Discrete Random Signals

• Signal :
  \[ \begin{cases} 
    \text{deterministic} : \text{unique value} \\
    \text{random signal} : \text{multiple value and described by probability density function}
  \end{cases} \]

• finite energy \quad \rightarrow \quad \text{deterministic}

• Infinite energy / nonperiodic signal
  \rightarrow \text{random signal} : \text{average / statistical properties.}
ensemble of sequences

Ex:
Bernoulli random process

\[ X(n) \]

\[ \cdots 0 1 2 \cdots n \]

a "sample" sequence

\[ \cdots \]

another "sample" sequence

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]

\[ X(n) \]

\[ \cdots n \]
to flip a coin random variable $X_n$

Head : $p$

Tail : $1 - p$

A given sequence $\{ X_n \mid -\infty < n < \infty \}$ is called “a realization of the random process, or a sample sequence of the random process.”

A random process is an indexed family of random variables $\{X_n\}$. The family of random variable is characterized by a set of prob. distribution functions that in general may be a function of the index $n$. 
prob. density function

\[ p_{X_n}(x_n, n) = \frac{\partial \overline{p}_{X_n}(x_n, n)}{\partial x_n} \]

prob. distribution function

\[ \overline{p}_{X_n}(x_n, n) = \int_{-\infty}^{x_n} p_{X_n}(x, n) \, dx \]

Joint prob. density function

\[ p_{X_n, X_m}(x_n, n, x_m, m) = \frac{\partial^2 \overline{p}_{X_n, X_m}(x_n, n, x_m, m)}{\partial x_n \partial x_m} \]

Joint prob. distribution function

\[ \overline{P}_{X_n, X_m}(x_n, n, x_m, m) = \text{prob}[X_n \leq x_n, X_m \leq x_m] \]

\[ = \int_{-\infty}^{x_n} \int_{-\infty}^{x_m} p_{X_n, X_m}(x_n, n, x_m, m) \, dx_n \, dx_m \]
If \( p_{X_n, X_m} = p_{X_n} \cdot p_{X_m} \)
\[ \implies \text{statistic independent} \]

If \( p_{X_{n+k}, X_{m+k}} = p_{X_n, X_m} \)
\[ \implies \text{stationary (in time)} \]
the prob. functions are independent of a shift of time-origin.
• Recall:
  average: \( m_{X_n} = E[X_n] = \int_{-\infty}^{\infty} x p_{X_n}(x, n) \, dx \) (mean): ensemble average
if \( g(\cdot) \) is a single-valued function, then \( g(x_n) \) is also a random variable, and the set of random variable \( \{g(x_n)\} \) defines a new random process.

\[
E[g(X_n)] = \int_{-\infty}^{\infty} g(x_n) p_{X_n}(x, n) \, dx
\]
if the random variables are quantized

\[ \Rightarrow E[g(x_n)] = \sum_{x} g(x) p_{X_n}(x, n) \]

\[ E[X_n + Y_m] = E[X_n] + E[Y_m] \]

\[ E[aX_n] = aE[X_n] \]
In general
\[ E[X_n \cdot Y_m] \neq E[X_n] \cdot E[Y_m] \]
if \[ E[X_n \cdot Y_m] = E[X_n] \cdot E[Y_m] \]
then \( X_n \) and \( Y_m \) are linearly independent

Statistically independent \( \iff \) Linearly independent

\[ E[X_n + iY_m] = E[X_n] + iE[Y_m] \]

Complex random variable
• \( E[X_n^2] \): mean square = \( \int_{-\infty}^{\infty} x^2 p_{X_n}(x, n) \, dx \):
  : average power

Variance = \( E[(X_n - m_{X_n})^2] = \sigma_{X_n}^2(n) \)

= \( E[X_n^2] - m_{X_n}^2 \)

= mean square – (mean)²
• Autocorrelation sequence:
  a measure of the dependence between values of the random process at different times.
  it describes the time variation of a random signal.

  \[ \phi_{xx}(n, m) = E \left[ X_n X_m^* \right] \]

  \[ = \int \int_{-\infty}^{\infty} x_n x_m^* p_{X_n, X_m}(x_n, n, x_m, m) \, dx_n \, dx_m \]

• Autocovariance

  \[ r_{xx}(n, m) = E \left[ (X_n - m_{x_n})(X_m - m_{x_m})^* \right] \]

  \[ = \phi_{xx}(n, m) - m_{x_n} m_{x_m} \]
• Cross – correlation

\[ \phi_{xy}(n, m) = E \left[ X_n Y_m^* \right] \]

\[ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x y^* \varphi(x, n, y, m) \, dx \, dy \]

• Cross – covariance

\[ r_{xy}(n, m) = E[(X_n - m_{x_n})(Y_m - m_{y_m})^*] \]

\[ = \phi_{xy}(n, m) - m_{x_n}^* m_{y_m} \]

For nonstationary process

\[ \phi_{xx}, r_{xx}, \phi_{xy}, r_{xy} : \text{two – dimensional sequences} \]
A stationary Random Process means the first order averages such as mean, \( E[X_n] \), and variance, \( E[(X_n - m_{x_n})^2] \), averaged only at one specific time \( n \) are independent of time; furthermore; the 2nd order average such as the autocorrelation, \( \phi_{xx}(n, m) = E[X_n, X_m^*] \) averaged at two specific time \( n \) & \( m \), are dependent of the time difference \( m-n \) only.

