Probability and Random Process

Sample Space, Events, and Probability Axioms:

When we deal with an experiment, the outcome at each trial is normally different. It is thus convenient to introduce a space which contains all the possible outcomes of trials. The sample space Ω is defined to be such a space and its elements are denoted by an outcome ω. Subsets of Ω are called events and are denoted by A. A probability is a function P that assigns a number to the evens A with the following properties:

$$\begin{aligned} &a) \quad 0 \le P(A) \le 1 \qquad \text{for any event } A \subseteq \Omega \\ &b) \quad P(\Omega) = 1 \\ &c) \quad \text{If } A \cap B = \phi, \quad \text{then } P(A \cup B) = P(A) + P(B) \end{aligned} \tag{p1}$$

where ϕ is the empty set. A probability space is a triple (Ω, A, P) where

Ω =sample space = set of outcomes
A =set of all events. (For Ω with n elements, there are 2^n events)
P = probability

Example:

E = toss two coins

$\Omega = \{\mathrm{HH},\ \mathrm{HT},\ \mathrm{TH},\ \mathrm{TT}\}$

$\mathrm{A} = 2^{\Omega}$ (has $2^4 = 16$ events)

Suppose that P(HH)=0.1, P(HT)=P(TH)=0.2. Then, P(TT)=0.5, and P({HH,TT})=0.6

Example:

E = measurement of some volume less than 1

$\Omega = [0,1)$

$\mathrm{A} = \{A:\ A = \text{finite union of intervals } [\mathrm{a}, \mathrm{b})\}$

$P([\mathrm{a}, \mathrm{b})) = b - a$

then, the above (Ω, A, P) is a probability space.

Conditioning and Independence

Let (Ω, A, P) be a probability space and let $A, B \in \mathrm{A}$ be two events with $P(B) \neq 0$. The *conditional probability* is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{p2}$$

Informally, this has the interpretation of the likelihood that event *A* will occur given the information that event *B* has occurred, which is graphically illustrated below.

A and *B* are *independent events* iff

$$P(A|B) = P(A) \qquad \text{or} \qquad P(A \cap B) = P(A)P(B) \tag{p3}$$

i.e., the information that *B* has occurred does not alter the probability that event A occurs.

A collection of events $\{A_1 \quad A_2 \quad \cdots\}$ is called *independent* if for every finite subcollection, we have

$$P(\bigcap_1^m A_{k_i}) = \prod_1^m P(A_{k_i}) \tag{p4}$$

Random Variables, Probability Distribution Function, and Probability Density Function

We are particular concerned with experiments whose outcome is a real number. For example, the measurement of a voltage, temperature, etc. Many experiments whose outcome is not a real number, such as coin-tossing can be made to look like a real number outcome experiment by assigning (perhaps unnaturally) numbers to outcomes. For example, in the two coin toss experiment, we can declare 1=HH, 2=HT, 3=TH, 4=TT.

- Let (Ω, A, P) be a probability space. A random variable X is a mapping $X: \Omega \to R$, i.e., an assignment of real numbers to every possible outcome $\omega \in \Omega$. When the outcomes of an experiment are already real numbers, the random variable X is the natural imbedding. The values that X takes are normally denoted x.

- The probability distribution function $P_X(x)$ associated with X is

$$P_X(x) = \Pr\{X \le x\} \tag{p5}$$

It is easy to see that

$$\begin{aligned} &a) && \lim_{x\to-\infty} P_X(x) = 0 \\ &b) && \lim_{x\to\infty} P_X(x) = 1 \\ &c) && P_X(x) \text{ is monotone nondescreasing in } x \\ &d) && P_X(x) \text{ is right continuous in } x \text{ (why?)} \end{aligned} \tag{p6}$$

- The probability density function $p_X(x)$ associated with X is

$$p_X(x) = \frac{d}{dx} P_X(x) \tag{p7}$$

i.e.,

$$p_X(x)dx = \Pr\{x < X \le x + dx\} \tag{p8}$$

It is easy to see that

$$\begin{aligned} &a) && p_X(x) \ge 0 \\ &b) && \int_{-\infty}^{x} p_X(\lambda)d\lambda = P_X(x) \end{aligned} \tag{p9}$$

Expectation

- Let X be a random variable. The expected value of X written as $E[X]$ is the "average value" of X, i.e.,

$$E[X] = \int_{-\infty}^{\infty} x p_X(x) dx \tag{p10}$$

- $E[X]$ is also written as $\bar{x}$ (or m_X) and is also called the mean of the random variable X.
- Let f be a real valued function and Y be a random variable which is a function of X, i.e., $Y = f(X)$. Analogous to (p10), for the random variable Y, we have

$$E[Y] = \int_{-\infty}^{\infty} y p_Y(y) dy \tag{p11}$$

which can be shown to be

$$E[Y] = E[f(X)] = \int_{-\infty}^{\infty} f(x) p_X(x) dx \tag{p12}$$

- $E[X^m]$ is called the mth moment of X.
- Of particular importance are the 1st and 2nd moments (mean and "variance")
- Variance of X is defined to be

$$\begin{aligned} E[(X-\bar{x})^2] &= \int_{-\infty}^{\infty} (x-\bar{x})^2 p_X(x) dx \\ &= E[X^2] - \bar{x}^2 \end{aligned} \tag{p13}$$

