- The terms "stochastic process" and "random process" are synonyms.

- We use functions to model deterministic signals. \( \text{Ex: } x(t) = \cos \omega_0 t \)

- We use stochastic processes to model random signals... a.k.a. statistical signals.

- Associated with a random process is an indexing set \( \mathcal{E} \).

  \[ \rightarrow \text{The indexing set can be uncountable, countable, or finite.} \]

  \( \text{Ex: } \mathcal{E} = \mathbb{R}, \mathcal{E} = \mathbb{Z}, \mathcal{E} = \{ -2, -1, 0, 1, 2 \}^3 \)

- We usually think of the indexing set as corresponding to time.

- A stochastic process is a collection of RVs, one for each element of \( \mathcal{E} \).

  \[ \rightarrow \text{Since the indexing set is ordered, the RVs are also ordered (in time).} \]
- Intuitively, the stochastic process \( X(t) \) has an RV at every \( t \in \mathbb{R} \).

- Likewise, the stochastic process \( X_k \) has an RV at every \( k \in \mathbb{Z} \).

- All of the RVs associated with a stochastic process have the same underlying set of experimental outcomes \( \mathcal{S} \).

\[ \Rightarrow \text{They also have the same "domain", \( \sigma \)-algebra, or "set of events" \( \mathcal{G} \) for their probability measure.} \]

\[ \Rightarrow \text{However, they may have different probability measures} \ P \ \text{in general.} \]

\[ \Rightarrow \text{For each trial of the experiment, each RV maps the experimental outcome} \ \omega \ \text{to a real number.} \]

\[ \Rightarrow \text{Thus, for experimental outcome} \ \omega, \ \text{the stochastic process is a function:} \]

**EX: \[ X(t) \]**

\[ \Rightarrow \text{The above function is called a "sample function" or "realization" of the process} \ X(t) \ \text{corresponding to the experimental outcome} \ \omega. \]
- Thus, we see that a stochastic process is actually a mapping from $\mathbb{R} \times S$ into $\mathbb{R}$.

- A stochastic process $X(t)$ with indexing set $\mathbb{R} = \mathbb{R}$ is called a "continuous-time" stochastic process.

- A stochastic process with indexing set $\mathbb{Z}$ is called a "discrete time" random process.

- To describe a random process completely, it is necessary to specify the joint density or distribution of all of the involved RVs.

**NOTE:** We have written $X(t)$ to denote a continuous-time process.

This is slightly misleading because the process $X(t)$ is not a function of $t$ in the usual sense.

$\Rightarrow$ However, any single sample function (realization) of the process is a function of $t$. 
- We will sometimes use the alternative notation $X_t$ to denote the process $X(t)$.

→ This gives us better consistency with the discrete-time notation $X_k$.

**Joint Densities for 2 Processes**

- Suppose $X_t$ and $Y_t$ are two continuous-time random processes.

- Let $t_1, t_2, \ldots, t_m, t_1', t_2', \ldots, t_m'$ be a set of $2M$ discrete time instants.

- Then we can consider the joint density function

  \[ f_{X_t, X_t', X_m, Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m}} (x_{t_1}, \ldots, x_{t_m}, y_{t_1'}, \ldots, y_{t_m'}) \]

→ This is often abbreviated

  \[ f_{x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_m} (x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_m) \]

- The processes $X_t$ and $Y_t$ are independent if

  \[ f_{x_1, \ldots, x_m, y_1, \ldots, y_m} (x_1, \ldots, y_m) \]

  \[ = f_{x_1, \ldots, x_m} (x_1, \ldots, x_m) f_{y_1, \ldots, y_m} (y_1, \ldots, y_m) \]

  for any choice of $t_1, \ldots, t_m, t_1', \ldots, t_m'$.
"Deterministic" Random Process

- This term is an oxymoron!
- The book uses it to describe a process like
  - \( X(t) = \Theta \), where \( \Theta \) is a \( N(m, \sigma^2) \) RV.
  - \( X(t) = \Theta(2 + \sin(\omega t)) \), \( \Theta \) an RV.

- For any realization, knowledge of the value of \( X(t) \) at any specific \( t \) gives us knowledge of the entire realization.

*Note:* In both cases above, there is really only a single RV involved.

*Note:* In both cases above, the correlation coefficient between \( X(t_1) \) and \( X(t_2) \) has magnitude 1 \( \forall t_1, t_2 \in \mathbb{R} \).

\Rightarrow \text{i.e., this is not a very interesting class of processes.}

\Rightarrow \text{The term "deterministic random process" is not standard.}
Let $X_k$ and $Y_k$ be discrete-time processes.

The "cross correlation" function of $X_k$ and $Y_k$ is defined by

$$R_{X,Y}(i,j) = E[X_i Y_j]$$

The "cross covariance" function is

$$C_{X,Y}(i,j) = Cov(X_i, Y_j) = E[(X_i - E[X_i])(Y_j - E[Y_j])]$$

$$= E[X_i Y_j] - E[X_i]E[Y_j]$$

Note: $R_{X,Y}(i,j) = Cov(X_i, Y_j) + E[X_i]E[Y_j]$.