For stationary process

\[
\Rightarrow
\]

\[
m_x = E[X_n]
\]

\[
\sigma_x^2 = E[(X_n - m_x)^2]
\]

\[
\phi_{xx}(n, n + m) = \phi_{xx}(m) = E[X_nX_{n+m}]
\]

\( \rightarrow \) 1-D sequence
• For only the above three equations hold, such random processes are called “stationary in wide sense”

• Not true stationary (strict sense stationary), i.e., their prob. distribution are not time invariant.

\[
\Rightarrow P_{X_{n+k}, X_{m+k}} (x_{n+k}, n + k, x_{m+k}, m + k) \\
\neq P_{X_n, X_m} (x_n, n, x_m, m)
\]
• Ex: Bernoulli Random Process:

\[ p_{X_n}(x_n, n) = \begin{cases} 
  p & , x_n = 1 \\
  1-p & , x_n = -1 \\
  0 & , \text{otherwise}
\end{cases} \]

\[ m_{x_n} = 1 \cdot p + (-1)(1 - p) = 2p - 1 \]

\[ E[X_n^2] = (1)^2 \cdot p + (-1)^2 (1 - p) = 1 \]

\[ \sigma^2 = E[X_n^2] - m_{x_n}^2 = 1 - (2p - 1)^2 = 4p(1 - p) \]
• If we assume statistical independent and stationary for \( m \neq 0 \)
\[
\phi_{xx}(m) = E[X_n X_{n+m}^*] = E[X_n] \cdot E[X_{n+m}^*] = m_x^2
\]

for \( m = 0 \)
\[
\phi_{xx}(m) = E[X_n X_n^*] = E[X_n^2] = 1
\]

\[\Rightarrow \phi_{xx}(m) = \begin{cases} 
1 & , \ m = 0 \\
 m_x^2 & , \ m \neq 0 
\end{cases}\]
• if \( p = 1 / 2 \), \( m_x = 2p - 1 = 0 \)
then \( \phi_{xx}(m) = \begin{cases} 1 , & m = 0 \\ 0 , & m \neq 0 \end{cases} = \delta(m) \)

we call such random process “White Noise Process”
due to its Fourier transform is unity for all frequencies.

\[
F[\delta(n)] = 1 \\
F[1] = \delta(f) = 2\pi\delta(w)
\]
Time average

\[ \langle X_n \rangle = \lim_{N \to \infty} \frac{1}{2N + 1} \sum_{n=-N}^{N} X_n \] : time average

\[ \langle X_n, X_{n+m} \rangle = \lim_{N \to \infty} \frac{1}{2N + 1} \sum_{n=-N}^{N} X_n X_{n+m}^* \] : time autocorrelation

the limit exist if \( \{X_n\} \) is a stationary process with finite mean.
Ergodic. Random Process:

The time average equals to the ensemble average.

\[ \langle X_n \rangle = E [ X_n ] = m_{x_n} = m_x \]

\[ \langle X_n, X_{n+m}^* \rangle = E [ X_n, X_{n+m}^* ] = \phi_{xx} (m) \]

For an ergodic process, \( m_x, \phi_{xx}(m), \ldots, \) etc., can be computed from a single finite segment of an infinite-energy sequence. i.e.,

\[
\begin{align*}
\langle X(n) \rangle_N &= \frac{1}{2N+1} \sum_{n=-N}^{N} X(n) \quad \leftarrow \text{sample data of } X_n \\
\langle X(n)X^*(n+m) \rangle_N &= \frac{1}{2N+1} \sum_{n=-N}^{N} X(n)X^*(n+m)
\end{align*}
\]

useful for practical power spectra estimation.
Although the Z-transform of an infinite energy signal does not exist, the auto covariance and auto correlation sequences of such a sequence are aperiodic sequence for which the Z-transform and fourier transform often do exist.

Properties of correlation and covariance sequences of random signals:

For two real, stationary random processes \( \{X_n\} \) & \( \{Y_n\} \)

\[
\begin{align*}
\varphi_{xx}(m) &= E[X_n, X_{n+m}], \quad r_{xx}(m) = E[(X_n - m_x)(X_{n+m} - m_x)] \\
\varphi_{xy}(m) &= E[X_n, Y_{n+m}], \quad r_{xy}(m) = E[(X_n - m_x)(Y_{n+m} - m_y)]
\end{align*}
\]

Property 1:

\[
\begin{align*}
r_{xx}(m) &= \varphi_{xx}(m) = m_x^2 \\
r_{xy}(m) &= \varphi_{xy}(m) = m_x m_y
\end{align*}
\]
if zero mean: then correlation = covariance

(if \( m_x = 0 \), then \( \phi_{xx}(m) = r_{xx}(m) \) and \( \phi_{xy}(m) = r_{xy}(m) \))

Property 2:

\[
\phi_{xx}(0) = E[X^2] \quad : \text{mean-square value}
\]

\[
r_{xx}(0) = \sigma_X^2 \quad : \text{variance}
\]

Property 3:

\[
\phi_{xx}(m) = \phi_{xx}(-m); \quad r_{xx}(m) = r_{xx}(-m)
\]

\[
\phi_{xy}(m) = \phi_{yx}(-m); \quad r_{xy}(m) = r_{yx}(-m)
\]

