Mathematical Topics Embraced by Signal Processing

Carrson C. Fung

Dept. of Electronics Engineering

National Chiao Tung University

Examples of Mathematical Models

- Linear signal models for discrete and continuous time, including transfer function and state space representations. Applications of these models to SP problems such as prediction, spectrum estimation, and so on
- Adaptive filtering models and applications to prediction, system identification, and so forth
- The Gaussian random variable, and other probability density functions, including the important idea of conditioning upon an observation
- Hidden Markov models
  - Model the dynamics of systems probabilistically

Why is modeling important?

- Our world is complicated
  - To describe it mathematically requires complicated mathematics
    - E.g. high-order differential equation
  - E.g. suppose you are ask to design a filter $h[n]$ satisfying some design specifications such as transition bandwidth, passband frequency, stopband frequency, filter order, …
    - Hence, design is usually done in frequency using $H(e^{j\omega})$
      - How many points in $H(e^{j\omega})$ do you need to design?
      - This is an impossible problem to solve as there are uncountable number of points in $[0,\pi]$

Problem Specifications and Variable Parametrization

Suppose the desired response is

$$D(\omega)=\begin{cases}1, & 0\le\omega\le\omega_p\\ 0, & \omega_s\le\omega\le\pi\\ \text{don't care}, & \omega_p<\omega\le\omega_s\end{cases}$$

Change the variable from $H(e^{j\omega})$ to amplitude response $H_0(e^{j\omega})$

$$H_0(e^{j\omega})=\sum_{n=0}^{N-1}h[n]e^{-j(n-M)\omega}$$
$$=\sum_{n=0}^{M}b[n]\cos n\omega,\quad M=\frac{(N-1)}{2}$$

assuming Type-I linear phase and

$$b[n]=\begin{cases}2h\left[\frac{(N-1)}{2}-n\right], & n\ne 0\\ h\left[\frac{(N-1)}{2}\right], & n=0\end{cases}$$

Problem Formulation

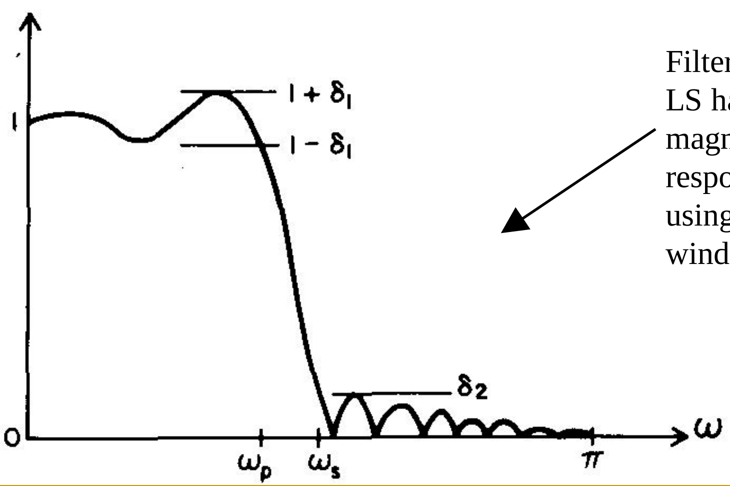
Then the filter design problem can be formulated as a LS problem

$$\min_{b[n]} \int_R \left[D(\omega) - H_0\left(e^{j\omega}\right) \right]^2 \frac{d\omega}{\pi},$$

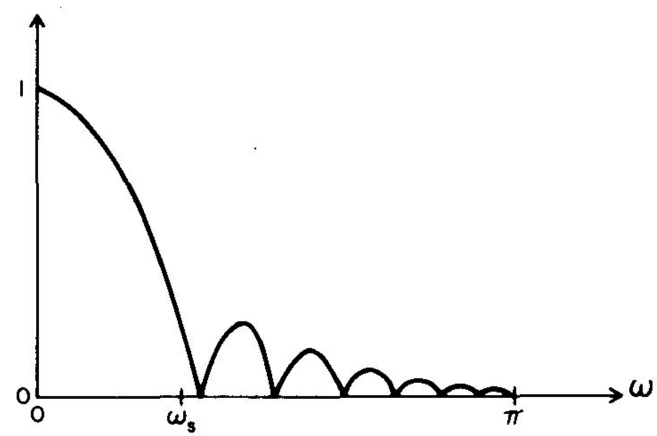
$R:\ 0 \le \omega \le \pi$, but excluding transition band.

Integration can be approximated by summation.

Now problem only needs to solve a finite number of variables.

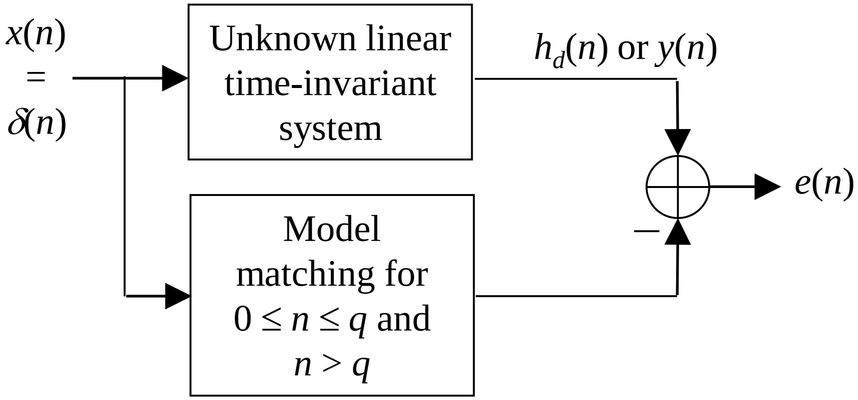


Filter designed using LS has much better magnitude frequency response than one using Kaiser windowing method



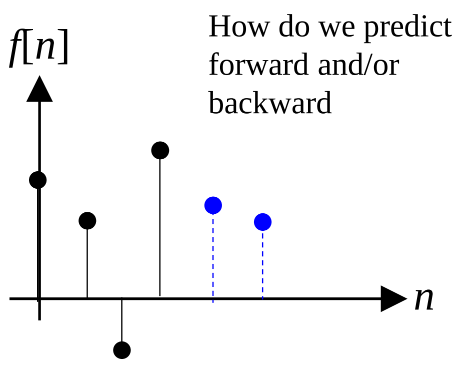
Other Motivations for Using Mathematics

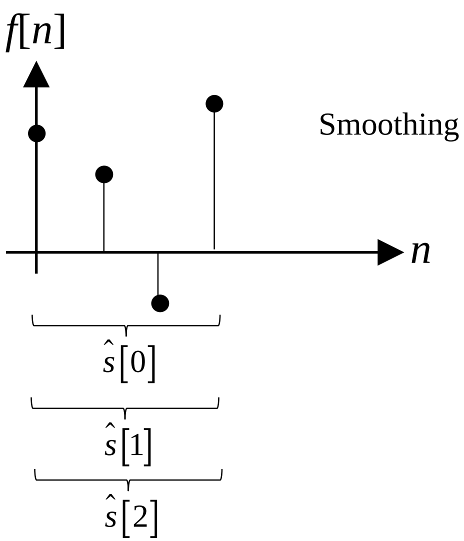
- Given a sequence of output data from a system, how can the parameters of the system be determined if the input signal is known
  - What if the input signal is not known?
  - What if system is nonlinear?



Other Motivations for Using Mathematics

- Determine a "minimal" representation of a system
- Given a signal from a system, determine a predictor for the signal
  - Forward and/or backward
- Determine an optimal and/or efficient smoothing method
  - E.g. Image smoothing
- Determine a means of efficiently coding (representing) a signal modeled as the output of an LTI system
- Develop computational efficient algorithms
- Develop adaptive technique to obtain desirable output of system





Complex-Valued Linear Discrete-Time Models: ARMA and MA

Autoregressive moving average (ARMA) model

$$y[n] = -a_1^* y[n-1] - a_1^* y[n-2] - \cdots - a_p^* y[n-p]$$
$$+ b_0^* f[n] + b_1^* f[n-1] + \cdots + b_q^* f[n-q]$$

$$\Leftrightarrow \sum_{k=0}^{p} a_k^* y[n-k] = \sum_{k=0}^{q} b_k^* f[n-k]$$

Moving average (MA) model

$$y[n] = b_0^* f[n] + b_1^* f[n-1] + \cdots + b_q^* f[n-q]$$

$$\Leftrightarrow y[n] = \sum_{k=0}^{q} b_k^* f[n-k]$$

Vector notation

Define $\mathbf{f}[n] = \begin{bmatrix} f[n] \\ f[n-1] \\ \vdots \\ f[n-q] \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_q \end{bmatrix}$

$$\Rightarrow y[n] = \mathbf{b}^H \mathbf{f}[n]$$

Complex-Valued Linear Discrete-Time Models: AR

Autoregressive (AR) model

$$y[n] = -a_1^* y[n-1] - a_1^* y[n-2] - \cdots - a_p^* y[n-p] + b_0^* f[n]$$

$$\Leftrightarrow y[n] = b_0^* f[n] - \sum_{k=1}^{p} a_k^* y[n-k]$$

Define

$$\mathbf{y}[n] = \begin{bmatrix} y[n-1] \\ y[n-2] \\ \vdots \\ y[n-p] \end{bmatrix} \text{ and } \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

$$\Leftrightarrow y[n] = b_0^* f[n] - \mathbf{a}^H \mathbf{y}[n]$$

System Function and Impulse Response

Assuming initial conditions are zero

$$\sum_{k=0}^{p} a_k^* z^{-k} Y(z) = \sum_{k=0}^{q} b_k^* z^{-k} F(z) \quad \Leftrightarrow \quad Y(z)A(z) = F(z)B(z)$$

ARMA System function

$$H(z) = \frac{Y(z)}{F(z)} = \frac{\sum_{k=0}^{q} b_k^* z^{-k}}{\sum_{k=0}^{p} a_k^* z^{-k}} = \frac{\sum_{k=0}^{q} b_k^* z^{-k}}{1 + \sum_{k=1}^{p} a_k^* z^{-k}} = \frac{B(z)}{A(z)}$$

(usually assume system is normalized so that $a_0 = 1$)

All-pole System function (IIR system)

$$H(z) = \frac{Y(z)}{F(z)} = \frac{\sum_{k=0}^{q} b_k^* z^{-k}}{1 + \sum_{k=1}^{p} a_k^* z^{-k}} = \frac{b_0^*}{A(z)}$$

All-zero system function (FIR system)

$$H(z) = \frac{Y(z)}{F(z)} = \sum_{k=0}^{q} b_k^* z^{-k} = B(z)$$

System Function and Impulse Response

$$H(z) = \sum_k f[k]h[n-k]$$

Factoring $H(z)$ into monomial factors using roots of numerator and denominator

$$H(z) = \frac{b_0^* \prod_{k=1}^{q} 1 - z_i z^{-1}}{\prod_{k=1}^{p} 1 - p_i z^{-1}} = \frac{B(z)}{A(z)}$$

Stochastic MA and AR Models

$f[n]$: assumed to be a white discrete-time random process, usually zero mean

b_0 : set to 1, with input power determined by the variance of the signal

$$E\left(f[n]\right) = 0, \quad \forall n$$

$$E\left(f[m]\,f^*[n]\right) = \begin{cases} \sigma_{ff}^2, & m = n \\ 0, & \text{otherwise} \end{cases}$$

SP often involves comparing two signals, one way for comparison is by correlation. When the signal is comparing with itself, the correlation is called autocorrelation function. For zero-mean WSS signal $y[n]$,

$r_{yy}[\ell - k] \triangleq E\left(y[n-k]\,y^*[n-\ell]\right)$ or $r_{yy}[k] = E\left(y[n]\,y^*[n-k]\right)$

(Note the convention: first argument minus second)

Autocorrelation Function

Note: $r_{yy}[k] = r_{yy}^*[-k]$ (more details later)

For real-valued random process, $r_{yy}[k] = r_{yy}[-k]$ (even function)

For MA process

$$y[n] = f[n] + b_1^* f[n-1] + \cdots + b_q^* f[n-q]$$

$$\Rightarrow r_{yy}[k] = E\left(y[n]y^*[n-k]\right)$$

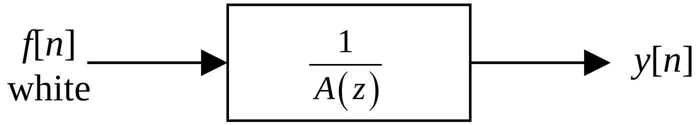
$$= E\left[\left(f[n] + b_1^* f[n-1] + \cdots + b_q^* f[n-q]\right)\left(f^*[n-k] + b_1 f^*[n-1-k] + \cdots + b_q f^*[n-q-k]\right)\right]$$

$$= r_{ff}[k] + |b_1|^2 r_{ff}[k] + \cdots + |b_q|^2 r_{ff}[k] = \sigma_{ff}^2 \sum_{k=1}^{q} |b_k|^2$$

For AR process

$$y[n] + a_1^* y[n-1] + \cdots + a_p^* y[n-p] = f[n]$$

Autocorrelation Function



For AR process

$$y[n]+a_1^* y[n-1]+\cdots+a_p^* y[n-p]=f[n]$$

Multply by $y^*[n-\ell]$ on both sides and take expectation:

$$E\left(\sum_{k=0}^{p} a_k^* y[n-k]\, y^*[n-\ell]\right)=\sum_{k=0}^{p} a_k^* r_{yy}[\ell-k]=E\left(f[n]\,y^*[n-\ell]\right)$$

$$=\begin{cases} r_{fy}[\ell], & \text{for } \ell=0 \\ 0, & \text{for } \ell>0 \end{cases}$$

(0 for $\ell>0$ because $f[n]$ is white-noise process)

For $\ell>0$

$$0=r_{fy}[\ell]=E\left[\left(y[n]+a_1^* y[n-1]+\cdots+a_p^* y[n-p]\right)y^*[n-\ell]\right]$$

$$=r_{yy}[\ell]+a_1^* r_{yy}[\ell-1]+\cdots+a_p^* r_{yy}[\ell-p]$$

$$\Rightarrow r_{yy}[\ell]=-a_1^* r_{yy}[\ell-1]-\cdots-a_p^* r_{yy}[\ell-p]$$

Yule-Walker Equations: Solving System ID Problem

$$r_{yy}[\ell] = -a_1^* r_{yy}[\ell-1] - \cdots - a_p^* r_{yy}[\ell-p]$$

Stacking $\ell = 1, 2, \ldots, p$ equations, we have

$$\begin{bmatrix} r_{yy}[0] & r_{yy}[-1] & \cdots & r_{yy}[-(p-1)] \\ r_{yy}[1] & r_{yy}[0] & \cdots & r_{yy}[-(p-2)] \\ \vdots & \vdots & \ddots & \vdots \\ r_{yy}[p-1] & r_{yy}[p-2] & \cdots & r_{yy}[0] \end{bmatrix} \begin{bmatrix} -a_1^* \\ -a_2^* \\ \vdots \\ -a_p^* \end{bmatrix} = \begin{bmatrix} r_{yy}[1] \\ r_{yy}[2] \\ \vdots \\ r_{yy}[p] \end{bmatrix}$$

Conjugating both sides:

$$\mathbf{w} \triangleq \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_p \end{bmatrix}, \quad \mathbf{r} \triangleq \begin{bmatrix} r_{yy}^*[1] \\ r_{yy}^*[2] \\ \vdots \\ r_{yy}^*[p] \end{bmatrix}$$

$$\begin{bmatrix} r_{yy}^*[0] & r_{yy}^*[-1] & \cdots & r_{yy}^*[-(p-1)] \\ r_{yy}^*[1] & r_{yy}^*[0] & \cdots & r_{yy}^*[-(p-2)] \\ \vdots & \vdots & \ddots & \vdots \\ r_{yy}^*[p-1] & r_{yy}^*[p-2] & \cdots & r_{yy}^*[0] \end{bmatrix} \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_p \end{bmatrix} = \begin{bmatrix} r_{yy}^*[1] \\ r_{yy}^*[2] \\ \vdots \\ r_{yy}^*[p] \end{bmatrix} \Leftrightarrow \mathbf{R}\mathbf{w} = \mathbf{r}$$

Observations about YW Equations

- $\mathbf{R}=\mathbf{R}^H$
  - Eigenvalues are real and eigenvectors corresponding to distinct eigenvalues are orthogonal/orthonormal. If **R** is real, then R is symmetric, i.e. $\mathbf{R}^T = \mathbf{R}$
- **R** is a Toeplitz matrix, i.e. $r_{ij} = r_{i-j}$
  - Values of **R** depend only on the difference between the index values
    - Has efficient algorithm to solve for solution
      - Power efficient in hardware implementation

Realization

A controller canonical form (from control) can be written by realizing that the transfer function can be written as

$$H(z)=\frac{\overbrace{Y(z)}^{B(z)}}{W(z)}\frac{W(z)}{F(z)}=\left(\sum_{k=0}^{q}b_k^* z^{-k}\right)\left(\frac{1}{\left(1+\sum_{k=1}^{p}a_k^* z^{-k}\right)}\right)=B(z)H_2(z)=H_1(z)H_2(z).$$

Since $W(z)\left(1+\sum_{k=1}^{p}a_k^* z^{-k}\right)=F(z)\Leftrightarrow w[n]+a_1^* w[n-1]+\cdots+a_p^* w[n-p]=f[n]$ or

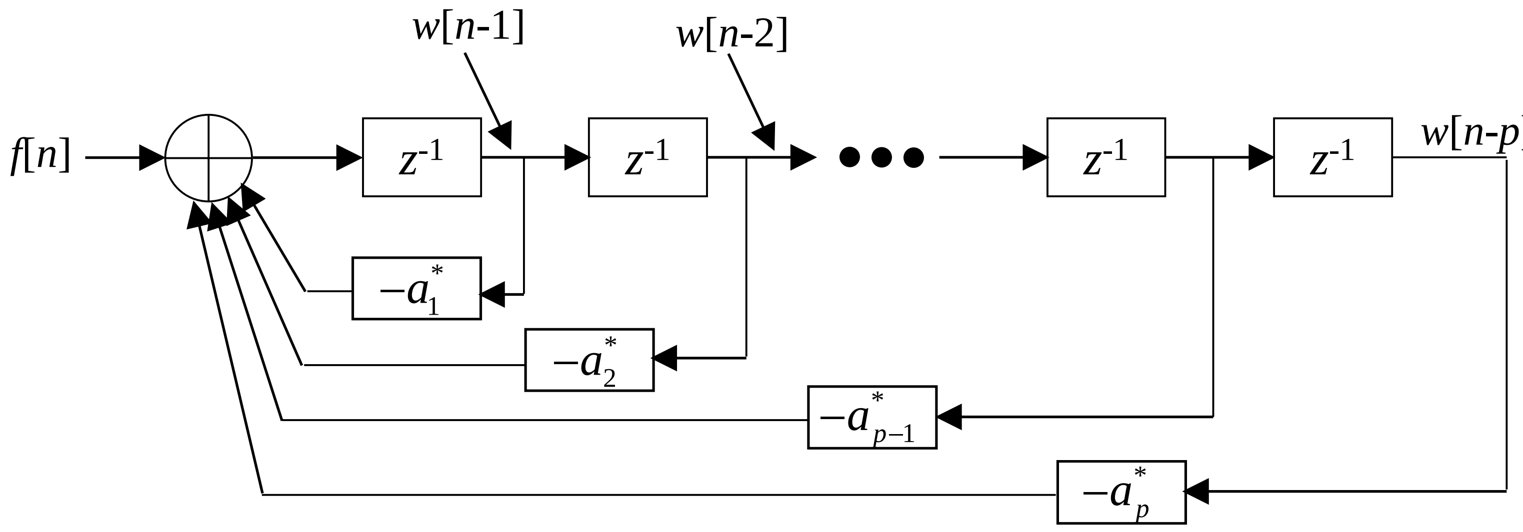
$$\Rightarrow w[n]=f[n]-a_1^* w[n-1]-\cdots-a_p^* w[n-p]$$

