Lecture 3: Review of Probability Theory

Connection with control systems
Random variable, distribution
Multiple random variables
Random process, filtering a random process

Big picture

why are we learning this:

We have been very familiar with deterministic systems:

$$x(k+1) = Ax(k) + Bu(k)$$

In practice, we commonly have:

$$x(k+1) = Ax(k) + Bu(k) + B_w w(k)$$

where $w(k)$ is the noise term that we have been neglecting. With the introduction of $w(k)$, we need to equip ourselves with some additional tool sets to understand and analyze the problem.

Sample space, events and probability axioms

- ▶ experiment: a situation whose outcome depends on chance
- ▶ trial: each time we do an experiment we call that a trial

Example (Throwing a fair dice)

possible outcomes in one trail: getting a ONE, getting a TWO, ...

- ▶ sample space Ω: includes all the possible outcomes

- ▶ probability: discusses how likely things, or more formally, events, happen
- ▶ an event S_i: includes some (maybe 1, maybe more, maybe none) outcomes of the sample space. e.g., the event that it won't rain tomorrow; the event that getting odd numbers when throwing a dice

Sample space, events and probability axioms

probability axioms
- ▶ $\Pr\{S_j\} \geq 0$
- ▶ $\Pr\{\Omega\} = 1$
- ▶ if $S_i \cap S_j = \varnothing$ (empty set), then $\Pr\{S_i \cup S_j\} = \Pr\{S_i\} + \Pr\{S_j\}$

Example (Throwing a fair dice)

the sample space:

$$\Omega = \{\underbrace{\text{getting a ONE}}_{\omega_1}, \underbrace{\text{getting a TWO}}_{\omega_2}, \ldots, \underbrace{\text{getting a SIX}}_{\omega_6}\}$$

the event S_1 of observing an even number:

$$S_1 = \{\omega_2, \omega_4, \omega_6\}$$
$$\Pr\{S_1\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

Random variables

to better measure probabilities, we introduce random variables (r.v.'s)

- ▶ r.v.: a real valued function $X(\omega)$ defined on Ω; $\forall x \in \mathbf{R}$ there defined the *(probability) cumulative distribution function (cdf)*

$$F(x) = \Pr\{X \le x\}$$

- ▶ cdf $F(x)$: non-decreasing, $0 \le F(x) \le 1$, $F(-\infty) = 0$, $F(\infty) = 1$

Example (Throwing a fair dice)

can define X: the obtained number of the dice

$$X(\omega_1) = 1, \ X(\omega_2) = 2, \ X(\omega_3) = 3, \ X(\omega_4) = 4, \ldots$$

can also define X: indicator of whether the obtained number is even

$$X(\omega_1) = X(\omega_3) = X(\omega_5) = 0, \ X(\omega_2) = X(\omega_4) = X(\omega_6) = 1$$

Probability density and moments of distributions

- ▶ probability density function (pdf):

$$p(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

$$\Pr(a < X \le b) = \int_a^b p(x)\,\mathrm{d}x, \ a < b$$

sometimes we write $p_X(x)$ to emphasize that it is for the r.v. X
- ▶ mean, or expected value (first moment):

$$m_X = \mathrm{E}[X] = \int_{-\infty}^{\infty} x p_X(x)\,\mathrm{d}x$$

- ▶ variance (second moment):

$$\mathrm{Var}[X] = \mathrm{E}\left[(X - m_X)^2\right] = \int_{-\infty}^{\infty} (x - m_X)^2 \, p_X(x)\,\mathrm{d}x$$

- ▶ standard deviation (std): $\sigma = \sqrt{\mathrm{Var}[X]}$
- ▶ exercise: prove that $\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2$

