Stochastic Processes

Topic 6

Random Sequence

nctuee09f

Summary

`Reading`: Textbook sec. 6.1 $\sim$ sec. 6.7

In this topic, I will discuss:

- Random Sequence
- Stationarity
- Wide-sense Stationary (WSS) Random Sequence
- Linear Time Invariant (LTI) System
- WSS in LTI
- Power Spectral Density
- Markov Chain
- Convergence of Random Sequence

Notation We will use the following notation rules, unless otherwise noted, to represent symbols during this course.

- Boldface upper case letter to represent **MATRIX**
- Boldface lower case letter to represent **vector**
- Superscript $(\cdot)^T$ and $(\cdot)^H$ to denote transpose and hermitian (conjugate transpose), respectively
- Upper case italic letter to represent *RANDOM VARIABLE*

1 Random Sequence

(1) In plain words, we can view a ***random sequence*** as follows:

→ A mathematical formulation of a probabilistic experiment that evolves in ***time***

→ A random sequence can be considered as an evolution in ***time*** of ***random variables***

* The outcomes constitute a sequence of numerical values
* The outcomes are measured in countable time instants, e.g. the time instants in the set $\mathcal{T} = \{0, 1, 2, \cdots\}$ or $\mathcal{T} = \{\cdots, -1, 0, 1, 2, \cdots\}$.

(2) For example, a random sequence can be used to model

→ the sequence of daily prices of a stock

→ the sequence of hourly traffic loads at a node of a network

→ the sequence of radar measurement of the position of an airplane

→ the sequence of failure times of a machine

→ the sequence of received and periodically sampled signal in a communication link

(3) Something of particular interests:

→ We tend to focus on the ***dependencies*** in the sequence. For example, how do future prices of a stock depend on past values?

→ We are often interested in ***long-term averages***, involving the entire sequence of generated values. For example, what is the fraction of time on average that a machine is idle?

→ We sometimes wish to characterize the likelihood or frequency of certain *boundary events*.

For example:

* What is the probability that within a given hour all circuits of some telephone system become simultaneously busy?

* What is the frequency with which some buffer in a computer network overflows with data?

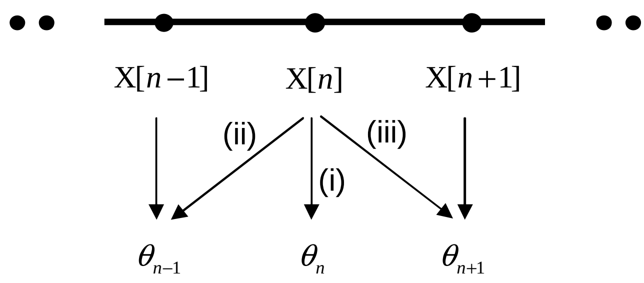

Figure 1: The (i) filtering, (ii) smoothing, and (iii) prediction operations for time-varying unknown parameters θ_n embedded in the random sequence $X[n]$ for all n.

(4) Things you may learn

- — Formulation of several probabilistic discrete time models
- — Filtering, smoothing, and prediction
- — Examine the behaviors, e.g. convergence, of the filtering, smoothing, and prediction operations shown in Fig. 2 as $n \to \infty$.

→ Filtering means that we estimate the parameter θ_n at the nth time based upon the observations up to time n

→ Smoothing means we go back to modify previously estimated parameter θ_i for $i < n$ when the nth observation becomes available

→ Prediction

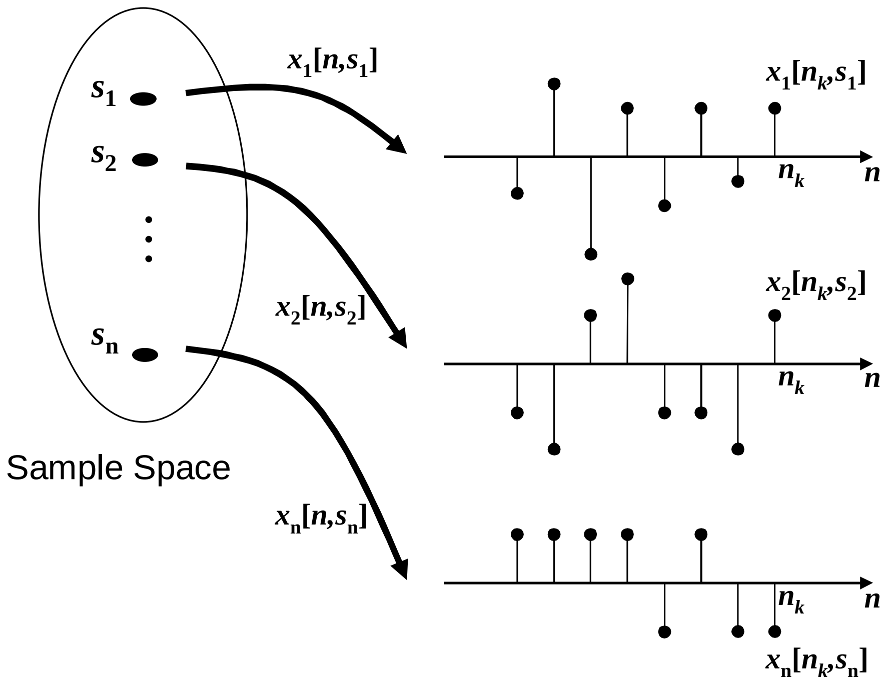

Figure 2: A mapping of a random sequence.

