e.

\[ f_{Z_{\ell+1}} (x_{\ell+1} | x_{\ell}) = f_{Z_\ell} (x_{\ell+1} | x_{\ell}) \]

for all \( \ell \geq 1 \) and \( \ell \geq 1 \), the Markov rand. seq. is called homogeneous.

- The mean and autocorrelation funs. of a Markov seq. can usually be expressed in recursive formulas, e.g.

\[ \mu_{Z_n} = E[Z_n] = E[E[Z_n | Z_{n-1}]] = E[ E[g(Z_{n-1})] ] \]

where \( g \) is some function.
Examples (i) iid random seqs are Markov (why?)

(ii) Indep. increment seqs are Markov:

Let \( \{X_n\}_{n=0}^{\infty} \) be an indep. increment seq.

\[
\begin{align*}
X_0 &= Z_0 \\
X_1 &= X_0 + Z_1 \\
X_2 &= X_1 + Z_2 \\
& \vdots \\
X_n &= X_{n-1} + Z_n
\end{align*}
\]

\[
\begin{align*}
I_0 &= Y_0 \\
I_1 &= Y_0 + Y_1 \\
I_2 &= Y_0 + Y_1 + Y_2 \\
& \vdots \\
I_n &= Y_0 + Y_1 + \cdots + Y_n
\end{align*}
\]

Note that \( Y_1, Y_2, \ldots, Y_n \) are indep. random variables.

Using the Jacobian method, we have

\[
f_{X_1, X_2, \ldots, X_n}(x_0, x_1, \ldots, x_n) = f_{I_0}(x_0)f_{I_1}(x_1-x_0) \cdots f_{I_n}(x_n-x_{n-1})
\]

It is then not hard to see (by induction) that

\[
f_{X_1|X_0}(x_1|x_0) = f_{Y_1}(x_1-x_0) \quad \text{for all } n \geq 1,
\]

for all \( n \geq 1 \).

Thus the rand. seq. \( \{X_n\}_{n=0}^{\infty} \) is Markov.

In particular, the random walk seq. is Markov.

(ii) Recall the first-order Gaussian random seq. \( Y_n = \lambda Y_{n-1} + Z_n \)

where \( |\lambda| < 1 \), \( \{Z_n\}_{n=0}^{\infty} \) is i.i.d. Gaussian random seq. with

mean \( \mu \) and variance \( \sigma^2 \), and \( Y_{-1} = 0 \).

\[
f_{Y_1, Y_2, \ldots, Y_0}(y_1, y_2, \ldots, y_0) = f_{Z_0}(y_0) f_{Z_1}(y_1-y_0) \cdots f_{Z_n}(y_n-y_{n-1})
\]

Thus \( \{Y_n\}_{n=0}^{\infty} \) is Markov.

\[
f_{Y_0, Y_1, \ldots, Y_n}(y_0, \ldots, y_n) = f_{Z_0}(y_0) \prod_{k=1}^{n} f_{Z_k}(y_k-y_{k-1}) = f_{Z_0}(y_0) \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_k - \mu)^2 \right\} \]

\[
\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_k - \mu)^2 + \frac{\lambda^2}{\sigma^2} (y_k - \lambda y_{k-1} - \mu)^2 \right\}
\]
\[
\phi(n) = E[\mathcal{Y}_n] = E[E[\mathcal{Y}_n | \mathcal{Y}_{n-1}]] = E[\lambda \mathcal{Y}_{n-1} + \mu] \\
= \lambda \phi(n-1) + \mu \\
\Rightarrow \phi(n) = \sum_{i=0}^{n} \lambda^i \mu = \frac{1 - \lambda^{n+1}}{1 - \lambda} \mu.
\]