Example distributions

uniform distribution

- ▶ a r.v. uniformly distributed between $x_{\min}$ and $x_{\max}$
- ▶ probability density function:

$$p(x) = \frac{1}{x_{\max} - x_{\min}}$$
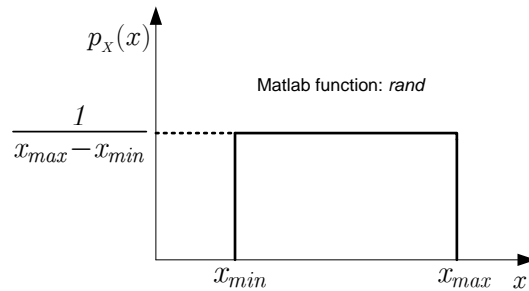
- ▶ cumulative distribution function:

$$F(x) = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad x_{\min} \leq x \leq x_{\max}$$
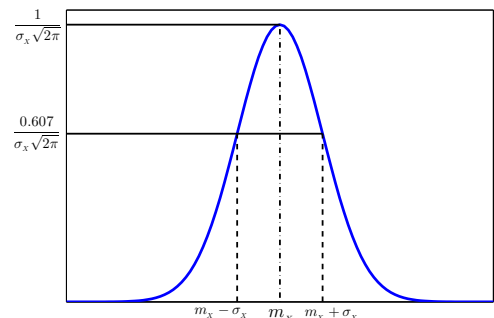
- ▶ mean and variance:

$$E[X] = \frac{1}{2}(x_{\max} + x_{\min}), \quad \text{Var}[X] = \frac{(x_{\max} - x_{\min})^2}{12}$$

Example distributions

Gaussian/normal distribution

- ▶ importance: sum of independent r.v.s $\to$ a Gaussian distribution
- ▶ probability density function:

$$p(x) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left(-\frac{(x - m_X)^2}{2\sigma_X^2}\right)$$

- ▶ pdf fully characterized by m_X and σ_X. Hence a normal distribution is usually denoted as $N(m_X, \sigma_X)$

- ▶ nice properties: if X is Gaussian and Y is a linear function of X, then Y is Gaussian

Example distributions

Central Limit Theorem: if X_1, X_2, ... are independent identically distributed random variables with mean m_X and variance σ_X^2, then

$$Z_n = \frac{\sum_{k=1}^{n}(X_k - m_X)}{\sqrt{n}\sigma_X^2}$$

converges in distribution to a normal random variable $X \sim N(0,1)$

example: sum of uniformly distributed random variables in $[0,1]$

```
X1 = rand(1,1e5);
X2 = rand(1,1e5);
X3 = rand(1,1e5);
Z = X1 + X2;
[fz,x] = hist(Z,100);
w_fz = x(end)/length(fz);
fz = fz/sum(fz)/w_fz;
figure, bar(x,fz)
xlabel 'x'; ylabel 'p_Z(x))';
Y = X1 + X2 + X3;
% ...
```
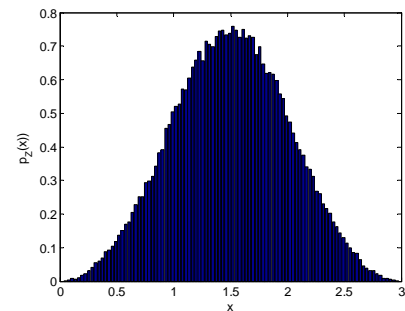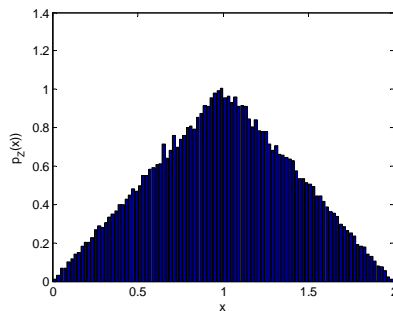
Multiple random variables

for the same sample space Ω, multiple r.v.'s can be defined

- ▶ joint probability: $\Pr(X = x, Y = y)$
- ▶ joint cdf:

$$F(x,y) = \Pr(X \le x, Y \le y)$$

- ▶ joint pdf: $p(x,y) = \frac{\partial^2}{\partial x \partial y}F(x,y)$
- ▶ covariance:

$$\mathrm{Cov}(X,Y) = \Sigma_{XY} = \mathrm{E}[(X - m_X)(Y - m_Y)] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - m_X)(y - m_Y)p(x,y)\,\mathrm{d}x\mathrm{d}y$$