(4) **Definition (Random Sequence)**

Let $\varepsilon \in \Omega$ be an outcome of the sample space Ω. Let $X[n, \varepsilon]$ be a mapping of the sample space Ω into a space of complex-valued sequence on some index set $\mathbb{Z}$. If for each fixed integer $n \in \mathbb{Z}$, $X[n, \varepsilon]$ is a random variable, then $X[n, \varepsilon]$ is a random sequence (also known as discrete-time random process).

Remarks:

(a) For a fixed outcome ε^*, the ***sample sequence*** $X[n, \varepsilon^*]$ is a non-random (deterministic) function. That is, once we know what the outcome ε^* is, the ***sample sequence*** $X[n, \varepsilon^*]$ associated with that ε is also determined.

(b) The randomness falls in that we cannot exactly know what the outcome is at each time instant before the experiment is conducted.

(c) We often write $X[n, \varepsilon]$ as $X[n]$ for notational simplicity.

(d) Conceptually, ***random sequence*** can be considered as ***a sequence of random variables***, or more generally, a sequence of random vectors.

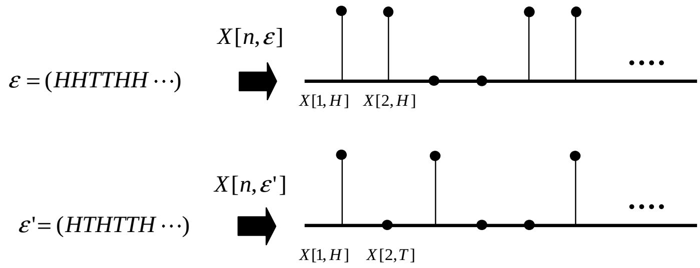

Figure 3: Two realizations, i.e. sample sequences, of the Bernoulli random sequence, one for $\varepsilon = (HHTTHH \cdots)$ and the other for $\varepsilon' = (HTHTTH \cdots)$.

Example: (Bernoulli Process)

Suppose X is a Bernoulli random variable modeling a success (E) or failure (E^c) of an event E with $X(\mathrm{E}) = 1$ and $X(\mathrm{E}^c) = 0$. For example, by flipping a coin, we can model $X(H) = 1$ and $X(T) = 0$ with H being the outcome of a head and T a tail.
The Bernoulli random sequence, or Bernoulli process, is defined as

$$X[n, \varepsilon] \triangleq X(\varepsilon_n),$$

where $\varepsilon_n \in \{H, T\}$ is the outcome of the nth flip and $\varepsilon = (\varepsilon_1 \varepsilon_2 \cdots \varepsilon_n \cdots)$ is an outcome, consisting of an infinite length sequence of events, in the sample space of the Bernoulli random sequence.

Two ***realizations***, also known as ***sample sequences***, of the Bernoulli random sequence are shown in Fig. 3, one for the event $\varepsilon = (HHTTHH \cdots)$ and the other for the event $\varepsilon' = (HTHTTH \cdots)$.

It should be noted that the sample space in this example consists of infinite outcomes, each with infinite length of events $(\varepsilon_1 \varepsilon_2 \cdots \varepsilon_n \cdots)$. ■

2 Statistical Description

(1) A random sequence is statistically specified by its Nth order probability distribution (or density) function

$$
\begin{aligned}
&F_X\Big(x_n, x_{n+1}, \cdots, x_{n+N-1}; n, n+1, \cdots, n+N-1\Big) \\
&\quad = P\Big[X[n] \le x_n, X[n+1] \le x_{n+1}, \cdots, X[n+N-1] \le x_{n+N-1}\Big]
\end{aligned}
$$

for all integers $N \ge 1$, and for all time instants $n, n+1, \cdots, n+N-1$.

(2) The ***mean*** function, ***autocorrelation*** function, and ***autocovariance*** function are defined as:

Mean function:

$$
\begin{aligned}
\mu_X[n] &\triangleq E\Big[X[n]\Big] \\
&= \int_{-\infty}^{\infty} x f_X(x;n)dx.
\end{aligned}
$$

Autocorrelation function: for all k and l

$$
\begin{aligned}
R_{XX}[k,l] &\triangleq E\Big[X[k]X^*[l]\Big] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_k x_l^* f_X(x_k, x_l; k, l)dx_k dx_l.
\end{aligned}
$$

Autocovariance function: for all k and l

$$
\begin{aligned}
K_{XX}[k,l] &\triangleq E\Big[\big(X[k] - \mu_X[k]\big)\big(X[l] - \mu_X[l]\big)^*\Big] \\
&= R_{XX}[k,l] - \mu_X[k]\mu_X^*[l].
\end{aligned}
$$

3 Independent Increments

(1) **Definition (Independent Increments)**

A random sequence is said to have ***independent increments*** if for all integers $n_1 < n_2 < \cdots < n_N$, the increments

$$X[n_1], X[n_2] - X[n_1], \cdots, X[n_N] - X[n_{N-1}]$$

are jointly independent for $N > 1$.

(2) The running sum

$$S[n] \triangleq \sum_{k=1}^{n} X[k]$$

of an independent random sequence $X[n]$ is also a random sequence, and has independent increments.