Since $B(z)=\frac{Y(z)}{W(z)} \quad\Rightarrow\quad Y(z)=W(z)B(z)$

$$\Leftrightarrow\; y[n]=w[n]*b[n]=b_0^* w[n]+b_1^* w[n-1]+\cdots+b_q^* w[n-q]$$

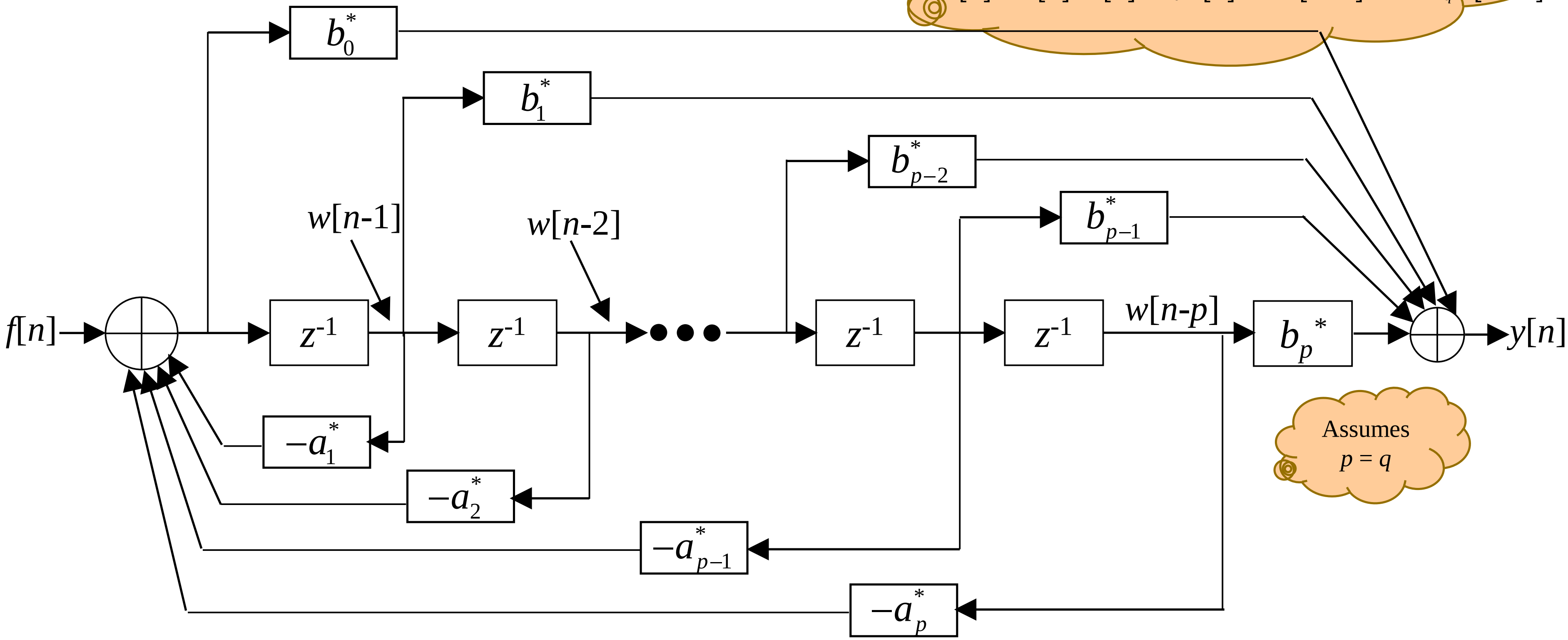
Realization: AR part of Transfer Function

$$w[n]+a_1^* w[n-1]+\cdots+a_p^* w[n-p]=f[n]$$



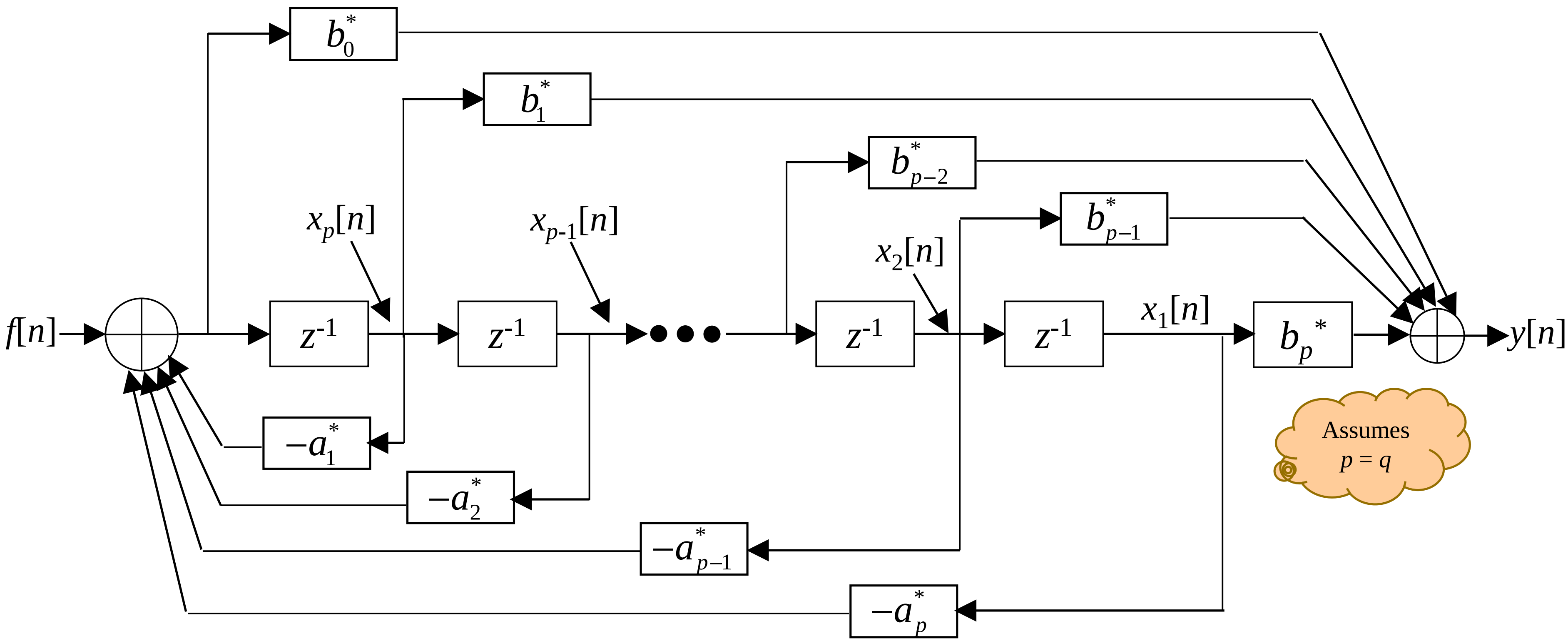
Realization of Complete Transfer Function

$$w[n]+a_1^* w[n-1]+\cdots+a_p^* w[n-p]=f[n]$$
$$y[n]=w[n]*b[n]=b_0^* w[n]+b_1^* w[n-1]+\cdots+b_q^* w[n-q]$$



- Signal processing practitioners usually attempt to analyze characteristics of a system by **ONLY** looking at the relationship between the input and output
  - Transfer function
- Imagine opening your system (a black box), which can now be modeled using a bunch of integrators (delay elements in discrete time) and putting a logic probe in each of the interconnect
  - Concatenation of these signals $\{w[n\text{-}k]\}$, $\forall k$ makes up the state of the system

State-Space Form



Consider relabeling the interconnect signals (states) as $\{x_k[n]\}$, for $k = 1, 2, \ldots, p$

State-Space Representation

$x_k[n]$'s are known as the state variables. Note that the transfer function can be written as

$$H(z)=\frac{Y(z)}{F(z)}=\frac{Y(z)}{X(z)}\frac{X(z)}{F(z)}=\left(\sum_{k=0}^{q}b_k z^{-k}\right)\left(\frac{1}{a_0+\sum_{k=1}^{p}a_k z^{-k}}\right)=H_1(z)H_2(z)$$

Assuming $p=q$, note that

$$x_1[n+1]=x_2[n]$$

$$x_2[n+1]=x_3[n]$$

$$\vdots$$

$$x_{p-1}[n+1]=x_p[n]$$

$$\Rightarrow\begin{cases} x_p[n+1]=f[n]-a_1^* x_p[n]-a_2^* x_{p-1}[n]-\cdots-a_{p-1}^* x_2[n]-a_p^* x_1[n] & \text{(state equation)} \\ y[n]=b_p^* x_1[n]+b_{p-1}^* x_2[n]+\cdots+b_2^* x_{p-1}[n]+b_1^* x_p[n] & \text{(input-output equation)} \\ \quad +b_0^*\left(f[n]-a_1^* x_p[n]-a_2^* x_{p-1}[n]-\cdots-a_p^* x_1[n]\right) & \end{cases}$$

State-Space Representation

Define the state vector $\mathbf{x}[n] \triangleq \begin{bmatrix} x_1[n] \\ \\ x_p[n] \end{bmatrix}$, containing state variables $x_k[n]$, $\forall k$,

$$\mathbf{b} \triangleq \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{c} \triangleq \begin{bmatrix} b_p^* - b_0^* a_p^* \\ b_{p-1}^* - b_0^* a_{p-1}^* \\ \vdots \\ b_1^* - b_0^* a_1^* \end{bmatrix}, \quad d \triangleq b_0^*, \quad \text{and } \mathbf{A} \triangleq \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ -a_p^* & -a_{p-1}^* & -a_{p-2}^* & -a_{p-3}^* & \cdots & -a_2^* & -a_1^* \end{bmatrix}$$

If $b_0 = 0$, $\mathbf{c}^T = \begin{bmatrix} b_p^* & b_{p-1}^* & & b_1^* \end{bmatrix}$, then

A is called a companion matrix

$$\Rightarrow \begin{cases} \mathbf{x}[n+1] = \mathbf{A}\mathbf{x}[n] + \mathbf{b}f[n] \\ y[n] = \mathbf{c}^T\mathbf{x}[n] + df[n] \end{cases}$$

State-space equation

Imagine opening your system (a black box), which can now be modeled using a bunch of integrators (delay elements in discrete time) and putting a logic probe in each of the interconnect

- Concatenation of these signals $\{x_k[n]\}$, $\forall k$ makes up the state of the system

Non-uniqueness of State-Space Equation

Let $\mathbf{x} = \mathbf{T}\mathbf{z}, \ \mathbf{T}: \ p \times p$ invertible matrix, then

$$\begin{aligned} \mathbf{T}\mathbf{z}[n+1] &= \mathbf{A}\mathbf{T}\mathbf{z}[n] + \mathbf{b}f[n] \\ y[n] &= \mathbf{c}^T\mathbf{T}\mathbf{z}[n] + df[n] \end{aligned} \Rightarrow \begin{aligned} \mathbf{z}[n+1] &= \mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{z}[n] + \mathbf{T}^{-1}\mathbf{b}f[n] \\ y[n] &= \mathbf{c}^T\mathbf{T}\mathbf{z}[n] + df[n] \end{aligned}$$

Terminologies (which will be explained later)

$\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ is a similarity transformation of $\mathbf{A}$, they share identical eigenvalues

Time-varying State-Space Model

When system is time-varying, the state-space representation becomes

$$\mathbf{x}[n+1] = \mathbf{A}[n]\mathbf{x}[n] + \mathbf{b}[n] f[n]$$
$$y[n] = \mathbf{c}^T[n]\mathbf{x}[n] + d[n] f[n]$$

so $\left(\mathbf{A}[n], \mathbf{b}[n], \mathbf{c}^T[n], d[n]\right)$ on the time index n is shown

Transformed State-Space Model

Taking the z-transform of the time-invariant SS model

$$z\mathbf{x}(z) = \mathbf{A}\mathbf{x}(z) + \mathbf{b}F(z)$$

$$Y(z) = \mathbf{c}^T\mathbf{x}(z) + dF(z)$$

Then the state equation becomes

$$(z\mathbf{I}_p - \mathbf{A})\mathbf{x}(z) = \mathbf{b}F(z)$$

$$\Rightarrow \mathbf{x}(z) = (z\mathbf{I}_p - \mathbf{A})^{-1}\mathbf{b}F(z).$$

Substituting, then the output equation

$$Y(z) = \mathbf{c}^T(z\mathbf{I}_p - \mathbf{A})^{-1}\mathbf{b}F(z) + dF(z)$$

$$= \left[\mathbf{c}^T(z\mathbf{I}_p - \mathbf{A})^{-1}\mathbf{b} + d\right]F(z).$$

Then the transfer function becomes

$$H(z) = \frac{Y(z)}{F(z)} = \mathbf{c}^T(z\mathbf{I}_p - \mathbf{A})^{-1}\mathbf{b} + d$$

Solution for State-Space Difference Equation

Recall the state-space difference equation

$$\mathbf{x}[n+1] = \mathbf{A}\mathbf{x}[n] + \mathbf{b}f[n]$$
$$y[n] = \mathbf{c}^T\mathbf{x}[n] + df[n]$$

Also initial condition $\mathbf{x}[-1]$, and for $n \geq 0$,

$$\mathbf{x}[0] = \mathbf{A}\mathbf{x}[-1] + \mathbf{b}f[0]$$

$$\begin{aligned}\mathbf{x}[1] &= \mathbf{A}\mathbf{x}[0] + \mathbf{b}f[1] \\ &= \mathbf{A}\left(\mathbf{A}\mathbf{x}[-1] + \mathbf{b}f[0]\right) + \mathbf{b}f[1] \\ &= \mathbf{A}^2\mathbf{x}[-1] + \mathbf{A}\mathbf{b}f[0] + \mathbf{b}f[1]\end{aligned}$$

$$\mathbf{x}[n] = \mathbf{A}^{n+1}\mathbf{x}[-1] + \sum_{k=0}^{n}\mathbf{A}^k\mathbf{b}f[n-k]$$

Solution for State-Space Difference Equation

$$y[n] = \mathbf{c}^T \mathbf{A}^{n+1}\mathbf{x}[-1] + \sum_{k=0}^{n} \mathbf{c}^T \mathbf{A}^k \mathbf{b} f[n-k] + df[n]$$

Quantities of $\mathbf{c}^T \mathbf{A}^k \mathbf{b}$ are known as the Markov parameters of the system.

- Note: $\mathbf{x}[n]$ is a linear function of $\mathbf{x}[-1]$ and $f[n-k]$, so it is also a Gaussian process (more on random process later)

State-Space Model: MIMO Extension

MIMO extension:

$$\mathbf{x}[n+1] = \mathbf{A}\mathbf{x}[n] + \mathbf{B}\mathbf{u}[n]$$

$$\mathbf{y}[n] = \mathbf{C}\mathbf{x}[n] + \mathbf{D}\mathbf{u}[n]$$

If there are p state variables and ℓ inputs and m outputs, then

$$\mathbf{A}: \; p \times p, \quad \mathbf{B}: \; p \times \ell, \quad \mathbf{C}: \; m \times p, \quad \mathbf{D}: \; m \times \ell$$

Simple algebra will show that

$$\mathbf{x}[n] = \mathbf{A}^{n+1}\mathbf{x}[-1] + \sum_{k=0}^{n} \mathbf{A}^{k}\mathbf{B}\mathbf{u}[k]$$

$$\mathbf{y}[n] = \mathbf{C}\mathbf{A}^{n+1}\mathbf{x}[-1] + \sum_{k=0}^{n} \mathbf{C}\mathbf{A}^{k}\mathbf{B}\mathbf{u}[k] + \mathbf{D}\mathbf{u}[n]$$

Quantities of $\mathbf{C}\mathbf{A}^{k}\mathbf{B}$ are known as the Markov parameters of the system.

State Equation Example: Two DC Power Supplies

Assume outputs are independent of each othen, then a reasonable model would be the scalar model for each output

$$x_1[n] = a_1 x_1[n-1] + u_1[n]$$
$$x_2[n] = a_2 x_2[n-1] + u_2[n],$$

where $x_1[-1] \sim \mathcal{N}\left(\mu_{x_1}, \sigma_{x_1}^2\right)$, $x_2[-1] \sim \mathcal{N}\left(\mu_{x_2}, \sigma_{x_2}^2\right)$, $u_1[n]$ and $u_2[n]$ are zero-mean WGN with variance $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$, respectively. All RVs are independent of each other.

Then $\begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} x_1[n-1] \\ x_2[n-1] \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1[n] \\ u_2[n] \end{bmatrix} \Leftrightarrow \mathbf{x}[n] = \mathbf{A}\mathbf{x}[n-1] + \mathbf{B}\mathbf{u}[n].$

Also, since $\mathbf{u}[n]$ is a vector WGN with zero mean and covariance

$$E\left(\mathbf{u}[m]\mathbf{u}^T[n]\right) = \begin{bmatrix} E\left(u_1[m]u_1[n]\right) & E\left(u_1[m]u_2[n]\right) \\ E\left(u_2[m]u_1[n]\right) & E\left(u_2[m]u_2[n]\right) \end{bmatrix} = \begin{bmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_2}^2 \end{bmatrix} \delta[m-n],$$

so $\mathbf{Q} = \begin{bmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_2}^2 \end{bmatrix}$ and $\mathbf{x}[-1] = \begin{bmatrix} x_1[-1] \\ x_2[-1] \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \end{bmatrix}, \begin{bmatrix} \sigma_{x_1}^2 & 0 \\ 0 & \sigma_{x_2}^2 \end{bmatrix} \right).$

State Equation Example: Vehicle Tracking

- Goal: Estimate and track range and bearing of vehicle (assuming $x-y$ Cartesian coordinates)
- Assume constant velocity, perturbed by only wind gusts, slight speed corrections
- Model these perturbations as noise inputs, leading to velocity equations

$$v_x[n] = v_x[n-1] + u_x[n]$$
$$v_y[n] = v_y[n-1] + u_y[n].$$

- Note that without the noise perturbations $u_x[n]$ and $u_y[n]$, the velocities would be constant, and the vehicle would be modeled as traveling in a straight line as indicated by the dashed line in Fig. 13.21
- The position equation at time n can then be written as

$$r_x[n] = r_x[n-1] + v_x[n-1]\Delta$$
$$r_y[n] = r_y[n-1] + v_y[n-1]\Delta,$$

where Δ is the sampling period.