- ▶ uncorrelated: $\Sigma_{XY} = 0$
- ▶ independent random variables satisfy:

$$F(x,y) = \Pr(X \le x, Y \le y) = \Pr(X \le x)\Pr(Y \le y) = F_X(x)F_Y(y)$$
$$p(x,y) = p_X(x)p_Y(y)$$

Multiple random variables

more about correlation

correlation coefficient:
$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

X and Y are uncorrelated if $\rho(X,Y) = 0$

- independent $\Rightarrow$ uncorrelated; uncorrelated $\not\Rightarrow$ independent
- uncorrelated indicates $\text{Cov}(X,Y) = E[XY] - E[X]E[Y] = 0$, which is weaker than X and Y being independent

Example

X–uniformly distributed on $[-1,1]$. Construct Y: if $X \leq 0$ then $Y = -X$; if $X > 0$ then $Y = X$. X and Y are uncorrelated due to

- $E[X] = 0$, $E[Y] = \frac{1}{2}$
- $E[XY] = 0$

however X and Y are clearly dependent

Multiple random variables

random vector

- vector of r.v.'s:
$$Z = \begin{bmatrix} X \\ Y \end{bmatrix}$$

- mean:
$$m_Z = \begin{bmatrix} m_X \\ m_Y \end{bmatrix}$$

- covariance matrix:

$$\Sigma = E\left[(Z - m_Z)(Z - m_Z)^T\right] = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \begin{bmatrix} (X - m_X)^2 & (X - m_X)(Y - m_Y) \\ (Y - m_Y)(X - m_X) & (Y - m_Y)^2 \end{bmatrix} p(x,y)\,dxdy$$

Conditional distributions

▸ joint pdf to single pdf:

$$p_X(x) = \int_{-\infty}^{\infty} p(x,y)\,dy$$

▸ conditional pdf:

$$p_X(x|y_1) = p_X(x|Y=y_1) = \frac{p(x,y_1)}{p_Y(y_1)}$$

▸ conditional mean:

$$E[X|y_1] = \int_{-\infty}^{\infty} x p_X(x|y_1)\,dx$$

▸ note: independent $\Rightarrow p_X(x|y_1) = p_X(x)$

▸ properties of conditional mean:

$$\underset{y}{E}[E[X|y]] = E[X]$$

Multiple random variables

Gaussian random vectors

Gaussian r.v. is particularly important and interesting as its pdf is mathematically sound

Special case: two independent Gaussian r.v. X_1 and X_2

$$p(x_1, x_2) = p_{X_1}(x_1) p_{X_2}(x_2) = \frac{1}{\sigma_{X_1}\sqrt{2\pi}} e^{-\left(x_1 - m_{X_1}\right)^2 / \left(2\sigma_{X_1}^2\right)} \frac{1}{\sigma_{X_2}\sqrt{2\pi}} e^{-\left(x_2 - m_{X_2}\right)^2 / \left(2\sigma_{X_2}^2\right)}$$

$$= \frac{1}{\sigma_{X_1}\sigma_{X_2}\left(\sqrt{2\pi}\right)^2} \exp\left\{ -\frac{1}{2} \begin{bmatrix} x_1 - m_{X_1} \\ x_2 - m_{X_2} \end{bmatrix}^T \begin{bmatrix} \sigma_{X_1}^2 & 0 \\ 0 & \sigma_{X_2}^2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - m_{X_1} \\ x_2 - m_{X_2} \end{bmatrix} \right\}$$

We can use the random vector notation: $X = [X_1, X_2]^T$

$$\Sigma = \begin{bmatrix} \sigma_{X_1}^2 & 0 \\ 0 & \sigma_{X_2}^2 \end{bmatrix}$$

and write

$$p_X(x) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{\det \Sigma}} \exp\left\{ -\frac{1}{2}[X - m_X]^T \Sigma^{-1} [X - m_X] \right\}$$

