Chapter 2 Random Signal Analysis

I. Discrete Random Processes

1) A discrete random signal, $\{x(n)\}$, is a sequence of random variables:

$$x(0), x(1), x(2), \ldots, x(i), \ldots \tag{1}$$

2) Averages

(a) First-order moments (mean)

$$\mu_n = E\{x(n)\} = \int_{-\infty}^{\infty} x(n)\, f(x(n))\, dx(n) \tag{2}$$

where $f(x(n))$ is the probability density function (pdf) of $x(n)$. Or

$$\mu_n = \lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N} x_i(n)\right] \tag{3}$$

Properties of expectation:

(i) $E\{ax(m)+bx(n)\} = aE\{x(m)\}+bE\{x(n)\}$, a, b are arbitrary constant values.

(ii) $E\{x(m)x(n)\} \neq E\{x(m)\}E\{x(n)\}$ in general.

(iii) $y = g(x) \Rightarrow E\{y\} = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$.

(b) Second-order moments

(i) Correlation or autocorrelation

$$\begin{aligned} R(m,n) &= E\{x(m)x(n)\} \\ R(n,n) &= E\{x^2(n)\} \end{aligned} \tag{4}$$

(ii) Covariance or autocovariance

$$\begin{aligned} C(m,n) &= E\{[x(m)-\mu_m][x(n)-\mu_n]\} \\ &= E\{x(m)x(n)\} - \mu_m\mu_n = R(m,n) - \mu_m\mu_n \\ C(n,n) &= \sigma_n^2 = E\{[x(n)-\mu_n]^2\} \\ &= E\{x^2(n)\} - \mu_n^2 = R(n,n) - \mu_n^2 \end{aligned} \tag{5}$$

(iii) Covariance and correlation matrices

$$\mathbf{x} = \begin{bmatrix} x(0) & x(1) & \cdots & x(L-1) \end{bmatrix}^T_{L\times 1} \tag{6}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_{L-1} \end{bmatrix}^T_{L\times 1} \tag{7}$$

$$\mathbf{R} = E\left\{\mathbf{x}\mathbf{x}^T\right\} = \begin{bmatrix} R(0,0) & R(0,1) & \cdots & R(0,L-1) \\ R(1,0) & R(1,1) & \cdots & R(1,L-1) \\ \vdots & \vdots & \ddots & \vdots \\ R(L-1,0) & R(L-1,1) & \cdots & R(L-1,L-1) \end{bmatrix}_{L\times L} \tag{8}$$

And $R(m,n) = R(n,m)$.

$$\begin{aligned} \mathbf{C} &= E\left\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\right\} \\ &= E\left\{\mathbf{x}\mathbf{x}^T\right\} - \boldsymbol{\mu}\boldsymbol{\mu}^T = \mathbf{R} - \boldsymbol{\mu}\boldsymbol{\mu}^T \end{aligned} \tag{9}$$

(c) Higher-order moments

$$\begin{aligned} R^{(N)}(l_1, l_2, \ldots, l_N) &= E\left\{x(l_1)x(l_2)\cdots x(l_N)\right\} \\ C^{(N)}(l_1, l_2, \ldots, l_N) &= E\left\{\left[x(l_1)-\mu_{l_1}\right]\left[x(l_2)-\mu_{l_2}\right]\cdots\left[x(l_N)-\mu_{l_N}\right]\right\} \end{aligned} \tag{10}$$

3) Independence, Correlation and Orthogonality

(a) $x(m)$ and $x(n)$ are said to be independent if

$$\begin{aligned} & f\left(x(m), x(n)\right) = f\left(x(m)\right) f\left(x(n)\right) \\ & \therefore E\left\{x(m)x(n)\right\} = \mu_m \mu_n \end{aligned} \tag{11}$$

(b) $x(m)$ and $x(n)$ are said to be uncorrelated if $C(m,n) = 0$.

Independent $\underset{\not\longleftarrow}{\longrightarrow}$ **Uncorrelated**

(**Except for the case of Gaussian pdf**)

(c) $x(m)$ and $x(n)$ are said to be orthogonal if $R(m,n) = 0$ (or if $x(m)$ and $x(n)$ are uncorrelated and if μ_m or μ_n is zero).

4) Stationary

(a) Strictly stationary or strongly stationary sequences

$F\left(x(n_1), x(n_2), \ldots, x(n_k)\right)$ is shift-invariant for any set $(n_1, n_2, \ldots, n_k)$ and for any *k*.

Or

$$\begin{aligned} & F\left(x(n_1), x(n_2), \ldots, x(n_k)\right) \\ & = F\left(x(n_1+n_0), x(n_2+n_0), \ldots, x(n_k+n_0)\right) \end{aligned} \tag{12}$$

for any *k* and n_0.

(b) Weakly or wide-sense stationary sequences

All the first and second-order moments are finite and independent of sample index.

(i) $E\{x(m)\} = E\{x(n)\} = \text{constant} = \mu$

(ii) $R(m,n) = E\{x(m)x(n)\} = r(n-m)$, $r(i) = E\{x(n)x(n+i)\}$

i: correlation lag.

Properties of $r(i)$:

- $r(i) = r(-i)$.
- $E\{x^2(n)\} = r(0) \geq |r(i)|$.
- $$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & \cdots & r(L-1) \\ r(1) & r(0) & \cdots & r(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(L-1) & r(L-2) & \cdots & r(0) \end{bmatrix}_{L\times L} \tag{13}$$

$\Rightarrow$ Toeplitz matrix

$\mathbf{R}$ is a positive semi-definite matrix, i.e., $\mathbf{a}^T\mathbf{R}\mathbf{a} \geq 0$ for any nonzero $L\times 1$ vector $\mathbf{a}$.

5) Some Important Random Signals

(a) Purely random signals (independent, identically distributed, iid, with zero mean)

(i) $f\left(x(0), x(1), \ldots\right) = f\left(x(0)\right) f\left(x(1)\right)\cdots$

(ii) Uncorrelated, orthogonal

$$r(n-m) = E\{x(m)x(n)\} = \sigma_w^2\delta(n-m),\ \ \sigma_w^2 = E\{x^2(n)\} \tag{14}$$

White sequence or white noise:

A signal with a correlation structure given by

$$E\{x(m)x(n)\} = \sigma_w^2\delta(n-m). \tag{15}$$

Note:

$$\begin{array}{c}\textbf{iid sequence}\\ \textbf{with zero mean}\end{array} \;\underset{\not\longleftarrow}{\longrightarrow}\; \textbf{White sequence (with zero mean)}$$

(Except for the case of Gaussian pdf)

- In the course, we assume that all white sequences encountered are purely random (iid with zero mean).
- Correlation matrix of a white sequence, $\mathbf{R} = \sigma_w^2 \mathbf{I}$.
- We can generate white noise, using say Gaussian or uniform densities.
- An iid sequence is always strictly stationary and is also weakly stationary except for sequences with unbounded moments. (Example: iid with Cauchy distribution $f(x) = \dfrac{1}{\pi\left(1+x^2\right)}$, $E\{x\} \to \infty$.)