Example:
Let $X[n]$ be the Bernoulli random sequence.

$$S[n] \triangleq \sum_{k=1}^{n} X[k]$$

is the random sequence, commonly known as the ***Binomial counting process***, used to model the ***number of successes (occurrences)*** of a certain event up to time n. The Binomial counting process has independent increments.

Example:
Let $X[n]$ be the Bernoulli random sequence. Define $Y[n] \triangleq 2X[n] - 1$. Then,

$$W[n] \triangleq \sum_{k=1}^{n} Y[k]$$

is the ***random walk*** sequence, which can be used to model the amount of money a gambler wins up to the nth trial, where he earns one unit with a win and gives one unit away with a lose. The random walk sequence $W[n]$ also has independent increments.

(3) Let $S[n]$ be a random sequence having independent increments. Its Nth order joint pdf can be written as products of the pdf's of its increments.

Proof:
Let $X_1, X_2, \cdots, X_N$ be the increments of the sequence $S[1], S[2], \cdots, S[N]$. That is,

$$X_1 \triangleq S[1],\ X_2 \triangleq S[2] - S[1], \cdots, X_N \triangleq S[N] - S[N-1].$$

Define $\mathsf{x} \triangleq [X_1, X_2, \cdots, X_N]$. We can obtain the joint pdf of

$$\mathsf{s} \triangleq [S[1], S[2], \cdots, S[N]]$$

from the joint pdf of x using the concept of linear transformation, which gives

$$\begin{aligned} f_{\mathsf{s}}(s_1, s_2, \cdots, s_N) &= \frac{1}{|\mathbf{J}|} f_{\mathsf{x}}(x_1, x_2, \cdots, x_N)\Big|_{x_1=s_1,\cdots,x_N=s_N-s_{N-1}} \\ &= f_{X_1}(s_1) f_{X_2}(s_2 - s_1) \cdots f_{X_N}(s_N - s_{N-1}) \\ &= f_{S[1]}(s_1) f_{S[2]-S[1]}(s_2 - s_1) \cdots f_{S[N]-S[N-1]}(s_N - s_{N-1}). \end{aligned}$$

■

Example: (Waiting Time)
Consider the random sequence $\tau[n]$ consisting of i.i.d. exponential random variables for all n with

$$f_\tau(t; n) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

Then, the running sum

$$T[n] \triangleq \sum_{k=1}^{n} \tau[k]$$

is the ***waiting time random sequence***, which can be used to model the ***waiting time*** to the nth occurrence of a certain event, *e.g.*, the total amount of time that the nth packet in a queue has to wait until being processed.

(a) What is the pdf of $T[n]$?

(b) What are the mean function and variance function of $T[n]$?

(c) What is the autocorrelation function?

(d) What is the Nth order joint pdf? (Use the independent increments property)

4 Stationarity

(1) **Definition (Stationary Random Sequence)**

A random sequence $X[n]$ is ***stationary*** *if for all $N \geq 1$*

$$F_X\left[x_n, x_{n+1}, \cdots, x_{n+N-1}; n, n+1, \cdots, n+N-1\right]$$
$$= F_X\left[x_{n+k}, x_{n+1+k}, \cdots, x_{n+N-1+k}; n+k, n+1+k, \cdots, n+N-1+k\right]$$

for all integer shift k and for all x_n through x_{n+N-1}.

Example:
The Bernoulli random sequence is stationary. But, the waiting time random sequence is NOT.

(2) **Definition (Wide-Sense Stationary)**

A random sequence $X[n]$ for $n \in \mathbb{Z}$ is ***wide-sense stationary*** (WSS) if

(i) The mean function $\mu_X[n]$ is constant for all integers n,

$$\mu_X[n] = \mu_X[0], \quad \text{and}$$

(ii) The correlation function is independent of any integer shift n.

$$R_{XX}[k, l] = R_{XX}[k+n, l+n], \quad \forall k, l \in \mathbb{Z}.$$

Example

Let $X[n]$ be a sequence of zero mean uncorrelated random variables with unit variance. Then, $X[n]$ is WSS by checking $R_{XX}[k, l] = \delta[k - l] = R_{XX}[k+n, l+n]$. This random sequence is known as ***white process***.

(3) All stationary random sequences are wide-sense stationary.

Remarks:

(1) The correlation function $R_{XX}[k,l]$ of a WSS random sequence $X[n]$ can be expressed in terms of the time shift $k-l$ only, instead of specifying two time instants k and l.

$$R_{XX}[k,l] = R_{XX}[k-l,0] \triangleq R_{XX}[k-l].$$

In particular, we write

$$R_{XX}[m] \triangleq R_{XX}[l+m,l]$$

to specify the correlation between two random variables in the random sequence with m time units apart.

(2) Due to the shift-invariant property of WSS, the output random sequence of a linear time invariant (LTI) system to a WSS random sequence input is also WSS.
WSS$\rightarrow$LTI$\rightarrow$WSS.

(3) Properties of $R_{XX}[m]$ for WSS $X[n]$:

(a) $|R_{XX}[m]| \leq R_{XX}[0]$ for arbitrary m.

(b) $|R_{XY}[m]|^2 \leq R_{XX}[0]R_{YY}[0]$ for WSS $X[n]$ and $Y[n]$.