Property 4:

\[
|\phi_{xy}(m)| \leq [\phi_{xx}(0) \phi_{yy}(0)]^{0.5}
\]

\[
|r_{xy}(m)| \leq [r_{xx}(0) r_{yy}(0)]^{0.5}
\]

\[
|\phi_{xx}(m)| \leq \phi_{xx}(0)
\]

\[
|r_{xx}(m)| \leq r_{xx}(0)
\]
Property 5: If \( Y_n = X_{n-n_0} \) (a time shift) and stationary process

\[
\begin{align*}
\varphi_{yy}(m) &= \varphi_{xx}(m) \\
r_{yy}(m) &= r_{xx}(m)
\end{align*}
\]

Property 6: For many random processes, the random variables become less correlated as they become more separated in time. Thus

\[
\begin{align*}
\lim_{m \to \infty} \varphi_{xx}(m) &= E[X_n X_{n+m}] = E[X_n] E[X_{n+m}] = m_x^2 \\
\lim_{m \to \infty} r_{xx}(m) &= E[(X_n-m_x)(X_{n+m}-m_x)] = 0 \\
\lim_{m \to \infty} \varphi_{xy}(m) &= E[X_n Y_{n+m}] = E[X_n] E[Y_{n+m}] = m_x m_y \\
\lim_{m \to \infty} r_{xy}(m) &= E[(X_n-m_x)(Y_{n+m}-m_y)] = 0
\end{align*}
\]
the correlation and covariance are aperiodic sequences which tend to die out for large values of \( m \). Thus it is often possible to represent these sequences in terms of their Z-transforms.

**Z-transform Representation:**

\[
\phi_{xx}(m) \leftrightarrow \Phi_{xx}(Z), \quad r_{xx}(m) \leftrightarrow R_{xx}(Z),
\]

\[
\phi_{xy}(m) \leftrightarrow \Phi_{xy}(Z), \quad r_{xy}(m) \leftrightarrow R_{xy}(Z).
\]

Note: the Z-transforms of \( \phi_{xx}(m) \) and \( \phi_{xy}(m) \) exist only when \( m_x=0 \), i.e., \( \phi_{xx}(m) = r_{xx}(m), \phi_{xy}(m) = r_{xy}(m) \). In which case:

\[
\begin{align*}
\Phi_{xx}(Z) &= R_{xx}(Z) \\
\Phi_{xy}(Z) &= R_{xy}(Z)
\end{align*}
\]

**Property 1:**

\[
R_{xx}(Z) = \sum_{m=-\infty}^{\infty} r_{xx}(m) Z^{-m}
\]
\[ \sigma_x^2 = r_{xx}(0) \]

\[ = \frac{1}{2\pi j} \oint_c R_{xx}(Z) Z^{-1} dZ \]

**Property 2:**

\[ R_{xx}(Z) = \sum_{m=-\infty}^{\infty} r_{xx}(m) Z^{-m} \]

\[ = \sum_{m=0}^{\infty} r_{xx}(m) Z^{-m} + \sum_{m=1}^{\infty} r_{xx}(-m) Z^m \]

\[ R_{xx}(Z^{-1}) = \sum_{m=0}^{\infty} r_{xx}(m) Z^m + \sum_{m=1}^{\infty} r_{xx}(m) Z^{-m} \]
\[
= \sum_{m=1}^{\infty} r_{xx}(m)Z^m + \sum_{m=0}^{\infty} r_{xx}(m)Z^{-m}
\]

\[\Rightarrow R_{xx}(Z^{-1}) = R_{xx}(Z)\]

Similarly \[R_{xy}(Z) = R_{yx}^*\left(1/Z^*\right)\]

The typical region of convergence of \(R_{xx}(Z)\) for two sided sequence is \(R_a < |Z| < 1/R_a, \ 0 < R_a < 1\)
Since the R.O.C contains the unit circle, then

\[ \sigma_x^2 = \frac{1}{(2\pi j)} \oint_c R_{xx}(Z) \, Z^{-1} \, dZ \]

\[ \equiv \frac{1}{(2\pi)} \int_{-\pi}^{\pi} P_{xx}(w) \, dw \]

where \( P_{xx}(w) = R_{xx}(Z) \mid_{Z=e^{jw}} \)

When \( m_x = 0 \), the variance is equal to the mean-square or average power. Thus the area under \( P_{xx}(w) \) for \(-\pi \leq w \leq \pi\) is proportional to the average power in the signal.

\( P_{xx}(w) \): power density spectrum (or simply spectrum)

\( P_{xx}(w) = P_{xx}(-w) \) and \( P_{xx}(w) \geq 0 \)

\[ P_{xx}(w) = R_{xx}(Z) \mid_{Z=e^{jw}} = F[r_{xx}(m)] \]

power density spectrum = Fourier transform of autocovariance

(autocorrelation if \( m_x = 0 \))
Note: the Fourier transform of the auto correlation sequence does not exist if $m_x \neq 0$. (impulse exist)

$$P_{xy}(w) = R_{xy}(Z) \mid_{z=e^{jw}} : \text{cross-power density spectrum}$$

$$P_{xy}(w) = P_{yx}^*(-w)$$

Response of Linear System to Random Signals:

$$X(n) : \text{real and “wide sense-stationary” random process}$$

$$Y(n) = \sum_{k=-\infty}^{\infty} h(n-k) X(k) = \sum_{k=-\infty}^{\infty} h(k) X(n-k)$$
If a stationary input is put into a linear shift-invariant system, is the output also stationary?

\[ m_y = E[Y(n)] = \sum_{k=-\infty}^{\infty} h(k)E[x(n-k)] = m_x \cdot \sum_{k=-\infty}^{\infty} h(k) = m_x \cdot H(e^{j0}) \]

\[ (H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} h(k)e^{-j\omega k}, \quad H(e^{j0}) = \sum_{k=-\infty}^{\infty} h(k), \quad E[x(n-k)] = m_x) \]