(b) First-order Markov signals

$$f\left(x(n) \mid x(n-1), x(n-2), \ldots, x(0)\right) = f\left(x(n) \mid x(n-1)\right) \tag{16}$$

"Conditioned on the previous sample alone."

(c) Gaussian random signals

Any set of L samples has a joint Gaussian density

$$f\left(x(i)\right) = \frac{1}{(2\pi)^{1/2}\sigma_i} \exp\left\{-\frac{1}{2\sigma_i^2}\left[x(i)-\mu_i\right]^2\right\} \tag{17}$$

$$\begin{aligned} & f\left(x(n_0), x(n_1), \ldots, x(n_{L-1})\right) = f(\mathbf{x}) \\ & = \frac{1}{(2\pi)^{L/2}\Delta^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \end{aligned} \tag{18}$$

where

$$\mathbf{x} = \left[x(n_0), x(n_1), \ldots, x(n_{L-1})\right]^T \tag{19}$$

$$\boldsymbol{\mu} = \left[\mu_{n_0}, \mu_{n_1}, \ldots, \mu_{n_{L-1}}\right]^T \tag{20}$$

$$\Delta = \text{Determinant of } \mathbf{C} \tag{21}$$

$$\mathbf{C} = Cov\{\mathbf{x}\}: \text{ covariance matrix for the elements of } \mathbf{x} \tag{22}$$

If $x(n)$ is zero-mean stationary, then

$$\mathbf{C} = \mathbf{R} \Rightarrow f(\mathbf{x}) = \frac{1}{(2\pi)^{L/2}\Delta^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T \mathbf{R}^{-1}\mathbf{x}\right\} \tag{23}$$

Further, assume that $x(n)$ is zero-mean Gaussian white. Then,

$$\mathbf{R} = diag\left(\left[\sigma_w^2,\ldots,\sigma_w^2\right]\right)_{L\times L},\ \ \sigma_w^2 = E\left\{x^2(n)\right\} \tag{24}$$

$$\mathbf{R}^{-1} = \frac{1}{\sigma_w^2}\mathbf{I},\ \ \Delta = \det(\mathbf{C}) = \det(\mathbf{R}) = \sigma_w^{2L},\ \ \Delta^{1/2} = \sigma_w^L \tag{25}$$

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{(2\pi)^{L/2}\sigma_w^L} \\ &\quad\times\exp\left\{-\frac{1}{2\sigma_w^2}\left[x(n_0),x(n_1),\ldots,x(n_{L-1})\right]\begin{bmatrix}1 & & \mathbf{0}\\ & \ddots & \\ \mathbf{0} & & 1\end{bmatrix}\begin{bmatrix}x(n_0)\\ x(n_1)\\ \vdots\\ x(n_{L-1})\end{bmatrix}\right\} \\ &= \frac{1}{(2\pi)^{L/2}\sigma_w^L}\exp\left\{-\frac{1}{2\sigma_w^2}\left[x^2(n_0)+x^2(n_1)+\ldots+x^2(n_{L-1})\right]\right\} \\ &= f(x(n_0))\cdot f(x(n_1))\cdot\ldots\cdot f(x(n_{L-1})) \end{aligned} \tag{26}$$

Properties of Gaussian random signals:

(i) Gaussian white noise signals are necessarily generated from zero-mean iid samples. $x(m)$ and $x(n)$ are samples of a Gaussian signal:
$x(m)$ and $x(n)$ are independent. $\Leftrightarrow$ $x(m)$ and $x(n)$ are uncorrelated.

(ii) Any linear operation (including linear filtering) applied to a Gaussian signal produces a result which is also Gaussian.

(iii) All higher order moments of a Gaussian distribution can be expressed in terms of the first- and second-order moments of the distribution only. $\Rightarrow$ Strict and weak stationarities are equivalent.

6) Complex Random Signals, $x(n) = x_R(n) + jx_I(n)$

Mean: $\mu_n = E\left\{x_R(n)\right\} + jE\left\{x_I(n)\right\}$.

Covariance: $C(m,n) = E\left\{\left[x(m)-\mu_m\right]^*\left[x(n)-\mu_n\right]\right\}$.

Correlation: $R(m,n) = E\left\{x^*(m)x(n)\right\}$.

For stationary cases,

$$r(i) = E\left\{x^*(n)x(n+i)\right\},\ \ r(0) = E\left\{\left|x(n)\right|^2\right\}. \tag{27}$$

II. Power Spectrum

1) A zero-mean stationary sequence $x(n)$ with a correlation sequence

$$r(i) = r(-i) = E\{x(m)x(m+i)\} \tag{28}$$

Z-transform of $r(i)$: $R(z) = \sum_{i=-\infty}^{\infty} r(i) z^{-i}$.

Fourier transform of $r(i)$:

$$R(e^{j\omega}) = \sum_{i=-\infty}^{\infty} r(i) e^{-j\omega i} \Leftrightarrow r(i) = \frac{1}{2\pi}\int_{-\pi}^{\pi} R(e^{j\omega}) e^{j\omega i} d\omega . \tag{29}$$

(a) $R(e^{j\omega})$ is called the power spectrum or power spectral density (PSD) of the random signal $x(n)$.

(b) $r(0) = E\{x^2(n)\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} R(e^{j\omega}) d\omega$, the signal power.

(c) $r(i)$ is a real and even function of i. $\Rightarrow$ $R(e^{j\omega})$ is a real, even and non-negative function of ω.

(d) $R(e^{j(\omega+2\pi k)}) = R(e^{j\omega}),\ k = 1, 2, \ldots$

(e) If $x(n)$ is white (i.e., $r(i) = \sigma_w^2 \delta(i)$), then $R(e^{j\omega}) = \mathcal{F}\{r(i)\} = \sigma_w^2$, i.e., the PSD is constant.

2) Cross-Correlation and Cross-Spectrum

(a) $R_{xy} = E\{x(m)y(n)\}$.

(b) Joint WSS

$x(n)$ and $y(n)$ are both WSS, and the cross-correlation between them depends only on the cross-correlation lags.

$$r_{xy}(i) = E\{x(n)y(n+i)\} = E\{y(n+i)x(n)\} = r_{yx}(-i) \tag{30}$$

Properties of $r_{xy}(i)$:

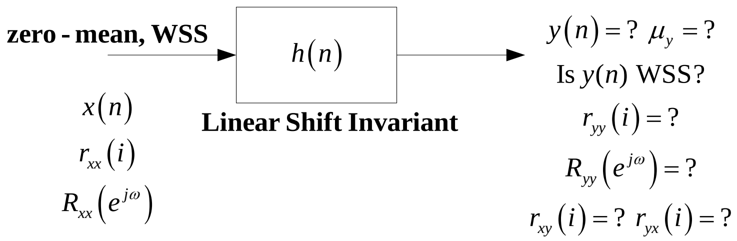
(i) $r_{xy}(i) = r_{yx}(-i)$.

(ii) $\left|r_{xy}(i)\right| \le \left[r_{xx}(0)r_{yy}(0)\right]^{1/2}$.