- The (discrete-time) velocity equations models the vehicle to be traveling at the velocity at $n-1$ and then changing abruptly at n. This is an approximation to the true continuous behavior

State Equation Example: Vehicle Tracking

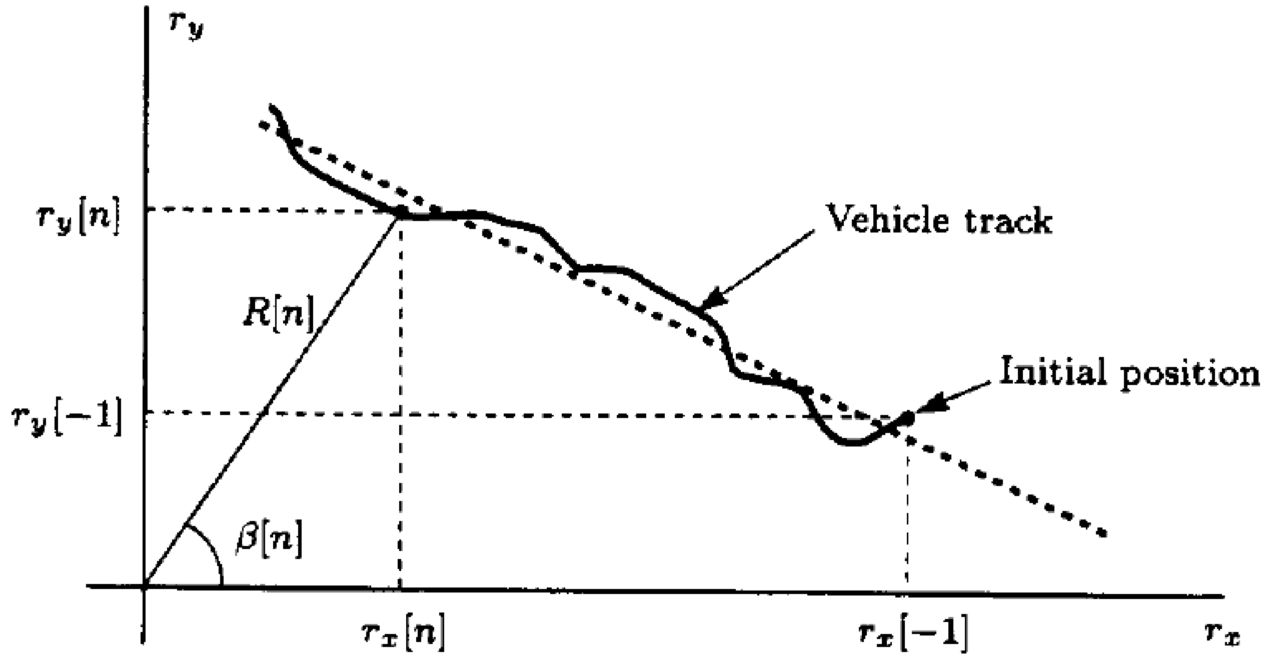


Figure 13.21 Typical track of vehicle moving in given direction at constant speed

State Equation Example: Vehicle Tracking

Define the signal vector as $\mathbf{x}[n] = \begin{bmatrix} r_x[n] \\ r_y[n] \\ v_x[n] \\ v_y[n] \end{bmatrix}$ from the velocity and position equations, we see that

$$\begin{bmatrix} r_x[n] \\ r_y[n] \\ v_x[n] \\ v_y[n] \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_x[n-1] \\ r_y[n-1] \\ v_x[n-1] \\ v_y[n-1] \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ u_x[n] \\ u_y[n] \end{bmatrix}$$

$$\Leftrightarrow \mathbf{x}[n] = \mathbf{A}\mathbf{x}[n-1] + \mathbf{u}[n].$$

The measurments are noisy observations of the range and bearing

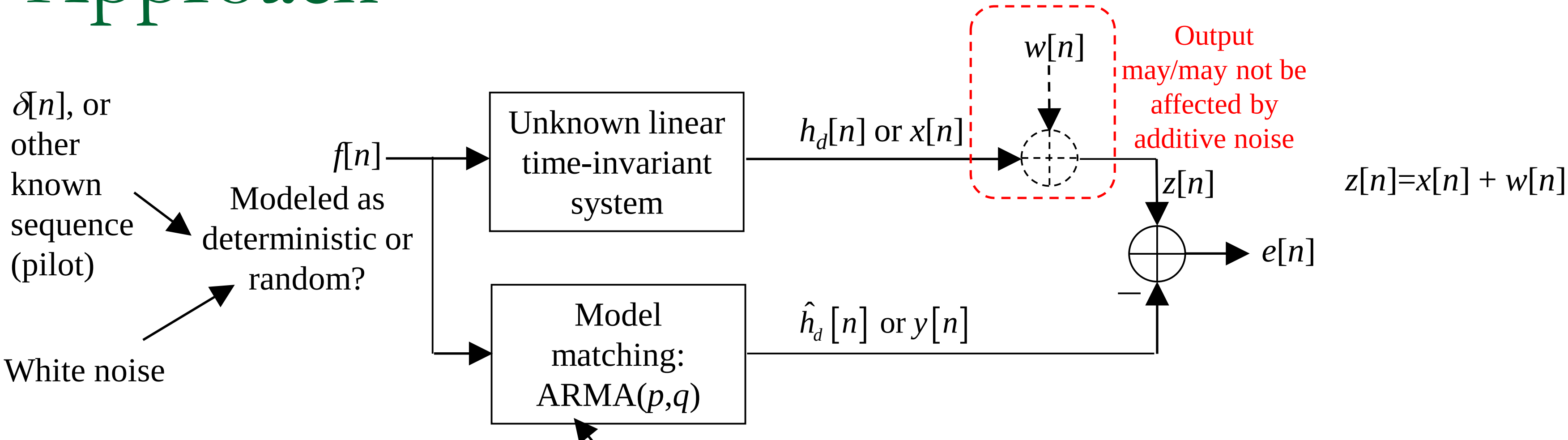
$$\hat{R}[n] = R[n] + w_R[n]$$

$$\hat{\beta}[n] = \beta[n] + w_\beta[n].$$

This can be written in general form as $\mathbf{y}[n] = \mathbf{h}(\mathbf{x}[n]) + \mathbf{w}[n],$ where

$$\mathbf{C}\mathbf{x}[n] \Rightarrow \mathbf{h}(\mathbf{x}[n]) = \begin{bmatrix} \sqrt{r_x^2[n] + r_y^2[n]} \\ \arctan \frac{r_y[n]}{r_x[n]} \end{bmatrix}$$

Example: System Estimation: One LS Approach



Using ARMA model to model the unknown system:

$$\hat{h}_d[n] \text{ or } y[n] = -a_1^* y[n-1] - a_1^* y[n-2] - \cdots - a_p^* y[n-p] + b_0^* f[n] + b_1^* f[n-1] + \cdots + b_q^* f[n-q]$$

$$\Leftrightarrow y[n] = -\sum_{k=1}^{p} a_k^* y[n-k] + \sum_{k=0}^{q} b_k^* f[n-k]$$

Assuming:

- $f[n]$ is known
- System: ARMA(p,q)

$\Rightarrow$ can setup equation $\mathbf{Ax} = \mathbf{b}$ to solve for parameters

Example: System Estimation: One LS Approach

Recall $y[n] = -\sum_{k=1}^{p} a_k^* y[n-k] + \sum_{k=0}^{q} b_k^* f[n-k]$

To ensure we deal with a causal $y[n]$

$$\mathbf{A} = \begin{bmatrix} y[p-1] & y[p-2] & \cdots & y[0] & f[p] & f[p-1] & \cdots & f[p-q] \\ y[p] & y[p-1] & \cdots & y[1] & f[p+1] & f[p] & \cdots & f[p+1-q] \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y[N-1] & y[N-2] & \cdots & y[N-p] & f[N] & f[N-1] & \cdots & f[N-q] \end{bmatrix} \begin{matrix} n=p \\ \\ \\ n=N \end{matrix}$$

$$\mathbf{x} = \begin{bmatrix} -a_1^* \\ -a_2^* \\ \vdots \\ -a_p^* \\ b_0^* \\ b_1^* \\ \vdots \\ b_q^* \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} z[p] \\ z[p+1] \\ \vdots \\ z[N] \end{bmatrix}$$

If N large $\Rightarrow$ over-determined system $\Rightarrow$ LS solution possible

E.g. Linear Prediction (Useful for Speech Coding and Recognition)

Assume we are told of an $\text{AR}(p)$ system $H(z) = \dfrac{1}{1+\sum_{k=1}^{p} a_k z^{-k}}$

Speech is often modeled as output of such system driven by either a zero-mean uncorrelated signal in the case of unvoiced speech (such as "f", "s" known as fricatives) or by a periodic pulse sequence in the case of voiced speech (vowels) due to the "peaky" nature of human speech signal (in time).

From $H(z)$ $\quad\Rightarrow y[n] = -\sum_{k=1}^{p} a_k y[n-k] + f[n] = -\mathbf{a}^T \mathbf{y}[n-1] + f[n] = -\mathbf{a}_a^T \mathbf{y}[n]$

$$\mathbf{a}_a \triangleq \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \text{ and } \mathbf{y}[n] \triangleq \begin{bmatrix} f[n] \\ y[n-1] \\ y[n-2] \\ \vdots \\ y[n-p] \end{bmatrix}$$

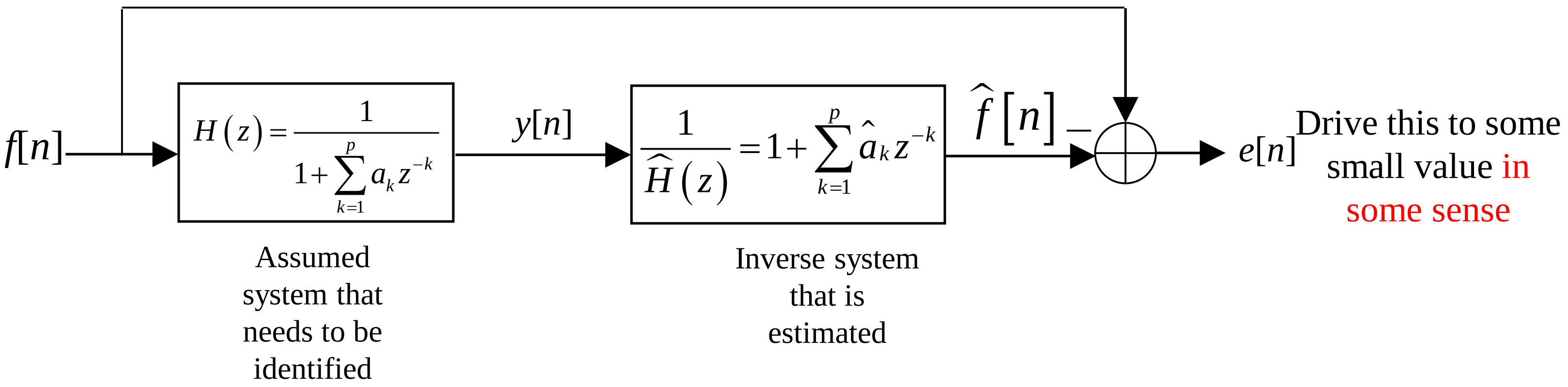
$\Rightarrow \hat{h}_d[n] = y[n] = -\hat{\mathbf{a}}_a^T \mathbf{y}[n]$

Goal is to find $\hat{\mathbf{a}}_a^T$ or $\hat{\mathbf{a}}$ so that $e[n] = z[n] - \hat{y}[n]$ is minimized

Application for Speech Recognition (big data example)

- Suppose there are several classes of signals to be distinguished (for example, several speech sounds to be recognized).
- Each signal will have its own set of prediction coefficients
  - Signal 1 has $\mathbf{a}_1$
  - Signal 2 has $\mathbf{a}_2$, …
- An unknown input signal can be reduced (by estimating the prediction coefficients that represent it) to a vector $\mathbf{a}$
  - Then $\mathbf{a}$ can be compared with $\mathbf{a}_1$, $\mathbf{a}_2$, and so on…to determine which signal the unknown input is most similar to

Inverse Problem: Another Perspective of Prediction



Inverse Problem: Another Perspective of Prediction

If $Y(z) = H(z)F(z) \quad \Rightarrow \quad F(z) = Y(z)\dfrac{1}{H(z)}$

$\Rightarrow f[n] = y[n] + \mathbf{a}^T \mathbf{y}[n-1]$

In this case, $y[n]$ is regarded as input, then $f[n]$ is output of an inverse system.

If we have an estimated system

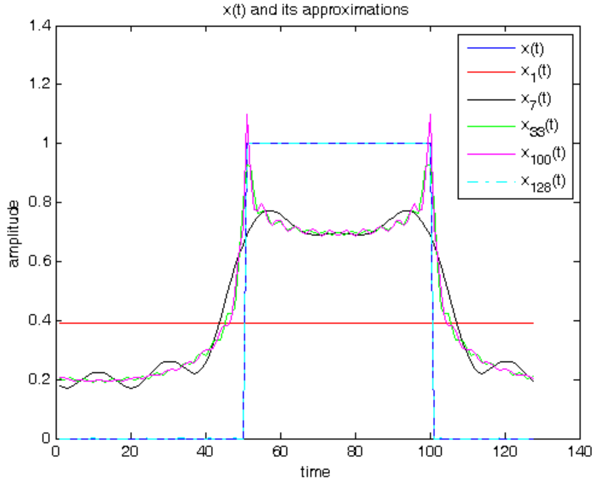
$$H(z) = \frac{1}{1 + \sum_{k=1}^{p} a_k z^{-k}} = \frac{Y(z)}{F(z)}$$

then choose $\qquad \widehat{f}[n] = y[n] + \mathbf{a}^T \mathbf{y}[n-1]$

so that is close to $f[n]$ in some sense. This is known as an inverse problem.

Nonparametric Spectrum Analysis

- From DSP, we know we can perform DFT on "any" signals to get a picture of the spectrum



"Analysis" Equation

$$X[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}$$

Why these equations are written this way?

"Synthesis" Equation

$$x[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k] e^{j\frac{2\pi}{N}kn}$$

- Not very accurate
  - Exploiting a priori knowledge of signal is better

Parameter Fourier Analysis

Assume we know the signal $s[n] = a\cos 2\pi f_0 n + b\sin 2\pi f_0 n,$ for $n = 0,1,\ldots,N-1,$ where $f_0 = k/N,$ with $k = 1,\ldots,N/2-1.$ Estimate $\boldsymbol{\theta} = [a \quad b]^T$.

$$\begin{bmatrix} s[0] \\ s[1] \\ \vdots \\ s[N-1] \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ \cos 2\pi f_0 & \sin 2\pi f_0 \\ \vdots & \vdots \\ \cos 2\pi f_0 (N-1) & \sin 2\pi f_0 (N-1) \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

It can shown that $\mathbf{H}$ is orthogonal, i.e.

$$\mathbf{h}_1^T\mathbf{h}_2 = \sum_{n=0}^{N-1}\cos\left(2\pi\frac{k}{N}n\right)\sin\left(2\pi\frac{k}{N}n\right) = 0$$

$$\mathbf{h}_1^T\mathbf{h}_1 = \mathbf{h}_2^T\mathbf{h}_2 = \frac{N}{2}$$

$$\Rightarrow \mathbf{H}^T\mathbf{H} = (N/2)\mathbf{I}_p$$

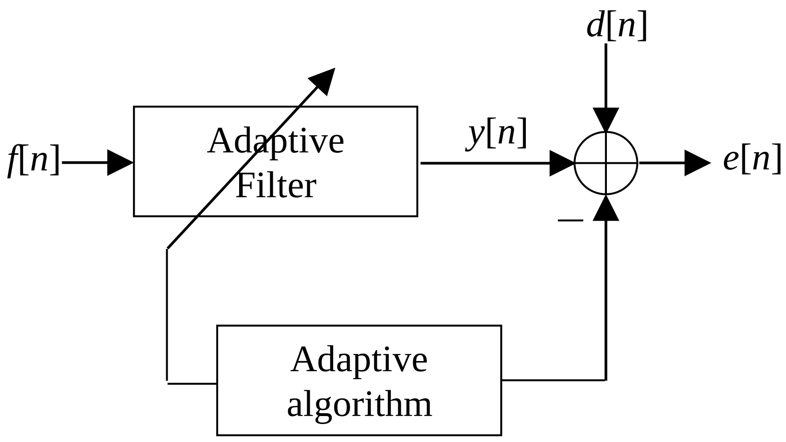
Parametric Fourier Analysis

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

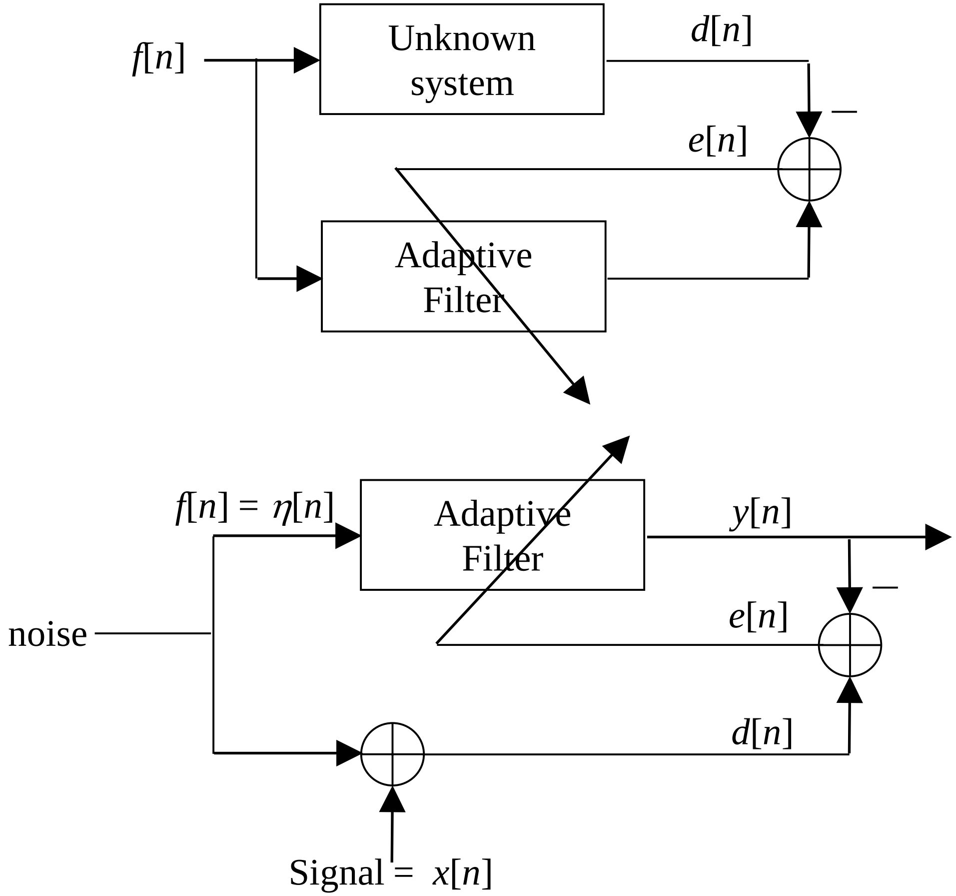
$$= \frac{2}{N}\mathbf{H}^T\mathbf{x} = \begin{bmatrix} \frac{2}{N}\sum_{n=0}^{N-1} x[n]\cos\left(2\pi\frac{k}{N}n\right) \\ \frac{2}{N}\sum_{n=0}^{N-1} x[n]\sin\left(2\pi\frac{k}{N}n\right) \end{bmatrix}$$