- Expectation is a linear operator, i.e., if X and Y are two random variables, then,

$$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y] \tag{p14}$$

note:

$$E[X^2] \ne E^2[X]$$

Examples:

Gaussian or Normal Distribution

A random variable has a Gaussian (or normal) distribution denoted by $N(m,\sigma^2)$ if its probability density function is

$$p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{p15}$$

where m is the mean and σ^2 is the variance. Note that the Guassian distribution is completely characterized by the mean and standard deviation σ, which is not the case for general distributions.

The Gaussian distribution is important and useful since

- A Gaussian distribution approximates the distribution of a random variable which is a sum of a large number of independent random variables
- Gaussian assumption is a good assumption in many practical situations.
- If Y is a linear function of X and if X is Gaussian, then Y is also Gaussian.

Joint Probability, Density Function, and Random Vectors

If we have more than one random variable defined on the same sample space Ω, we have to consider the joint density (and distribution) function. If we have two random variables, X and Y, their joint distribution function is

$$P_{XY}(x,y)=\Pr\{X\le x, Y\le y\} \tag{p16}$$

If $P_{XY}(x,y)$ is differentiable, the joint density function is

$$p_{XY}(x,y)=\frac{\partial^2}{\partial x\partial y}P_{XY}(x,y) \tag{p17}$$

If the joint distribution function satisfies

$$P_{XY}(x,y)=\Pr\{X\le x, Y\le y\}=\Pr\{X\le x\}\Pr\{Y\le y\}=P_X(x)P_Y(y) \tag{p18}$$

then X and Y are said to be ***independent***. When X and Y are independent, their joint density function satisfies

$$p_{XY}(x,y)=p_X(x)p_Y(y) \tag{p19}$$

Random variables may form a random vector. For example, random variables, X and Y, may be put in the form

$$Z=\begin{bmatrix}X\\Y\end{bmatrix} \tag{p20}$$

For this random vector, the mean is

$$m_Z=\begin{bmatrix}m_X\\m_Y\end{bmatrix}=E[Z]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\begin{bmatrix}x\\y\end{bmatrix}p_{XY}(x,y)dxdy \tag{p21}$$

and the covariance (matrix) is

$$\begin{aligned}\Lambda_{ZZ}&=E\left[(Z-m_Z)(Z-m_Z)^T\right]=\begin{bmatrix}\Lambda_{XX}&\Lambda_{XY}\\\Lambda_{YX}&\Lambda_{YY}\end{bmatrix}\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\begin{bmatrix}(x-m_X)^2&(x-m_X)(y-m_Y)\\(y-m_Y)(x-m_X)&(y-m_Y)^2\end{bmatrix}p_{XY}(x,y)dxdy\end{aligned} \tag{p22}$$

Notice that the covariance matrix is symmetric, i.e., $\Lambda_{XY}=\Lambda_{YX}$. The off-diagonal elements indicate the correlation between X and Y. If $\Lambda_{XY}=\Lambda_{YX}=0$, X and Y are said to be ***uncorrelated.*** If X and Y are independent, they are uncorrelated. However, uncorrelated X and Y are not necessarily independent.

When X and Y are jointly distributed, the probability density function for X is

$$p_X(x)=\int_{-\infty}^{\infty}p_{XY}(x,y)dy \tag{p23}$$

since all the possible values of Y must be taken into account. When Y takes a particular value, say *y*, the probability density function of X is given by the ***conditional density function,***

$$p_X(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)} \tag{p24}$$

where $p_Y(y)$ in the denominator is the scaling factor to assure

$$P_X(\infty|y) = \int_{-\infty}^{\infty} p_X(x|y)dx = 1 \tag{p25}$$

If X and Y are *independent*, then,

$$p_X(x|y) = p_X(x) \tag{p26}$$

When X and Y are dependent, the conditional density function of X given Y=y is different from the unconditioned density function, i.e., $p_X(x|y) \neq p_X(x)$. The conditional mean and covariance are obtained by using the conditional density function. For example, the conditional mean is

$$E[X|y] = \int_{-\infty}^{\infty} x p_X(x|y)dx \tag{p27}$$

Random Vectors (n-dimensional)