(c) Cross-spectrum

$$\begin{aligned} R_{xy}\left(e^{j\omega}\right) &= \sum_{i=-\infty}^{\infty} r_{xy}(i)e^{-j\omega i} = \sum_{i=-\infty}^{\infty} r_{yx}(-i)e^{-j\omega i} \\ &= \sum_{l=-\infty}^{\infty} r_{yx}(l)e^{j\omega l}, \ (l=-i) \\ &= \left[\sum_{l=-\infty}^{\infty} r_{yx}(l)e^{-j\omega l}\right]^* = R_{yx}^*\left(e^{j\omega}\right) \end{aligned} \tag{31}$$

III. Response of Linear Systems to Random Signals



■ **Figure 2.1** A linear shift invariant system [1].

1)

$$\begin{aligned} \mu_y &= E\left\{y(n)\right\} = E\left\{x(n)*h(n)\right\} \\ &= E\left\{\sum_{k=-\infty}^{\infty} x(k)h(n-k)\right\} \\ &= \sum_{k=-\infty}^{\infty} E\left\{x(k)\right\}h(n-k) = 0 \end{aligned} \tag{32}$$

2)

$$\begin{aligned} r_{yy}(n,n+m) &= E\left\{y(n)y(n+m)\right\} \\ &= E\left\{\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty} h(k_1)h(k_2)x(n-k_1)x(n+m-k_2)\right\} \\ &= \sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty} h(k_1)h(k_2)E\left\{x(n-k_1)x(n+m-k_2)\right\} \\ &= \sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty} h(k_1)h(k_2)r_{xx}(m+k_1-k_2) = r_{yy}(m) \end{aligned} \tag{33}$$

$\Rightarrow$ depends on lags only (independent on n)

From 1) and 2), we can see that $y(n)$ is a WSS process with

$$\begin{gathered} r_{yy}(m)=\sum_{k=-\infty}^{\infty} r_{xx}(m-k)\sum_{k_1=-\infty}^{\infty} h(k_1)h(k+k_1) \\ \left(\text{let } k=k_2-k_1 \Rightarrow k_2=k+k_1\right) \end{gathered} \tag{34}$$

Let $g(k)=\sum_{k_1=-\infty}^{\infty} h(k_1)h(k+k_1)=h(k)*h(-k)$.

$$\begin{aligned} r_{yy}(m) &= \sum_{k=-\infty}^{\infty} r_{xx}(m-k)g(k) \\ &= g(m)*r_{xx}(m) \\ &= h(m)*h(-m)*r_{xx}(m) \end{aligned} \tag{35}$$

Note:

-

$$\begin{aligned} & g(m)=h(m)*h(-m) \\ & \Rightarrow G(z)=H(z)H\left(z^{-1}\right) \\ & \Rightarrow G\left(e^{j\omega}\right)=H\left(e^{j\omega}\right)H\left(e^{-j\omega}\right) \overset{h(n)\text{ is real.}}{=} H\left(e^{j\omega}\right)H^*\left(e^{j\omega}\right)=\left|H\left(e^{j\omega}\right)\right|^2 \end{aligned} \tag{36}$$

-

$$R_{yy}(z)=G(z)R_{xx}(z)=H(z)H\left(z^{-1}\right)R_{xx}(z) \tag{37}$$

$$R_{yy}\left(e^{j\omega}\right)=\left|H\left(e^{j\omega}\right)\right|^2 R_{xx}\left(e^{j\omega}\right) \tag{38}$$

$$\begin{aligned} r_{xy}(m) &= E\{x(n)y(n+m)\} \\ &= E\left\{x(n)\sum_{k=-\infty}^{\infty} h(k)x(n+m-k)\right\} \\ &= \sum_{k=-\infty}^{\infty} h(k)r_{xx}(m-k) \\ &= r_{xx}(m)*h(m) \end{aligned} \tag{39}$$

$$R_{xy}\left(e^{j\omega}\right)=R_{xx}\left(e^{j\omega}\right)H\left(e^{j\omega}\right) \tag{40}$$

$$r_{yx}(m)=r_{xy}(-m)=r_{xx}(m)*h(-m) \tag{41}$$

$$R_{yx}\left(e^{j\omega}\right)=R_{xx}\left(e^{j\omega}\right)H\left(e^{-j\omega}\right) \tag{42}$$

$$\Rightarrow r_{yy}(m)=r_{xy}(m)*h(-m)=r_{yx}(m)*h(m) \tag{43}$$

- Special case: white noise, $x(n) \to w(n)$

$$r_{xx}(m) = \sigma_w^2 \delta(m) \tag{44}$$

$$R_{xx}\left(e^{j\omega}\right) = \sigma_w^2 \tag{45}$$

$$r_{yy}(m) = r_{xx}(m) * g(m) = \sigma_w^2 \delta(m) * g(m) = \sigma_w^2 g(m) \tag{46}$$

$$r_{yy}(0) = \sigma_w^2 g(0) = \sigma_w^2 \sum_{k=-\infty}^{\infty} h^2(k) \tag{47}$$

$$R_{yy}\left(e^{j\omega}\right) = R_{xx}\left(e^{j\omega}\right)\left|H\left(e^{j\omega}\right)\right|^2 = \sigma_w^2 \left|H\left(e^{j\omega}\right)\right|^2 \tag{48}$$

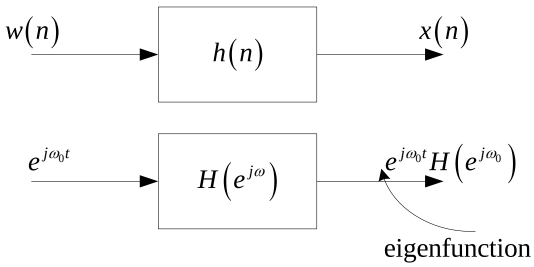
$$R_{xy}\left(e^{j\omega}\right) = R_{xx}\left(e^{j\omega}\right) H\left(e^{j\omega}\right) = \sigma_w^2 H\left(e^{j\omega}\right) \tag{49}$$

$$R_{yx}\left(e^{j\omega}\right) = R_{xx}\left(e^{j\omega}\right) H\left(e^{-j\omega}\right) = \sigma_w^2 H\left(e^{-j\omega}\right) = R_{xx}\left(e^{j\omega}\right) H^*\left(e^{j\omega}\right) \tag{50}$$

IV. Random Signal Models

1) Wold's Theorem

Any stationary sequence without deterministic components can be represented as the output of a stable, causal, and shift-invariant linear filter with a white noise input.



$x(n)$ is called a linear process.

$$R_{xx}\left(e^{j\omega}\right) = \sigma_w^2 \left|H\left(e^{j\omega}\right)\right|^2$$

■ **Figure 2.2** A white noise is input to a stable, causal, and linear shift invariant system [1].

2) Modeling the Linear Process

(a) Autoregressive Moving Average (ARMA) Models

$$H(z) = \frac{b_0 + b_1 z^{-1} + \cdots + b_N z^{-N}}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_M z^{-M}}$$
$$\begin{cases} \text{pole of } H(z): p_i,\ i = 1, 2, \ldots, M \\ \left|p_i\right| < 1 \end{cases} \tag{51}$$

(b) Autoregressive (AR) Models

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_M z^{-M}}, \text{ IIR} \tag{52}$$

AR(M): AR process of order M.

All poles of $H(z)$ are inside the unit circle.