The "autocorrelation" of $X_k$ is the cross-correlation of $X_k$ with itself:

$$R_X(i,j) = R_{X,X}(i,j) = E[X_i X_j]$$

Note: $R_X(i,j) = Cov(X_i, X_j) + E[X_i]E[X_j]$.
Note: When \( X_k \) is a stochastic process, \( \text{Cov}(X_i, X_j) \) is often called the "Autocovariance" of \( X_k \).

Correlation and Covariance for Continuous-Time Processes

Let \( X_t \) and \( Y_t \) be continuous-time stochastic processes.

Let \( s, t \) be real variables.

The "cross correlation" function of \( X_t \) and \( Y_t \) is given by

\[
R_{X,Y}(s,t) = E[X_s Y_t].
\]

The "cross covariance" function is given by

\[
C_{X,Y}(s,t) = \text{Cov}(X_s, Y_t) = E[(X_s - E[X_s])(Y_t - E[Y_t])].
\]

Note:

\[
R_{X,Y}(s,t) = \text{Cov}(X_s, Y_t) + E[X_s]E[Y_t].
\]
- The "Autocorrelation" function of $X_t$ is the cross-correlation of $X_t$ with itself:

$$R_X(s,t) = R_{X,X}(s,t) = \mathbb{E}[X_s X_t] = \text{Cov}(X_s, X_t) + \mathbb{E}[X_s] \mathbb{E}[X_t].$$

- When $X_t$ is a stochastic process, $\text{Cov}(X_s, X_t)$ is often called the "Autocovariance" function of the process.

**STATIONARY PROCESSES**

- A stochastic process is called "strict sense stationary" if the joint density of any number of the involved RVs is invariant under time translation.

$$\mathbb{E}_X: f_{X_{t_1}, X_{t_2}}(x_{t_1}, x_{t_2}) = f_{X_{t_1+\Delta}, X_{t_2+\Delta}}(x_{t_1+\Delta}, x_{t_2+\Delta})$$

$\Rightarrow$ For strict sense stationarity, or "SSS", this must be true for all joint densities of all order and $\forall \Delta \in \mathbb{R}$.

($\forall \Delta \in \mathbb{Z}$ if discrete-time).
A stochastic process is called "wide sense stationary" (WSS) if the mean and autocorrelation are invariant under time translation:

\[
\begin{align*}
\text{continuous time} & \quad \begin{cases} \\
E[X_{t1}] = E[X_{t_1 + \hat{t}}] & \forall t_1, t_2, \hat{t} \in \mathbb{R} \\
R_X(t_1, t_2) = R_X(t_1 + \hat{t}, t_2 + \hat{t}) 
\end{cases} \\
\text{discrete time} & \quad \begin{cases} \\
E[X_i] = E[X_{i + \hat{k}}] & \forall i, j, \hat{k} \in \mathbb{Z} \\
R_X(i, j) = R_X(i + \hat{k}, j + \hat{k}) 
\end{cases}
\end{align*}
\]

For WSS processes, the mean is constant and the autocorrelation depends only on the "shift amount" \( \hat{t} \) or \( \hat{k} \), displacement \( l_{t_2 - t_1} \) or \( l_{j - i} \).

We write:

\[
\begin{align*}
R_X(t_1, t_2) &= R_X(l) = R_X(l_{t_2 - t_1}) \\
R_X(i, j) &= R_X(k) = R_X(l_{j - i})
\end{align*}
\]

for WSS processes.
Intuition:

- For an SSS process, all joint moments are invariant under time translation.
- For a WSS process, the first and second order moments are invariant under time translation.

NOTE:

1. SSS implies WSS.
2. The converse is FALSE.

Ergodicity

- The concept of ergodicity is related to but distinct from that of stationarity.
- Let $X(t)$ be a random process with the underlying set of experimental outcomes $\mathcal{S}$. 
- Any given trial of the experiment results in a specific outcome \( \omega \in \Omega \), which the process maps to a sample function (realization) \( X_\omega(t) \):

\[
X_\omega(t) \quad \xrightarrow{t} 
\]

- We can compute the moments of \( X_\omega(t) \):

  \( \rightarrow \) First moment (sample mean):

\[
\overline{X}_\omega = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} X_\omega(t) \, dt
\]

  \( \rightarrow \) Second central moment:

\[
\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} [X_\omega(t) - \overline{X}_\omega]^2 \, dt
\]

- These moments are usually referred to as "time averages".

- Computing them involves fixing the experimental outcome and averaging over time.
- Alternatively, we could fix time and 
  average over experimental outcomes.
- Let $t$ be fixed at $t_0$.