General Gaussian random vectors

pdf for a n-dimensional jointly distributed Gaussian random vector X:

$$p_X(x) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}[X - m_X]^T \Sigma^{-1} [X - m_X]\right\} \quad (1)$$

joint pdf for 2 Gaussian random vectors X (n-dimensional) and Y (m-dimensional):

$$p(x,y) = \frac{1}{\left(\sqrt{2\pi}\right)^{n+m} \sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}\begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}\right\}$$
$$(2)$$

$$\Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}$$

where Σ_{XY} is the cross covariance (matrix) between X and Y

$$\Sigma_{XY} = \mathsf{E}\left[(X - m_X)(Y - m_Y)^T\right] = \mathsf{E}\left[(Y - m_Y)(X - m_X)^T\right]^T = \Sigma_{YX}^T$$

General Gaussian random vectors

conditional mean and covariance
important facts about conditional mean and covariance:

$$m_{X|y} = m_X + \Sigma_{XY}\Sigma_{YY}^{-1}[y - m_Y]$$
$$\Sigma_{X|y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$$

proof uses $p(x,y) = p(x|y)p(y)$, (1), and (2)

▶ getting $\det \Sigma$ and the inverse Σ^{-1}: do a transformation

$$\begin{bmatrix} I & -\Sigma_{XY}\Sigma_{YY}^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}\begin{bmatrix} I & 0 \\ -\Sigma_{YY}^{-1}\Sigma_{YX} & I \end{bmatrix}$$
$$= \begin{bmatrix} \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} & 0 \\ 0 & \Sigma_{YY} \end{bmatrix} \quad (3)$$

hence
$$\det \Sigma = \det \Sigma_{YY} \det\left(\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right) \quad (4)$$

General Gaussian random vectors

inverse of the covariance matrix
computing the inverse Σ^{-1}:
−(3) gives

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} I & 0 \\ -\Sigma_{YY}^{-1}\Sigma_{YX} & I \end{bmatrix} \begin{bmatrix} \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} & 0 \\ 0 & \Sigma_{YY} \end{bmatrix}^{-1} \begin{bmatrix} I & -\Sigma_{XY}\Sigma_{YY}^{-1} \\ 0 & I \end{bmatrix}$$

−hence in (2):

$$\begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}$$

$$= \begin{bmatrix} \star \end{bmatrix}^T \begin{bmatrix} \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} & 0 \\ 0 & \Sigma_{YY} \end{bmatrix}^{-1} \underbrace{\begin{bmatrix} x - (m_X + \Sigma_{XY}\Sigma_{YY}^{-1}[y - m_Y]) \\ y - m_Y \end{bmatrix}}_{[\star]}$$

$$(5)$$

General Gaussian random vectors

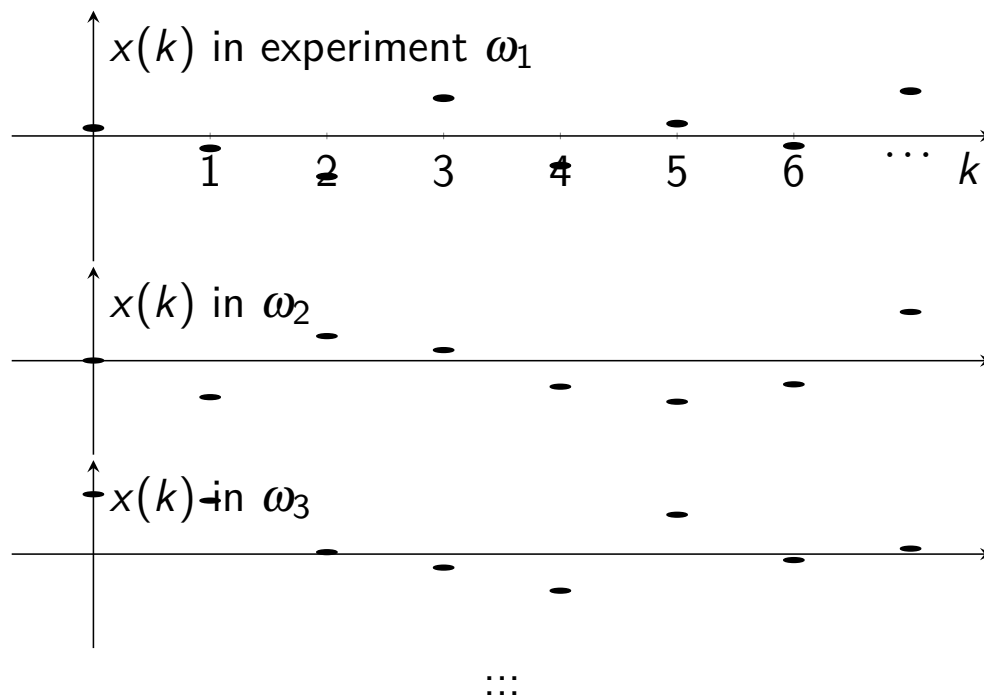$p(x,y) = p(x|y)p(y) \Rightarrow p(x|y) = p(x,y)/p(y)$