\[K_p(m,n) = E[(\mathcal{Y}_m - \mu_p(m))(\mathcal{Y}_n - \mu_p(n))] \\
\text{For } m > n, \quad E[(\mathcal{Y}_m - \mu_p(m))(\mathcal{Y}_n - \mu_p(n))] = E[E[(\mathcal{Y}_m - \mu_p(m))(\mathcal{Y}_n - \mu_p(n)) | \mathcal{Y}_n]] \\
= E[E(\mathcal{Y}_m - \mu_p(m) | \mathcal{Y}_n)] (\text{by Markov}) \\
= E[\lambda \mathcal{Y}_{m-1} + \mu - \mu_p(m) | \mathcal{Y}_n] \\
= \lambda E[\mathcal{Y}_{m-1} - \mu_p(m-1) | \mathcal{Y}_n] \\
= \lambda^{m-n} (\mathcal{Y}_n - \mu_p(n)) \quad (\text{by induction}).
\]
Thus \[K_p(m,n) = \lambda^{m-n} E[(\mathcal{Y}_n - \mu_p(n))^2] \quad \text{for } m > n.
\]
Similarly, for \( m < n \), \[K_p(m,n) = \lambda^{n-m} E[(\mathcal{Y}_n - \mu_p(n))^2].
\]
Here we need only to calculate \( \text{var}(\mathcal{Y}_n) \):
\[\text{var}(\mathcal{Y}_n) = E[E[(\mathcal{Y}_n - \mu_p(n))^2 | \mathcal{Y}_{n-1}]].
\]
But \[E[(\mathcal{Y}_n - \mu_p(n))^2 | \mathcal{Y}_{n-1}] = E[(\lambda \mathcal{Y}_{n-1} + \mu - \mu_p(m-1))^2 | \mathcal{Y}_{n-1}] \\
= \lambda^2 (\mathcal{Y}_{n-1} - \mu_p(n-1))^2 + \sigma^2.
\]
Thus \[\text{var}(\mathcal{Y}_n) = E[\lambda^2 (\mathcal{Y}_{n-1} - \mu_p(n-1))^2 + \sigma^2]
\]
\[= \lambda^2 \text{var}(\mathcal{Y}_{n-1}) + \sigma^2 \\
= \sum_{k=0}^{n-1} \lambda^2 \sigma^2 \frac{1 - \lambda^{2(k+1)}}{1 - \lambda^2} (\text{by induction}) \\
= \sigma^2 \frac{1 - \lambda^{2(n+1)}}{1 - \lambda^2}. \]
\[ K(m,n) = \sum_{l=0}^{\min(m,n)} X^{m+n-2l} \frac{\sigma^2}{1 - X^{-2}} \]

Since \( X_n \) is Gaussian (because \( f_{x_{m+n}}(y_n \mid y_{m+n-1}) \) is jointly Gaussian and \( \mu(Y) \to \frac{Y}{1-X} \), \( \var(y) \to \frac{\sigma^2}{1-X} \)), \( Y_n \sim \mathcal{N}\left(\frac{\mu}{1-X}, \frac{\sigma^2}{1-X} \right) \).

### Markov Chains

- Let \( \{Z_n\}_{n=0}^{\infty} \) be a seq. of discrete r.v.'s defined on the same prob. space. It is called a Markov chain if
  \[ P(Z_n = z_n \mid Z_{n-1} = z_{n-1}, \ldots, Z_0 = z_0) = P(Z_n = z_n \mid Z_{n-1} = z_{n-1}), \]
  for all \( n \geq 1 \).

- Although it is not necessary, we will assume that the range of all \( Z_n \)'s are identical as \( \{1, 2, \ldots, M\} \) for simplicity of discussion. With this assumption, the Markov chain is usually referred to as a finite-state Markov chain. It is often employed to model the operation of a state. In this case, we can use a \((M \times M)\) matrix to represent the conditional pdf \( P(Z_n = z_n \mid \cdot) \),

\[
P(\cdot | z_{n-1}) = \begin{bmatrix}
P(1_{z_n = 1} | z_{n-1}) & P(1_{z_n = 2} | z_{n-1}) & \cdots & P(1_{z_n = M} | z_{n-1}) \\
P(2_{z_n = 1} | z_{n-1}) & P(2_{z_n = 2} | z_{n-1}) & \cdots & P(2_{z_n = M} | z_{n-1}) \\
\vdots & \vdots & \ddots & \vdots \\
P(M_{z_n = 1} | z_{n-1}) & P(M_{z_n = 2} | z_{n-1}) & \cdots & P(M_{z_n = M} | z_{n-1})
\end{bmatrix}
\]

This matrix is often called the state transition matrix.

- A homogeneous Markov chain is one that with shift-invariant state \( P(z_n = z \mid \cdot) \) or transition matrix, i.e., \( P^n(\cdot | z_{n-1}) = P(\cdot | z_{n-1}) \) for all \( n \geq 1 \) (or \( P^n = P \)).
Obviously as before the joint pmf

\[ P_{X_1, \ldots, X_n}(X_1, X_2, \ldots, X_n) = P_{X_0}(X_0) \prod_{k=1}^{n} P_{X_k | X_{k-1}}(X_k | X_{k-1}) \]

But most of the time, we are interested in the marginal pmf

\[ P_{X_n}(X_n) = \sum_{X_{n-1}=1}^{M} P_{X_n, X_{n-1}}(X_n, X_{n-1}) = \sum_{X_{n-1}=1}^{M} P_{X_n | X_{n-1}}(X_n | X_{n-1}) P_{X_{n-1}}(X_{n-1}) \]