  $\Rightarrow$ First moment (mean of $X(t_0)$):
  $$E[X(t_0)] = \int_{-\infty}^{\infty} \theta f_{X(t_0)}(\theta) \, d\theta$$

  $\Rightarrow$ Second Central Moment:
  $$E[(X(t_0) - E[X(t_0)])^2]$$
  $$= \int_{-\infty}^{\infty} (\theta - E[X(t_0)])^2 f_{X(t_0)}(\theta) \, d\theta$$

  $\Rightarrow$ The above are immediately recognized as the 
  mean and variance of the RV $X(t_0)$.

  $\Rightarrow$ In the context of stochastic processes,
  moments computed by averaging over 
  experimental outcomes for a fixed time
  are called "ensemble averages".
- An "ensemble" is the collection of sample functions generated as the experimental outcome varies over all of \( S \).

- Thus:

  \[ X_d(t) \]

  \[ \text{average} \]

  \[ \text{Time averaging: fix } d : \]

  \[ \text{Ensemble averaging: fix } t : \]

\[ \text{ensemble} \]

\[ \text{average} \]

\[ \downarrow \]
- If all time averages are equal to their corresponding ensemble averages, then the process $X(t)$ is called \underline{ergodic}.

- For an ergodic process, all of the moments -- including autocorrelation and autocovariance -- can be calculated from a single sample function.

**EX:** For a WSS ergodic process $X(t)$, the autocorrelation function can be computed from a single realization $X_d(t)$ according to

$$R_X(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} X_d(t) X_d(t+\tau) \, dt$$
SUPPLEMENT TO
PAGES 2.10-2.14: Ergodicity.

- Ergodicity is discussed in Section 2.4 of the book.

  > The book is quite informal and vague:

  "A random process is said to be "ergodic" if time averaging is equivalent to ensemble averaging."


- Let \( \{ \omega \} \) be an indexed set of experimental outcomes, where each \( \omega \in \Omega \).

- Let \( X(t) \) be a random process and \( X_{\omega_i}(t) \) be the sample function obtained when \( \omega_i \) is the experimental outcome.

- Suppose we wish to find the mean of the process \( \eta(t) \).

- Given \( N \) sample functions \( X_{\omega_i}(t) \), we can approximate the mean by the ensemble average

\[
\eta(t) \approx \eta_N(t) = \frac{1}{N} \sum_i X_{\omega_i}(t).
\]

- In the limit as \( N \to \infty \),

\[
\lim_{N \to \infty} \eta_N(t) = \eta(t).
\]
- Given a single sample function \( X_d(t) \), we can also compute the time average (a number)

\[
\bar{X} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} X_d(t) \, dt,
\]

- If \( X \) is not stationary, so that \( \eta(t) \) varies with \( t \), then \( \bar{X} \) is clearly a poor estimate of \( \eta(t) \). In this case, the process \( X(t) \) is not ergodic.

- If the process is stationary, so that \( \eta(t) = \eta \) (a number), then it may be that \( \bar{X} = \eta \).

\( \rightarrow \) Ergodicity describes the conditions under which time averages are equal to the corresponding ensemble averages.

**DEF**: A process \( X(t) \) is called "ergodic" if its ensemble averages equal the appropriate time averages.

\( \rightarrow \) This means that, with probability 1, any statistic of \( X(t) \) can be determined from a single sample function \( X_d(t) \).

\( \rightarrow \) Often, we are interested only in a specific set of statistics. This leads to less restricted, specialized types of ergodicity.
**DEF**: Given a process $X(t)$ with constant mean $\mathbb{E}[X(t)] = \eta$ and the time average

$$\eta_T = \frac{1}{2T} \int_{-T}^{T} x(t) \, dt,$$

the process is called mean-ergodic if

$$\lim_{T \to \infty} \eta_T = \eta$$

with probability $1$.

Note: $\eta_T$ is an RV. Whether $X(t)$ is mean ergodic or not, we have that

$$\mathbb{E}\{\eta_T^2\} = \frac{1}{2T} \int_{-T}^{T} \mathbb{E}\{X(t)^2\} \, dt = \eta.$$

Thus, $\mathbb{E}\{\eta_T^2\} = \eta$ is not a sufficient test for mean ergodicity.

The variance $\text{Var}\{\eta_T^2\} = \sigma_T^2$, however, does provide a sufficient test.

$X(t)$ is mean ergodic iff

$$\lim_{T \to \infty} \sigma_T^2 = 0.$$
**Theorem**: A process $X(t)$ with constant mean $\eta$ is mean ergodic iff its autocovariance $c_X(t_1, t_2)$ satisfies
\[
\lim_{T \to \infty} \frac{1}{4T^2} \int_{-T}^{T} \int_{-T}^{T} c_X(t_1, t_2) \, dt_1 \, dt_2 = 0.
\]

**Corollary**: A WSS process $X(t)$ with constant mean $\eta$ is mean ergodic if its autocovariance $c_X(t) = R_X(t) - \eta^2$ satisfies
\[
\lim_{T \to \infty} \frac{1}{2T} \int_{-2T}^{2T} c_X(t) \left(1 - \frac{|t|}{2T}\right) \, dt = 0.
\]

**Sufficient Condition**: If $X(t)$ is WSS and if
\[
\int_{-\infty}^{\infty} |c_X(t)| \, dt < \infty,
\]
then $X(t)$ is mean ergodic.