▶ using (4) and (5) in (2), we get

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \underbrace{(\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX})}_{[\star\star]}}}$$

$$\times \exp\left\{ -\frac{1}{2} \begin{bmatrix} \underline{\ldots} \end{bmatrix}^T [\star\star]^{-1} \begin{bmatrix} \underline{x - (m_X + \Sigma_{XY}\Sigma_{YY}^{-1}[y - m_Y])} \end{bmatrix} \right\}$$

hence $X|y$ is also Gaussian, with

$$m_{X|y} = m_X + \Sigma_{XY}\Sigma_{YY}^{-1}[y - m_Y]$$

$$\Sigma_{X|y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$$

Random process

- discrete-time random process: a random variable evolving with time $\{x(k),\ k=1,2,\dots\}$
- a stack of random vectors: $x(k)=[x(1),x(2),\dots]^T$

$x(k)$ in experiment ω_1

$x(k)$ in ω_2

$x(k)$ in ω_3

:::

Random process

$x(k)=[x(1),x(2),\dots]^T$:

- complete probabilistic properties defined by the joint pdf $p(x(1),x(2),\dots)$, which is usually difficult to get
- usually sufficient to know the mean $\mathsf{E}[x(k)]=m_x(k)$ and *auto-covariance*:

$$\mathsf{E}[(x(j)-m_x(j))(x(k)-m_x(k))]=\Sigma_{xx}(j,k) \qquad (6)$$

- sometimes $\Sigma_{xx}(j,k)$ is also written as $X_{xx}(j,k)$

Random process

let $x(k)$ be a 1-d random process

- time average of $x(k)$:

$$\overline{x(k)} = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{j=-N}^{N} x(j)$$

- ensemble average:

$$E[x(k)] = m_{x(k)}$$

- *ergodic* random process: for all moments of the distribution, the ensemble averages equal the time averages

$$E[x(k)] = \overline{x(k)}, \ \Sigma_{xx}(j,k) = \overline{[x(j)-m_x][x(k)-m_x]}, \ldots$$

- ergodicity: not easy to test but many processes in practice are ergodic; extremely important as large samples can be expensive to collect in practice
- one necessary condition for ergodicity is stationarity

Random process

stationarity: tells whether the statistics characteristics changes w.r.t. time

- *stationary in the strict sense*: probability distribution does not change w.r.t. time

$$Pr\{x(k_1) \le x_1, \ldots, x(k_n) \le x_n\} = Pr\{x(k_1+l) \le x_1, \ldots, x(k_n+l) \le x_n\}$$

- *stationary in the week/wide sense*: mean does not dependent on time

$$E[x(k)] = m_x = \text{costant}$$

and the auto-covariance (6) depends only on the time difference $l = j - k$
- can hence write

$$E[(x(k)-m_x)(x(k+l)-m_x)] = \Sigma_{xx}(l) = X_{xx}(l)$$

- for stationary and ergodic random processes:

$$\Sigma_{xx}(l) = E[(x(k)-m_x)(x(k+l)-m_x)] = \overline{(x(k)-m_x)(x(k+l)-m_x)}$$

Random process
covariance and correlation for stationary ergodic processes

- ▶ we will assume stationarity and ergodicity unless otherwise stated
- ▶ *auto-correlation*: $R_{xx}(l) = E[x(k)x(k+l)]$.