(c) Moving Average (MA) Models

$$H(z) = b_0 + b_1 z^{-1} + \cdots + b_N z^{-N}, \text{ FIR} \tag{53}$$

3) Process Synthesis

(a) ARMA cases

$$H(z) = \frac{X(z)}{W(z)} \quad \text{(just an imaginary form!)} \tag{54}$$

$$\begin{aligned} x(n) = {} & a_1 x(n-1) + a_2 x(n-2) + \cdots + a_M x(n-M) \\ & + b_0 w(n) + b_1 w(n-1) + \cdots + b_N w(n-N) \end{aligned} \tag{55}$$

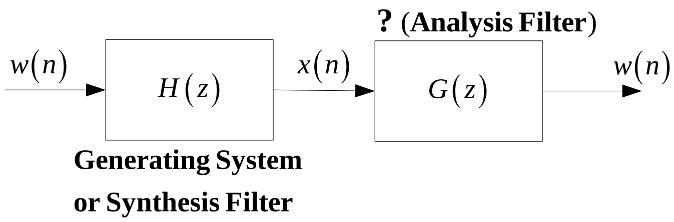
(b) AR cases

$$x(n) = a_1 x(n-1) + a_2 x(n-2) + \cdots + a_M x(n-M) + w(n) \tag{56}$$

(c) MA cases

$$x(n) = b_0 w(n) + b_1 w(n-1) + \cdots + b_N w(n-N) \tag{57}$$

4) Process Analysis



■ **Figure 2.3** Process synthesis and analysis – A white noise signal $w(n)$ is input to a system $H(z)$ and produces output $x(n)$. Then, $x(n)$ is input to the corresponding inverse system $G(z)$ and produces the white noise $w(n)$ [1].

$$G(z) = \frac{1}{H(z)} \tag{58}$$

$G(z)$ is referred to as the analysis filter.

$H(z)$ is causal and stable. Is $G(z)$ causal and stable?

5) The Stationarity of the Linear Process and the Stability of the Generating System

A linear process is stationary if and only if the generating system $H(z)$ is stable and then only after the transient effects have decayed.

Example: AR(1) process

$$x(n) = a_1 x(n-1) + w(n)$$

(a)

$$R_{xx}(n, n+m) = E\{x(n)x(n+m)\}$$

$$\begin{aligned} x(n) &= a_1\left[a_1 x(n-2) + w(n-1)\right] + w(n) \\ &= a_1^2 x(n-2) + a_1 w(n-1) + w(n) \\ &\vdots \\ &= a_1^n w(0) + a_1^{n-1} w(1) + \cdots + a_1^2 w(n-2) + a_1 w(n-1) + w(n) \end{aligned}$$

$$\begin{aligned} R_{xx}(n, n+m) = E\Big\{&\left[w(n) + a_1 w(n-1) + \cdots + a_1^{n-1} w(1) + a_1^n w(0)\right] \\ &\times\left[w(n+m) + a_1 w(n+m-1) + \cdots + a_1^m w(n)\right. \\ &\left. + a_1^{m+1} w(n-1) + \cdots + a_1^{n+m-1} w(1) + a_1^{n+m} w(0)\right]\Big\} \end{aligned}$$

$$\begin{aligned} &= \left[a_1^m + a_1^{m+2} + \cdots + a_1^{m+2n-2} + a_1^{m+2n}\right]\sigma_w^2 \\ &= a_1^m \frac{1 - a_1^{2(n+1)}}{1 - a_1^2}\sigma_w^2 \end{aligned}$$

As $n \to \infty$, $R_{xx}(n, n+m) \to \dfrac{a_1^m}{1 - a_1^2}\sigma_w^2$ for $|a_1| < 1$.

$\Rightarrow$ $R_{xx}(n, n+m)$ depends only on m.

$\Rightarrow$ Stationary, $r_{xx}(m) = \dfrac{a_1^{|m|}}{1 - a_1^2}\sigma_w^2, \; -\infty < m < \infty$.

Note: An alternative method for finding $r_{xx}(m)$

$$\begin{aligned} r_{xx}(m) &= E\{x(n)x(n-m)\} \\ &= E\{\left[a_1 x(n-1) + w(n)\right]x(n-m)\} \\ &= a_1 E\{x(n-1)x(n-m)\} + E\{w(n)x(n-m)\} \end{aligned}$$

$$E\{w(n)x(n-m)\}=0$$

$(\because x(n-m)$ depends only on $w(n-m), w(n-m-1), \ldots, w(0))$

$$r_{xx}(m)=a_1 r_{xx}(m-1) \Rightarrow r_{xx}(m)=a_1^m r_{xx}(0)$$

$$\begin{aligned}
r_{xx}(0) &= E\{x^2(n)\} \\
&= a_1^2 E\{x^2(n-1)\}+2a_1 E\{w(n)x(n-1)\}+E\{w^2(n)\} \\
&= a_1^2 r_{xx}(0)+\sigma_w^2 \\
&\Rightarrow r_{xx}(0)=\frac{\sigma_w^2}{1-a_1^2}
\end{aligned}$$

$$\therefore r_{xx}(m)=\frac{a_1^{|m|}\sigma_w^2}{1-a_1^2}$$

(b) Power spectrum of the AR(1) process

$$\begin{aligned}
R_{xx}(e^{j\omega}) &= \sum_{m=-\infty}^{\infty} r_{xx}(m)e^{-j\omega m} \\
&= \frac{\sigma_w^2}{1-a_1^2}\sum_{m=-\infty}^{\infty} a_1^{|m|}e^{-j\omega m} \\
&= \frac{\sigma_w^2}{1-a_1^2}\left[\sum_{m=0}^{\infty} a_1^m e^{-j\omega m}+\sum_{m=-\infty}^{-1} a_1^{-m}e^{-j\omega m}\right] \\
&= \frac{\sigma_w^2}{1-a_1^2}\left[\sum_{m=0}^{\infty}\left(a_1 e^{-j\omega}\right)^m+\sum_{m=0}^{\infty} a_1^m e^{j\omega m}-1\right] \\
&= \frac{\sigma_w^2}{1-a_1^2}\left[\frac{1}{1-a_1 e^{-j\omega}}+\frac{1}{1-a_1 e^{j\omega}}-1\right] \\
&= \frac{\sigma_w^2}{1-2a_1\cos\omega+a_1^2}
\end{aligned}$$

Note: An alternative method

$$H(z)=\frac{1}{1-a_1 z^{-1}} \Rightarrow H(e^{j\omega})=\frac{1}{1-a_1 e^{-j\omega}}$$

$$\therefore R_{xx}(e^{j\omega})=\left|H(e^{j\omega})\right|^2\sigma_w^2=\frac{\sigma_w^2}{1-2a_1\cos\omega+a_1^2}$$

■

V. The PSD of the Linear Process and the Phase of the Generating System

1) Ambiguity and the Power Spectrum

(a) $R_{xx}\left(e^{j\omega}\right)$ is real, even, nonnegative, and hence contains no phase information.

(b) The generating system associated with $R_{xx}\left(e^{j\omega}\right)$ is not unique (same amplitude spectrum, but different phases).