Thus if we write the pmf into a row vector:

\[ \Pi_{n} = \begin{bmatrix} P_{X_1}(1) & P_{X_1}(2) & \cdots & P_{X_1}(M) \end{bmatrix} \]

which is usually called the state distribution at time \( n \).

then \[ \Pi_{n} = \Pi_{0} \prod_{k=1}^{n} P_{k}(X_k) \]

(by induction)

For homogeneous Markov chain, \[ \Pi_{n} = \Pi_{0} P^{n} \]

Example: Consider a laptop computer operating in one of the three states: Sleep(1), Inactive(2), Active(3), described by the state diagram below:

```
```

Let \( X_n \) be the r.v. that describes the state that the laptop is in at time instant \( n \). Then \( \{X_n\}_{n=0}^{\infty} \) is a homogeneous Markov chain with state transition matrix \( P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.5 & 0.2 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{bmatrix} \)
\[ T(n) = T(0) \cdot P^n \]

For instance, if at time 0, the laptop starts at the sleep state, i.e., \( T(0) = [1\ 0\ 0] \), then

\[ T(1) = [1\ 0\ 0] \cdot \begin{bmatrix} 0.9 & 0 & 0.1 \\ 0.5 & 0.2 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{bmatrix} = [0.9\ 0\ 0.1] \]

\[ T(2) = [1\ 0\ 0] \cdot \begin{bmatrix} 0.9 & 0 & 0.1 \\ 0.5 & 0.2 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}^2 = [0.82\ 0.04\ 0.14] \]

\[ T(3) = [1\ 0\ 0] \cdot \begin{bmatrix} 0.9 & 0 & 0.1 \\ 0.5 & 0.2 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}^3 = [0.772\ 0.064\ 0.164] \]

In general, one can perform eigen-decomposition on \( P \)

\[ P = VDV^{-1} \]

where \( V = \begin{bmatrix} 0.3774 & -0.3015 & 0.1702 \\ 0.3774 & 0.3015 & 0.7720 \\ 0.3774 & 0.9045 & -0.6317 \end{bmatrix} \]

\[ D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0 & 0 \end{bmatrix} \]

\[ T(n) = T(0) \cdot P^n = T(0) \cdot (VDV^{-1})^n = T(0) \cdot (VDV^{-1}) \cdot (VDV^{-1}) \cdot \ldots \cdot (VDV^{-1}) \]

\[ = T(0) \cdot VDV^{-1} \]

In this case, \( T(n) = [1\ 0\ 0] \cdot V \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.6^n & 0 \\ 0 & 0 & 0 \end{bmatrix} V^{-1} \)

\[ = \begin{bmatrix} 0.7 + 0.3333 \cdot 0.6^n & 0.1 - 0.1667 \cdot 0.6^n & 0.2 - 0.1667 \cdot 0.6^n \end{bmatrix} \]

Thus at time \( n \), the probabilities of the laptop in the sleep, inactive, and active states are \( 0.7 + 0.3333 \cdot 0.6^n \), \( 0.1 - 0.1667 \cdot 0.6^n \), and \( 0.2 - 0.1667 \cdot 0.6^n \), respectively.
Suppose the current load at each of the states is conditionally
\[ \text{unif distributed as below:} \]

sleep : \[ \text{In} \sim \text{unif } [0,0.02] A \]
active : \[ \text{In} \sim \text{unif } [0.05,0.15] A \]
怠 : \[ \text{In} \sim \text{unif } [0.1,0.3] A . \]

Then the current load at time \( n \),
\[ E[\text{In}] = E[E[\text{In} | \mathcal{F}_n]] = 0.01 \pi_1(n) + 0.1 \pi_2(n) + 0.2 \pi_3(n) \]
\[ = 0.057 - 0.0467 \times 0.6^n A \]

Assuming a resistive load of 1Ω, the mean power consumption
at time \( n \) is
\[ E[\text{In}^2] = E[E[\text{In}^2 | \mathcal{F}_n]] = 2 \times 10^{-4} \pi_1(n) + 0.01 \pi_2(n) + 0.04 \pi_3(n) \]
\[ = 0.09 - 0.0083 \times 0.6^n W \]

Interestingly, note that as \( n \to \infty \),
\[ \pi(n) \to \pi = [0.7 \ 0.1 \ 0.2], \quad E[\text{In}] \to 0.057 A \text{ and} \]
\[ E[\text{In}^2] \to 0.09 / W. \]

This is like the state machine (Markov chain) reaches a
steady state!