- ▶ *cross-covariance*:

$$\Sigma_{xy}(l) = X_{xy}(l) = E[(x(k) - m_x)(y(k+l) - m_y)]$$

- ▶ property (using ergodicity):

$$\Sigma_{xy}(l) = X_{xy}(l) = \overline{(x(k) - m_x)(y(k+l) - m_y)}$$
$$= \overline{(y(k+l) - m_y)(x(k) - m_x)} = X_{yx}(-l) = \Sigma_{yx}(-l)$$

Random process
white noise

- ▶ *white* noise: a purely random process with $x(k)$ not correlated with $x(j)$ at all if $k \neq j$:

$$X_{xx}(0) = \sigma_{xx}^2, \ X_{xx}(l) = 0 \ \forall l \neq 0$$

- ▶ non-stationary zero mean white noise:

$$E[x(k)x(j)] = Q(k)\delta_{kj}, \ \delta_{kj} = \begin{cases} 1 & , \ k = j \\ 0 & , \ k \neq j \end{cases}$$

Random process

auto-covariance and spectral density

- ▸ *spectral density*: the Fourier transform of auto-covariance

$$\Phi_{xx}(\omega) = \sum_{l=-\infty}^{\infty} X_{xx}(l) e^{-j\omega l}, \quad X_{xx}(l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega l} \Phi_{xx}(\omega) \, d\omega$$

- ▸ *cross spectral density*:

$$\Phi_{xy}(\omega) = \sum_{l=-\infty}^{\infty} X_{xy}(l) e^{-j\omega l}, \quad X_{xy}(l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega l} \Phi_{xy}(\omega) \, d\omega$$

properties:

- ▸ the variance of x is the area under the spectral density curve

$$\text{Var}[x] = E\left[(x - E[x])^2\right] = X_{xx}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{xy}(\omega) \, d\omega$$

- ▸ $X_{xx}(0) \geq |X_{xx}(l)|, \ \forall l$

Filtering a random process

passing a random process $u(k)$ through an LTI system (convolution) generates another random process:

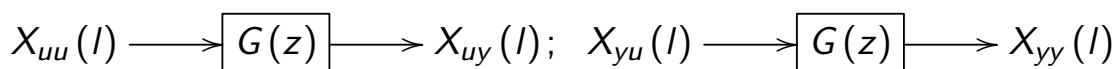$$y(k) = g(k) * u(k) = \sum_{i=-\infty}^{\infty} g(i) u(k-i)$$

- ▸ if u is *zero mean* and ergodic, then

$$X_{uy}(l) = \overline{u(k) \sum_{i=-\infty}^{\infty} u(k+l-i) g(i)}$$

$$= \sum_{i=-\infty}^{\infty} \overline{u(k) u(k+l-i)} g(i) = \sum_{i=-\infty}^{\infty} X_{uu}(l-i) g(i) = g(l) * X_{uu}(l)$$

similarly

$$X_{yy}(l) = \sum_{i=-\infty}^{\infty} X_{yu}(l-i) g(i) = g(l) * X_{yu}(l)$$

- ▸ in pictures:

$$X_{uu}(l) \longrightarrow \boxed{G(z)} \longrightarrow X_{uy}(l); \quad X_{yu}(l) \longrightarrow \boxed{G(z)} \longrightarrow X_{yy}(l)$$

Filtering a random process

input-output spectral density relation
for a general LTI system

$$u(k) \longrightarrow \boxed{G(z) = \frac{b_n z^n + b_{n-1} z^{n-1} + \cdots + b_0}{z^n + a_{n-1} z^{n-1} + \cdots + a_0}} \longrightarrow y(k)$$

$$Y(z) = G(z) U(z) \Leftrightarrow Y(e^{j\omega}) = G(e^{j\omega}) U(e^{j\omega})$$

▶ auto-covariance relation in the last slide:

$$X_{uu}(l) \longrightarrow \boxed{G(z)} \longrightarrow X_{uy}(l); \quad X_{yu}(l) \longrightarrow \boxed{G(z)} \longrightarrow X_{yy}(l)$$

$$X_{yu}(l) = X_{uy}(-l) = g(-l) * X_{uu}(-l) = g(-l) * X_{uu}(l)$$

hence

$$\boxed{\Phi_{yy}(\omega) = G(e^{j\omega}) G(e^{-j\omega}) \Phi_{uu}(\omega) = \left| G(e^{j\omega}) \right|^2 \Phi_{uu}(\omega)}$$