<u>*Example*</u>:

$$H(z)=1+h(1)z^{-1}+h(2)z^{-2}+\cdots+h(N)z^{-N}$$
$$R_{xx}(z)=H(z)H\left(z^{-1}\right)\sigma_w^2$$

Let

$$H(z)=\left(1-z_1z^{-1}\right)\left(1-z_2z^{-1}\right)\cdots\left(1-z_Nz^{-1}\right)$$
$$\text{roots }\left(z=z_1,\ z_2,\ldots,\ z_N\right)$$
$$H\left(z^{-1}\right)=\left(1-z_1z\right)\left(1-z_2z\right)\cdots\left(1-z_Nz\right)$$
$$\text{roots }\left(z=z_1^{-1},\ z_2^{-1},\ldots,\ z_N^{-1}\right)$$
$$R_{xx}(z)=H(z)H\left(z^{-1}\right)\sigma_w^2$$
$$\therefore z=z_1,\ z_2,\ldots,\ z_N,\ z_1^{-1},\ z_2^{-1},\ldots,\ z_N^{-1}$$

Consider another generating system, $h'(n)$:

$$H'(z)=\left(z^{-1}-z_1\right)\left(z^{-1}-z_2\right)\cdots\left(z^{-1}-z_N\right)=z^{-N}H\left(z^{-1}\right)$$
$$H'\left(z^{-1}\right)=\left(z-z_1\right)\left(z-z_2\right)\cdots\left(z-z_N\right)=z^{N}H(z)$$
$$R'_{xx}(z)=H'(z)H'\left(z^{-1}\right)\sigma_w^2=H(z)H\left(z^{-1}\right)\sigma_w^2$$

(i) $R_{xx}(z)$ and $R'_{xx}(z)$ are identical, although they are generated using different generating systems.

(ii) The generating sequences with transforms $H(z)$ and $H'(z)$ may be said to be spectrally equivalent.

(iii) $h'(n)$ is simply a time-reversal and delay version of $h(n)$.

(iv) There are many other spectrally equivalent generating systems.

■

2) Minimum and Maximum Phase Sequences

(a) Consider a causal (N+1)-point sequence with transform

$$\begin{aligned} H(z) &= h(0) + h(1)z^{-1} + h(2)z^{-2} + \cdots + h(N)z^{-N} \\ &= K\left(1 - z_1 z^{-1}\right)\left(1 - z_2 z^{-1}\right)\cdots\left(1 - z_N z^{-1}\right) \end{aligned} \tag{59}$$

where $z_1,\ z_2, \ldots,\ z_N$ are zeros of $H(z)$.

(i) $H(z)$ is said to be minimum phase if all zeros satisfy $|z_i| < 1$, i = 1, 2,…, N. (i.e., all zeros are inside the unit circle.)

(ii) $H(z)$ is said to be maximum phase if all zeros satisfy $|z_i| > 1$, $i = 1,\ 2, \ldots, N$. (i.e., all zeros are outside the unit circle.)

(iii) $H(z)$ is said to be mixed phase if neither (1) nor (2) is satisfied.

(b) Two factorizations of $H(z)$

Assume $|z_i| < 1$ for $i = 1,\ 2, \ldots,\ N_1$

$|z_i| > 1$ for $i = N_1 + 1, \ldots,\ N$

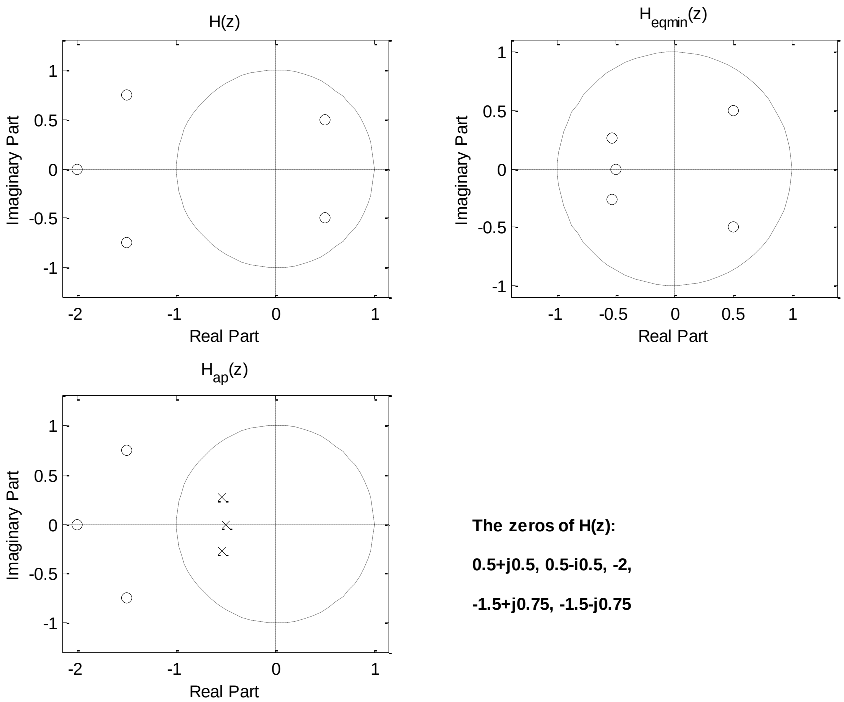
(i) minimum/maximum phase

$$\begin{aligned} H(z) &= K H_{\min}(z) H_{\max}(z) \\ &= K\left(1 - z_1 z^{-1}\right)\left(1 - z_2 z^{-1}\right)\cdots\left(1 - z_{N_1} z^{-1}\right) \\ &\quad \left(1 - z_{N_1+1} z^{-1}\right)\left(1 - z_{N_1+2} z^{-1}\right)\cdots\left(1 - z_N z^{-1}\right) \end{aligned} \tag{60}$$

(ii) equivalent minimum phase/all-pass

$$\begin{aligned} H(z) &= K\left(1 - z_1 z^{-1}\right)\left(1 - z_2 z^{-1}\right)\cdots\left(1 - z_{N_1} z^{-1}\right) \\ &\quad \left(1 - z_{N_1+1} z^{-1}\right)\left(1 - z_{N_1+2} z^{-1}\right)\cdots\left(1 - z_N z^{-1}\right) \\ &= K\left(1 - z_1 z^{-1}\right)\left(1 - z_2 z^{-1}\right)\cdots\left(1 - z_{N_1} z^{-1}\right) \\ &\quad \cdot\left[\frac{\left(z^{-1} - z_{N_1+1}\right)\left(z^{-1} - z_{N_1+2}\right)\cdots\left(z^{-1} - z_N\right)}{\left(z^{-1} - z_{N_1+1}\right)\left(z^{-1} - z_{N_1+2}\right)\cdots\left(z^{-1} - z_N\right)}\right] \\ &\quad \cdot\left(1 - z_{N_1+1} z^{-1}\right)\left(1 - z_{N_1+2} z^{-1}\right)\cdots\left(1 - z_N z^{-1}\right) \\ &= K H_{eq\min}(z) \frac{\left(1 - z_{N_1+1} z^{-1}\right)\left(1 - z_{N_1+2} z^{-1}\right)\cdots\left(1 - z_N z^{-1}\right)}{\left(z^{-1} - z_{N_1+1}\right)\left(z^{-1} - z_{N_1+2}\right)\cdots\left(z^{-1} - z_N\right)} \\ &= K H_{eq\min}(z) H_{ap}(z) \end{aligned} \tag{61}$$