General Adaptive Filter Configuration

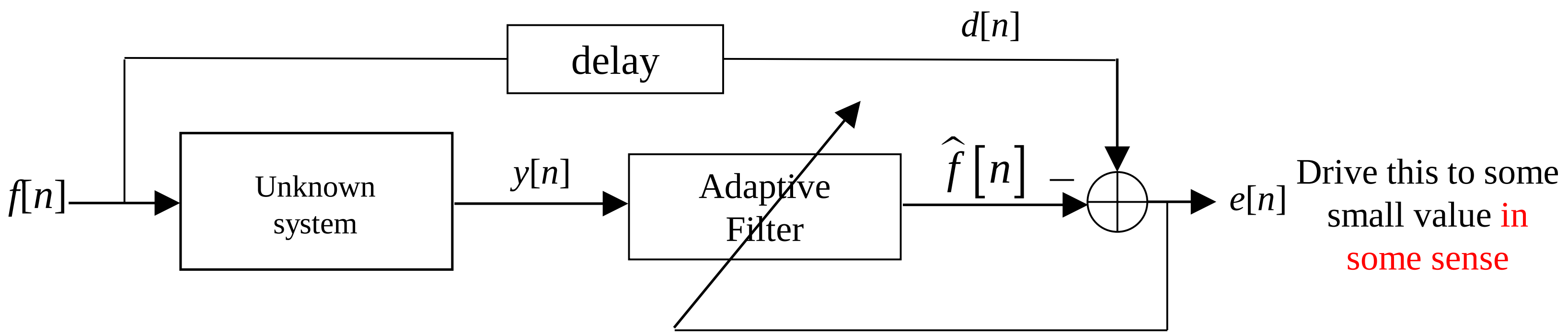
- Select parameters to achieve the "best" match between the desired signal $d[n]$ and filter output – optimizing the performance function such as
  - Least-squares error
  - Mean-squared error
- Characteristics of AF
  - Can automatically adjust (or adapt) in the face of changing environments and changing system requirements
  - Can be trained to perform specific filtering or decision-making tasks
  - Should have some "adaptation algorithm" (learning algorithm) for adjusting system's parameters

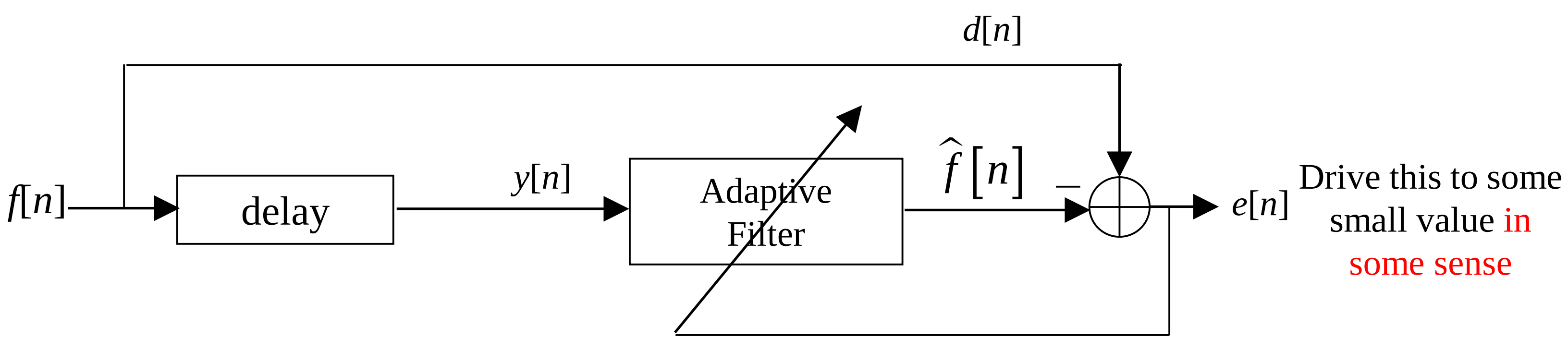


Applications of AF: System Identification and Interference Cancellation



Applications of AF: Inverse Modeling and Predictors





Random Variable (RV)

- A random variable is a function that assigns a numerical value each possible outcome in *S*, i.e. $S \rightarrow \Re$ (field of real number)
  - More convenient to work with a numerical value than nonnumerical value
- Can be discrete or continuous (example of discrete RV on top right, continuous RV on bottom right)
- Convention
  - Capital letters denote RVs
  - Lowercase letters denote values the RVs take on
    - E.g. $f_X(x)$ distribution function for RV X with value x

Table 5.2 Possible Random Variables (RV)

Outcome: S_i	RV No. 1: $X_1(S_i)$	RV No. 2: $X_2(S_i)$
S_1 = heads	$X_1(S_1) = 1$	$X_2(S_1) = \pi$
S_2 = tails	$X_1(S_2) = -1$	$X_2(S_2) = \sqrt{2}$

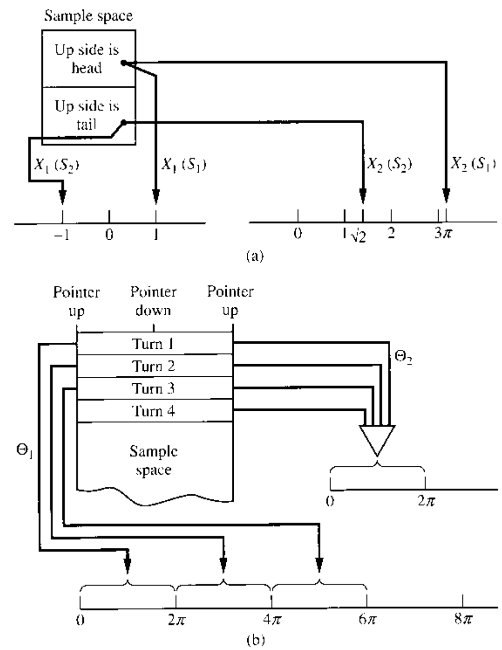


Figure 5.4
Pictorial representation of sample spaces and random variables.
(a) Coin-tossing experiment.
(b) Pointer-spinning experiment.

CDF and PDF

- Functions which relates the probability of an event to a numerical value assigned to an event
- Parameter vs. nonparameteric
  - There are several different parametric PDFs
  - Nonparametric
    - Estimated directly from data
    - Easily adaptable

Probability (Cumulative) Distribution Functions

- A way to probabilistically describe an RV

$$F_X(x) \triangleq P(X \le x)$$

Properties of $F_X(x)$

1. $0 \le F_X(x) \le 1$, with $F_X(-\infty) = 0$, $F_X(\infty) = 1$
2. $F_X(x)$ is continuous from the right, that is,

$$\lim_{x \to x_0^+} F_X(x) = F_X(x_0)$$

3. $F_X(x)$ is a nondecreasing function of x, i.e.

$$F_X(x_1) \le F_X(x_2) \text{ if } x_1 < x_2$$

From 2., $F_X(x)$ is continuous from right, so the jump amount = P_0

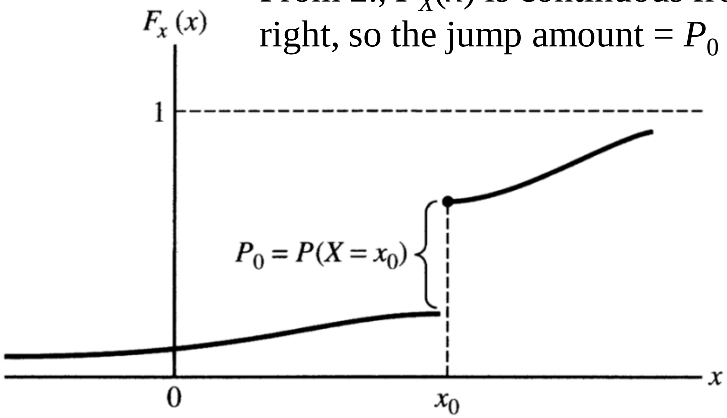


Figure 5.5
Illustration of the jump property of $F_X(x)$.

Probability Density Functions (PDF)

More convenient to express statistical averages using PDFs

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Properties of $f_X(x)$

1. $F_X(x) = \int_\eta f_X(\eta)\,d\eta \quad \Rightarrow \; f_X(x) = \frac{dF_X(x)}{dx} \geq 0$

2. $\int_x f(x)\,dx = 1$

3. $P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx$

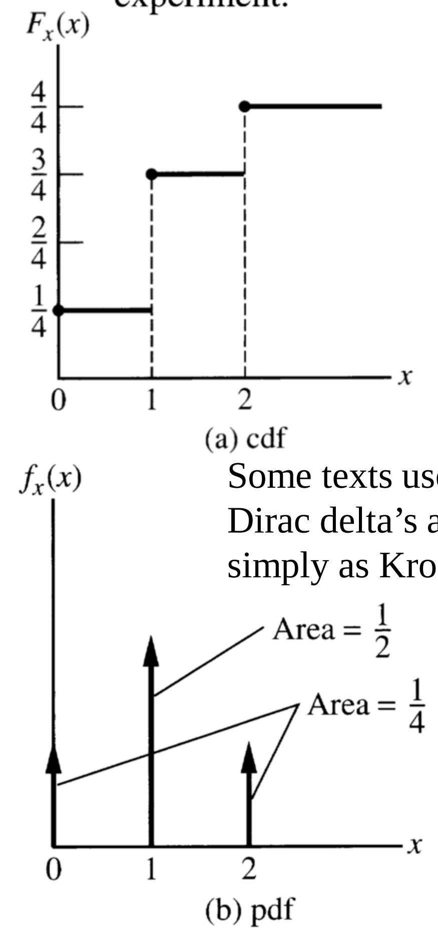
4. $f_X(x)\,dx = P(x - dx < X \leq x)$

Example: Discrete PDF and CDF

- 2 fair coins are tossed
- X: # of heads

Outcome	X	$P(X=x_j)$
TT	$x_1=0$	¼
TH HT		
HH	$x_3=3$	¼

Figure 5.6
The cdf and pdf for a coin-tossing experiment.



Some texts use pmf where the Dirac delta's are represented simply as Kronecker delta's

Example: Cont. PDF and CDF

Consider the pointer-spinning experiment. Assume any one stopping point is not favored over any other and that the RV Θ is defined as the angle that the pointer makes with the vertical, modulo 2π. Thus Θ is limited to $[0,2\pi)$ and for any two angles θ_1 and θ_2 in $[0,2\pi)$, we have

$$P(\theta_1 - \Delta\theta < \Theta \le \theta_1) = P(\theta_2 - \Delta\theta < \Theta \le \theta_2) \quad \text{(equally likely assumption)}$$

$$\Rightarrow f_\Theta(\theta_1) = f_\Theta(\theta_2),\ 0 \le \theta_1, \theta_2 < 2\pi.$$

$$\Rightarrow f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi}, & 0 \le \theta < 2\pi, \\ 0, & \text{otherwise} \end{cases}$$

Area under PDF curve is the probability.

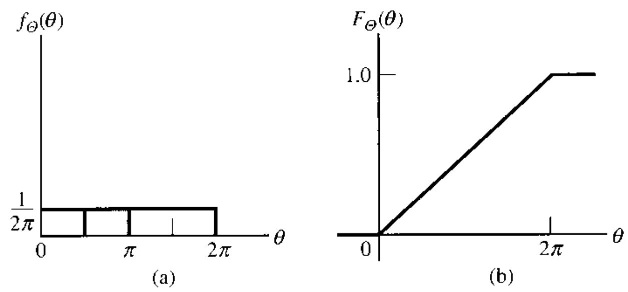


Figure 5.7
The (a) pdf and (b) cdf for a pointer-spinning experiment.

Joint CDFs and PDFs

Characterized by two or more RVs

$$F_{XY}(x,y) = P(X \le x, Y \le y)$$

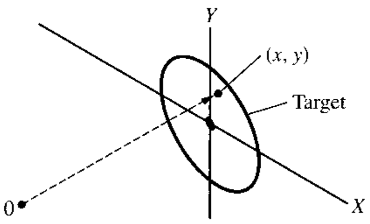
$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

$$P(x_1 < X \le x_2, y_1 < Y \le y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{XY}(x,y)\,dxdy$$

$$\Rightarrow F_{XY}(\infty,\infty) = \int_y\int_x f_{XY}(x,y)\,dxdy = 1$$

$$\Rightarrow f_{XY}(x,y)\,dxdy = P(x - dx < X \le x, y - dy < Y \le y)$$

Figure 5.8
The dart-throwing experiment.



Marginal CDFs and PDFs

Can obtain cdf or pdf of one of the RVs from joint RVs

$$F_X(x,y) = P(X \le x, Y \le \infty) = F_{XY}(x,\infty)$$

$$F_Y(x,y) = P(X \le \infty, Y \le y) = F_{XY}(\infty, y)$$

$$F_X(x) = \int_{y'} \int_{-\infty}^{x} f_{XY}(x',y')\,dx'dy'$$

$$F_Y(y) = \int_{-\infty}^{y} \int_{x'} f_{XY}(x',y')\,dx'dy'.$$

Since $f_X(x) = \dfrac{dF_X(x)}{dx}$ and $f_Y(y) = \dfrac{dF_Y(y)}{dy}$

$$\Rightarrow f_X(x) = \int_{y'} f_{XY}(x,y')\,dy' \quad \text{and} \quad f_Y(y) = \int_{x'} f_{XY}(x',y)\,dx'$$

Conditional CDFs and PDFs

Conditional RV:

$$F_{X|Y}\left(x|Y\right)=F_{X|Y}\left(x|Y\le y\right)=\frac{F_{XY}\left(x,y\right)}{F_{Y}\left(y\right)}$$

$$f_{X|Y}\left(x|y\right)=\frac{\partial F_{X|Y}\left(x|Y=y\right)}{\partial x}=\frac{f_{XY}\left(x,y\right)}{f_{Y}\left(y\right)}$$

Bayes Theorem:

$$f_{X|Y}\left(x|y\right)=\frac{f_{XY}\left(x,y\right)}{f_{Y}\left(y\right)}=\frac{f_{Y|X}\left(y|X=x\right)f_{X}\left(x\right)}{f_{Y}\left(y\right)}=\frac{f_{Y|X}\left(y|x\right)f_{X}\left(x\right)}{f_{Y}\left(y\right)}$$

where $f_{Y|X}\left(y|x\right)dx=P\left(y-dy<Y\le y \text{ given } X=x\right)$.

Statistical Independence

Two RVs are stat. independent if values one takes on do not influence the values that the other takes on.

$$\Rightarrow P(X \le x, Y \le y) = P(X \le x) P(Y \le y) \quad \text{or}$$

$$F_{XY}(x, y) = F_X(x) F_Y(y)$$

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

If X and Y are not independent, then using Bayes' rule

$$f_{XY}(x, y) = f_X(x) f_{Y|X}(y|x) = f_Y(y) f_{X|Y}(x|y).$$

Example: Statistical Independence

Two RVs X and Y have joint pdf

$$f_{XY}(x,y)=\begin{cases} Ae^{-(2x+y)}, & x,y\ge 0 \\ 0, & \text{otherwise.}\end{cases}$$

A can be found by noting that

$$F_{XY}(\infty,\infty)=\int_y\int_x f_{XY}(x,y)\,dxdy=1$$

Since $\int_0^\infty\int_0^\infty Ae^{-(2x+y)}dxdy=1 \;\Rightarrow\; A=2$

$$f_X(x)=\int_y f_{XY}(x,y)\,dy=\begin{cases}\int_0^\infty 2e^{-(2x+y)}dy, & x\ge 0\\ 0, & x<0\end{cases}=\begin{cases}2e^{-2x}, & x\ge 0\\ 0, & x<0\end{cases}$$

$$f_Y(y)=\int_x f_{XY}(x,y)\,dx=\begin{cases}e^{-y}, & y\ge 0\\ 0, & y<0\end{cases}$$

$$f_{X|Y}(x|y)=\frac{f_{XY}(x,y)}{f_Y(y)}=\begin{cases}2e^{-2x}, & x\ge 0\\ 0, & x<0\end{cases}$$

$$f_{Y|X}(y|x)=\frac{f_{XY}(x,y)}{f_X(y)}=\begin{cases}e^{-y}, & y\ge 0\\ 0, & y<0\end{cases}$$

Conditional prob's are equal to respective marginals ➔ X and Y are independent.