(c) The sequence $R_{XX}[m]$ is complex-conjugate symmetric, i.e.

$$R_{XX}[m] = R^*_{XX}[-m].$$

(d) (Positive semidefinite) For all $N \geq 1$ and all complex $a_1, \cdots, a_N$, we must have

$$\sum_{n=1}^{N}\sum_{k=1}^{N} a_n a_k^* R_{XX}[n-k] \geq 0.$$

Recall: (Page 254 in the text)

A square matrix $\mathbf{R}$ is positive semi-definite if for any vector $\mathbf{a}$

$$\mathbf{a}^H \mathbf{R} \mathbf{a} \geq 0.$$

5 Linear Time Invariant System

(1) Definition of a system
A system is a **mapping** that transforms an arbitrary **input** sequence into an **output** sequence.

(2) A system is ***linear*** if to a linear combination of inputs corresponds the same linear combination of outputs. That is, suppose we know

$$x_1[n] \rightarrow \boxed{\mathrm{L}\{\}} \rightarrow L\{x_1[n]\} \quad \text{and} \quad x_2[n] \rightarrow \boxed{\mathrm{L}\{\}} \rightarrow L\{x_2[n]\}.$$

Then, the linear system with operator $L\{\}$ guarantees

$$a_1x_1[n] + a_2x_2[n] \rightarrow \boxed{\mathrm{L}\{\}} \rightarrow a_1L\{x_1[n]\} + a_2L\{x_2[n]\}.$$

(3) The ***impulse response*** $h[n]$ of a linear system is the output sequence when the input is an impulse $\delta[n]$,

$$h[n] \triangleq L\{\delta[n]\}.$$

For any input sequence $x[n]$, we can write

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k].$$

If the system is linear, then the output sequence $y[n]$ is

$$\begin{aligned} y[n] &= L\{x[n]\} \\ &= L\left\{\sum_{k=-\infty}^{\infty} x[k]\delta[n-k]\right\} = \sum_{k=-\infty}^{\infty} x[k]L\{\delta[n-k]\} \\ &= \sum_{k=-\infty}^{\infty} x[k]h[n,k], \end{aligned}$$

where we define $h[n,k] \triangleq L\{\delta[n-k]\}$ as the output response at time n to an impulse applied at time k.

(4) A system is called ***time-invariant*** if shifting the input by k time units results in shifting the output by k time unit. That is, suppose we know

$$x[n] \rightarrow \boxed{\quad \mathrm{L}\{\} \quad} \rightarrow y[n]$$

with $L\{\cdot\}$ being a time-invariant operator. Then, the time-invariant property tells us

$$x[n-k] \rightarrow \boxed{\quad \mathrm{L}\{\} \quad} \rightarrow y[n-k].$$

So, for a linear time-invariant (LTI) system with impulse response $h[n]$, the output sequence is

$$\begin{aligned} y[n] &= \sum_{k=-\infty}^{\infty} x[k] L\{\delta[n-k]\} \\ &= \sum_{k=-\infty}^{\infty} x[k] h[n-k] \\ &= x[n] * h[n], \end{aligned}$$

which is the ***convolution*** between the input and the impulse response.

(5) (**Fourier Transform**)
The Fourier transform, or more precisely the discrete-time Fourier transform, for a sequence $x[n]$ is defined by

$$X(\omega) \triangleq \sum_{k=-\infty}^{\infty} x[n]e^{-j\omega n} \quad \text{for} \quad -\pi \leq \omega \leq \pi.$$

The inverse Fourier transform is

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\omega)e^{j\omega n}d\omega.$$

(6) The Fourier transform of $y[n] = x[n] * h[n]$ is

$$Y(\omega) = X(\omega)H(\omega).$$

Proof:

$$\begin{aligned}
Y(\omega) &= \sum_n y[n]e^{-j\omega n} = \sum_n \Big(x[n] * h[n]\Big)e^{-j\omega n} \\
&= \sum_n \sum_k x[k]h[n-k]e^{-j\omega n} \\
&= \sum_n \sum_k x[k]h[n-k]e^{-j\omega(n-k+k)} \\
&= \sum_k x[k]e^{-j\omega k}\left(\sum_n h[n-k]e^{-j\omega(n-k)}\right) = X(\omega)H(\omega).
\end{aligned}$$

■

6 Random Sequences in LTI Systems

(1) When the input $X[n]$ to a linear system is a random sequence, the output $Y[n] = L\{X[n]\}$ is also a random sequence in the sense that for each outcome $\varepsilon \in \Omega$, the deterministic sample $X[n, \varepsilon]$ is mapped to

$$Y[n, \varepsilon] = L\{X[n, \varepsilon]\}.$$

That is, overall, $Y[n, \varepsilon]$ is yet another ***mapping*** from each outcome ε of the sample space to a ***new*** sample sequence.

(2) We are usually interested in the behavior of random sequences in LTI systems.
Consider the following bounded LTI system with an impulse response $h[n]$. Suppose the input random sequence $X[n]$ is WSS.

$$X[n] \rightarrow \boxed{\quad h[n] \quad} \rightarrow Y[n]$$

Questions:

- $\rightarrow$ What is the correlation function $R_{YY}[k, l]$ of $Y[n]$?
- $\rightarrow$ Is $Y[n]$ WSS?