Pole-zero diagrams of $H(z)$, $H_{eq\min}(z)$, and $H_{ap}(z)$:



■ **Figure 2.4** Factorization of $H(z)$ into equivalent minimum phase and all-pass [1]. (×: poles; ○: zeros)

(c) Spectrally equivalent generating sequences

$$w(n),\sigma_w^2 \longrightarrow \boxed{\begin{matrix} H(z) \\ h(n) \end{matrix}} \longrightarrow \begin{matrix} x(n) \\ R_{xx}(z)=\sigma_w^2 H(z)H(z^{-1}) \\ h(n)=Z^{-1}\{H(z)\} \end{matrix}$$

$$H(z)=K(1-z_1 z^{-1})(1-z_2 z^{-1})\cdots(1-z_N z^{-1})$$

zero pairs

$$\text{zeros,}\ \begin{cases} H(z): z_1,\ z_2, \ldots,\ z_N \\ H(z^{-1}): 1/z_1,\ 1/z_2, \ldots,\ 1/z_N \\ R_{xx}(z): z_1,\ z_2, \ldots,\ z_N,\ 1/z_1,\ 1/z_2, \ldots,\ 1/z_N \end{cases} \tag{62}$$

Choose

$$\begin{cases} H_1(z): 1/z_1,\ z_2,\ z_3, \ldots,\ z_N \\ H_1(z^{-1}): z_1,\ 1/z_2,\ 1/z_3, \ldots,\ 1/z_N \end{cases} \tag{63}$$

$$\begin{cases} H_2(z): z_1,\ 1/z_2,\ z_3, \ldots,\ z_N \\ H_2(z^{-1}): 1/z_1,\ z_2,\ 1/z_3, \ldots,\ 1/z_N \end{cases} \tag{64}$$

(i) Any system $H(z)$ constructed using any combination of N zeros obtained by choosing one zero from each pair of zeros has the same spectrum.

(ii) For an (N+1)-point sequence, there are a total 2^N distinct sequences which have the same spectrum.

(iii) All of the 2^N spectrally equivalent sequences are related by

$$H(z) = \underbrace{H_i(z)}_{H_{eq\min}(z)} H_{ap}(z) \tag{65}$$

(d) Comparison of spectrally equivalent generating sequences

(i) Same total energy

Parseval's relation

$$\sum_{n=0}^{N} \left|h(n)\right|^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left|H\left(e^{j\omega}\right)\right|^2 d\omega \tag{66}$$

(ii) Different partial energy

partial energy: $P_l = \sum_{n=0}^{l} \left|h(n)\right|^2,\ l < N$.

- P_l will generally differ for each spectrally equivalent generating sequence for any given $l < N$.
- The minimum phase sequence has the highest partial energy among the 2^N spectrally equivalent sequences.

<u>*Example*</u>: A 3-point sequence

$$h(0) = 1,\ h(1) = 1.3,\ h(2) = 0.4$$
$$H(z) = 1 + 1.3z^{-1} + 0.4z^{-2} = \left(1 + 0.8z^{-1}\right)\left(1 + 0.5z^{-1}\right)$$

Corresponding spectrally equivalent sequences:

$$H_1(z) = \left(1 + 0.8z^{-1}\right)\left(1 + 0.5z^{-1}\right) = 1 + 1.3z^{-1} + 0.4z^{-2}$$
$$H_2(z) = \left(z^{-1} + 0.8\right)\left(1 + 0.5z^{-1}\right) = 0.8 + 1.4z^{-1} + 0.5z^{-2}$$
$$H_3(z) = \left(1 + 0.8z^{-1}\right)\left(z^{-1} + 0.5\right) = 0.5 + 1.4z^{-1} + 0.8z^{-2}$$
$$H_4(z) = \left(z^{-1} + 0.8\right)\left(z^{-1} + 0.5\right) = 0.4 + 1.3z^{-1} + z^{-2}$$

Partial energy for each sequence:

l	h_1	h_2	h_3	h_4
0	1	0.64	0.25	0.16
1	2.69	2.60	2.21	1.85
2	2.85	2.85	2.85	2.85

■

(e) Extension to the case with both poles and zeros

$$
\begin{aligned}
H(z) &= \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_N z^{-N}}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_M z^{-M}} \\
&= K \frac{\left(1 - z_1 z^{-1}\right)\left(1 - z_2 z^{-1}\right)\cdots\left(1 - z_N z^{-1}\right)}{\left(1 - p_1 z^{-1}\right)\left(1 - p_2 z^{-1}\right)\cdots\left(1 - p_M z^{-1}\right)}
\end{aligned}
\tag{67}
$$

(i) $H(z)$ is called minimum phase if all zeros z_i and all poles p_i satisfy $|z_i| < 1$ and $|p_i| < 1$.

(ii) $H(z)$ is called maximum phase if all zeros z_i and all poles p_i satisfy $|z_i| > 1$ and $|p_i| > 1$.

(iii) $H(z)$ is called mixed phase if neither (1) nor (2) is satisfied.

VI. Estimation of Moments

$$
\begin{gathered}
\mu_n = E\left\{x(n)\right\} = \int_{-\infty}^{\infty} x(n)\, f\left(x(n)\right) dx(n) \\
\text{or} \\
\mu_n = \lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N} x_i(n)\right]
\end{gathered}
\tag{68}
$$

1) Problems

(a) We usually do not know a priori the probability density function associated with any sample $x(n)$.

(b) Usually only one member of the ensemble (a single experiment result) is available.

(c) The available experiment is likely to be finite in length.

2) Estimation Procedure

A procedure for estimating a deterministic parameter θ, say, derived from M points of a random data sequence $x(n)$, produces an estimate $\hat{\theta}$ that is itself a random variable

$$
\hat{\theta} = T\left(x(n)\right) \tag{69}
$$

(a) The random variable $\hat{\theta}$ is referred to as an estimator.

(b) Any particular value taken by the random variable $\hat{\theta}$ is an estimate.

3) Criteria for Assessing the Quality of an Estimator

(a) Bias

$$B\left(\hat{\theta}\right) = E\left\{\hat{\theta} - \theta\right\} = E\left\{\hat{\theta}\right\} - \theta = \bar{\theta} - \theta . \tag{70}$$

If $B\left(\hat{\theta}\right) = 0$, then $E\left\{\hat{\theta}\right\} = \theta$ and the estimation procedure is unbiased.

Asymptotically unbiased:

$$\lim_{M\to\infty} E\left\{\hat{\theta} - \theta\right\} \to 0 \quad \text{or} \quad \lim_{M\to\infty} B\left(\hat{\theta}\right) \to 0 . \tag{71}$$

(b) Efficiency

The efficiency of an estimation procedure is determined by the variance associated with $\hat{\theta}$. The lower the variance, the more efficient the estimator.

$$R_e \equiv \text{Relative efficiency} = \text{Var}\left\{\hat{\theta}_1\right\} \big/ \text{Var}\left\{\hat{\theta}_2\right\} \times 100\% \tag{72}$$

It's not clear which of the two estimation procedures is better.