Example: Statistical Independence

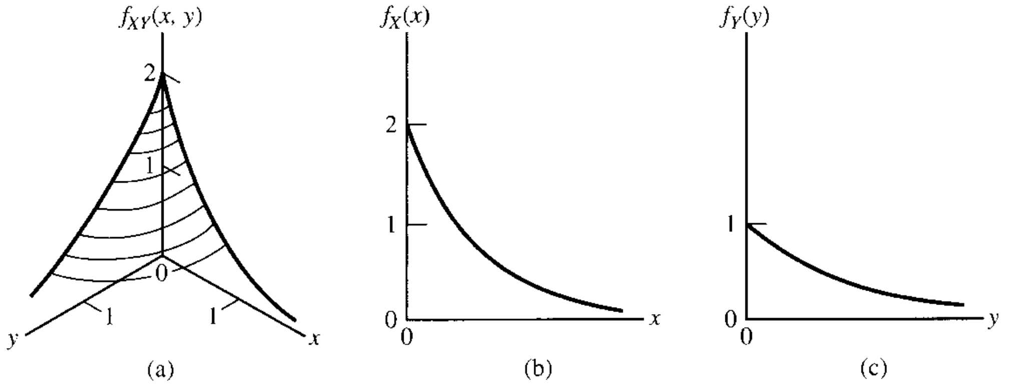


Figure 5.9
Joint and marginal pdfs for two random variables. (a) Joint pdf. (b) Marginal pdf for X. (c) Marginal pdf for Y.

$$f_{XY}(x,y)=\begin{cases} Ae^{-(2x+y)}, & x,y\ge 0 \\ 0, & \text{otherwise.} \end{cases}$$

$$f_X(x)=\int_y f_{XY}(x,y)\,dy=\begin{cases} \int_0^\infty 2e^{-(2x+y)}, & x\ge 0 \\ 0, & x<0 \end{cases}=\begin{cases} 2e^{-2x}, & x\ge 0 \\ 0, & x<0 \end{cases}$$

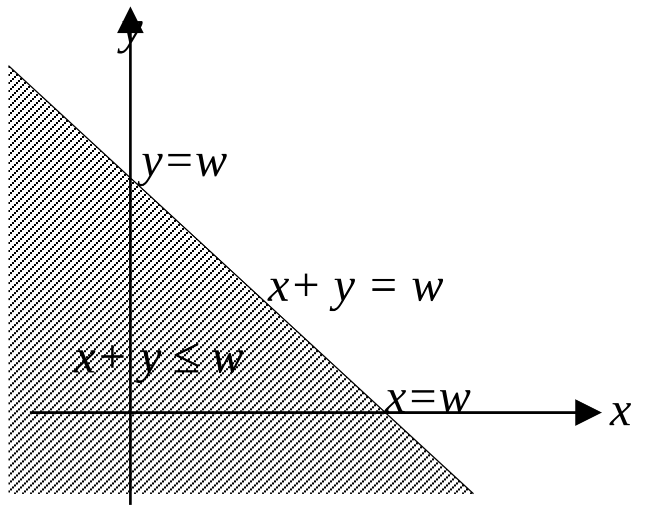
$$f_Y(y)=\int_x f_{XY}(x,y)\,dx=\begin{cases} e^{-y}, & y\ge 0 \\ 0, & y<0 \end{cases}$$

Sum of Two Statistically Indep. RVs

- The density of the sum of two statistically independent RVs is the convolution of their individual density functions.
- Suppose *X*, and *Y* are three independent RVS where $W = X + Y$, then

$$f_W(w) = \int_y f_Y(y) f_X(w-y) dy$$

$f_W(w)$, $f_X(x)$, and $f_Y(y)$ are pdfs of *W*, *X*, and *Y*, respectively



$$F_W(w) = P(W \le w) = P(X + Y \le w)$$
$$= \int_y \int_{x=-\infty}^{w-y} f_{X,Y}(x,y)\, dxdy$$
$$= \int_y f_Y(y) \int_{x=-\infty}^{w-y} f_X(x)\, dxdy \quad \text{(stat. indep.)}$$

Differentiating we get the result

Statistical Averages

- Sometimes full description of RVs, i.e. knowing its CDF or PDF are not required
- Sometimes only partial information is needed
  - One type of partial information of a set of RVs statistical average or mean value

Average of Discrete RV

Expectation of M RVs, $x_1, \ldots, x_M$ with respective probabilities $P_1, \ldots, P_M$

$$\mu_x \triangleq E[X] = \sum_{j=1}^{M} x_j P_j$$

Justification:

Let experiment be perform N number of time, **with N large**

Arithmetic mean: $\dfrac{n_1 x_1 + \cdots + n_m x_m}{N} = \sum_{j=1}^{M} x_j \dfrac{n_j}{N}$

By relative frequency interpretation: $\lim_{N \to \infty} \dfrac{n_j}{N} = P_j$

$$\Rightarrow \frac{n_1 x_1 + \cdots + n_m x_m}{N} = \sum_{j=1}^{M} x_j P_j$$

Average of Cont. RV

Expectation of x_0 to x_M with pdf $f_X(x)$. Suppose we break up this interval into subintervals of size Δx (assume small). The probability that X lies between $x_i - \Delta x$ to x_i is

$$P(x_i - \Delta x < X \le x_i) \approx f_X(x_i)\Delta x, \text{ for } i = 0, \ldots, M.$$

Hence, approximated X by a discrete RV that takes on values x_0 to x_M with probabilities $f_X(x_0)\Delta x, \ldots, f_X(x_M)\Delta x.$

$$\Rightarrow \mu_x \triangleq E[X] \approx \sum_{i=1}^{M} x_i f_X(x_i)\Delta x \overbrace{=}^{\lim\limits_{\Delta x \to 0}} \int_x x f_X(x)\, dx$$

Properties of Expectation

- $E[\cdot]$ is a linear operator
  - Sometimes need to perform $E(tr(\cdot))$. $tr(\cdot)$ is also linear operator ➔ $E(tr(\cdot)) = tr(E(\cdot))$
  - Additive
    - $E[X+Y] = E[X] + E[Y]$ for any 2 RVs
  - Homogeneity
    - $E[cX] = cE[X]$, for any constant c

Average of a Function of a RV

Let $Y = g(X)$.

$$\mu_Y \triangleq E[Y] = \begin{cases} \sum_i y_i P(y_i), & \text{discrete RV} \\ \int_y y f_Y(y)\,dx, & \text{cont. RV} \end{cases}.$$

r^{th} moment of X, for $r = 0,1,2,\ldots$. Let $Y = g(X) = X^r$

$$\xi_r \triangleq E\left[X^r\right] = \begin{cases} \sum_i x_i^r P(x_i), & \text{discrete RV} \\ \int_x x^r f_X(x)\,dx, & \text{cont. RV} \end{cases}$$

r^{th} central moment of X, for $r = 0,1,2,\ldots$. Let $Y = g(X) = (X - \mu_X)^r$

$$m_r \triangleq E\left[(X - \mu_X)^r\right]$$

Special case: variance: $r = 2$

$$\mathrm{var}[X] \triangleq m_2 \triangleq E\left[(X - \mu_X)^2\right] = E\left[X^2\right] - \mu_x^2 \triangleq \sigma_X^2$$

Average of a Function of a RV

r^{th} joint moment of X and Y, for $i, j = 0,1,2,\ldots$

$$\xi_{ij} \triangleq E\left[X^i Y^j\right] = \begin{cases} \sum_{i,j} x_\ell^i y_m^j P\left(x_i, y_m\right), & \text{discrete RV} \\ \int_{x,y} x^i y^j f_{XY}\left(x, y\right) dxdy, & \text{cont. RV} \end{cases}$$

Correlation: $\xi_{11} \triangleq E\left[XY\right]$

Note:

Independent: $E_{XY}\left(XY\right) = E_X\left(X\right) E_Y\left(Y\right)$

Uncorrelated: $E_{XY}\left[\left(X - \mu_X\right)\left(Y - \mu_Y\right)\right] = 0$

Orthogonal: $E\left(XY\right) = 0$

Implications:

- If X and Y are independent and have zero mean, implies X and Y are uncorrelated and orthogonal.
- If X and Y are uncorrelated and have zero mean, implies they are orthogonal.
- Hence, independence is the strongest of the three properties.

Average of a Function of a RV

r^{th} joint central moment of X and Y, for $i, j = 0,1,2,\ldots$

$$m_{ij} \triangleq E\left[\left(X - \mu_X \right)^i \left(Y - \mu_Y \right)^j \right]$$

Covariance:

$$Cov\left[X, Y \right] \triangleq m_{11} \triangleq E\left[\left(X - \mu_X \right)\left(Y - \mu_Y \right) \right] = E\left[XY \right] - \mu_X \mu_Y$$

Correlation coefficient for X and Y:

$$\rho \triangleq \frac{m_{11}}{\sqrt{m_{20} m_{02}}} = \frac{Cov\left[X, Y \right]}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Conditional Expectation

Conditional expectation of X given $Y = y$

$$E\left[X \middle| Y\right] = E\left[X \middle| Y = y\right] = \int_x x f_{X|Y}\left(x \middle| Y = y\right) dx$$

Expectation of functions of X : $Y = g\left(X\right)$

$$E\left[Y\right] = E\left[g\left(X\right)\right] = \int_x g\left(x\right) f_X\left(x\right) dx$$

Removing Conditional Expectation Via Expectation

Since $E_{X|Y}\left(X|Y\right)$ is a function of Y, it is also a RV.

$$
\begin{aligned}
E_Y\left[E_{X|Y}\left(X|Y\right)\right] &= \int_y \int_x x f_{X|Y}\left(x|y\right) dx\, f_Y\left(y\right)\, dy \\
&= \int_x x \int_y f_{X|Y}\left(x|y\right) f_Y\left(y\right) dy\, dx \\
&= \int_x x \int_y f_{XY}\left(XY\right) dy\, dx \\
&= \int_x x f_X\left(X\right) dx \\
&= E_X\left[X\right]
\end{aligned}
$$

Conditional Expectation

This is an "expectation" version of the total probability theorem.
In many cases, we can simplify a problem by conditioning or "fixing" one RV and performing an expectation. Then remove the conditioning in a second step by taking the expectation w.r.t. the conditioning RV.

More generally:

$$E\left[g(X)\right] = E_Y\left[E_{X|Y}\left(g(X)\middle|Y\right)\right]$$

Special Average: Characteristic Function

Let $g(X) = e^{j\omega X}$

$$\Phi(\omega) \triangleq E\left[e^{j\omega X}\right] = \int_x f_X(x) e^{j\omega x} dx$$

$$f_X(x) = \frac{1}{2\pi} \int_v \Phi(\omega) e^{-j\omega x} dv$$

Note:

- This is Fourier transform of $f_X(x)$ if we have $e^{-j\omega X}$
- Sometimes it is more convenient to use the variable s in place of $j\omega$, the result becomes **moment generating function**.

Obtaining moments of a RV:

$$\frac{\partial \Phi(\omega)}{d\omega} = j \int_x x f_X(x) e^{jvx} dx$$

Set $\omega = 0$: $$\Rightarrow E[X] = (-j) \left.\frac{\partial \Phi(\omega)}{d\omega}\right|_{\omega=0}$$

$$\Rightarrow E\left[X^n\right] = (-j)^n \left.\frac{\partial^n \Phi(\omega)}{d\omega^n}\right|_{\omega=0}$$

Chebyshev Inequality and the Law of Large Numbers

Let X be a RV with mean μ_X and finite variance σ_X^2. Then for any $\delta > 0$,

$$P\left(\left|X - \mu_X\right| \geq \delta\right) \leq \frac{\sigma_X^2}{\delta^2} \quad \text{(Chebyshev Inequality)}$$

Let $X_1, X_2, \ldots, X_N$ be i.i.d. (independent and identically distributed) RVs with mean μ_X and variance σ_X^2 each. Let the sample mean be

$$\hat{\mu}_X = \frac{1}{N}\sum_{i=1}^{N} X_i.$$

Then, for any fixed $\delta > 0$,

$$\lim_{N\to\infty} P\left(\left|\mu_X - \hat{\mu}_X\right| \geq \delta\right) = 0. \quad \text{(LLN)}$$

Intuitively, this means the estimator, $\hat{\mu}_X$, will converge to μ_X in probability.

If the above limit equals 0, $\hat{\mu}_X$ is called a consistent estimator of μ_X.

Useful PDFs

- Discrete RVs
  - Binomial distribution
    - Related to chance experiments with two mutually exclusive outcomes with probability p and $1-p$
    - Model number of times event A has occurred in n trials (events are indep)
  - Poisson distribution
    - Related to chance experiment in which an event whose probability of occurrence in a very small time interval ΔT is $P=\alpha\Delta T$, where α is a constant
    - Model the probability of k events occurring in time T
    - Commonly used to model arrival time of packets in packet switching networks
- Continuous RVs
  - Normal (Gaussian) distribution
    - Commonly used to model large number of indep. random events when distribution of each event is unknown
    - Sum of large number of independent RVs converges to a Gaussian distribution
  - Rayleigh distribution
    - (see above)
  - Rician distribution
    - Commonly used to model distribution of power profile of wireless channel when direct line-of-sight (LOS) exists
    - x = sqrt($x_1^2+x_2^2$), where $x_1 \sim N(\mu_1,\sigma^2)$, $x_2 \sim N(\mu_2,\sigma^2)$ are indep. RV

Useful PDFs

- Continuous RVs
  - Chi-Squared (central and noncentral)
    - Commonly encounter in detector design

$$\chi_\nu^2 \text{ with } \nu \text{ degrees of freedom}$$

$$x = \sum_{i=1}^{\nu} x_i^2, \;\; x_i \sim N\left(0 \text{ or } \mu_i, 1\right) \text{ and indep.}$$

  - F-distribution (central and noncentral)
    - Commonly encounter in detector design

$$F \text{ PDF: ratio of 2 indep. } \chi_\nu^2 \text{ RVs}$$

$$x = \frac{x_1 / \nu_1}{x_2 / \nu_2}, \;\; x_1 \sim \chi_{\nu_1}^2\left(\lambda\right), \;\; x_2 \sim \chi_{\nu_2}^2 \text{ and indep.}$$

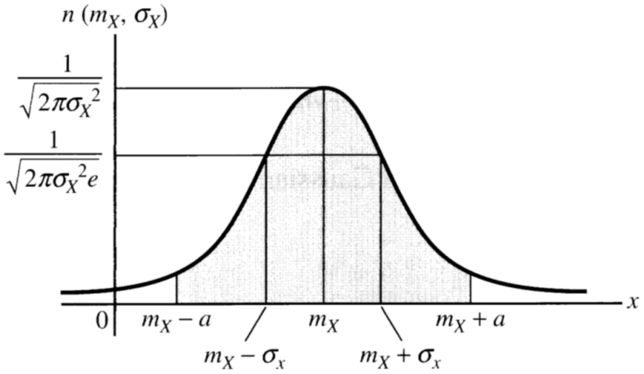
$$\lambda = 0: \text{ central } F-\text{dist.}$$

Gaussian (Normal) Distribution

1 – dimensional:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{\sqrt{2\sigma^2}}(x-\mu)^2\right]$$

where $\mu \triangleq E[X]$, $\sigma^2 \triangleq E\left[(X-\mu)^2\right]$



Joint CDFs and PDFs:

$$F_{XY}(x,y) = P(X \le x, Y \le y)$$

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

$$P(x_1 \le X \le x_2, y_1 \le Y \le y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{XY}(x,y)\,dxdy$$

Marginal distribution:

$$F_X(x) = F_{XY}(x,\infty) = F_{XY}(x, Y \le \infty)$$

$$F_Y(y) = F_{XY}(\infty, y) = F_{XY}(X \le \infty, y)$$

$$f_X(x) = \int_x f_{XY}(x,y)\,dy$$

2-D (Bivariate) Gaussian Distribution

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{\left[\left(x-\mu_x\right)/\sigma_x\right]^2-2\rho\left[\left(x-\mu_x\right)/\sigma_x\right]\left[\left(y-\mu_y\right)/\sigma_y\right]+\left[\left(y-\mu_y\right)/\sigma_y\right]^2}{2\left(1-\rho^2\right)}\right)$$

where

$$\mu_x=E[X],\ \ \mu_y=E[Y],\ \ \sigma_x^2=\mathrm{var}[X],\ \ \sigma_y^2=\mathrm{var}[Y]$$

$$\rho=\frac{E\left[\left(X-\mu_x\right)E\left(Y-\mu_y\right)\right]}{\sigma_x\sigma_y}=\frac{Cov[X,Y]}{\sqrt{\sigma_x^2\sigma_y^2}}$$

2-D (Bivariate) Gaussian Distribution

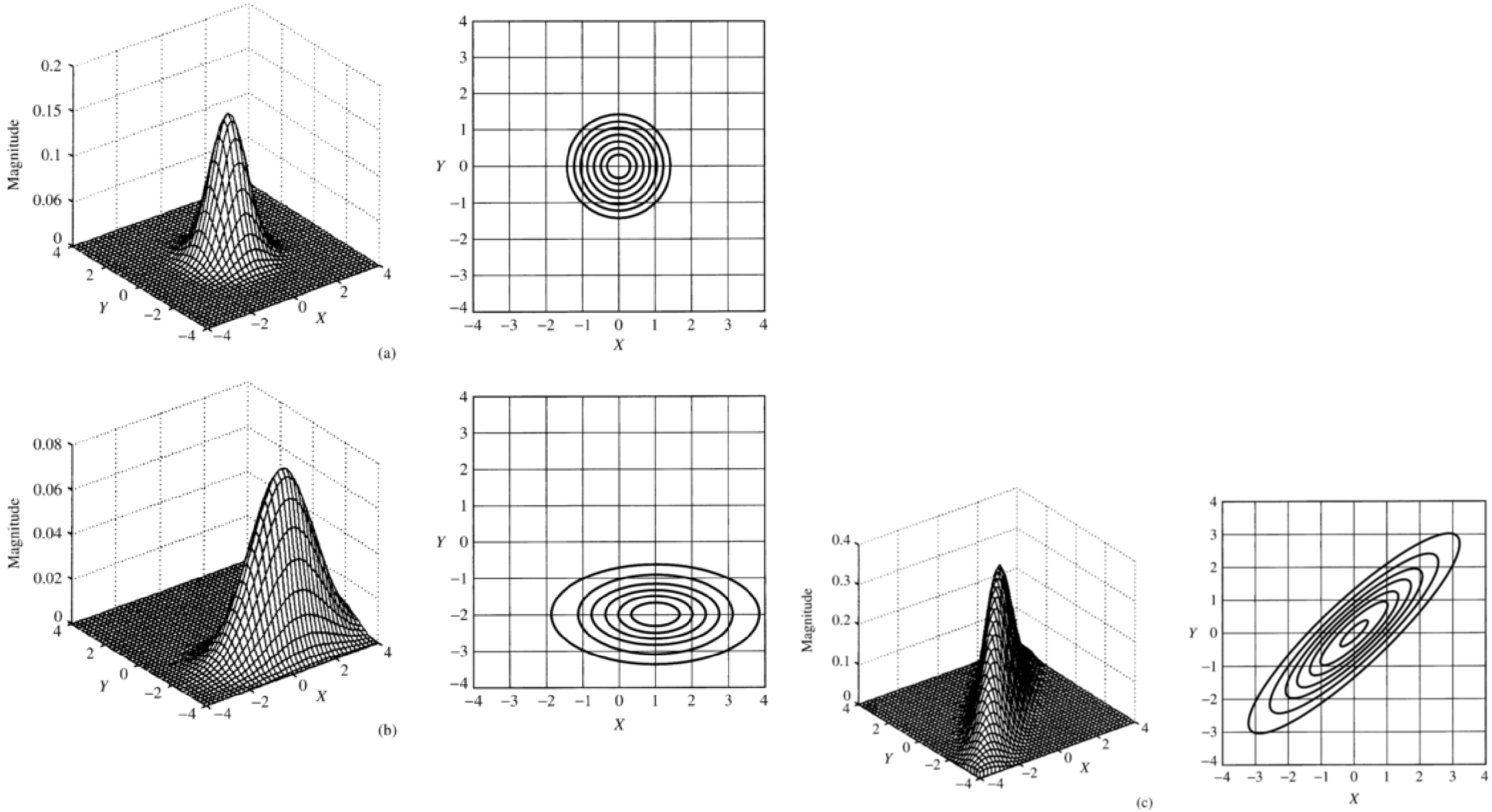


Figure 5.18

Bivariate Gaussian pdfs and corresponding contour plots. (a) $m_x = 0$, $m_y = 0$, $\sigma_x^2 = 1$, $\sigma_y^2 = 1$ and $\rho = 0$. (b) $m_x = 1$, $m_y = -2$, $\sigma_x^2 = 2$, $\sigma_y^2 = 1$, and $\rho = 0$.(c) $m_x = 0$, $m_y = 0$, $\sigma_x^2 = 1$, $\sigma_y^2 = 1$, and $\rho = 0.9$.