Filtering a random process

MIMO case:

▶ if u and y are vectors, $G(z)$ becomes a transfer function matrix
▶ dimensions play important roles:

$$X_{uy}(l) = E\left[(u(k) - m_u)(y(k+l) - m_y)^T \right] = X_{yu}(-l)^T$$

$$X_{uu}(l) \longrightarrow \boxed{G(z)} \longrightarrow X_{uy}(l); \quad X_{yu}(l) \longrightarrow \boxed{G(z)} \longrightarrow X_{yy}(l)$$

$$\begin{aligned} X_{yy}(l) &= g(l) * X_{yu}(l) = g(l) * X_{uy}^T(-l) \\ &= g(l) * [g(-l) * X_{uu}(-l)]^T \end{aligned}$$

$$\boxed{\Phi_{yy}(e^{j\omega}) = G(e^{j\omega}) \cdot \Phi_{uu}(e^{j\omega}) G^T(e^{-j\omega})}$$

Filtering a random process in state space

consider: $w(k)$–zero mean, white, $E[w(k)w(k)^T] = W(k)$ and

$$x(k+1) = A(k)x(k) + B_w(k)w(k) \qquad (7)$$

assume random initial state $x(0)$ (uncorrelated to $w(k)$):

$$E[x(0)] = m_{x_0}, \; E\left[(x(0) - m_{x_0})(x(0) - m_{x_0})^T\right] = X_0$$

▸ mean of state vector $x(k)$:

$$\boxed{m_x(k+1) = A(k)m_x(k)}, \; m_x(0) = m_{x_0} \qquad (8)$$

▸ covariance $X(k) = X_{xx}(k,k)$: (7)-(8)$\Rightarrow$

$$\boxed{X(k+1) = A(k)X(k)A^T(k) + B_w(k)W(k)B_w^T(k)}, \; X(0) = X_o$$

▸ intuition: covariance is a "second-moment" statistical property

Filtering a random process in state space

dynamics of the mean:

$$\boxed{m_x(k+1) = A(k)m_x(k)}, \; m_x(0) = m_{x_0}$$

dynamics of the covariance:

$$\boxed{X(k+1) = A(k)X(k)A^T(k) + B_w(k)W(k)B_w^T(k)}, \; X(0) = X_o$$

▸ (steady state) if $A(k) = A$ and is stable, $B_w(k) = B_w$, and $w(k)$ is stationary $W(k) = W$, then

$$m_x(k) \to 0, \; X(k) \to \text{a steady state } X_{ss}$$

$$X_{ss} = AX_{ss}A^T + B_wWB_w^T: \text{ discrete-time Lyapunov Eq.} \qquad (9)$$

$$X_{ss}(l) = E\left[x(k)x^T(k+l)\right] = X_{ss}\left(A^T\right)^l$$

$$X_{ss}(-l) = X_{ss}(l)^T = A^l X_{ss}$$

Filtering a random process in state space

Example

first-order system

$$x(k+1) = ax(x) + \sqrt{1-a^2}\, w(k), \quad \mathsf{E}[w(k)] = 0, \quad \mathsf{E}[w(k)w(j)] = W\delta_{kj}$$

with $|a| < 1$ and $x(0)$ uncorrelated with $w(k)$.
steady-state variance equation (9) becomes