**Sufficient Condition**: If $X(t)$ is WSS, if $c_X(0) < \infty$, and if $\lim_{|t| \to \infty} c_X(t) = 0$, then $X(t)$ is mean ergodic.
Distribution Ergodic Processes

- Let $X(t)$ be a SSS process. Then all the RVs that make up $X(t)$ have the same distribution (cdf) $F(x) = P\{X(t) \leq x\}$.

- We wish to determine $F(x)$ from a single sample function $X_d(t)$.

- For each value $x$, we form the new process
  $$y(t) = \begin{cases} 1, & X(t) \leq x \\ 0, & X(t) > x \end{cases}.$$

Note: $P\{y(t) = 1\} = P\{X(t) \leq x\} = F(x)$

$P\{y(t) = 0\} = P\{X(t) > x\} = 1 - F(x)$

$$\Rightarrow E\{y(t)\} = P\{y(t) = 1\} \cdot 1 + P\{y(t) = 0\} \cdot 0 = P\{y(t) = 1\} = F(x).$$

- The question is: can we time average a sample function of $y(t)$ to determine $F(x)$?

- The answer is YES, provided $y(t)$ is mean ergodic. In this case we say that the process $X(t)$ is "distribution ergodic".
More formally, for a sample function of $X(t)$ let $\tau_1, \tau_2, \ldots, \tau_n$ be the lengths of the time intervals where the sample function is $\leq x$.

Define the time average

$$y_T = \frac{1}{2T} \int_{-T}^{T} y(t) \, dt = \frac{\sum_{i=1}^{n} \tau_i}{2T}$$

**DEF**: if $\lim_{T \to \infty} y_T = F(x)$, then we say that $X(t)$ is "distribution ergodic."

**Theorem**: Let $F(x_1, x_2; \tau) = P\{ X(t+\tau) \leq x_1, X(t) \leq x_2 \}$.

Then an SSS process $X(t)$ is distribution ergodic iff

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\tau|}{2T}\right) \left[F(x, x; \tau) - F^2(x)\right] \, d\tau = 0.$$
Correlation Ergodic Processes

- Let \( X(t) \) be WSS.
- Given a single sample function of \( X(t) \), we want to find the autocorrelation
  \[
  R_X(\tau) = E\{X(t)X(t+\tau)\}.
  \]
- For each value \( \tau \), we form the new process
  \[
  Z(t) = X(t)X(t+\tau).
  \]

Note: \( E\{Z(t)\} = E\{X(t)X(t+\tau)\} = R_X(\tau) \).

- The question is: can we time average a sample function of \( Z(t) \) to determine \( R_X(\tau) \)?
  
  \( \rightarrow \) The answer is YES, provided \( Z(t) \) is mean ergodic. In this case we say that the process \( X(t) \)
  is "correlation ergodic!"

- More formally, define the time average
  \[
  R_T = \frac{1}{2T} \int_{-T}^{T} Z(t) \, dt = \frac{1}{2T} \int_{-T}^{T} X(t)X(t+\tau) \, dt.
  \]

**DEF:** if \( \lim_{T \to \infty} R_T = R_X(\tau) \), then we say that \( X(t) \)
  is "correlation ergodic"
Note: \[ R_z(\lambda) = E\{z(t)z(t+\lambda)\}^2 \]
\[ = E\{x(t)x(t+\lambda)x(t+\lambda)x(t+\lambda+\lambda)\}^2. \]

\[ \rightarrow \text{The Autocovariance of } z(t) \text{ is} \]
\[ C_z(\lambda) = R_z(\lambda) - R^2_x(\lambda). \]

\[ \Rightarrow x(t) \text{ is correlation ergodic iff } C_z(\lambda) \]
\[ \text{satisfies the corollary on PAGE 2.14.4;} \]
\[ \text{that is, iff} \]
\[ \lim_{T \to \infty} \frac{1}{2T} \int_{-2T}^{2T} C_z(\lambda)(1 - \frac{1|\lambda|}{2T}) \, d\lambda = 0. \]

Note: if \( x(t) \) is correlation ergodic, then
\[ \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x^2(t) \, dt = E\{x^2(t)\}^2 \]
\[ \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \{x(t+\lambda) + x(t)\}^2 \, dt = 2[R_x(0) + R_x(\lambda)]. \]
Stationary Autocorrelation Function

- Let $X(t)$ be a WSS process with autocorrelation $R_X(\tau)$.

1. $R_X(0) = E[X_t^2] \geq 0$. (commutativity of multiplication)

2. $R_X(-\tau) = E[X_t X_{t-\tau}] = E[X_{t-\tau} X_t]$
   
   $= E[X_t X_{t+\tau}] = R_X(\tau)$ (WSS)

3. $|R_X(\tau)| \leq R_X(0)$.