N-dimensional Gaussian Distribution

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} (\det \mathbf{C})^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \right]$$

$$\boldsymbol{\mu}_{\mathbf{x}} \triangleq E[\mathbf{x}] = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_N) \end{bmatrix}$$

$$\mathbf{C} \triangleq E\left[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \right] \text{ (applied element-wise)}$$

Central Limit Theorem

Let $X_1, X_2, \ldots, X_N$ be indep. RVs with zero mean and variance $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$. Let $s_N^2 \triangleq \sigma_1^2 + \cdots + \sigma_N^2$. If for any fixed $\varepsilon > 0$, there exists a sufficient large N such that

$$\sigma_k^2 < \varepsilon s_N, \quad \text{for } k = 1, \ldots, N,$$

then the normalized RV

$$Z_N \triangleq \frac{X_1 + X_2 + \cdots + X_N}{s_N}$$

converges to the standard normal (Gaussian) PDF.

Q-Function

Gaussian Q-Function:

Normalized Normal distribution of $N\left(\mu_x, \sigma_x^2\right)$

Consider $P\left(\mu_x - a \le X \le \mu_x + a\right) = \int_{\mu_x - a}^{\mu_x + a} \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{1}{2\sigma_x^2}\left(x-\mu_x\right)^2\right] dx$

(let $y = \frac{x-\mu_x}{\sigma_x}$)
$$= \int_{-a/\sigma_x}^{a/\sigma_x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

$$= 2\int_{0}^{a/\sigma_x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

(since area under PDF=1)
$$= 1 - 2\int_{a/\sigma_x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

$$= 1 - 2Q\left(\frac{a}{\sigma_x}\right)$$

where $Q(u) \triangleq \int_{u}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \approx \frac{1}{u\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right), \quad \text{for } u \gg 1$

has been computed numerically.

Normalized Distribution Function: $F(x)$ and $Q(x)$

Normalized cumulative distribution function: $\mu_x = 0, \ \sigma_x = 1$

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\xi^2/2} d\xi$$

$$F(-x) = 1 - F(x)$$

A related function: $F(x) = 1 - Q(x)$

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\xi^2/2} d\xi$$

$$Q(-x) = 1 - Q(x)$$

TABLE B-1
Values of $F(x)$ for $0 \leq x \leq 3.89$ in steps of 0.01

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9773	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	1.0000	1.0000	1.0000

Normalized cumaltive distribution function

$$F(x)$$

$$F(x) = 1 - Q(x)$$

Stochastic Process

- Random Processes (Stochastic Processes)
  - Informal definition
    - The outcomes (events) of a chance experiment are mapped into functions of time (waveforms)
    - Cf. Random variables: outcomes are mapped into numbers
  - Each waveform is called a sample function, or a realization. The totality of all sample functions is called an ensemble
  - Chance experiment that gives rise to this ensemble is called a random/stochastic process
  - Formal definition
    - Every outcome ζ we assign, according to a certain rule, a time function $X(t,\zeta)$. $X(t,\zeta_i)$ signifies a single time function
    - $X(t_j,\zeta)$ denotes a single RV
    - $X(t_j,\zeta_i)$ is a number

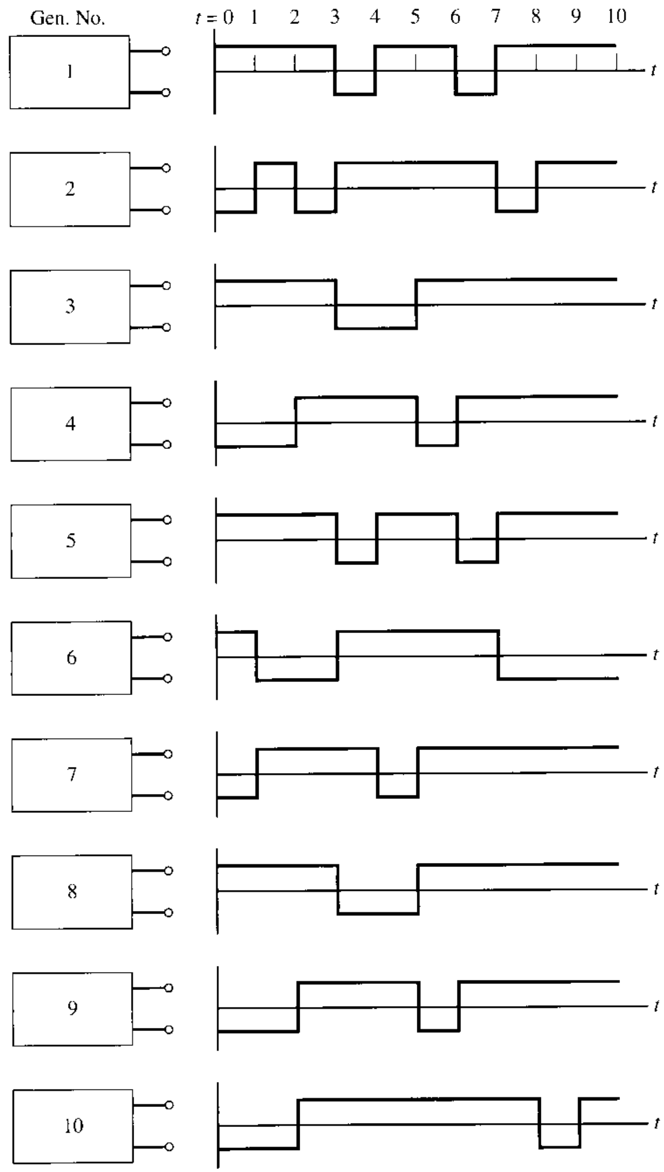


Figure 6.1
A statistically identical set of binary waveform generators with typical outputs.

Voltage at the terminals of a noise generator. 10 ensemble experiments

Statistical Description of Random Process

- A random process is statistically specified by its N^{th} order joint pdf's that describes a typical sample function at times $t_N > t_{N-1} > \ldots > t_1$, for any N where

$$F_{X1X2\ldots XN}(x_1,t_1;x_2,t_2;\ldots;x_N,t_N) = P(x_1 - dx_1 < X_1 \le x_1 \text{ at time } t_1,\ x_2 - dx_2 < X_2 \le x_2 \text{ at time } t_2,\ \ldots,\ x_N - dx_N < X_N \le x_N \text{ at time } t_N)$$

where $X_n \equiv X(t_n,\zeta)$, for $n=1,\ldots N$

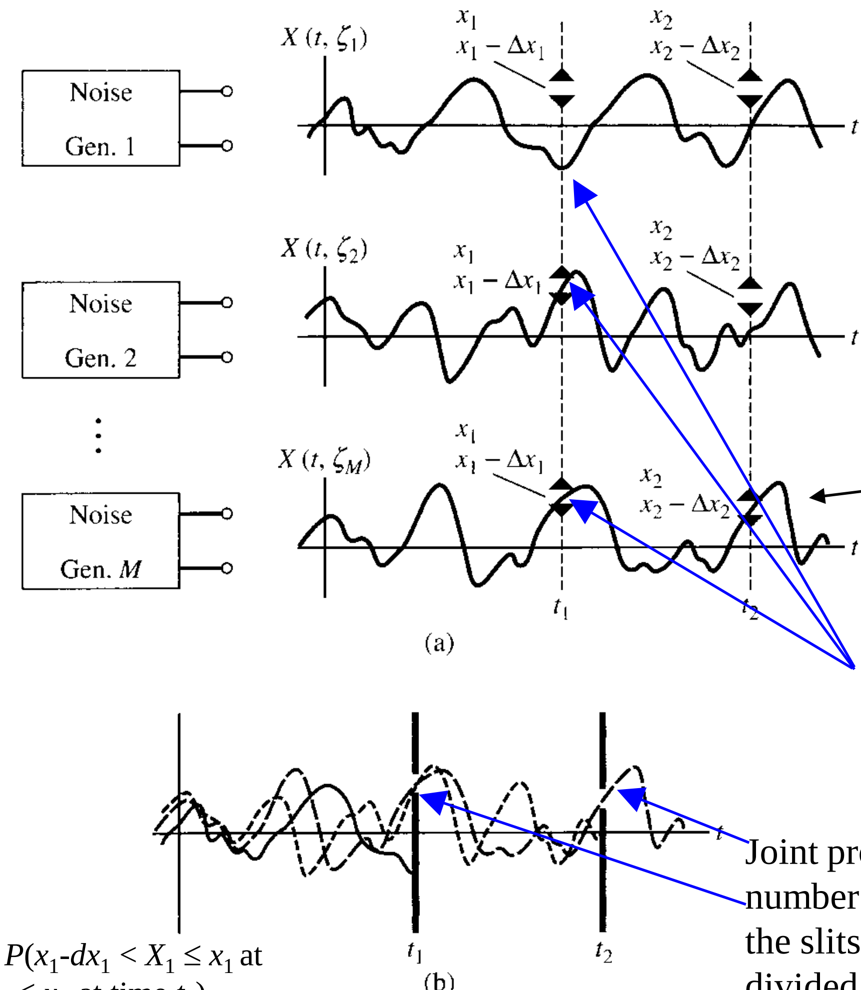


Figure 6.2
Typical sample functions of a random process and illustration of the relative-frequency interpretation of its joint pdf. (a) Ensemble of sample functions. (b) Superposition of the sample functions shown in (a).

Random process from realization ζ_M.

$X(t_j,\zeta)$ is a random variable

$F_{X1X2}(x_1,t_1;x_2,t_2) = P(x_1\text{-}dx_1 < X_1 \le x_1$ at time t_1, $x_2\text{-}dx_2 < X_2 \le x_2$ at time $t_2)$

Joint probability (from relative frequency) is the number of sample functions that pass through the slits placed at $t=t_1$ and $t=t_2$ in both barriers divided by the total number of M of sample functions as M becomes large w/o bound

Stationarity and Wide-Sense Stationarity

- Statistical stationarity in the strict sense or stationarity
  - Joint pdfs depend only on the time differences t_2-t_1, t_3-t_1, …, t_N-t_1
    - Not dependent on time origin
  - Mean and variance independent of time
  - Correlation coefficient or covariance depends only on difference, e.g. t_2-t_1
- Wide-sense stationarity (WSS)
  - Joint pdfs are dependent on time origin
  - Mean and variance independent of time
  - Correlation coefficient or covariance depends only on difference, e.g. t_2-t_1
- Stationarity ➔ WSS
  - Converse is not necessarily true
    - Exception: Gaussian random process (Why?)

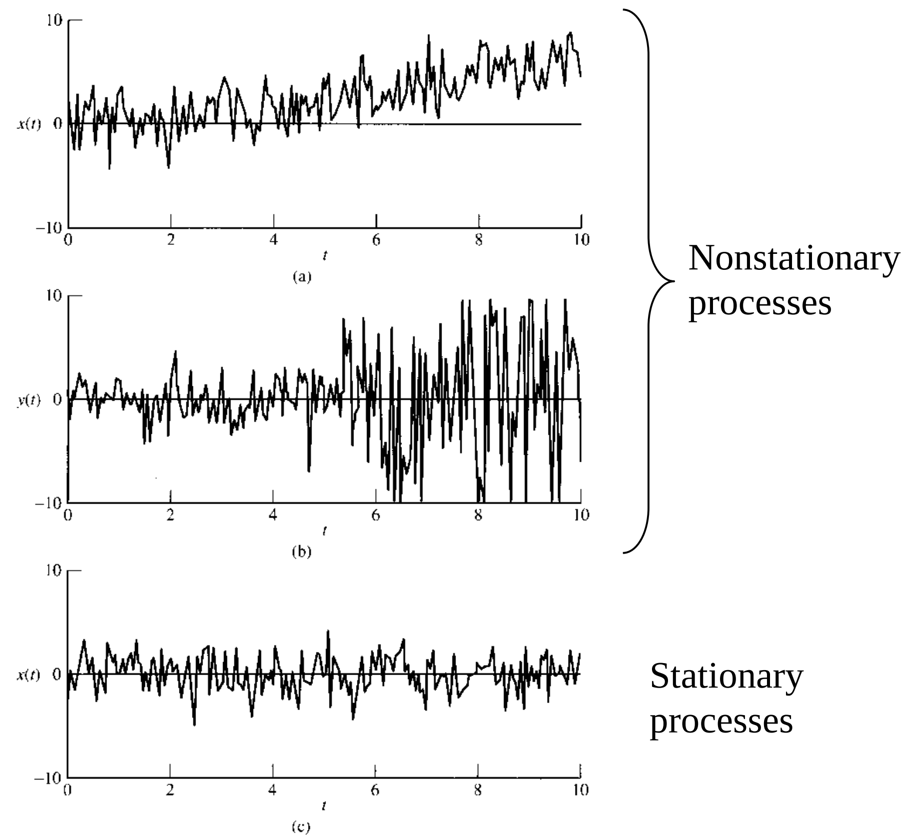


Figure 6.3
Sample functions of nonstationary processes contrasted with a sample function of a stationary process. (a) Time-varying mean. (b) Time-varying variance. (c) Stationary.

Ensemble Average (Expectation)

Mean: $\mu_x(t) = E\left[x(t)\right] = \overline{x(t)} = \int_\alpha \alpha\, f_X(\alpha, t)\, d\alpha$

Variance: $\sigma_{xx}^2(t) = E\left\{\left|x(t) - \overline{x(t)}\right|^2\right\} = E\left[\left|x(t)\right|^2\right] - \left|\overline{x(t)}\right|^2$

Covariance:

$$c_{xx}(t_1, t_2) = E\left\{\left[x(t_1) - \overline{x(t_1)}\right]\left[x(t_2) - \overline{x(t_2)}\right]^*\right\}$$

$$= E\left[x(t_1)x^*(t_2)\right] - \overline{x(t_1)}\,\overline{x(t_2)}^*$$

$$c_{xx}(t_2, t_1) = E\left\{\left[x(t_2) - \overline{x(t_2)}\right]\left[x(t_1) - \overline{x(t_1)}\right]^*\right\}$$

$$= E\left[x(t_2)x^*(t_1)\right] - \overline{x(t_2)}\,\overline{x(t_1)}^*$$

$\Rightarrow c_{xx}(t_1, t_2) = c_{xx}^*(t_2, t_1)$

Autocorrelation:

$$r_{xx}(t_1, t_2) = E\left[x(t_1)x^*(t_2)\right]$$

$$= \int_{\alpha_2}\int_{\alpha_1} \alpha_1\alpha_2\, f_{X_1X_2}(\alpha_1, t_1; \alpha_2, t_2)\, d\alpha_1 d\alpha_2$$

Ensemble Average (Vector Random Process)

Mean: $\boldsymbol{\mu}_x(t) = E\left[\mathbf{x}(t)\right] = \overline{\mathbf{x}(t)}$

Variance: $\sigma_{xx}^2(t) = E\left\{\left[\mathbf{x}(t) - \overline{\mathbf{x}(t)}\right]^H \left[\mathbf{x}(t) - \overline{\mathbf{x}(t)}\right]\right\}$

$$= E\left[\left|x(t)\right|^2\right] - 2\,\mathrm{Re}\left\{\mathbf{x}^H(t)\overline{\mathbf{x}(t)}\right\} + \left|\overline{x(t)}\right|^2$$

Covariance:

$$\mathbf{C}_{xx}(t_1, t_2) = E\left\{\left[\mathbf{x}(t_1) - \overline{\mathbf{x}(t_1)}\right]\left[\mathbf{x}(t_2) - \overline{\mathbf{x}(t_2)}\right]^H\right\}$$

$$= E\left[\mathbf{x}(t_1)\mathbf{x}^H(t_2)\right] - E\left[\mathbf{x}(t_1)\overline{\mathbf{x}(t_2)}^H\right] - E\left[\overline{\mathbf{x}(t_1)}\mathbf{x}^H(t_2)\right] + \overline{\mathbf{x}(t_1)}\,\overline{\mathbf{x}(t_2)}^H$$

Autocorrelation:

$$\mathbf{R}_{xx}(t_1, t_2) = E\left[\mathbf{x}(t_1)\mathbf{x}^H(t_2)\right]$$

Ensemble Average (Expectation) for WSS Process

WSS:

Mean: $\mu_x(t) = E\left[x(t)\right] = \text{constant}$

Variance: $\sigma_{xx}^2(t) = \text{constant}$

Covariance:

$$c_{xx}(\tau) \triangleq E\left\{\left[x(t) - \overline{x(t)}\right]\left[x(t-\tau) - \overline{x(t-\tau)}\right]^*\right\}$$

$$= E\left[x(t)x^*(t-\tau)\right] - \overline{x(t)}\,\overline{x(t-\tau)}^*$$

Autocorrelation:

$$r_{xx}(\tau) \triangleq E\left[x(t)x^*(t-\tau)\right]$$

$$\Rightarrow r_{xx}^*(\tau) \triangleq E\left[x^*(t)x(t-\tau)\right]$$

$$\Rightarrow r_{xx}^*(-\tau) \triangleq E\left[x^*(t)x(t+\tau)\right] = E\left[x(t+\tau)x^*(t)\right]$$

$$= E\left[x(p)x^*(p-\tau)\right] = r_{xx}(\tau)$$

Ensemble Average for Vector WSS Process

WSS:

Mean: $\boldsymbol{\mu}_x(t) = E\left[\mathbf{x}(t)\right] = \text{constant}$

Variance: $\sigma_{xx}^2(t) = E\left[\mathbf{x}^H(t)\mathbf{x}(t)\right] = \text{constant}$

Covariance:

$$\mathbf{C}_{xx}(\tau) \triangleq E\left\{\left[\mathbf{x}(t) - \overline{\mathbf{x}(t)}\right]\left[\mathbf{x}(t-\tau) - \overline{\mathbf{x}(t-\tau)}\right]^H\right\}$$

$$= E\left[\mathbf{x}(t)\mathbf{x}^H(t-\tau)\right] - \overline{\mathbf{x}(t)}\,\overline{\mathbf{x}(t-\tau)}^H$$

Autocorrelation:

$$\mathbf{R}_{xx}(\tau) \triangleq E\left[\mathbf{x}(t)\mathbf{x}^H(t-\tau)\right]$$

Ergodicity

Ergodic processes are processes for which time and ensemble averages are interchangeable.