If a stationary input is put into a linear shift-invariant system, is the output also stationary?

\[ \phi_{yy}(m) = E[y(n), y(n+m)] = E[\sum_{k=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} h(k)h(r)x(n-k)x(n+m-r)] \]

\[ = \sum_{k=-\infty}^{\infty} h(k) \sum_{r=-\infty}^{\infty} h(r)E[x(n-k)x(n+m-r)] = \sum_{k=-\infty}^{\infty} h(k)\phi_{xx}'(m+k) \]

where \( E[x(n-k)x(n+m-r)] = \phi_{xx}(m+k-r) \)

\[ \phi_{xx}'(m+k) \equiv \sum_{r=-\infty}^{\infty} h(r)\phi_{xx}(m+k-r) \]
The output auto correlation sequence also depends only on the time difference \( m \).

The output is also stationary.

\[
\phi_{yy}(m) = \sum_{k=-\infty}^{\infty} h(k) \sum_{r=-\infty}^{\infty} h(r) \phi_{xx}(m+k-r)
\]

Let \( \ell = r-k \)

\[
\phi_{yy}(m) = \sum_{k=-\infty}^{\infty} h(k) \sum_{k=\ell}^{\infty} h(k+\ell) \phi_{xx}(m-\ell) = \sum_{k=-\infty}^{\infty} h(k) \sum_{l=-\infty}^{\infty} h(k+\ell) \phi_{xx}(m-l) \}
\]

\[
= \sum_{l=-\infty}^{\infty} \phi_{xx}(m-\ell)v(\ell) = \phi_{xx}(m)v(m)
\]

convolution
Where \( v(\ell) = \sum_{k=-\infty}^{\infty} h(k) h(k+\ell) = h(\ell) * h(-\ell) \) \( \leftarrow \) Fourier transform can apply

\[
\Phi_{yy}(Z) = \Phi_{xx}(Z) . V(Z) = \Phi_{xx}(Z) H(Z) H(Z^{-1})
\]

\( V(Z) = Z[h(\ell) * h(-\ell)] = H(Z) . H(Z^{-1}) \)

\( P_{yy}(m) = \Phi_{yy}(Z) \bigg|_{Z=e^{jw}} = \Phi_{xx}(e^{jw}) H(e^{jw}) H(e^{-jw}) \quad \left( H(e^{-jw})=H^{*}(e^{jw}) \right) \)

\( \Rightarrow P_{yy}(w) = P_{xx}(w) \left| H(e^{jw}) \right|^2 \)  

\( \Rightarrow \) Output power density spectra \( \Rightarrow \) input power density spectra

Output auto correlation = input auto correlation * impulse response autocorrelation

Output auto correlation = input auto correlation * impulse response autocorrelation
\[ \sigma_y^2 = \varphi_{yy}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{yy}(w) \, dw \quad \text{: average power in the output} \]

\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{xx}(w) \left| H(e^{jw}) \right|^2 dw \]

Now, suppose that \( H(e^{jw}) \) is an ideal bandpass filter, as shown below.

Note: \( P_{xx}(w) = P_{xx}(-w) \) \hspace{1cm} \text{even function of } w

\[ \left| H(e^{jw}) \right|^2 = \left| H(e^{-jw}) \right|^2 \]
\[ \sigma_y^2 = \phi_{yy}(0) = \frac{2}{2\pi} \int_0^\pi |H(e^{j\omega})|^2 P_{xx}(\omega) \, d\omega \]

\[ = \frac{1}{\pi} \int_{W_a}^{W_b} P_{xx}(\omega) \, d\omega : \text{average power in the output} \]

since \( P_{xx}(0) \geq 0 \Rightarrow \sigma_y^2 \geq 0 \)

\[ \phi_{xy}(m) = E[x(n)y(n+m)] : \text{cross-correlation between input and output} \]

\[ = \int_{-\infty}^{\infty} \sum_{k=-\infty}^{\infty} h(k) x(n+m-k) \, d\omega \]

\[ = \sum_{k=-\infty}^{\infty} h(k) \phi_{xx}(m-k) = \phi_{xx}(m) * h(m) \quad \text{take } z\text{-transform} \]
\[ \Phi_{xy}(Z) = H(Z) \Phi_{xx}(Z) \Rightarrow P_{xy}(w) = H(e^{jw}) P_{xx}(w) \quad (z=e^{jw}) \]

\[ \varphi_{xy}(m) = \varphi_{xx}(m) \ast h(m) \]

Now if, \( \varphi_{xx}(m) = \sigma^2_x \delta(m) \) : the input is a white random process

\[ \Rightarrow \varphi_{xy}(m) = h(m) \ast \sigma^2_x \delta(m) = \sigma^2_x h(m) \]

\[ \Rightarrow \] for a white-noise input, the cross-correlation between input and output of a linear system is proportional to the impulse of the system.

\[ h(m) = \varphi_{xy}(m) / \sigma^2_x \] : system identification

\[ \Rightarrow \] for a white-noise input, the cross-correlation between input and output of a linear system is proportional to the impulse of the system.