   Follows from Cauchy–Schwartz inequality and the fact that $|\rho_{X_t X_{t+\tau}}| \leq 1$. 
Cross-Correlation for WSS Processes

- Let $X_t$ and $Y_t$ be two WSS processes.
- If the cross-correlation function $R_{X,Y}(s,t)$ is invariant to time shifts, then it is a function of $|s-t| = \tau$.
- In this case, we write

$$R_{X,Y}(\tau) = R_{X,Y}(s,t) = R_{X,Y}(|s-t|)$$

$\rightarrow$ The following properties then hold:

1. $R_{X,Y}(0) = R_{Y,X}(0)$
2. $R_{X,Y}(\tau) = R_{Y,X}(-\tau)$
3. $|R_{X,Y}(\tau)| \leq \sqrt{[R_X(0)R_Y(0)]}$

$\rightarrow$ The processes $X_t$ and $Y_t$ are called "jointly wide sense stationary."
Power Spectral Density

- For a WSS continuous-time process $X(t)$, the "power spectral density" is the Fourier Transform of the autocorrelation $R_X(\tau)$:

$$S_X(\omega) = \mathcal{F}[R_X(\tau)] = \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega \tau} d\tau$$

- The book writes $S_X(j\omega)$ instead of $S_X(\omega)$.

- The power spectral density is also known as:
  
  power spectrum
  spectral density
  PSD

1. Because $R_X(\tau)$ is real and even, $S_X(\omega)$ is also real, even, and non-negative.

2. The autocorrelation can be recovered from the PSD using the inverse Fourier transform

$$R_X(\tau) = \mathcal{F}^{-1}[S_X(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) e^{j\omega \tau} d\omega$$

- Plugging in $\tau = 0$, we obtain

$$R_X(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) d\omega = \mathbb{E} [X^2(t)],$$

  the mean power of the process.

- Thus, $S_X(\omega)$ admits an interpretation as power per unit frequency for the process $X(t)$.

* non-negative is shown in section 2.7 of the book.
- Suppose there is a strong correlation between
  the RVs $X(t_0)$ and $X(t_0 + \tau_0)$ for some fixed
  $t_0$ and $\tau_0$.

- Because $X(t)$ is WSS, this implies that there is strong
  correlation between $X(t)$ and $X(t + \tau_0)$ for any $t$.

- In particular, there is strong correlation between
  $X((t+\tau_0))$ and $X((t+\tau_0) + \tau_0) = X(t+2\tau_0)$.

- Generalizing, we see that the collection of variables
  $X(t+k\tau_0)$, $k \in \mathbb{Z}$

  are all highly correlated with one another.

- This implies that the process $X(t)$ possesses a sort
  of "pseudo-periodicity" in a statistical sense:

  ![Diagram showing all variables highly correlated]

  - In this case, the autocorrelation $R_x(t)$ will
    generally exhibit peaks at $R_x(k\tau_0)$, $k \in \mathbb{Z}$.

  - The power spectrum $S_x(\omega)$ will generally also
    have a peak at $\omega_0 = \frac{2\pi}{\tau_0}$. 
If the WSS process $X(t)$ varies rapidly, i.e., there is correlation between the RV's $X(t)$ and $X(t+\tau)$ only for $|\tau|$ small,

$\rightarrow$ Then $R_x(\tau)$ falls off rapidly away from the origin:

$R_x(\tau)$

$\rightarrow$ By the reciprocal spreading principle, this implies that $S_x(\omega)$ falls off slowly away from the origin:

$S_x(\omega)$

$\rightarrow$ Interpretation: the rapidly varying process $X(t)$ has significant high frequency content.

$\rightarrow$ We say that this process has a "short correlation length".

PAGE 219
- Likewise, if the WSS process $X(t)$ varies slowly, then there is appreciable correlation between $X(t)$ and $X(t+\tau)$ for $|\tau|$ large.

$\rightarrow$ Then $R_x(\tau)$ falls off slowly away from the origin.

$\rightarrow$ This implies that $S_x(\omega)$ falls off rapidly away from the origin.

$\rightarrow$ Intuitively, this means that the process $X(t)$ has substantial low frequency content.

$\rightarrow$ We say that this process has a "long correlation length".
- Let $X(t)$ be a WSS process.
- Truncate $X(t)$ to a time interval of length $T$ and denote the truncated process $X_T(t)$.
- Let $X_{\omega,T}(t)$ be a particular realization (sample function) of $X_T(t)$.

$\Rightarrow$ Then $X_{\omega,T}(t)$ is a function in the usual deterministic sense.