For example, for real-valued WSS processes:

$$\mu_x = E\left[x(t)\right] = \left\langle x(t) \right\rangle$$

$$\sigma_{xx}^2 = E\left\{\left[x(t) - \overline{x(t)}\right]^2\right\} = \left\langle \left[x(t) - \left\langle x(t) \right\rangle\right]^2 \right\rangle$$

$$r_{xx}(\tau) = E\left[x(t)x(t+\tau)\right] = \left\langle x(t)x(t+\tau) \right\rangle,$$

where $\left\langle v(t) \right\rangle \triangleq \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} v(t)\,dt.$

Note:

- All time and ensemble averages are interchangeable, not just the above.
- Ergodicity $\Rightarrow$ strict-sense stationarity

Example 1: Ergodicity

Consider a random process with sample function

$$n(t) = A\cos(2\pi f_0 t + \theta),$$

where f_0 is a constant and Θ is a RV with pdf

$$f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi}, & |\theta| \le \pi \\ 0, & \text{otherwise} \end{cases}.$$

Calculate its ensemble and time-average.

$$E[n(t)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} A\cos(2\pi f_0 t + \theta)\, d\theta = 0$$

$$\sigma_{nn}^2(t) = E[n^2(t)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} [A\cos(2\pi f_0 t + \theta)]^2\, d\theta$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} A^2\cos^2(2\pi f_0 t + \theta)\, d\theta$$

$$= \frac{A^2}{4\pi}\int_{-\pi}^{\pi} [1 + \cos(4\pi f_0 t + 2\theta)]\, d\theta$$

$$= \frac{A^2}{2}$$

$$\langle n(t)\rangle = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} A\cos(2\pi f_0 t + \theta)\, dt = 0$$

$$\langle n^2(t)\rangle = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} A^2\cos^2(2\pi f_0 t + \theta)\, dt$$

$$= \frac{A^2}{2}$$

$E[n(t)] = \langle n(t)\rangle$=constant and $\sigma_{nn}^2(t) = \langle n^2(t)\rangle =$ constant.

It may be stationary and ergodic.

Example 2: Ergodicity

Suppose $f_\Theta(\theta) = \begin{cases} \dfrac{2}{\pi}, & |\theta| \le \dfrac{\pi}{4} \\ 0, & \text{otherwise} \end{cases}$.

Calculate its ensemble and time-average.

$$E[n(t)] = \frac{2}{\pi}\int_{-\pi/4}^{\pi/4} A\cos(2\pi f_0 t + \theta)\, d\theta$$

$$= \frac{2}{\pi} A\sin(2\pi f_0 t + \theta)\Big|_{-\pi/4}^{\pi/4} = \frac{2\sqrt{2}A}{\pi}\cos(2\pi f_0 t)$$

$$r_{nn}^2(0) = E\left[n^2(t)\right] = \frac{2}{\pi}\int_{-\pi/4}^{\pi/4}\left[A\cos(2\pi f_0 t + \theta)\right]^2 d\theta$$

$$= \frac{A^2}{\pi}\int_{-\pi/4}^{\pi/4}\left[1 + \cos(4\pi f_0 t + 2\theta)\right] d\theta$$

$$= \frac{A^2}{2} + \frac{A^2}{\pi}\cos(4\pi f_0 t)$$

Process is not stationary as first and second moment depends on t, hence it is for different time origin.

Summary for Ergodic Process

1. Mean: $\mu_x(t) = E\left[x(t)\right] = \langle x(t) \rangle$ is the DC component
2. $\overline{x(t)}^2 = \langle x(t) \rangle^2$ is the DC power
3. $r_{xx}(0) = \overline{x^2(t)} = \langle x^2(t) \rangle$ is the total power
4. $\sigma_{xx}^2(t) = \overline{x^2(t)} - \overline{x(t)}^2 = \langle x^2(t) \rangle - \langle x(t) \rangle^2$ is the power in the alternating current (time-varying) component
5. Total power $\overline{x^2(t)} = \sigma_{xx}^2(t) + \langle x(t) \rangle^2$ is the AC power plus the DC power

Power Spectral-Density Functions (PSD) and Cross-Spectral Density

The PSD of a wide-sense stationary random process is the Fourier transform of the autocorrelation function. For continuous-time random process

$$S_{xx}(j\Omega) = \int_{\tau} r_{xx}(\tau) e^{-j\Omega\tau} d\tau.$$

Since $r_{xx}(\tau)$ is symmetric, the PSD is a real-valued function of Ω. Since real-valued power cannot be negative, the PSD must satisfy $S_{xx}(\Omega) \geq 0$, $\forall\Omega$. Then average power of a random process is

$$r_{xx}(0) = E\left[x(t)x^*(t)\right] = E\left[\left|x(t)\right|^2\right]$$

$$\Leftrightarrow \frac{1}{2\pi}\int_{\Omega} S_{xx}(j\Omega) e^{j\Omega\tau} d\Omega\bigg|_{\tau=0} = \frac{1}{2\pi}\int_{\Omega} S_{xx}(j\Omega) d\Omega$$

Cross-Spectral Density: $S_{xy}(j\Omega) = \int_{\tau} r_{xy}(\tau) e^{-j\Omega\tau} d\tau$

Bilateral Laplace Transform of the Autocorrelation Function

Note: $s = \sigma + j\Omega$ (entire complex plane). Define

$$S_{xx}(s) \triangleq \int_{\tau} r_{xx}(\tau) e^{-s\tau} d\tau$$

$$S_{xy}(s) \triangleq \int_{\tau} r_{xy}(\tau) e^{-s\tau} d\tau$$

For real-valued random process, since autocovariance is real and even, its Laplace transform will be even

$$S_{xx}(s) = S_{xx}(-s).$$

If $s = j\Omega$ $$S_{xx}(-j\Omega) = S_{xx}^{*}(j\Omega).$$

Discrete-Time PSD and its Laplace Transform Representation

For discrete-time, PSD: $$S_{xx}\left(e^{j\omega}\right)=\sum_{k} r_{xx}[k]e^{-j\omega k}$$

Cross-Spectral Density: $$S_{xy}\left(e^{j\omega}\right)=\sum_{k} r_{xy}[k]e^{-j\omega k}$$

Define $$S_{xx}(z)\triangleq\sum_{k} r_{xx}[k]z^{-k}$$

$$S_{xy}(z)\triangleq\sum_{k} r_{xy}[k]z^{-k}$$

For real-valued process

$$S_{xx}\left(\frac{1}{z}\right)=S_{xx}(z)$$

and $$S_{xx}\left(e^{-j\omega}\right)=S_{xx}^{*}\left(e^{j\omega}\right)$$

Uncorrelated, Orthogonal, Independent Random Processes

Given two random processes $X(t)$ and $Y(t)$

(1) Uncorrelated

if $R_{XY}(t_1,t_2) = m_X(t_1)m_Y^*(t_1), \quad \forall t_1,t_2$

(2) Orthogonal

if $R_{XY}(t_1,t_2) = 0, \quad \forall t_1,t_2$

(3) Independence: if

$$f_{XY}(x_1,y_1,t_1;x_2,y_2,t_2;\ldots;x_n,y_n,t_n)$$
$$= f_X(x_1,t_1;x_2,t_2;\ldots;x_n,t_n)\, f_Y(y_1,t_1;y_2,t_2;\ldots;y_n,t_n)$$

Remarks:

(1) Independence $\Rightarrow$ Uncorrelated

(2) Uncorrelated $\Rightarrow$ $\left(X(t)-m_X(t)\right)$ and $\left(Y(t)-m_Y(t)\right)$ are orthogonal

(3) (Uncorrelated and either $m_X(t)=0$ or $m_Y(t)=0$) $\Rightarrow$ orthogonal

(4) Uncorrelated and Gaussian $\Rightarrow$ Independent

Linear Systems and Random Processes

Given $h(t)$ is LTI, and $y(t)=h(t)*x(t)$

Mean of $y(t)$:

$$\mu_y(t)=E\left[h(t)*x(t)\right]=E\left[\int_u h(u)x(t-u)\,du\right]=\int_u h(u)E\left[x(t-u)\right]du$$

$$=\mu_x(t)\int_u h(u)\,du=\mu_x(t)H(0)$$

Cross-correlation

$$r_{xy}(t_1,t_2)=E\left[x(t_1)y^*(t_2)\right]=E\left[x(t_1)\int_u h^*(u)x^*(t_2-u)\,du\right]$$

$$=\int_u h^*(u)E\left[x(t_1)x^*(t_2-u)\right]du$$

$$=\int_u h^*(u)r_{xx}(t_1-t_2+u)\,du$$

If $x(t)$ is WSS, let $\tau=t_1-t_2$

$$r_{xy}(\tau)=\int_u h^*(u)r_{xx}(\tau+u)\,du=h^*(-\tau)*r_{xx}(\tau)$$

Linear Systems and Random Processes

Similarly

$$r_{yx}(t_1,t_2)=E\left[y(t_1)x^*(t_2)\right]=E\left[\int_u h(u)x(t_1-u)du\; x^*(t_2)\right]$$

$$=\int_u h(u)E\left[x(t_1-u)x^*(t_2)\right]du$$

$$=\int_u h(u)r_{xx}(t_1-t_2-u)du$$

If $x(t)$ is WSS, let $\tau=t_1-t_2$

$$r_{yx}(\tau)=\int_u h(u)r_{xx}(\tau-u)du=h(\tau)*r_{xx}(\tau)$$

$$r_{yy}(\tau)=E\left[y(t)y^*(t-\tau)\right]=E\left[y(t)\int_u h^*(u)x^*(t-u-\tau)du\right]$$

$$=\int_u h^*(u)E\left[y(t)x^*(t-u-\tau)\right]du$$

$$=\int_u h^*(u)r_{yx}(u+\tau)du$$

$$=h^*(-\tau)*r_{yx}(\tau)$$

$$=h^*(-\tau)*h(\tau)*r_{xx}(\tau)$$

Linear Systems and Power Spectral Densities

$$r_{xy}(\tau) = h^*(-\tau) * r_{xx}(\tau) \quad \Leftrightarrow S_{xy}(j\Omega) = H^*(j\Omega) S_{xx}(j\Omega)$$

$$r_{yx}(\tau) = h(\tau) * r_{xx}(\tau) = r_{xy}^*(-\tau) \quad \Leftrightarrow S_{yx}(j\Omega) = H(j\Omega) S_{xx}(j\Omega)$$

Since $r_{xx}^*(-\tau) = r_{xx}(\tau)$ and $\mathcal{F}\left\{r_{xx}^*(-\tau)\right\} = S_{xx}^*(j\Omega) = S_{xx}(-j\Omega)$

$$\Leftrightarrow S_{yx}(j\Omega) = H(j\Omega) S_{xx}(j\Omega) = H(j\Omega) S_{xx}(-j\Omega)$$

$$r_{yy}(\tau) = h^*(-\tau) r_{yx}(\tau) \quad \Leftrightarrow S_{yy}(j\Omega) = H^*(j\Omega) S_{yx}(j\Omega)$$

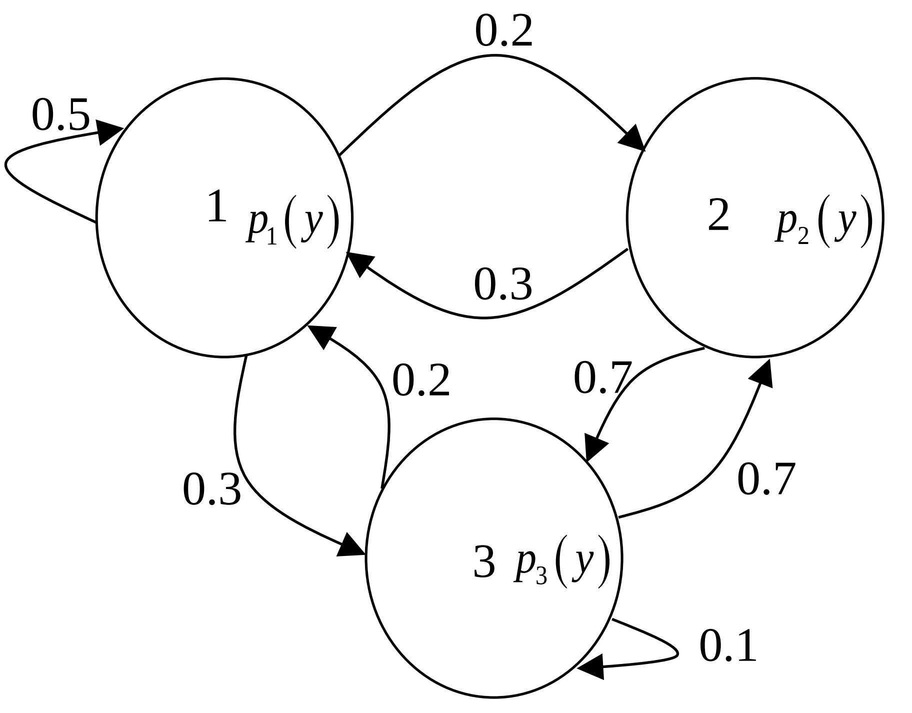
$$= h^*(-\tau) * h(\tau) * r_{xx}(\tau) \quad \Leftrightarrow \quad = H^*(j\Omega) H(j\Omega) S_{xx}(j\Omega) = \left|H(j\Omega)\right|^2 S_{xx}(j\Omega)$$

Markov and Hidden Markov Models (HMM)

- HMM is a stochastic model that is used to model time-varying random phenomena
  - E.g. speech signal, video sequence
  - Can be understood in terms of state-space models

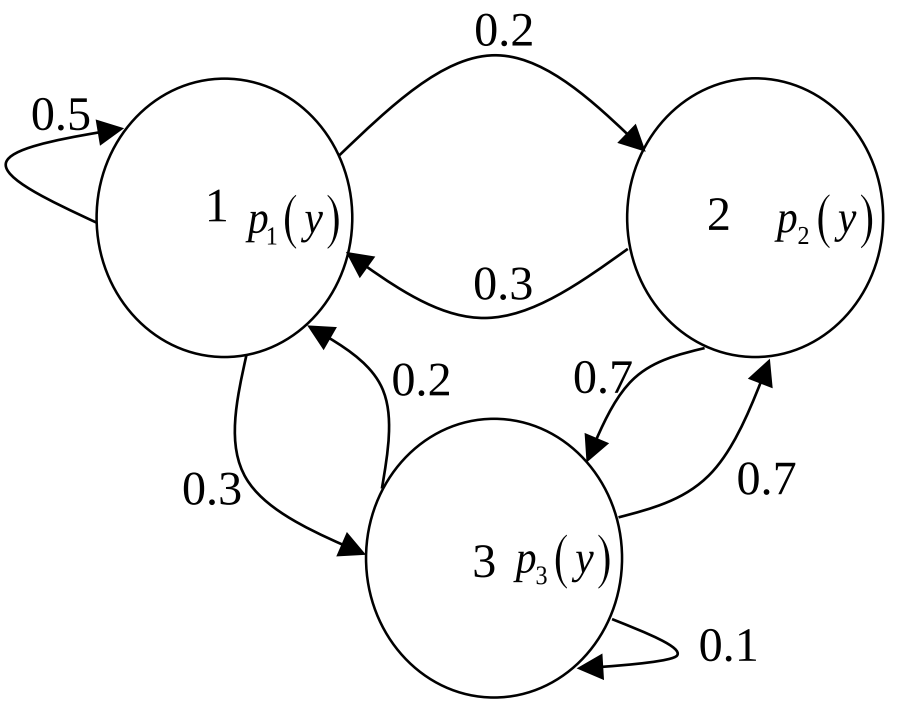
Markov Models

- Used to model evolution of random phenomena that can be in discrete states as a function to time,
  - Transition from one state to the next is random
- E.g. A system can be in one of the *S* distinct states
  - At each step of discrete time it can move to another state at random, with probability of the transition at the time *t* dependent only upon the state of the system at time *t*
    - i.e. only the previous state is relevant



Markov Models

- From state 1 to state 1 is possible with probability 0.5
- Denote $S[t]$ denote the state at time t, where it takes on one of the values 1,2, …, S.
- Initial state is selected according to a probability π
- $\pi_i = P(S[1] = i)$, $i = 1, 2, \ldots, S$



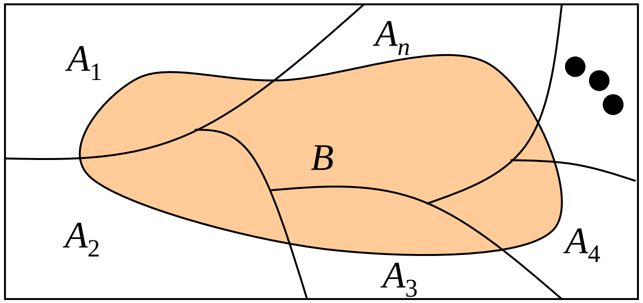
Markov Models

- Probability of transition depends ONLY upon the current state
  - $P(S[t+1] = j \mid S[t] = i, S[t-1] = k, S[t-2] = \ell, \ldots) = P(S[t+1] = j \mid S[t] = i)$
- This structure of probability is called the *Markov property*, and the random sequence of state values $S[0]$, $S[1]$, $S[2]$, … is called a *Markov sequence* or a Markov chain
- Sequence is the output of the Markov model
- Can determine the probability of arriving in the next state by adding up all the probabilities of the ways of arriving there, i.e.

$$P\left(S\left[t+1\right]=j\right)=P\left(S\left[t+1\right]=j \middle| S[t]=1\right)P\left(S[t]=1\right)+P\left(S\left[t+1\right]=j \middle| S[t]=2\right)P\left(S[t]=2\right)$$
$$+\cdots+P\left(S\left[t+1\right]=j \middle| S[t]=S\right)P\left(S[t]=S\right)$$

- Note that this is just the law of total probability

Partitions and Total Probability



Suppose the events $A_1, A_2, \ldots, A_n$ form a partition of a sample space S, that is, the events A_i's are mutually exclusive and their union is S. Suppose B is any other event. Then

$$B = S \cap B = \left(\bigcup_{i=1}^{n} A_i \right) \cap B$$

$$= (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B),$$

where $A_i \cap B$ are also mutually exclusive. Then

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B).$$

From the multiplication theorem,

$$P(B) = P(A_1) P(B | A_1) + P(A_2) P(B | A_2) + \cdots + P(A_n) P(B | A_n).$$

This is known as the **law of total probability**.