\[ \varphi_{xx}(m) = \sigma^2_x \delta(m) \]

\[ \Phi_{xx}(Z) = \sigma^2_x \cdot 1 \]

\[ \Rightarrow P_{xx}(w) = \Phi_{xx}(Z) \bigg|_{z=e^{jw}} = \sigma^2_x, \quad -\pi \leq w \leq \pi \]

\[ \Rightarrow P_{xy}(w) = \sigma^2_x H(e^{jw}) \]

: the cross-power spectrum is proportional to the frequency response of the system.
Power Spectrum Estimation

one of the most important areas of application for digital signal processing

1. Basic Principles of Estimation Theory

x(n): signal --- random process (ergodic process)

For each sample sequence generated by the random process,

\[ m_x = \langle x(n) \rangle = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=N}^{N+1} X(n) \]

\[ \hat{m}_x = \frac{1}{N} \sum_{n=0}^{N-1} X(n) \] : an estimate of \( m_x \)

and \( \hat{m}_x \to m_x \) as \( N \) “large enough”
The branch of statistical theory that pertains to such situation is called "estimation theory". Let’s consider a stationary ergodic process \( \{X_n\} \), \(-\infty < n < \infty\).

\[
\sigma_x^2 = E[(X-m_x)^2] = \langle (X_n-m_x)^2 \rangle \\
r_{xx}(m) = E[(X_n-m_x)(X_{n+m}*-m_x*)] = \langle (X_n-m_x)(X_{n+m}*-m_x*) \rangle \\
\text{and } P_{xx}(m) = \sum_{m=-\infty}^{\infty} r_{xx}(m) e^{-jwm}
\]

The estimation of a parameter of the random process is based on a finite segment of a single sample sequence, i.e., we have \( N \) values \( x(n) \), \( 0 \leq n \leq N-1 \), from which to estimate some parameters which initially we will denote "\( \alpha \)."

The estimate \( \hat{\alpha} \) of \( \alpha \) is a function of the r.v. \( X_n \), \( 0 \leq n \leq N-1 \), i.e.,

\[
\hat{\alpha} = F[X_0,X_1,\ldots,X_{N-1}] \ \Rightarrow \ \hat{\alpha} \text{ is also a random variable.}
\]
Prob. Density function $P_{\hat{\alpha}}(\hat{\alpha})$ will depend on

(i) The choice of the estimator $F[]$

(ii) The prob. Density functions of $X_n$

- Estimate 2 is superior to estimate 1 because the prob. Density of estimate 2 is more concentrated about the true value $\alpha$
Confidence interval:

to measure or characterize the concentration of the prob. Density function of an estimator

\[ \alpha - \Delta_2 \leq \hat{\alpha} \leq \alpha + \Delta_1 \]
Generally speaking, for a good estimator the prob. Density function $P_\hat{\alpha}(\hat{\alpha})$ should be narrow and concentrated around the true value.

Two properties of estimators that are commonly used as a basis for comparison

(I) bias

(II) variance

The bias of an estimator is defined as the true value of the parameter minus the expected value of the estimate, i.e.,

$$\text{bias} = \alpha - E[\hat{\alpha}] \triangleq B$$

$B = 0 \implies$ unbiased estimator.

The variance of the estimator in effect measures the width of the prob. Density and is defined by

$$\text{var}[\hat{\alpha}] = E[(\hat{\alpha} - E[\hat{\alpha}])^2] = \sigma_{\hat{\alpha}}^2$$
A small variance suggests that the prob. density $P_{\hat{\alpha}}(\hat{\alpha})$ is concentrated around its mean value, which, if the estimator is also unbiased, will be the true value of the parameter.

$$\text{mean square error} = E[(\alpha - \hat{\alpha})^2] = \sigma_{\hat{\alpha}}^2 + B^2$$

An estimate is said to be consistent if as the number of observations becomes large, the bias and the variance both tend to zero.
Detection and Estimation

Basic concept of detection theory:
detection process:
   to make a decision regarding the presence or absence of a particular signal in a noisy environment.

Hypothesis Testing: the fundamental of yes-no decision

\[
\begin{aligned}
    r(n) &= s(n) + w(n) \\
    r(n) &= 0 + w(n)
\end{aligned}
\]

By observing \( r(n) \) we must decide if the source signal \( s(n) \) is present or not. Intuitively, we will expect to make the decision based on some amplitude comparison.
Questions:
What do we compare?
How do we make the comparison?
such that the decision process will be optimum (and implicitly,”
what do we mean by optimum?”)

To begin, we designate two mutually exclusive hypotheses
$H_0$ and $H_1$ as follows :

$H_0 : r(n) = w(n)$

$H_1 : r(n) = s(n) + w(n)$
If we observe $r(n)$ and make an arbitrary decision on the presence or absence of $s(n)$ we can have 4 possible outcomes:

2 correct decisions:
- choose $H_1$ and $H_1$ is true
- choose $H_0$ and $H_0$ is true

2 incorrect decisions:
- choose $H_1$ and $H_0$ is true
- choose $H_0$ and $H_1$ is true

There are 4 common approaches which lead to rules for choosing a hypothesis. Each depends on an increasing level of statistical knowledge about the environment.

1. Maximum a Posteriori (MAP) Criterion:

  choose the hypothesis that is most likely to have the result presented in the given observation.
Consider the a posteriori (after the fact) conditional properties
\[ P \{ H_0 | r \} \text{ and } P \{ H_1 | r \} \]

decision rule:
if \[ P \{ H_1 | r \} > P \{ H_0 | r \} \] then choose \( H_1 \)

\[ \frac{P \{ H_1 | r \}}{P \{ H_0 | r \}} > 1 \]

This approach demands the minimum knowledge about the statistics of
the environment.
2. Maximum Likelihood (ML) Criterion

The previous decision rule considers only the a posteriori probabilities for $H_0$ and $H_1$ with no consideration of the actual a priori probabilities of the source signal $s(n)$ being present or not.