$\Rightarrow$ The periodogram of $X_{\omega,T}(t)$ is given by

$$\frac{1}{T} \left| \mathcal{F}[X_{\omega,T}(t)] \right|^2.$$ 

- The ensemble average of the periodogram is given by

$$E \left[ \frac{1}{T} \left| \mathcal{F}\{x_T(t)\} \right|^2 \right].$$

- In the book, it is shown that

$$\lim_{T \to \infty} E \left[ \frac{1}{T} \left| \mathcal{F}\{x_T(t)\} \right|^2 \right] = \mathcal{F}\{R_X(T)\} = S_X(\omega).$$

- It is sometimes useful to define the PSD in terms of the Laplace transform:

$$S_X(s) = \mathcal{L}\{R_X(T)\} = \int_{-\infty}^{\infty} R_X(T) e^{-st} dT$$

$\Rightarrow$ $S_X(\omega)$ and $S_X(s)$ are both called "power spectral density".
- Let \( X_k \) be a WSS discrete-time process with autocorrelation \( R_x(k) \).

- As in the continuous-time case, the PSD is the Fourier transform of the autocorrelation:

\[
S_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} R_x(k) e^{-j\omega k}
\]

\[
R_x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(e^{j\omega}) e^{j\omega k} \, d\omega
\]

- The interpretation of \( R_x(k) \) is analogous to that of \( R_x(t) \).

- It is sometimes useful to define the PSD of a discrete-time WSS process in terms of the Z-transform

\[
S_x(z) = \sum_{k=-\infty}^{\infty} R_x(k) z^{-k}
\]

- \( S_x(e^{j\omega}) \) and \( S_x(z) \) are both referred to as the "power spectral density" of \( X_k \).

- The symmetry of \( S_x(z) \) is

\[
S_x(z) = S_x\left(\frac{1}{z}\right)
\]
Cross Power Spectrum

- Let \( X(t) \) and \( Y(t) \) be jointly WSS processes with cross correlation functions \( R_{XY}(\tau) \) and \( R_{YX}(\tau) \).

  **Note:** \( R_{YX}(\tau) = R_{XY}(-\tau) \).

- The cross power spectra of \( X(t) \) and \( Y(t) \) are given by

  \[
  S_{XY}(\omega) = \mathcal{F}[R_{XY}(\tau)] = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j\omega \tau} d\tau
  \]

  \[
  S_{YX}(\omega) = \mathcal{F}[R_{YX}(\tau)] = \int_{-\infty}^{\infty} R_{YX}(\tau) e^{-j\omega \tau} d\tau
  \]

- The relationship between \( S_{XY}(\omega) \) and \( S_{YX}(\omega) \) is

  \[
  S_{XY}(\omega) = S_{YX}^*(\omega).
  \]

- Analogous to the correlation coefficient of two RVs, we define the "coherence function" of the WSS processes \( X(t) \) and \( Y(t) \) according to

  \[
  \gamma_{XY}^2(\omega) = \frac{|S_{XY}(\omega)|^2}{S_X(\omega) S_Y(\omega)}
  \]

- The coherence function may be interpreted as a frequency domain correlation coefficient for \( X(t) \) and \( Y(t) \).
The magnitude of the coherence function is always less than or equal to 1.

In the maximum correlation case, we have

\[ Y_{xx}^2(w) = \frac{|S_{xx}(w)|^2}{S_x(w)S_x(w)} = 1 \]

The minimum correlation occurs when \( X(t) \) and \( Y(t) \) have a zero cross-correlation function. In this case, \( Y_{xy}^2(w) = 0 \).

**PSD Example:** Suppose \( X(t) \) and \( Y(t) \) are zero mean jointly WSS processes and \( Z(t) = X(t) + Y(t) \).

Then \( S_Z(w) = S_x(w) + S_{xy}(w) + S_{yx}(w) + S_y(w) \).

→ If \( X(t) \) and \( Y(t) \) have a zero cross-correlation function, then this reduces to

\[ S_Z(w) = S_x(w) + S_y(w). \]
For two jointly WSS discrete-time processes \( X_k \) and \( Y_k \), the cross spectral density is given by

\[
S_{XY}(e^{j\omega}) = \mathcal{F}[R_{XY}(k)]
\]

\[
= \sum_{k=-\infty}^{\infty} R_{XY}(k) e^{-j\omega k}
\]

or

\[
S_{XY}(z) = \mathcal{Z}[R_{XY}(k)]
\]

\[
= \sum_{k=-\infty}^{\infty} R_{XY}(k) z^{-k}
\]

White Noise

- A continuous-time WSS process \( X(t) \) is called a "white noise process" if \( R_X(\tau) = \alpha \delta(\tau) \), \( \alpha \) a constant.

- In this case, the PSD is constant:

\[
S_X(\omega) = \alpha.
\]

- Light containing all frequencies in equal amounts is white \( \Rightarrow \) A process containing all frequencies in equal amounts is a "white noise".
The RVs that make up a white noise process are mutually uncorrelated for different times; e.g.,

\[ R_x(t) = \delta(t). \]

→ For this reason, a white noise process is also called an "uncorrelated process".

**NOTE:** A process that is not white is called "colored noise".

→ Light containing different amounts of different frequencies has color.

→ If a white noise is input to a linear filter, then the filter generally causes correlation in the output process. In this sense, such a filter is often referred to as a "coloring filter" or "coloration filter".