$\hat{\theta}_1$ is unbiased but has a high variance;

$\hat{\theta}_2$ is biased but has a low variance.

(c) Mean-squared error

$$\begin{aligned} mse\left(\hat{\theta}\right) &= E\left\{\left(\hat{\theta} - \theta\right)^2\right\} \\ &= E\left\{\left[\hat{\theta} - \left(\bar{\theta} - B\left(\hat{\theta}\right)\right)\right]^2\right\} \\ &= E\left\{\left[\hat{\theta} - \bar{\theta} + B\left(\hat{\theta}\right)\right]^2\right\} \\ &= E\left\{\left(\hat{\theta} - \bar{\theta}\right)^2\right\} + E\left\{2B\left(\hat{\theta}\right)\left[\hat{\theta} - \bar{\theta}\right]\right\} + E\left\{B^2\left(\hat{\theta}\right)\right\} \\ &= Var\left(\hat{\theta}\right) + 2B\left(\hat{\theta}\right) \cdot 0 + B^2\left(\hat{\theta}\right) \\ &= Var\left(\hat{\theta}\right) + B^2\left(\hat{\theta}\right) \end{aligned} \tag{73}$$

The mean-squared error provides a more complete measure of the quality of $\hat{\theta}$ than either measure alone.

(d) Consistency

The more observations used, the closer the parameter estimate $\hat{\theta}$ should be to the parameter θ. An estimation procedure producing an estimate $\hat{\theta}$ is consistent if $\hat{\theta}$ converges to θ as $M \to \infty$ (M: the number of observations).

$$\lim_{M\to\infty} \hat{\theta} \to \theta \text{ in probability.} \tag{74}$$

(For any $\varepsilon > 0$ and any δ such that $0 < \delta < 1$, we can find an M_0 such that $P\left\{\left|\hat{\theta}-\theta\right| \geq \varepsilon\right\} < \delta$ for all $M \geq M_0$.)

※ $\lim_{M\to\infty} \left\{mse\left(\hat{\theta}\right)\right\} \to 0 \Rightarrow \lim_{M\to\infty} \hat{\theta} \to \theta$ in probability.

<u>Proof</u>:

$$\lim_{M\to\infty} \left\{mse\left(\hat{\theta}\right)\right\} = \lim_{M\to\infty} \left\{Var\left(\hat{\theta}\right) + B^2\left(\hat{\theta}\right)\right\} \to 0 \tag{75}$$

$$\Rightarrow \lim_{M\to\infty} \left\{Var\left(\hat{\theta}\right)\right\} \to 0 \text{ and } \lim_{M\to\infty} \left\{B^2\left(\hat{\theta}\right)\right\} \to 0 \tag{76}$$

(e) Sufficiency

$\hat{\theta} = T\left(x(n)\right)$ is a sufficient statistic for θ if the distribution of the observed data $x(0), x(1), \ldots, x(M-1)$ conditioned on $\hat{\theta}$ is not dependent on θ.

(i) $f\left(x(0), x(1), \ldots, x(M-1) \middle| \hat{\theta}\right)$ is not dependent on θ.

(ii) $\hat{\theta}$ contains all of the information in the observation which is relevant to θ.

4) Estimation of the Mean by a Time Average

Estimator:

$$\begin{aligned} \hat{\mu} &= \frac{1}{M}\sum_{n=0}^{M-1} x(n) \\ E\{\hat{\mu}\} &= \frac{1}{M}\sum_{n=0}^{M-1} E\{x(n)\} = \mu \end{aligned} \tag{77}$$

(a) This is an unbiased estimator, i.e., $B(\hat{\mu}) = 0$.

(b) A stationary sequence is ergodic in the mean if the estimator is consistent, i.e., $\lim_{M\to\infty}\left\{\frac{1}{M}\sum_{n=0}^{M-1} x(n)\right\} \to \mu$, in probability.

(c) This estimator is consistent if $\lim_{M\to\infty}\left\{Var\left(\hat{\theta}\right)\right\} \to 0$.

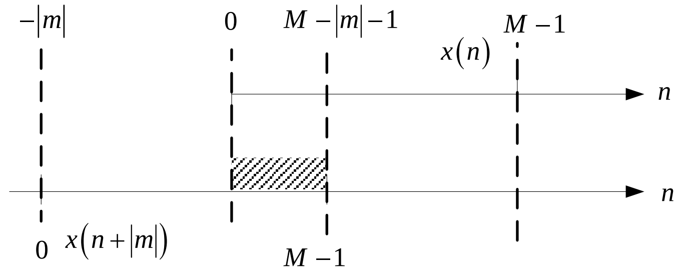
$$\left(\lim_{M\to\infty}\left\{\underbrace{C(i)}_{\text{Covariance}}\right\} \to 0 \Rightarrow \lim_{M\to\infty}\left\{Var\left(\hat{\theta}\right)\right\} \to 0\right) \quad \text{(see Appendix)} \tag{78}$$

$$C(i) = C(n-m) = E\{x(n)x(m)\} - \mu_n\mu_m \tag{79}$$

5) Estimation of the Correlation by a Time Average

(a) Estimator I:

$$\hat{r}(m) = \frac{1}{M-|m|} \sum_{n=0}^{M-|m|-1} x(n)x(n+|m|) \tag{80}$$



■ **Figure 2.5** The data used for the estimation process.

(i) unbiased: $E\{\hat{r}(m)\} = r(m)$.

(ii) As $|m|$ increases, $\hat{r}(m)$ is based on less data and the estimate becomes less and less reliable.

(iii) For Gaussian data and sufficiently large M,

$$Var\{\hat{r}(m)\} \cong \frac{1}{M-|m|} \sum_{i=-\infty}^{\infty} \left[r^2(i) + r(i-m)r(i+m) \right] \tag{81}$$

$M \to \infty \Rightarrow Var\{\hat{r}(m)\} \to 0$. This estimator is consistent.

(iv) $M > 4|m|$ is an empirical rule often given.

(v) The positive semi-definite property of the correlation matrix is lost when its elements are the $\hat{r}$'s.

(b) Estimator II:

$$r'(m) = \frac{1}{M} \sum_{n=0}^{M-|m|-1} x(n)x(n+|m|) \tag{82}$$

(i) Biased:

$$E\{r'(m)\} = \frac{M-|m|}{M} r(m) \tag{83}$$

But asymptotically unbiased:

$$\lim_{M \to \infty} E\{r'(m)\} \to r(m) \tag{84}$$

(ii) For Gaussian data and sufficiently large M,

$$Var\left\{r'(m)\right\} \cong \frac{1}{M}\sum_{i=-\infty}^{\infty}\left[r^2(i)+r(i-m)r(i+m)\right] \tag{85}$$

$M \to \infty \Rightarrow Var\left\{r'(m)\right\} \to 0$.

- This estimator is consistent.
- $Var\left\{r'(m)\right\}$ is lower than $Var\left\{\hat{r}(m)\right\}$.

(iii) This estimator has the advantage of retaining the positive semi-definite property for the correlation matrix.