From the Bayes Theorem, we obtain

$$P\{H_1|r\} = P\{r|H_1\}P\{H_1\}/P\{r\}$$

and

$$P\{H_0|r\} = P\{r|H_0\}P\{H_0\}/P\{r\} = P\{r|H_0\}(1-P\{H_1\})/P\{r\}$$

where $P\{r|H_1\}$ represents the prob. of an observation $r$ given $H_1$ is true.

decision rule:

choose $H_1$ if $P\{r|H_1\}/P\{r|H_0\} > (1-P\{H_1\})/P\{H_1\}$ : likelihood ratio test

from the result of MAP

otherwise choose $H_0$ .
The ratio $P \{ r | H_1 \} / P \{ r | H_o \}$ is known as the “likelihood ratio”
The effect of the likelihood ratio test is to shift the level at which the
decision changes based on the a priori probability $P \{ H_1 \}$.
Note that, the ML and MAP criterion are equivalent when $P \{ H_1 \}= P \{ H_o \} = \frac{1}{2}$.
Both the preceding criteria ignored the cost or penalty of being wrong.
In many instances some errors are more costly then others. This will
be considered next.

3. Bayes Criterion:
   based on the assignment of a cost to each possible outcome of the
decision process and performing a likelihood ratio test that minimizes
the average cost.
Consider the joint probabilities for each possible outcome of the
decision process, $P_1, P_2, P_3$ and $P_4$, i.e.
$P_1 = \{ \text{choose } H_1 \text{, and } H_1 \text{ is true} \} = P \{ h_1, H_1 \}$
$= P \{ h_1 | H_1 \} P \{ H_1 \}$
\[ \Rightarrow P\{h_1|H_1\} = P\{h_1, H_1\}/P\{H_1\} \]

Similarly
\[ P_2 = P\{h_o|H_o\}P\{H_o\} \]
\[ P_3 = P\{h_1|H_o\}P\{H_o\} \]
\[ P_4 = P\{h_o|H_1\}P\{H_1\} \]

By assigning costs $c_1, c_2, c_3$ and $c_4$ to each joint prob. $P_1, P_2, P_3$ and $P_4$ we form the “Bayes risk function”

\[ R = c_1P_1 + c_2P_2 + c_3P_3 + c_4P_4 \]
\[ = P\{H_1\}[c_1P\{h_1|H_1\} + c_4P\{h_o|H_1\}] + P\{H_o\}[c_2P\{h_o|H_o\} + c_3P\{h_1|H_o\}] \]
\[ = P\{H_1\}c_4 + P\{H_o\}c_2 - P\{H_1\}[c_4 - c_1]P\{h_1|H_1\} + P\{H_o\}[c_3 - c_2]P\{h_1|H_o\} \]

The first two terms of $R$ are independent of the decision. Thus, to minimize the risk we must minimize the sum of the last two terms.
We assume that the cost of a correct decision is less than that of an incorrect decision; \( \Rightarrow c_4 - c_1 \geq 0 \) and \( c_3 - c_2 \geq 0 \)

to minimize the cost, we wish to make our decision such that
\[
P\{H_1\}[c_4-c_1]P\{h_1|H_1\} \geq P\{H_0\}[c_3-c_2]P\{h_1|H_0\}
\]

or
\[
\frac{P\{h_1|H_1\}}{P\{h_1|H_0\}} \geq \frac{P\{H_0\}[c_3-c_2]}{P\{H_1\}[c_4-c_1]}
\]

The prob. That the observation is one that decision rule assign to \( H_1 \) given that \( H_1 \) is true.

decision rule: choose \( H_1 \) if
\[
\frac{P\{h_1|H_1\}}{P\{h_1|H_0\}} \geq \frac{P\{H_0\}[c_3-c_2]}{P\{H_1\}[c_4-c_1]}
\]

Otherwise, choose \( H_0 \)
4. Neyman-Pearson Criterion

The Neyman-Pearson criterion for optimum detection is based on specifying a fixed probability for false detection (i.e., fixed $P\{h_1 \mid H_0\}$ which is often denoted $P_f$ for prob. of false alarm) and maximizing the prob. of detection $P_d = P\{h_1 \mid H_1\}$

Since $P\{h_1 \mid H_1\} = 1 - P\{h_0 \mid H_1\}$ and $P\{h_1 \mid H_0\}$ is fixed, we can minimize the quantity $Q = P\{h_0 \mid H_1\} + \alpha P\{h_1 \mid H_0\}$ which is equivalent to maximizing $P\{h_1 \mid H_1\}$, where $\alpha$ represents an arbitrary constant. If we assume zero cost for a correct decision i.e., $c_1 = c_2 = 0$. Then from 3, we obtain

$$R = P\{H_1\} c_4 - P\{H_1\} c_4 P\{h_1 \mid H_1\} + P\{H_0\} c_3 P\{h_1 \mid H_0\}$$

$$= P\{H_1\} [1 - P\{h_1 \mid H_1\}] c_4 + P\{H_0\} c_3 P\{h_1 \mid H_0\}$$

Thus, if $P\{H_1\} c_4 = 1$ and $P\{H_0\} c_3 = \alpha$

We have $R = Q$
From the Bayes criterion we know that minimizing $Q$ will give the decision rule: choose $H_1$ if
\[ P\{r|H_1\}/P\{r|H_0\} \geq (P\{H_0\}c3/P\{H_1\}c4) = \alpha \]

The cost associated with false alarms.