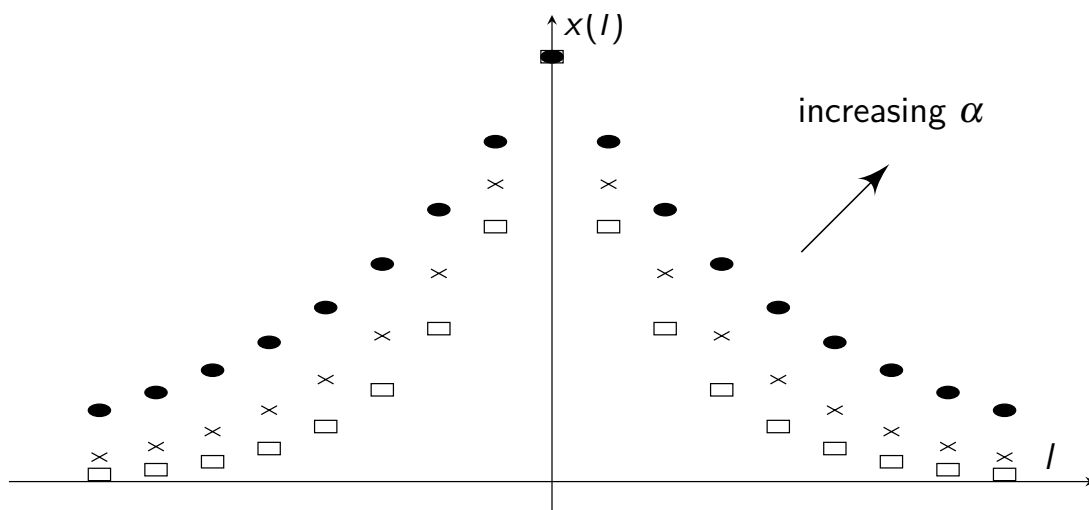$$X_{ss} = a^2 X_{ss} + (1-a^2)\, W \Rightarrow X_{ss} = W$$

and

$$X(l) = X(-l) = a^l X_{ss} = a^l W$$

Filtering a random process in state space

Example

$$x(k+1) = ax(x) + \sqrt{1-a^2}\, w(k), \quad \mathsf{E}[w(k)] = 0, \quad \mathsf{E}[w(k)w(j)] = W\delta_{kj}$$
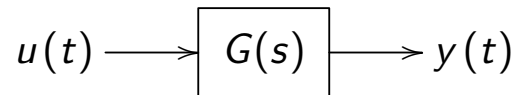
$$X(l) = X(-l) = a^l X_{ss} = a^l W$$

Filtering a random process

similar results hold in the continuous-time case:

$$u(t) \longrightarrow \boxed{G(s)} \longrightarrow y(t)$$

▶ spectral density (SISO case):

$$\Phi_{yy}(j\omega) = G(j\omega)\, G(-j\omega)\, \Phi_{uu}(j\omega) = |G(j\omega)|^2\, \Phi_{uu}(j\omega)$$

▶ mean and covariance dynamics:

$$\frac{dx(t)}{dt} = Ax(t) + B_w w(t), \ \mathsf{E}[w(t)] = 0, \ \mathsf{Cov}[w(t)] = W$$

$$\frac{dm_x(t)}{dt} = Am_x(t), \ m_x(0) = m_{x_0}$$

$$\frac{dX(t)}{dt} = AX + XA^T + B_w W B_w^T$$

▶ steady state: $X_{ss}(\tau) = X_{ss}e^{A^T\tau}$; $X_{ss}(-\tau) = e^{A\tau}X_{ss}$ where

$$AX_{ss} + X_{ss}A^T = -B_w W B_w^T : \text{ continuout-time Lyapunov Eq.}$$

Appendix: Lyapunov equations

▶ discrete-time case:
$$A^T P A - P = -Q$$

has the following unique solution iff $\lambda_i(A)\lambda_j(A) \neq 1$ for all $i,j = 1,\dots,n$:

$$P = \sum_{k=0}^{\infty} \left(A^T\right)^k Q A^k$$

▶ continuous-time case:

$$A^T P + P A = -Q$$

has the following unique solution iff $\lambda_i(A) + \bar{\lambda}_j(A) \neq 0$ for all $i,j = 1,\dots,n$:

$$P = \int_0^{\infty} e^{A^T t} Q e^{At} dt$$

Summary

1. Big picture

2. Basic concepts: sample space, events, probability axioms, random variable, pdf, cdf, probability distributions

3. Multiple random variables
   - random vector, joint probability and distribution, conditional probability
   - Gaussian case

4. Random process