To answer these questions, we need to figure out the cross-correlation $R_{XY}[k, l]$ first.

$$\begin{aligned}
R_{XY}[k, l] &= E[X[k]Y^*[l]] \\
&= E\left[X[k]\Big(\sum_{n=-\infty}^{\infty} h[n]X[l-n]\Big)^*\right] \\
&= \sum_{n=-\infty}^{\infty} h^*[n]E\Big[X[k]X^*[l-n]\Big] \\
&= \sum_{n=-\infty}^{\infty} h^*[n]R_{XX}[k-l+n].
\end{aligned}$$

It is obvious that $R_{XY}[k, l]$ depends on the relative shift $k - l$ only, instead of two separate k and l. We can write the correlation function in a more compact form as

$$R_{XY}[m] \triangleq R_{XY}[k, l] = h^*[-m] * R_{XX}[m],$$

where $m \triangleq l - m$.

Then, the auto-correlation function of Y is

$$\begin{aligned} R_{YY}[k,l] &= E[Y[k]Y^*[l]] \\ &= E\left[\Big(\sum_{n=-\infty}^{\infty} h[n]X[k-n]\Big)Y^*[l]\right] \\ &= \sum_{n=-\infty}^{\infty} h[n]R_{XY}[k-l-n], \end{aligned}$$

which also depends on the relative shift $k-l$ only. It is therefore evident that the correlation function of $Y[n]$ is shift-invariant. A more compact form for $R_{YY}[k,l]$ is

$$\begin{aligned} R_{YY}[m] &= \sum_{n=-\infty}^{\infty} h[n]R_{XY}[m-n] \\ &= h[m] * R_{XY}[m] \\ &= h[m] * h^*[-m] * R_{XX}[m]. \end{aligned}$$

This suggests that we can compute the correlation function of $Y[n]$ from the impulse response and the correlation function of the input random sequence. Here, it would be easier to work the problem in the frequency domain, giving

$$\begin{aligned} S_{YY}(\omega) &= H(\omega)H^*(\omega)S_{XX}(\omega) \\ &= |H(\omega)|^2 S_{XX}(\omega). \end{aligned}$$

where $S_{YY}(\omega)$, $H(\omega)$, and $S_{XX}(\omega)$ are the discrete-time Fourier transform of $R_{YY}[m]$, $h[m]$, and $R_{XX}[m]$, respectively. The correlation function can be found by the inverse Fourier transform

$$R_{YY}[m] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{YY}(\omega)e^{j\omega m}d\omega. \tag{1}$$

Remarks:

- The output random sequence $Y[n]$ is WSS.

- $S_{YY}(\omega)$ and $S_{XX}(\omega)$ are called the ***power spectral density*** (PSD) of $Y[n]$ and $X[n]$, respectively. From the definition of a density, we can say the average power of the random sequence $X[n]$ is

$$\begin{aligned} \mathrm{P}_{\text{ave}} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{XX}(\omega)d\omega \\ &= R_{XX}[0] \\ &= E[|X[n]|^2]. \end{aligned}$$

*So, for a **zero mean WSS** random sequence, the **average power** is its **variance**.*
We see this statement very often in the literature.

– **Question:**
Why can we say the Fourier transform of $R_{XX}[m]$ is the PSD of $X[n]$?

7 Power Spectral Density

(1) Deterministic signals can be classified into (i) power signals, and (ii) energy signals.

(i) For power signal, the power of a signal $x[n]$ is defined by

$$\mathrm{P} \triangleq \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} |x[n]|^2. \tag{2}$$

(ii) For energy signal, the energy of a signal $x[n]$ is defined by

$$\mathrm{E} \triangleq \sum_{n=-\infty}^{\infty} |x[n]|^2.$$

The Parseval's theorem states the following relation

$$\begin{aligned}
\sum_{n=-\infty}^{\infty} |x[n]|^2 &= \sum_{n=-\infty}^{\infty} x[n]x^*[n] \\
&= \sum_{n=-\infty}^{\infty} x[n] \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{j\omega m} d\omega \right)^* \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} X^*(\omega) \left(\sum_{n=-\infty}^{\infty} x[n] e^{-j\omega m} \right) d\omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)|^2 d\omega, \qquad (3)
\end{aligned}$$

where, by the definition of a density, we can take $|X(\omega)|^2$ as the energy spectral density.

(2) For a power signal $x[n]$, its energy is infinity. We can define a truncated version of $x[n]$ as

$$x_T[n] = \begin{cases} x[n] & -N \le n \le N \\ 0 & \text{otherwise}, \end{cases}$$

which has a finite energy

$$\begin{aligned}
\mathrm{E} &= \sum_{n=-\infty}^{\infty} |x_T[n]|^2 = \sum_{n=-N}^{N} |x[n]|^2 \qquad (4) \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} |X_T(\omega)|^2 d\omega,
\end{aligned}$$

where the last equality comes from the Parseval's relation and $X_T(\omega)$ is the Fourier transform of $x_T[n]$.