- For a white noise \( X(t) \), if the marginal probability density at each time is Gaussian, then \( X(t) \) is called a "Gaussian white noise".
- If the WSS discrete-time process $X_k$ is a sequence of zero mean mutually uncorrelated RVs, then $X_k$ is called a (discrete-time) white noise process.

- In this case,

$$R_x(k) = \delta[k]$$
$$S_x(e^{j\omega}) = \alpha$$

Band Limited White Noise

- If $X(t)$ is a WSS process with auto correlation

$$R_x(\tau) = \alpha \omega_0 \frac{\sin(\omega_0 \tau)}{\pi \omega_0 \tau}$$

then the PSD is constant in the baseband interval $\omega \in [-\omega_0, \omega_0]$, and zero outside this interval:

$\rightarrow$ In this case, $X(t)$ is called "band limited white noise".
Bandpass White Noise

- If $X(t)$ is a WSS process with autocorrelation

$$R_X(\tau) = 2\alpha \omega_m \frac{\sin \omega_m \tau}{\pi \omega_m \tau} \cos \omega_0 \tau$$

Then the PSD $S_X(\omega)$ is constant in a pair of passbands symmetrically located about the frequency origin:

$$S_X(\omega)$$

$\omega_0 - \omega_m \quad -\omega_0 \quad \omega_0 + \omega_m$

$\omega_0 - \omega_m \quad -\omega_0 \quad \omega_0 + \omega_m$

$\rightarrow$ In this case, the process $X(t)$ is called "Bandpass White Noise".
IID Process

- If the RVs comprising the process $X(t)$ or $X_k$ are all mutually independent and all have the same pdf, then the process is called an "independent, identically distributed" process or "IID" process.

$\Rightarrow$ IID implies zero mean; otherwise there would be nonzero correlation between the RVs at different times.

$\Rightarrow$ IID implies that the process is a white noise.

$\Rightarrow$ IID implies strict sense stationarity (SSS).

Gaussian Process

- A random process is called "Gaussian" if any finite set of RVs selected from the process are jointly Gaussian.

$\Rightarrow$ The joint density is then completely specified by the collection of first and second order moments.
Markov Process

- Let $X_t$ be a stochastic process.
- Pick a finite number of ordered time indices from the indexing set $\mathcal{D} = \mathbb{R}$:
  
  $$t_1 < t_2 < \ldots < t_{n+1}$$

- Associated with this set of times is a finite collection of RVs $X_{t_1}, X_{t_2}, \ldots, X_{t_{n+1}}$.

- The process $X_t$ is called "Markovian" or a "Markov Process" if

$$f_{X_{t_{n+1}} \mid x_t \ldots x_{t_1}} (x_{t_{n+1}} \mid x_{t_n} \ldots, x_{t_1}) = f_{X_{t_{n+1}} \mid x_{t_n}} (x_{t_{n+1}} \mid x_{t_n}).$$

→ This means that the conditional density of $X_{t_{n+1}}$ conditioned on a series of prior realizations $X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n}$ is the same as the conditional density of $X_{t_{n+1}}$ conditioned on $X_{t_n} = x_{t_n}$ alone.

→ All of the information about $X_{t_{n+1}}$ that is contained in $X_{t_1}, \ldots, X_{t_n}$ is captured by $X_{t_n}$ itself.
- For a particular sample function, let the realizations of the RV's be $X_{t_1} = X_{t_1}, \ldots, X_{t_{n+1}} = X_{t_{n+1}}$.

→ We often say that the process is in the "state" $X_{t_k}$ at time $t_k$. We call the change from $X_{t_k} = x_{t_k}$ to $X_{t_{k+1}} = x_{t_{k+1}}$ a "state transition" of the process.

- A Markov process is completely characterized by an initial marginal density

$$f_{X_{t_0}}(x_{t_0})$$

and the set of "transition probabilities"

$$f_{X_{t_{k+1}}|X_{t_k}}(x_{t_{k+1}}|x_{t_k}), \quad t_k < t_{k+1}.$$ 

- The transition probabilities are conditional pdf's that must satisfy the "Chapman-Kolmogorov" equation

$$f_{X_{t_3}|X_{t_1}}(x_{t_3}|x_{t_1}) = \int_{-\infty}^{\infty} f_{X_{t_3}|X_{t_2}}(x_{t_3}|x_{t_2}) f_{X_{t_2}|X_{t_1}}(x_{t_2}|x_{t_1}) \, dx_{t_2},$$

$$t_1 < t_2 < t_3.$$ 

→ This is a sort of transitivity property.
The definition of a discrete-time Markov process is completely analogous.

**Gauss-Markov Process**

- A Gauss-Markov process is both Gaussian and Markovian.
- A Gaussian process is Gauss-Markov if and only if the covariances of the involved RVs all satisfy the separability condition
  \[
  \text{cov}(X_{t_3}, X_{t_1}) = \frac{\text{Cov}(X_{t_3}, X_{t_2}) \text{Cov}(X_{t_2}, X_{t_1})}{\sigma_{X_{t_2}}^2}
  \]
  \[\forall \ t_1 < t_2 < t_3.\]
- The book considers only WSS Gauss-Markov processes.