(iv) The biased estimator improves the numerical condition of the matrix R, and this leads to generally better results.

Appendix – A Sufficient Condition for an Unbiased Estimator to Be Consistent

1) Assume that the mean μ of a stationary sequence $x(n)$ is estimated by

$$\hat{\mu} = \frac{1}{M}\sum_{n=0}^{M-1} x(n) \tag{A1}$$

Given the unbiased nature of $\hat{\mu}$, the estimator is consistent, i.e.,

$$\lim_{M\to\infty}\{\frac{1}{M}\sum_{n=0}^{M-1} x(n)\} \to \mu \tag{A2}$$

This equation is satisfied provided

$$\lim_{M\to\infty} Var\{\hat{\mu}\} \to 0 \tag{A3}$$

A sufficient condition for this is

$$\lim_{i\to\infty}\{C(i)\} \to 0 \tag{A4}$$

where $C(n-m) = E\{x(n)x(m)\} - \mu_m\mu_n$ is the covariance of stationary $x(n)$.

2) <u>*Proof*:</u> (Ref.: W. W. S. Wei, Time Series Analysis: Univariate and Multivariate Methods. Addison-Wesley, 1990.)

The variance of $\hat{\mu}$ is

$$\begin{aligned} Var\{\hat{\mu}\} &= E\{\hat{\mu}^2\} - E^2\{\hat{\mu}\} \\ &= E\left\{\left(\frac{1}{M}\sum_{n=0}^{M-1} x(n)\right)^2\right\} - \mu^2 = \frac{1}{M^2}\sum_{n=0}^{M-1}\sum_{m=0}^{M-1} E\{x(n)x(m)\} - \mu^2 \\ &= \frac{1}{M^2}\sum_{n=0}^{M-1}\sum_{m=0}^{M-1} C(n,m) \end{aligned} \tag{A5}$$

Since $x(n)$ is stationary, this may be written as

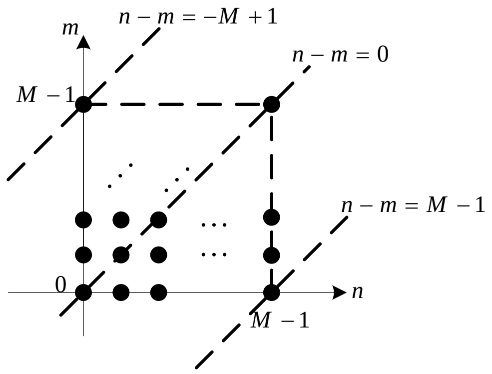
$$Var\{\hat{\mu}\} = \frac{1}{M^2}\sum_{n=0}^{M-1}\sum_{m=0}^{M-1} C(n-m) \tag{A6}$$

The correlation coefficient between $x(n)$ and $x(m)$ is

$$\rho(i) = \frac{C(n,m)}{\sqrt{Var\{x(n)\}}\sqrt{Var\{x(m)\}}} = \frac{C(i)}{C(0)} \tag{A7}$$

With the following figure and substituting (A7) into (A6), the variance of $\hat{\mu}$ becomes

$$
\begin{aligned}
Var\{\hat{\mu}\} &\overset{i=n-m}{=} \frac{1}{M^2}\sum_{n=0}^{M-1}\sum_{m=0}^{M-1} C(i) \\
&= \frac{1}{M^2}\sum_{i=-M+1}^{M-1} C(i)\left(M-|i|\right) = \frac{C(0)}{M}\sum_{i=-M+1}^{M-1}\rho(i)\left(1-\frac{|i|}{M}\right)
\end{aligned} \tag{A8}
$$



■ **Figure A1** The region of summation in Eq. (A8).

Thus, if

$$
\lim_{M\to\infty}\left[\sum_{i=-M+1}^{M-1}\rho(i)\left(1-\frac{|i|}{M}\right)\right] \tag{A9}
$$

is finite, then $Var\{\hat{\mu}\}\to 0 \;\; as \;\; M\to\infty$. A sufficient condition for this result to hold is that $\rho(i)\to 0 \;\; as \;\; i\to\infty$. This is so because $\rho(i)\to 0 \;\; as \;\; i\to\infty$ implies that for any $\varepsilon>0$, we can choose an *N* such that $|\rho(i)|<\varepsilon/4$ for all $i>N$. Hence, for $i>(N+1)$, we have

$$
\begin{aligned}
\left|\frac{1}{M}\sum_{i=-M+1}^{M-1}\rho(i)\right| &\le \frac{2}{M}\sum_{i=0}^{M-1}|\rho(i)| \\
&\le \frac{2}{M}\sum_{i=0}^{N}|\rho(i)| + \frac{2}{M}\sum_{i=N+1}^{M-1}|\rho(i)| \\
\left(\frac{2}{M}\sum_{i=N+1}^{M-1}|\rho(i)| \le \frac{2}{M}\cdot M\frac{\varepsilon}{4}\right) \quad &\le \frac{2}{M}\sum_{i=0}^{N}|\rho(i)| + \frac{1}{2}\varepsilon \\
\left(\begin{array}{l} M \text{ is large enough so that} \\ \dfrac{2}{M}\displaystyle\sum_{i=0}^{N}|\rho(i)| \le \dfrac{1}{2}\varepsilon \\ (\text{Cesáro mean}) \end{array}\right) \quad &\le \varepsilon
\end{aligned} \tag{A10}
$$

This shows that when $\rho(i) \to 0$ *as* $i \to \infty$, we have

$$\lim_{M\to\infty} \frac{1}{M} \sum_{i=-M+1}^{M-1} \rho(i) \to 0 \tag{A11}$$

which implies that $\lim_{M\to\infty} Var\{\hat{\mu}\} \to 0$.

Intuitively, the above results simply say that if values of the process sufficiently far apart are almost uncorrelated, then some useful new information can be continually added so that the time average will approach the ensemble average.

Note: Cesáro mean

Given the sequence $\{S_j\}_{j=1}^{\infty}$, the $\{C_n\}$ sequence defined by $C_n = \frac{1}{n}\sum_{j=1}^{n} S_j$. If the sequence $\{S_j\}$ converges to *s*, then the sequence $\{C_n\}$ converges to *s*.

Proof: Given $\varepsilon > 0$, we may choose an *N* such that $|S_i - s| < \frac{1}{2}\varepsilon$ for all *j* > *N*. For *n* > *N*, we have

$$\begin{aligned} \left|\frac{1}{n}\sum_{j=1}^{n} S_j - s\right| &\le \frac{1}{n}\sum_{j=1}^{n} |S_j - s| = \frac{1}{n}\sum_{j=1}^{N} |S_j - s| + \frac{1}{n}\sum_{j=N+1}^{n} |S_j - s| \\ &\le \frac{1}{n}\sum_{j=1}^{N} |S_j - s| + \frac{1}{2}\varepsilon \end{aligned} \tag{A12}$$

Since we can choose an *n* large enough so that the first term is less than $\frac{1}{2}\varepsilon$, the result follows.

■

Reference

[1] P. M. Clarkson, *Optimal and Adaptive Signal Processing*. Boca Raton, FL: CRC, 1993.