By dividing equation (4) by $2N+1$ and taking the limit, we have

$$\begin{aligned} \mathrm{P} &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} |x[n]|^2 \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \lim_{N\to\infty} \frac{1}{2N+1} |X_T(\omega)|^2 d\omega. \end{aligned}$$

We can see from the above that the power spectral density of the signal $x[n]$ is

$$\mathrm{PSD} = \lim_{N\to\infty} \frac{1}{2N+1} |X_T(\omega)|^2.$$

(3) For random sequence $x[n]$, the sample sequence $X[n, \varepsilon_i]$ for each outcome $\varepsilon_i \in \Omega$ is deterministic and can be plugged into the above relation. That is, the PSD for $X[n, \varepsilon_i]$ is

$$\lim_{N\to\infty} \frac{1}{2N+1} |X_T(\omega, \varepsilon_i)|^2,$$

where $X_T(\omega, \varepsilon_i)$ is the Fourier transform of the similarly truncated $X_T[n, \varepsilon_i]$.

By averaging all realizations of sample sequences, we have the *average power*

$$\begin{aligned} \mathrm{P}_{\text{ave}} &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} E\left[|X[n, \varepsilon_i]|^2\right] \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \lim_{N\to\infty} \frac{1}{2N+1} E\left[|X_T(\omega, \varepsilon_i)|^2\right] d\omega. \end{aligned}$$

So, the power spectral density $S_{XX}(\omega)$ of the random sequence $X[n]$ that bears more physical meanings is from the above

$$\boxed{S_{XX}(\omega) = \lim_{N\to\infty} \tfrac{1}{2N+1} E\left[|X_T(\omega, \varepsilon_i)|^2\right].} \qquad (5)$$

Next, we will show that the equation (5) is indeed the Fourier transform of the correlation function of $R_{XX}[m]$.

(4) (**Wiener-Khinchine Theorem**)
The PSD

$$S_{XX}(\omega) = \lim_{N\to\infty} \frac{1}{2N+1} E\left[|X_T(\omega,\varepsilon_i)|^2\right],$$

if it exists, of a random sequence $X[n]$ is the Fourier transform of $R_{XX}[m]$.
Proof:

$$\begin{aligned}
\frac{1}{2N+1} E\left[|X_T(\omega,\varepsilon_i)|^2\right] &= \frac{1}{2N+1} \sum_{k=-N}^{N} \sum_{l=-N}^{N} R_{XX}[k-l] e^{-j\omega(k-l)} \\
&= \sum_{m=-2N}^{2N} R_{XX}[m] \left(1 - \frac{|m|}{2N+1}\right) e^{-j\omega m}
\end{aligned}$$

Then, it follows, by taking the limit of the above,

$$\begin{aligned}
S_{XX}(\omega) &= \lim_{N\to\infty} \frac{1}{2N+1} E\left[|X_T(\omega,\varepsilon_i)|^2\right] \\
&= \lim_{N\to\infty} \sum_{m=-2N}^{2N} R_{XX}[m] \left(1 - \frac{|m|}{2N+1}\right) e^{-j\omega m} \\
&= \sum_{m=-\infty}^{\infty} R_{XX}[m] e^{-j\omega m}.
\end{aligned}$$

■

(4) Properties of PSD $S_{XX}(\omega)$:

- $S_{XX}(\omega)$ is real.
- $S_{XX}(\omega)$ is an even function if $X[n]$ is real.
- $S_{XX}(\omega) \geq 0$ for all ω.

8 Autoregressive Moving Average (ARMA) Model

(1) What is ARMA model?

— A model used to describe ***time-varying*** phenomena, e.g. channel fading, with correlation at different time instants

— Linear constant coefficient difference equation model having the form [p. 364 in textbook]

$$X[n] = \sum_{k=1}^{M} c_k X[n-k] + \sum_{k=0}^{L} d_k W[n-k],$$

where $W[n]$ is the input independent random sequence with zero mean and unit variance, and $X[n]$ is the output sequence.

$\Rightarrow$ $X[n]$ is used to model a ***time-varying*** parameter that evolves from $X[m]$ with $m \leq n$ and perturbed by noise $W[m]$ with $m = n-L, \ldots, n$.

— When $L = 0$, the model is called **autoregressive** (AR) model.

$$X[n] = \sum_{k=1}^{M} c_k X[n-k] + d_0 W[n]$$

— When $M = 0$, the linear equation is

$$X[n] = \sum_{k=0}^{L} d_k W[n-k],$$

and is called **moving average**.

(2) Can use the concept of system to view ARMA.

$$W[n] \rightarrow \boxed{\quad h[n] \quad} \rightarrow X[n]$$

If the model is stable, we have the PSD of the output sequence

$$\begin{aligned} S_{XX}(\omega) &= |H(\omega)|^2 S_{WW}(\omega) \\ &= \frac{\left|\sum_{k=0}^{L} d_k e^{-j\omega k}\right|^2}{\left|1 - \sum_{k=0}^{M} c_k e^{-j\omega k}\right|^2} \end{aligned}$$

(3) Consider the problem of generating a random sequence with a specified PSD or correlation function. (Used in computer simulation) See textbook p. 355 and example 6.4-5 and 6.4-6.