Ex: Consider a unit value signal in zero mean Gaussian noise with unit variance.
If the signal is present then the observed signal will be the additive result of signal plus noise and will appear as a Gaussian distribution with unit mean.
If the signal is not present the noise alone will appear as a Gaussian distribution with zero mean, i.e.,
\[ P\{r|H_1\} = \frac{1}{\sqrt{2\pi}} e^{-r^2/2} \quad \text{and} \quad P\{r|H_0\} = \frac{1}{\sqrt{2\pi}} e^{-r^2/2} \]
The likelihood ratio, \( \lambda (r) \), is
\[
\frac{P\{r|H_1\}}{P\{r|H_0\}} = \lambda (r) = e^{\frac{r-1}{2}}
\]
and our decision rule becomes: choose \( H_1 \) if \( e^{\frac{r-1}{2}} \geq T \), where \( T \) is a threshold whose determination depends on which optimality criterion is being used.

It is common to use a log likelihood ratio test derived from the fact that taking logarithms do not alter the threshold position but gives a linear comparison of the form:
choose \( H_1 \) if \( r \geq \frac{1}{2} + \ln T \)

The observation \( r \) (possibly a voltage level measurement) represents a sufficient statistics on which to base our decision.
Effect of threshold placement on the prob. of false alarm and missed detection:

\[ P_{\text{miss}} = P\{h_0 \mid H_1 \} \]

\[ 1/2 + \ln T \]

\[ P_f = P\{h_1 \mid H_0 \} \]

choose \( H_0 \)

choose \( H_1 \)
The matched Filter : (The correlation filter)

\( w(n) \) : white noise

the detection of a signal in white noise will be more reliable if we make several observations rather than a single observation. If we were to average the observations over a period of time and then make the threshold comparison we would smooth out the sample by sample fluctuations due to the noise, and essentially have a measure of the mean value over the observation time with which to make our comparison.

\( r(t) = s(t) + w(t) \)

the optimum detection receiver is a correlation receiver based on the decision rule: choose \( H_1 \) if

\[
\int_0^T r(t)s(t)dt > T \quad \text{a predetermined threshold level}
\]

This operation can be obtained by applying \( r(t) \) to a linear filter with impulse response \( S(\tau - t) \) : a time reversed replica of the signal to be detected.
digital matched filter:

\[ r(n) \rightarrow h(n) \rightarrow \]

\( h(n) \): the time inverse of the signal to be detected.

\( r(n) = \{ s(n) + w(n) \)

\[ \Rightarrow r(n) * s(-n) = \sum_{k=-\infty}^{\infty} s(-k) r(n-k) \]

\[ = \sum_{i=-\infty}^{\infty} s(i) r(n+i) \quad : \text{correlation} \]

\[ = \sum_{i=-\infty}^{\infty} s(i) (s(n+i) + w(n+i)) \]

\[ = R_{ss}(n) + R_{sw}(n) \quad \text{(with signal)} \]

or \( r(n) * s(-n) = R_{sw}(n) \)

\( \) (without signal)
Thus, for detecting a known signal the output of a matched filter represents the sufficient statistic that is compared to a threshold value.

**Decision rule:** at each time $\tau$ choose $H_1$ if

$$r(n)*s(-n) \geq T$$

The matched filter problem can be proved by considering the following optimization problem:

$$\text{LTI}$$

$$h(\tau) : \frac{y_s^2(t_0)}{y_w^2(t_0)} \text{(max)}$$

$$y_s(t) + y_w(t)$$

$$y_s(t_0) = \int_{-\infty}^{\infty} s(\tau)h(t_0 - \tau)d\tau$$

$$y_w(t_0) = \int_{-\infty}^{\infty} w(\tau)h(t_0 - \tau)d\tau$$

$$\left(\frac{S}{N}\right)_{t_0} = \frac{y_s^2(t_0)}{y_w^2(t_0)} : \text{Signal to noise ratio maximized}$$
Basic Concepts of Estimation Theory

\[ r(n) = s(n) + w(n) \]

If we wish to determine only the presence or absence of \( s(n) \) we have a detection problem. If we wish to measure (estimate) some parameter of \( s(n) \) we generally refer to this as an estimation problem.

Recall: the Quality of an Estimation

\( \hat{\alpha} \)

\[ E[\hat{\alpha}] = \alpha \quad : \text{unbiased estimate} \]

\[ E[\hat{\alpha}] = \alpha + c \quad : \text{biased estimate} \]

where \( c \) is a constant

\[ E[\hat{\alpha}] = \alpha + f(\alpha) \quad : \text{the estimate has an unknown bias} \]

where \( f(\alpha) \) is some function of \( \alpha \)
\[ \text{var}[\hat{\alpha}] = E\{[\hat{\alpha} - E[\hat{\alpha}]]^2\} \] : how close a single estimate is to the mean

\[ \text{var}\{(\alpha - \alpha)^2\} = E\{[\hat{\alpha} - \alpha - E\{\hat{\alpha} - \alpha\}]^2\} \]

To have the greatest degree of confidence in an estimate we would like it to be unbiased and a minimum variance. If both the variance and bias approaches zero as the number of estimation trials approaches infinity the estimation is said to be “consistent.”
The projection Theorem:

Recall: 2-D case

\( \vec{v} \): the vector from origin to the point \((x,y)\)

The shortest distance from the point to the line \(L\) through the origin is the vector \(\vec{v} - \hat{v}\) which is perpendicular or orthogonal to the line \(L\).