**Steady State Distribution of Markov Chain**

In general, not all Markov Chain reaches a “steady state”.
A common sufficient condition for a Markov Chain to reach
a “steady state” is that the Markov Chain is primitive, i.e.
the state transition matrix \( P \) satisfies the following property:
There exists a \( k \) s.t. \( P^k \) has only positive elements.
Intuitively, this means that we can go from any state to any
other state with positive prob. in a finite #s of steps.
It turns out that if the Markov chain is primitive, then $P$ admits the eigen-decomposition $P = V D V^{-1}$ with 1 as one of its eigenvalues and all other eigenvalues must have magnitude strictly smaller than 1. In addition, there is one and only one eigenvector (with positive elements) associated with the eigenvalue 1.

With these results, it is not hard to see that

$$P^n = V D^n V^{-1} \rightarrow V \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} V^{-1} \quad \text{as} \quad n \rightarrow \infty$$

Hence

$$\pi(n) = \pi(0) P^n = \pi(0) V D^n V^{-1} \rightarrow \pi(0) V \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} V^{-1}$$

Thus, the steady state distribution $\pi$ exists and is unique, given by

$$\pi = \pi(0) V \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} V^{-1} \quad \text{--- (x)}$$

This expression for $\pi$ is not very convenient. To obtain a more convenient way to find $\pi$, first let's notice that $P e = e$, where $e = [1 \cdots 1]^T$ the all-one column vector (Why?). Thus $e$ must be the unique eigenvector of $P$ associated with the eigenvalue 1. As a result, from (x)

$$\pi V = \pi(0) e = 0 \ 0 \cdots 0 = [1 \ 0 \cdots 0] \quad \text{--- (xx)}$$

or

$$\pi = [1 \ 0 \cdots 0] V^{-1} \quad \text{(doesn't depend on } \pi(0)!)$$

Right-multiplying both sides of (xx) by $DV^{-1}$, we have

$$\pi V D V^{-1} = [1 \ 0 \cdots 0] D V^{-1} = [1 \ 0 \cdots 0] V^{-1}$$

Thus

$$\pi P = \pi \quad \text{--- (xxx)}$$
The equation (**) is usually called the balance equation of the primitive Markov chain and the steady state distribution \( \pi \) is the left-eigenvector of \( \mathbf{P} \) associated with the eigenvalue 1.

Notice that the balance equation doesn't completely specify \( \pi \) (why?). To solve for \( \pi \) using the balance equation, we need to add the additional constraint that \( \pi e = 1 \) (Why this constraint makes sense?). Putting things together we have the linear equations:

\[
\pi \left[ \begin{array}{cccc}
-0.5 & 0.2 & -0.3 & 0.1 \\
0.1 & -0.4 & 0.5 & 0.1 \\
0.1 & -0.4 & 0.5 & 0.1 \\
-0.5 & 0.2 & -0.3 & 0.1 \\
0.1 & -0.4 & 0.5 & 0.1 \\
0.1 & -0.4 & 0.5 & 0.1 \\
\end{array} \right] = [0 \ 0 \ 1] 
\]

Example: Consider the previous laptop example. The Markov chain is primitive and hence we can find the steady state distribution (it exists) by solving the balance equation:

\[
\pi \left[ \begin{array}{cccc}
0.1 & -0.1 & 0.1 & 0.1 \\
0.1 & -0.1 & 0.1 & 0.1 \\
0.1 & -0.1 & 0.1 & 0.1 \\
0.1 & -0.1 & 0.1 & 0.1 \\
0.1 & -0.1 & 0.1 & 0.1 \\
0.1 & -0.1 & 0.1 & 0.1 \\
\end{array} \right] = [0 \ 0 \ 1] 
\]

Solving it gives, \( \pi = [0.7 \ 0.1 \ 0.1 \ 0.2] \) exactly as before. Also, it is the same regardless of the initial distribution \( \pi(0) \). (Go back to the previous development and try different \( \pi(0) \) to verify this fact.)

Finally, since a steady state distribution exists, we have \( \mathbf{X}_n \xrightarrow{d} \mathbf{X} \) where \( \mathbf{X} \approx \mathbb{N} \). Since \( \mathbf{X} \) and \( \mathbf{Z}_n \) have the same discrete finite range and \( \mathbf{X}_n \xrightarrow{d} \mathbb{N}, \mathbf{Z}_n \xrightarrow{d} \mathbb{N} \),

Thus \( \mathbf{X}_n \xrightarrow{d} \mathbb{N} \) and since \( \mathbf{X}_n \xrightarrow{d} \mathbb{N} \), \( \mathbf{X}_n \xrightarrow{d} \mathbb{N} \).