9 Markov Random Sequence

(1) **Definition (Markov Property)**
If the pdf or pmf of a random sequence $X[n]$ has the property

$$f_X\left(x_{n+1}\middle|x_n, x_{n-1}, \cdots, x_0\right) = f_X\left(x_{n+1}\middle|x_n\right)$$

for all $x_0, \cdots, x_n, x_{n+1}$ with $n \geq 0$, then we say $X[n]$ is a ***Markov*** *random sequence.*

Example:
Consider the ***Binomial*** random sequence

$$S[n] = \sum_{k=1}^{n} X[k]$$

with $X[n]$ being the Bernoulli random sequence. The Binomial random sequence is Markov since $S[n]$ directly depends on $S[n-1]$, from the recursive relation

$$S[n] = S[n-1] + X[n].$$

Remarks:

— If the random sequence $X[n]$ is Markov, the above relation can be generalized to

$$f_X\left(x_{n+k}\middle|x_n, x_{n-1}, \cdots, x_0\right) = f_X\left(x_{n+k}\middle|x_n\right)$$

for all positive integer k.

— The Markov property allows us to specify a random sequence more efficiently.

We can find the Nth order joint pdf for a Markov random sequence $X[n]$ using

$$f_X\left(x_0, \cdots, x_{N-1}\right) = f_X(x_0) \prod_{k=1}^{N-1} f_X(x_k|x_{k-1}).$$

We can see this from

$$
\begin{aligned}
f_X\Big(x_0,\cdots,x_{N-1}\Big) &= \underbrace{f_X\Big(x_{N-1}\Big|x_0,\cdots,x_{N-2}\Big)}_{f_X(x_{N-1}|x_{N-2})}\cdot f_X\Big(x_0,\cdots,x_{N-2}\Big) \\
&= f_X\left(x_{N-1}|x_{N-2}\right) f_X\left(x_{N-2}|x_{N-3}\cdots x_0\right) f_X\Big(x_0,\cdots,x_{N-3}\Big) \\
&= \quad \cdots \\
&= f_X\left(x_{N-1}|x_{N-2}\right) f_X\left(x_{N-2}|x_{N-3}\right)\cdots f_X\left(x_1|x_0\right) f_X(x_0),
\end{aligned}
$$

which is based on repeated use of conditioning and on the Markov property.

– The ***discrete-valued*** Markov random sequence is called ***Markov chain***.

(2) **Definition (Markov Chain)**

A discrete-time ***Markov chain*** is a random sequence $X[n]$ whose Nth order conditional pmfs satisfy

$$
P_X\Big[x[n]\Big|x[n-1],\cdots,x[n-N]\Big] = P_X\Big[x[n]\Big|x[n-1]\Big]
$$

for all n, and all integers $N>1$. The value of $X[n]$ at time n is called the ***state*** at time n.

(3) We are particularly interested in the applications of Markov chains which have a finite state space, i.e. $X[n]$ takes on a finite set of values, typically integer.

A state transition diagram can represent the transition probability matrix. Such a diagram shows the states, and the probabilities are represented by numbers on arrows between states.

Example:
Consider the Binomial counting process $S[n] = \sum_{k=1}^{n} X[k]$. In each step, $S[n]$ can either stay the same or increase by one. The state transition matrix is given by

$$
\mathbf{P} = \begin{bmatrix} 1-p & p & 0 & 0 & \cdots \\ 0 & 1-p & p & 0 & \cdots \\ 0 & 0 & 1-p & p & \cdots \\ .. & & \cdots & & \cdots \end{bmatrix}.
$$

Example:
Suppose $X[n]$ is a finite state Markov chain taking on values from $\{1, 2, \cdots, M\}$. The probability that $X[n]$ is in state j at time n is

$$P_X[X[n] = j] = \sum_{i=1}^{M} \underbrace{P_X\left[X[n] = j \middle| X[n-1] = i\right]}_{\triangleq p_{ij}} P_X\left[X[n-1] = i\right].$$

Writing it in a matrix representation, we have

$$\boxed{\mathbf{p}^T[n] = \mathbf{p}^T[n-1] \cdot \mathbf{P}_n} \tag{6}$$

where

$$\mathbf{p}[n] \triangleq \left[P_X[X[n] = 1], \cdots, P_X[X[n] = M]\right]^T$$

and the (i, j)th component of the matrix $\mathbf{P}_n$ is

$$\mathbf{P}_n(i, j) = P_X\left[X[n] = j \middle| X[n-1] = i\right] = p_{ij}.$$

The matrix $\mathbf{P}_n$ is referred to as the ***state transition matrix*** (or ***transition probability matrix***). If $\mathbf{P}_n$ does not change over time, that is $\mathbf{P}_n$ does not depend on the time instant n, we say $X[n]$ is a ***homogeneous*** Markov chain.

(4) For a homogeneous Markov chain $X[n]$, the state probability at time n can be given by

$$\mathbf{p}^T[n] = \mathbf{p}^T[0] \cdot \mathbf{P}^n,$$

which requires the knowledge of initial state probability $\mathbf{p}[0]$. If the matrix $\mathbf{P}$ is symmetric, we can resort to eigenvalue decomposition to find $\mathbf{P}^n$ efficiently. Otherwise, we may need to solve the difference equation (See p. 370 in textbook)

$$\mathbf{p}^T[n] = \mathbf{p}^T[n-1] \cdot \mathbf{P}.$$

(5) The finite dimensional distributions of $X[n]$ can be completely specified by the initial distribution $\mathbf{p}[0]$ and the transition matrix $\mathbf{P}$.

$$\begin{aligned} P_X\Big[\quad & x[n] = x_n, X[n-1] = x_{n-1}, \cdots, X[0] = x_0\Big] \\ &= P_X\left[X[n] = x_n | X[n-1] = x_{n-1}\right] \cdots P_X\left[X[1] = x_1 | X[0] = x_0\right] \\ &\times P\left[X[0] = x_0\right]. \end{aligned}$$

10 Convergence

We have seen that a random sequence is actually a sequence of functions (random variables).