→ In this case, the autocorrelation is exponential:
  \[
  R_X(\tau) = \sigma^2 e^{-\beta |\tau|}
  \]

→ The PSD takes the form
  \[
  S_X(\omega) = \frac{2\sigma^2 \beta}{\omega^2 + \beta^2} \quad S_X(s) = \frac{2\sigma^2 \beta}{-s^2 + \beta^2}
  \]
  where \(\sigma\) and \(\beta\) are parameters that completely specify the process.
- Study the Gauss-Markov and Markov examples in Ex. 2.11 and Section 2.11 of the book.

- Study the narrowband Gaussian process example in Section 2.12 of the book.

- Study the Brownian-Motion process described in Section 2.13 of the book.

- Experimental determination of the autocorrelation and PSD is discussed in section 2.15 of the book.

**Stochastic Differentiation**

- Let $X(t)$ be a stochastic process and $X_d(t)$, $t \in \mathcal{S}$, be a sample function.

- Clearly, $X_d(t)$ has a derivative in the usual sense:

$$
\dot{X}_d(t) = \frac{d}{dt} X_d(t) = \lim_{\varepsilon \to 0} \frac{X_d(t+\varepsilon) - X_d(t)}{\varepsilon}
$$

- We can also define the derivative $\dot{X}(t)$ of the process $X(t)$ itself. It is the stochastic process $\dot{X}(t)$ satisfying

$$
\lim_{\varepsilon \to 0} E \left\{ \left[ \frac{X(t+\varepsilon) - X(t)}{\varepsilon} - \dot{X}(t) \right]^2 \right\} = 0.
$$

→ This is known as "mean-square" differentiability.
Properties of the m.s. derivative:

\[ E[\dot{x}(t)] = \frac{d}{dt} E[x(t)] \]

\[ R_{\dot{x}\dot{x}}(t_1, t_2) = \frac{\partial R_{xx}(t_1, t_2)}{\partial t_2} \]

\[ R_{\dot{x}\dot{x}}(t_1, t_2) = \frac{\partial R_{xx}(t_1, t_2)}{\partial t_1} \]

\[ R_{\dot{x}\dot{x}}(t_1, t_2) = \frac{\partial^2 R_{xx}(t_1, t_2)}{\partial t_1 \partial t_2} \]

\[ E \left[ \frac{d^n}{dt^n} x(t) \right] = \frac{d^n}{dt^n} E[x(t)] \]

\[ R_{x^{(n)}x^{(m)}}(t_1, t_2) = \frac{\partial^{n+m} R_{xx}(t_1, t_2)}{\partial t_1^n \partial t_2^m} \]

For a WSS process, these reduce to

\[ R_{x\dot{x}}(\tau) = -\frac{d}{d\tau} R_x(\tau) \]

\[ R_{\dot{x}\dot{x}}(\tau) = \frac{d}{d\tau} R_{x\dot{x}}(\tau) = -\frac{d^2}{d\tau^2} R_x(\tau) \]

\[ R_{x^{(n)}}(\tau) = (-1)^n \frac{d^{2n}}{d\tau^{2n}} R_x(\tau) \]

\[ E[\dot{x}(t)] = \frac{d}{dt} E[x(t)] = 0 \]

\[ E \left[ \frac{d^n}{dt^n} x(t) \right] = \frac{d^n}{dt^n} E[x(t)] = 0 \]
- If \( X(t) \) and \( Y(t) \) are stochastic processes, then
\[
R_{X(n)Y(m)}(t_1, t_2) = \frac{d^{n+m}}{dt_1^n dt_2^m} R_{XY}(t_1, t_2)
\]

In the case where \( X(t) \) and \( Y(t) \) are jointly WSS, this reduces to
\[
R_{X(n)Y(m)}(I) = (-1)^m \frac{d^{n+m}}{dI^{n+m}} R_{XY}(I).
\]

**Stochastic Integration**

- As with the derivative, sample functions \( X_d(t) \) of a process \( X(t) \) may be integrated in the usual sense:
\[
Z_d = \int_a^b X_d(t) \, dt.
\]

Considered over the ensemble \( \{X_d(t)\}_{d \in \mathcal{D}} \), \( Z_d \) is an ordinary RV.

- We can also consider the mean-square integral of the process \( X(t) \) itself. By definition, it is the RV \( Z \) satisfying
\[
\lim_{\Delta t_i \to 0} E \left\{ \left[ Z - \sum_i X(t_i) \Delta t_i \right]^2 \right\} = 0.
\]

- A process \( X(t) \) is m.s. integrable if
\[
\int_a^b \int_a^b |R_{X(t_1, t_2)}| \, dt_1 \, dt_2 < \infty.
\]