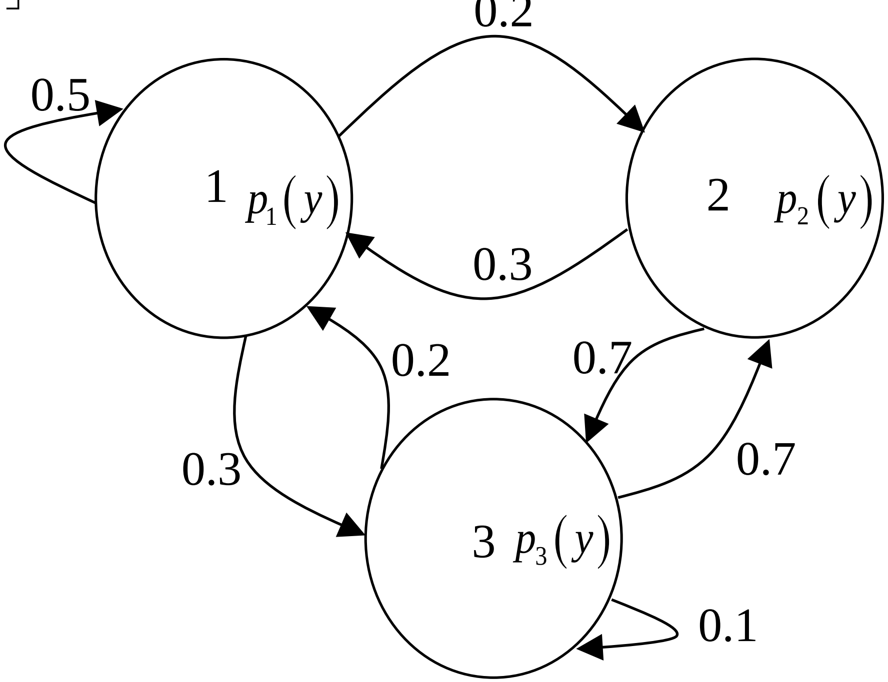
Markov Models

Can be written in matrix form. Define

$$\mathbf{p}[n] \triangleq \begin{bmatrix} P(S[n]=1) \\ P(S[n]=2) \\ \vdots \\ P(S[n]=S) \end{bmatrix}, \quad \mathbf{A} \triangleq \begin{bmatrix} P(1|1) & P(1|2) & \cdots & P(1|S) \\ P(2|1) & P(2|2) & \cdots & P(2|S) \\ \vdots & \vdots & \ddots & \vdots \\ P(S|1) & P(S|2) & \cdots & P(S|S) \end{bmatrix}, \text{ with } a_{ij} = P(i|j) \triangleq P(S[t+1]=i|S[t]=j).$$

From the previous example:

$$\mathbf{A} = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0 & 0.7 \\ 0.3 & 0.7 & 0.1 \end{bmatrix}$$



Markov Models

- A steady-state probability assignment is one that does not change from one time step to the next, so the probability must satisfy the equation

$$\mathbf{Ap}=\mathbf{p}$$

- This is an eigenequation, with eigenvalue = 1.
- By law of total probability, each column of **A** sum to 1
- Definition: An $m \times m$ matrix **P**, such that $\sum_{j=1}^{m} p_{ij} = 1$ (each row sums to 1) and each element of **P** is nonnegative, is called a *stochastic matrix*. If the rows and columns each sum to 1, then P is *doubly stochastic*

Markov Models

$$\mathbf{A} \triangleq \begin{bmatrix} P(1|1) & P(1|2) & \cdots & P(1|S) \\ P(2|1) & P(2|2) & \cdots & P(2|S) \\ \vdots & \vdots & \ddots & \vdots \\ P(S|1) & P(S|2) & \cdots & P(S|S) \end{bmatrix}$$ is the transpose of a stochastic matrix. The vector $\boldsymbol{\pi}$ contains the initial probabilities. Thus, we can write the probabilistic update equation is

$$\mathbf{p}[t+1] = \mathbf{A}\mathbf{p}[t], \quad \text{with } \mathbf{p}[0] = \boldsymbol{\pi}.$$

Or,

$$\mathbf{p}[t+1] = \mathbf{A}\mathbf{p}[t] + \boldsymbol{\pi}\delta_t,$$

with $\mathbf{p}[t] = 0$ for $t \leq 0$. Note that the above is similar to the state equation

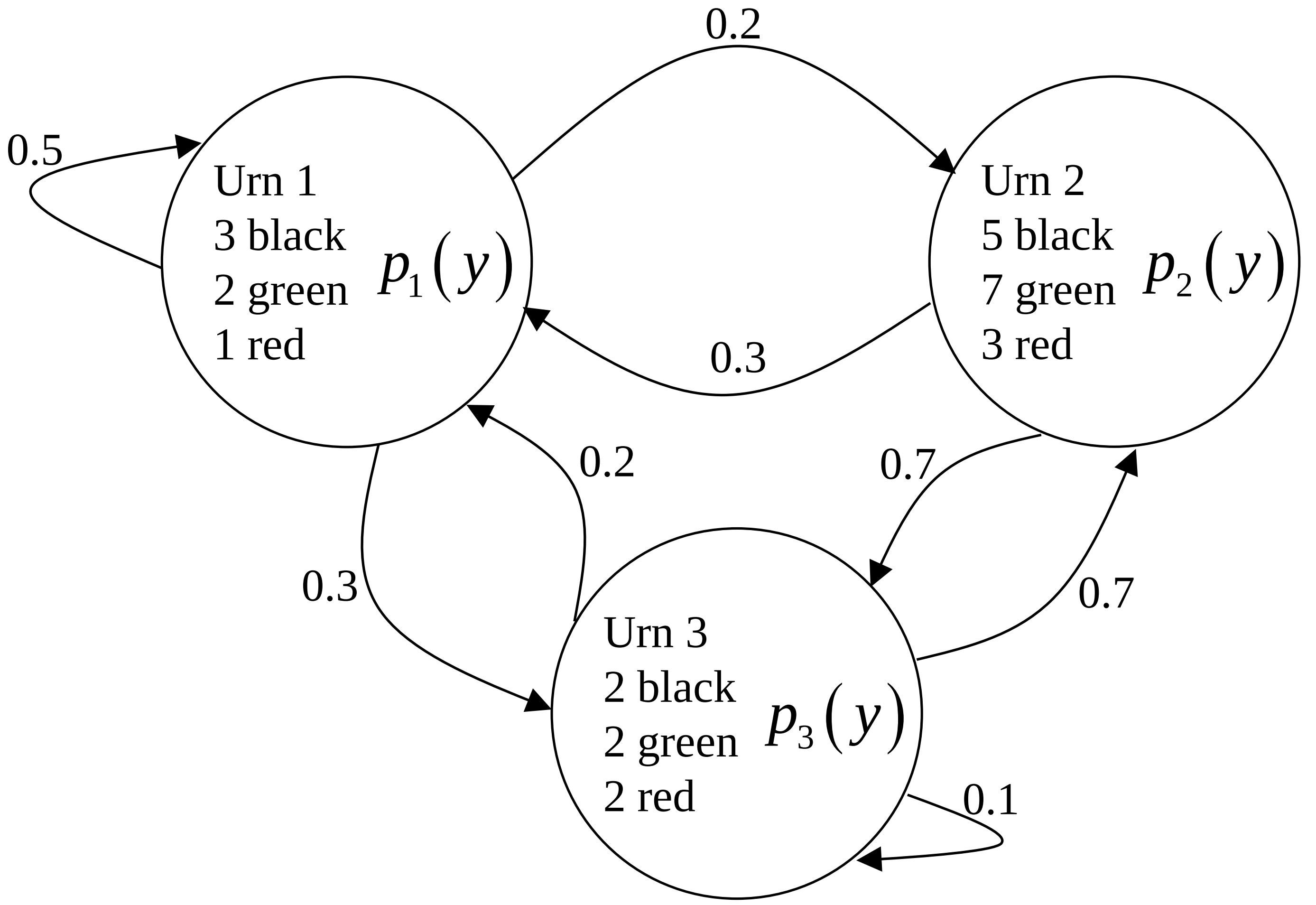
$$\mathbf{x}[n+1] = \mathbf{A}\mathbf{x}[n] + \mathbf{b}f[n].$$

Note that the "state" represented by $\mathbf{p}[t+1] = \mathbf{A}\mathbf{p}[t] + \boldsymbol{\pi}\delta_t$ is actually the vector of probabilities $\mathbf{p}[t]$, not the state of the Markov sequence $S[t]$

Relationship to Markov Models and HMM

- Pick a ball from 3 urns
- Each urn contains 3 types of colored balls: black green, and red
- At each instant of time, an urn is selected by genie at random according to the state it was in at the previous time instant
- Genie – magic creature which could do everything
- Ball is then drawn at random from the urn at time *t*
- Observation = ball selected
- Actual state is hidden
  - State of the system before the ball was chosen ➔ the state of the system after

Relationship to Markov Models and HMM: State Diagram



Relationship to Markov Models and HMM

- To further clarify the relationship,

$$\mathbf{p}[t+1] = \mathbf{A}\mathbf{p}[t] + \boldsymbol{\pi}\delta_t$$

  provides for the state update of the Markov system.
- However, in most linear system, the state vector $\mathbf{x}[t]$ is not directly observable, instead, it is observed only through the observation matrix $\mathbf{C}$ (assuming $\mathbf{D} = \mathbf{0}$), i.e. $\mathbf{y}[t] = \mathbf{C}\mathbf{x}[t]$
- In an HMM, the state is hidden from direct observation
- Instead, each state has a probability distribution associated with it

Relationship to Markov Models and HMM

- In the HMM, we do not observe the "state" $p[t]$
  - Instead, each state has a probability distribution associated with it
- When HMM moves into state $s[t]$ at time t, the observed output $y[t]$ is an outcome of a random variable $Y[t]$ that is selected according to the distribution $f(y[t]|S[t] = s)$, which we will represent using the notation $f(y|S[t]=s) = f_s(y)$
- In the urn example, the output probabilities depend on the contents of the urns
- A sequence of outputs from an HMM is $y[0]$, $y[1]$, $y[2]$, …
- The underlying state information is hidden
- Distribution in each state can be of any type
  - Each state could have its own distribution
  - In practice, distribution of each state is the same, but with different parameters

Summary: HMM

Denote the state at time t as $S[t]$.

Initial state is selected according to probability $\pi_i = P(S[1]=i), \quad i=1,2,\ldots,S$

$(\text{assume } P(S[t]=i)=0, \text{ for } t \leq 0)$.

Transition probability depends ONLY on current state:

$$P(S[t+1]=j \mid S[t]=i, S[t-1]=k, S[t-2]=\ell, \ldots) = P(S[t+1]=j \mid S[t]=i)$$

Then, the probability of arriving in the next state is

$$P(S[t+1]=j) = P(S[t+1]=j \mid S[t]=1) P(S[t]=1) + P(S[t+1]=j \mid S[t]=2) P(S[t]=2)$$
$$+ \cdots + P(S[t+1]=j \mid S[t]=S) P(S[t]=S)$$

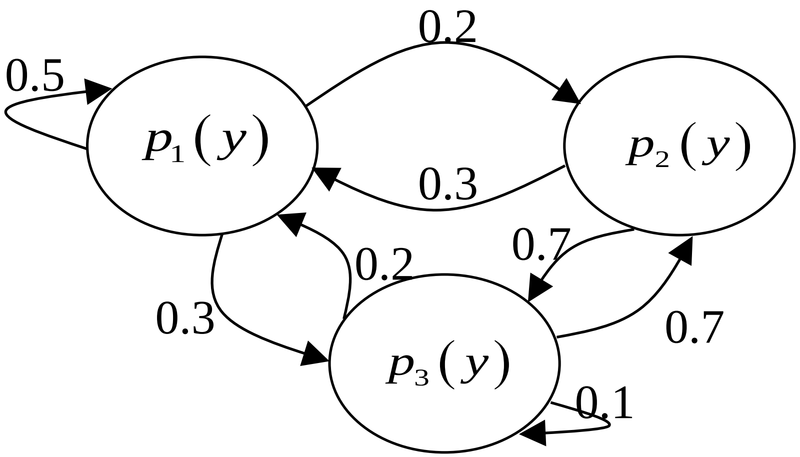
Summary: HMM State Transition

Can be written in matrix form. Define

$$\mathbf{p}[t] \triangleq \begin{bmatrix} P(S[t]=1) \\ P(S[t]=2) \\ \vdots \\ P(S[t]=S) \end{bmatrix}, \quad \mathbf{A} \triangleq \begin{bmatrix} P(1|1) & P(1|2) & \cdots & P(1|S) \\ P(2|1) & P(2|2) & \cdots & P(2|S) \\ \vdots & \vdots & \ddots & \vdots \\ P(S|1) & P(S|2) & \cdots & P(S|S) \end{bmatrix}, \quad \text{with } a_{ij} = P(i|j) \triangleq P(S[t+1]=i|S[t]=j).$$

From urn example:

$$\mathbf{A} = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0 & 0.7 \\ 0.3 & 0.7 & 0.1 \end{bmatrix}$$



$$\Rightarrow \mathbf{p}[t+1] = \mathbf{A}\mathbf{p}[t] + \boldsymbol{\pi}\delta_t,$$

$$\text{with } \boldsymbol{\pi} \triangleq \mathbf{p}[0]$$

Summary: HMM Input-Output

Let M denote the number of possible outcomes from all states

Let $Y[t]$ be the random variable output at time t, with outcome $y[t]$

Then probability of each possible output is

$$P\left(Y[t]=j\right)=P\left(Y[t]=j\,\middle|\,S[t]=1\right)P\left(S[t]=1\right)+P\left(Y[t]=j\,\middle|\,S[t]=2\right)P\left(S[t]=2\right)$$
$$+\cdots+P\left(Y[t]=j\,\middle|\,S[t]=S\right)P\left(S[t]=S\right)$$

Summary: HMM Input-Output

Can be written in matrix form. Define

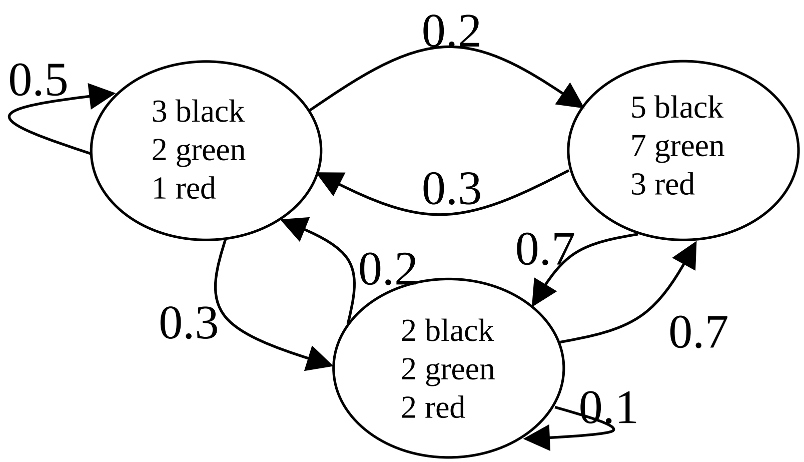
$$\mathbf{q}[t] \triangleq \begin{bmatrix} P(Y[t]=1) \\ P(Y[t]=2) \\ \vdots \\ P(Y[t]=M) \end{bmatrix}, \quad \mathbf{C} \triangleq \begin{bmatrix} P(Y[t]=1|S[t]=1) & P(Y[t]=1|S[t]=2) & \cdots & P(Y[t]=1|S[t]=S) \\ P(Y[t]=2|S[t]=1) & P(Y[t]=2|S[t]=2) & \cdots & P(Y[t]=2|S[t]=S) \\ \vdots & \vdots & \ddots & \vdots \\ P(Y[t]=M|S[t]=1) & P(Y[t]=M|S[t]=2) & \cdots & P(Y[t]=M|S[t]=S) \end{bmatrix},$$

with $c_{ij} = P(Y[t]=i|S[t]=j)$.

From urn example, with $S=1$ (black), $=2$ (green), $=3$ (red):

$$\mathbf{C} = \begin{bmatrix} 1/2 & 1/3 & 1/3 \\ 1/3 & 7/15 & 1/3 \\ 1/6 & 1/5 & 1/3 \end{bmatrix}$$

$$\Rightarrow \mathbf{q}[n] = \mathbf{C}\mathbf{p}[n]$$



State-Space vs. HMM

State-space:

$$\mathbf{x}[n+1] = \mathbf{A}\mathbf{x}[n] + \mathbf{B}u[n]$$

$$\mathbf{y}[n] = \mathbf{C}\mathbf{x}[n] + \mathbf{D}\mathbf{u}[n]$$

HMM:

$$\mathbf{p}[n+1] = \mathbf{A}\mathbf{p}[n] + \boldsymbol{\pi}\delta_n, \quad \text{with } \boldsymbol{\pi} \triangleq \mathbf{p}[0]$$

$$\mathbf{q}[n] = \mathbf{C}\mathbf{p}[n]$$

Recall solution for state-space is

$$\mathbf{x}[n] = \mathbf{A}^{n+1}\mathbf{x}[-1] + \sum_{k=0}^{n} \mathbf{A}^{k}\mathbf{B}\mathbf{u}[n-k]$$

$$\mathbf{y}[n] = \mathbf{C}\mathbf{A}^{n+1}\mathbf{x}[-1] + \sum_{k=0}^{n} \mathbf{C}\mathbf{A}^{k}\mathbf{B}\mathbf{u}[n-k] + \mathbf{D}\mathbf{u}[n].$$

$\Rightarrow$ Hence, $\mathbf{A}^k$ models dynamics of system by treating the system as a Markov process.

Example: Speech Modeling for Speech Recognition

- Patterns in speech signal occurring sequentially in time
- Each word or sound (phoneme) to be recognized is represented by an HMM
  - Output is some feature vector that is derived from the speech data
  - Random variability in the feature vector and the amount of time each feature is produced is modeled by the HMM
  - Variability in the duration of the word is modeled by the Markov model
  - Variability in the outputs is modeled by the random selection from within each state

Example

- Given a small vocabulary system with N words
  - There are N HMMs: (A_i, π_i, C_i)
  - i denotes a particular state
- Training phase
  - Each is trained to represent the parameters for that word
- Testing phase
  - Sequence of feature vectors is computed (front end part)
  - The likelihood (probability) that this sequence of feature vectors was produced by the HMM (A_i, π_i, C_i) is computed for each i
  - HMM that produces the highest probability selects the recognized word

Issues for HMM

- Training:
  - How can the parameters (**A**, **π**, **C**) be estimated based upon observations of the data?
    - In other words, how can we train the parameters of the models in the pattern recognition problem?
- Testing
  - How can we determine how well the observed data fits the model that has been trained?
  - How can we determine the sequence of states of the underlying Markov model?
    - I.e. How do we discover the hidden states?