That is, the dot product of \((\vec{v} - \hat{v})\) with any vector along \(L\) is zero.

The vector \(\hat{v}\) along \(L\) such that \((\vec{v} - \hat{v})\) is orthogonal to \(L\) is called the projection of \(\vec{v}\) on \(L\).
The 2-D case can be extended to an arbitrary number of dimensions. ⇒ “Hilbert space”

The vectors of a Hilbert space may be real or complex sequences, functions or random variables and infinite in number (e.g. a Fourier series).

The dot product of Euclidean space is generalized to an inner product in Hilbert space defined as a function that assigns to every pair of vectors \( x \) and \( y \) the scalar value \( (x, y) \) such that:

1. \( (x, y) = (x^*, y^*) \) where \( x^* \) is the complex conjugate of \( x \).
2. \( c(x, y) = (cx, y) = (x, cy) \); \( c \) is a scalar
3. \( (x + z, y) = (x, y) + (z, y) \)
4. \( (x, x) \geq 0 \)
5. \( (x, x) = 0 \) iff \( x = \phi \) : null vector.
6. \( (x, x) = \|x\|^2 \) : norm of \( x \)
Any two vectors $\vec{x}$ and $\vec{y}$ in a Hilbert space $H$ are orthogonal if their inner product $(\vec{x}, \vec{y}) = 0$.

The norm is a direct extension of the concept of the magnitude or length of a vector in Euclidean space.

The projection theorem assures us that for any vector $\vec{v}$ in a Hilbert space $H$, there is a unique vector $\vec{v}^\wedge$ defined on a complete subspace $S$ of $H$ called the projection of $\vec{v}$ on $S$ such that norm of $\vec{v} - \vec{v}^\wedge$ is minimum if $\vec{v} - \vec{v}^\wedge$ is orthogonal to $S$. 
Given a set of K vectors \( S_1, S_2, \ldots, S_k \) each of dimension N, they define a subspace S of the N-dimension Hilbert space H. We wish to determine the projection \( \hat{v} \) of an arbitrary vector \( v \) in H on the subspace S. We know from the projection theorem that \( v - \hat{v} \) must be orthogonal to S, that is the inner-product of \( v - \hat{v} \) with every vector in S must be zero.

\[
\Rightarrow \quad v - \hat{v} \text{ is orthogonal to each vector } S_j, j=1,2,\ldots,k, \text{ i.e.}
\]

\[
(\nu - \hat{v}, S_j) = (\nu, S_j) - (\hat{v}, S_j) = 0
\]

Now \( \hat{v} \) is a vector in S and can be represented as a linear combination of the vectors \( S_j \), i.e.,

\[
\hat{v} = \sum_{j=1}^{k} c_j s_j = S_c
\]

Where \( S_c \) is the n x k matrix whose columns are the vectors \( S_j \) and \( C \) is a k element vector of coefficients.
\[(v, S_j) = (\hat{v}, S_j) = \sum_{j=1}^{k} c_i (S_i, S_j)\]

In matrix form,

\[
\begin{pmatrix}
(v, S_1) \\
\cdot \\
\cdot \\
(v, S_k)
\end{pmatrix} =
\begin{pmatrix}
(S_1, S_1)(S_1, S_2)\ldots(S_1, S_k) \\
\cdot \\
\cdot \\
(S_k, S_1)(S_k, S_2)\ldots(S_k, S_k)
\end{pmatrix}
\begin{pmatrix}
C_1 \\
\cdot \\
\cdot \\
C_k
\end{pmatrix}
\]

or \(A = GC\) : “normal equations” with the obvious matrix equivalences.

If the inner product depends only on the difference of the indices.

\(\Rightarrow\) Toeplitz matrix.
G: Gram matrix: is non-singular iff \( S_j \) are linearly independent
(i.e., the set of vectors \( S_j \) from a basis set of the subspace \( S \))
\[ C = G^{-1}A \]

If the norm of \( \vec{x} - \hat{x} \) is minimized, then \( \hat{x} \) is a good estimate of \( \vec{x} \) in some sense.

Note that the Gram matrix contains only information about the subspace onto which the projection is being made. All the information regarding the vector to be estimated is contained in the \( A \) matrix. Thus, when formulating estimation problems to be solved using the projection theorem, the known or observed data from which the estimation is to be made will define the Gram matrix. Specification of the \( A \) matrix may be based on observed data as well in some problems or it may be derived from a model or an assumption regarding the relationship of the quantity to be estimated to the data from which it is to be estimated.

\[
\hat{v} = \sum_{j=1}^{k} c_j s_j
\]
\( \hat{\nu} = \sum_{j=1}^{k} c_j s_j \)

Represents the basic model of a linear estimation problem, expressing the desired estimate as a linear combination of the input data.

ex: eigen space decomposition — estimation
   — data compression
Spectral Estimation:

digital filtering

Dsp

power spectral estimation

1. Periodogram: the Fourier transform of the autocorrelation / autocovariance estimates
   . Autocorrelation estimate: consistent
   . Periodogram: not consistent estimate of spectral density

2. Average/smoothed periodograms
   : consistent

3. Autoregressive spectral estimation
   . Maximum likelihood method (MLM)
   . Maximum entropy method (MEM)