Thus, before plunging into the convergence concepts of random sequences, we need to review the basic definitions of the convergence of a sequence of numbers and a sequence of functions.

(1) **Definition (Convergence of A Sequence of Numbers)**

A sequence of complex numbers x_n converges to the complex number x if given an $\varepsilon > 0$, there exists an integer n_0 such that for $n > n_0$, we have

$$|x_n - x| < \varepsilon.$$

Notationally, we can write

$$\lim_{n\to\infty} x_n = x \quad (\text{or} \quad x_n \to x \quad \text{as} \quad \text{n} \to \infty).$$

(2) **Definition (Convergence of A Sequence of Functions)**

The sequence of functions $f_n(x)$ converges to the function $f(x)$ if the sequence of complex numbers $f_n(x_0)$ converges to $f(x_0)$ for each x_0.

11 Convergence of Random Sequences

(1) **Definition** (**Almost-Sure Convergence**)

The random sequence $X[n]$ converges ***almost surely*** *to the random variable X if the sequence $X[n, \varepsilon]$ converges $X[\varepsilon]$ for all $\varepsilon \in \Omega$ except possibly on a set of probability 0.*

Example:
Consider a sample space $\Omega = [0, 1]$, the closed interval between 0 and 1. Assume that each sample point $\varepsilon \in \Omega$ has a uniform distribution. Then, the random sequence

$$X[n, \varepsilon] = \exp\left(-n^2(\varepsilon - n)\right)$$

converges to $X = 0$ almost surely.

(2) **Definition** (**Mean-square Convergence**)

A random sequence $X[n]$ converges in the ***mean-square sense*** *to the random variable X if*

$$E\left[|X[n] - X|^2\right] = 0$$

as $n \to \infty$.

Remarks:
The convergence in mean square sense has the physical concept of power.

(3) **Definition** (**Convergence in Probability**)

A random sequence $X[n]$ converges ***in probability*** *to the random variable X if for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} P[|X[n] - X| > \varepsilon] = 0.$$

Remarks:

— Convergence in mean square sense implies the convergence in probability. This can be shown by the Chebyshev inequality.

— Convergence almost surely implies convergence in probability.

Example:
(Difference between a.s. convergence and convergence in prob.)
Let the sample space be the closed interval $\Omega = [0, 1]$ with the uniform probability distribution. Define the sequence as follows:

$$
\begin{aligned}
&X[1, s] = s + \mathrm{I}_{[0,1]}(s), \quad \text{(group 1)}\\
&X[2, s] = s + \mathrm{I}_{[0,1/2]}(s), \quad X[3, s] = s + \mathrm{I}_{[1/2,1]}(s), \quad \text{(group 2)}\\
&X[4, s] = s + \mathrm{I}_{[0,1/3]}(s), \quad X[5, s] = s + \mathrm{I}_{[1/3,2/3]}(s), \quad X[6, s] = s + \mathrm{I}_{[2/3,1]}(s),\\
&X[7, s] = s + \mathrm{I}_{[0,1/4]}(s), \quad X[8, s] = s + \mathrm{I}_{[1/4,2/4]}(s), \cdots\\
&\quad \vdots
\end{aligned}
$$

Define a random variable X as $X(s) = s$.

— Considering any sample sequence generated by picking a sample point s in Ω, say $s = 3/8$. It is clear that $X[n]$ does not converges to X with probability one, since, for any sample point, there always exists one or two jumps in each group as n grows to a very large number. So, we do not see the concept of convergence of a regular (deterministic) sequence for every sample sequence.

— The random sequence $X[n]$ does converges in probability. We have to examine the probability measure

$$P\Big[|X[n] - X| > \alpha\Big]$$

for any $\alpha > 0$ as n approaches infinity.
Let n be the time instant located at the kth point of the lth group. Mathematically,

$$n = 1 + 2 + \ldots + (l - 1) + k = \frac{l(l - 1)}{2} + k.$$

With this, we know

$$
\begin{aligned}
X[n] &= s + \mathrm{I}_{[\frac{k-1}{l}, \frac{k}{l}]}(s), \quad \text{yielding}\\
X[n] - X &= \mathrm{I}_{[\frac{k-1}{l}, \frac{k}{l}]}(s).
\end{aligned}
$$

It follows that

$$
P[|X[n] - X| > \alpha] = \begin{cases} 0, & \text{if } s \notin [\frac{k-1}{l}, \frac{k}{l}] \text{ for any } \alpha > 0;\\ 0, & \text{if } s \in [\frac{k-1}{l}, \frac{k}{l}] \text{ for } \alpha > s;\\ 1/l, & \text{if } s \in [\frac{k-1}{l}, \frac{k}{l}] \text{ for } 0 < \alpha < s\ . \end{cases}
$$

As $n \to \infty$, we see the above probability goes to zero.