The above notions of random vector, joint distribution function, and joint density function for two random variables can be generalized to *n* random variables. Specifically, let $X_1, \quad \cdots \quad X_n$ be random variables over the same sample space. We can form the n-dimensional random vector and its specific value vector as

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{p28}$$

The joint distribution function is then defined as

$$P_X(x) = \Pr\{ X_1 \le x_1, \quad \cdots \quad X_n \le x_n \} \tag{p29}$$

If $P_X(x)$ is differentiable, the joint density function is

$$p_X(x) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} P_X(x) \tag{p30}$$

If the joint distribution function satisfies

$$P_X(x) = \Pr\{X_1 \le x_1\} \cdots \Pr\{X_n \le x_n\} = P_{X_1}(x_1) \cdots P_{X_n}(x_n) \tag{p31}$$

or equivalently

$$p_X(x) = p_{X_1}(x_1) \cdots p_{X_n}(x_n) \tag{p32}$$

then, $X_1, \quad \cdots \quad X_n$ are said to be ***independent***.

let X and Y be random n-vectors. The covariance matrix of a random vector X is defined as

$$\begin{aligned} \Lambda_{XX} &= E[(X - m_X)(X - m_X)^T] \quad \in R^{n \times n} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x - m_X)(x - m_X)^T p_X(x)dx \end{aligned} \tag{p33}$$

The cross-covariance matrix of two random vectors X and Y is defined as

$$\begin{aligned} \Lambda_{XY} &= E[(X - m_X)(Y - m_Y)^T] \quad \in R^{n \times n} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x - m_X)(y - m_Y)^T p_{XY}(x,y)dxdy \end{aligned} \tag{p34}$$

Gaussian Random Vectors (n-dimensional)

The joint density function of a *n*-dimensional Gaussian random vector X is :

$$p_X(x) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\Lambda_{XX}|}} e^{-\frac{1}{2}(x - m_X)^T \Lambda_{XX}^{-1} (x - m_X)} \tag{p35}$$

where m_X and Λ_{XX} are the mean and covariance matrix of X respectively, i.e.,

$$m_x = \mathrm{E}[X], \quad \Lambda_{XX} = \mathrm{E}\left[(X-m_x)(X-m_x)^T\right] \tag{p36}$$

In particular, if X_1 and X_2 are independent Gaussian random variables, the joint density function is

$$\begin{aligned} p_X(x) &= p_{X_1}(x_1)p_{X_2}(x_2) = \frac{1}{\sigma_{X_1}\sqrt{2\pi}} e^{-\frac{(x_1-m_{X_1})^2}{2\sigma_{X_1}^2}} \frac{1}{\sigma_{X_2}\sqrt{2\pi}} e^{-\frac{(x_2-m_{X_2})^2}{2\sigma_{X_2}^2}} \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^2 \sigma_{X_1}\sigma_{X_2}} e^{-\frac{1}{2}\begin{bmatrix} x_1-m_{X_1} \\ x_2-m_{X_2}\end{bmatrix}^T \begin{bmatrix} \sigma_{X_1}^2 & \\ & \sigma_{X_2}^2\end{bmatrix}^{-1} \begin{bmatrix} x_1-m_{X_1} \\ x_2-m_{X_2}\end{bmatrix}} \end{aligned} \tag{p37}$$

Note that $\sigma_{X_1}\sigma_{X_2}$ is the $\sqrt{|\Lambda_{XX}|}$. The off-diagonal elements of the covariance matrix Λ_{XX} are zero because X_1 and X_2 are independent.

A particular property of Gaussian random vectors is given below, which is very useful in the construction of Kalman filters. Let X and Y are jointly distributed *n* and *m* dimensional Gaussian random vectors. Then, their joint density function is

$$p_{XY}(x,y) = \frac{1}{\left(\sqrt{2\pi}\right)^{n+m}\sqrt{|\Lambda|}} e^{-\frac{1}{2}\begin{bmatrix} x-m_X \\ y-m_Y\end{bmatrix}^T \Lambda^{-1} \begin{bmatrix} x-m_X \\ y-m_Y\end{bmatrix}} \tag{p38}$$

where Λ is an $(n+m)\times(n+m)$ covariance matrix written as

$$\Lambda = \begin{bmatrix} \Lambda_{XX} & \Lambda_{XY} \\ \Lambda_{YX} & \Lambda_{YY}\end{bmatrix} \tag{p39}$$

The conditional density function of X given $Y = y$ is also Gaussian with mean and covariance matrix given by

$$\begin{aligned} m_{X|y} &= m_X + \Lambda_{XY}\Lambda_{YY}^{-1}\left[y-m_Y\right] \\ \Lambda_{X|y} &= \Lambda_{XX} - \Lambda_{XY}\Lambda_{YY}^{-1}\Lambda_{YX} \end{aligned} \tag{p40}$$

Notice that the conditional mean depends on y while the conditional covariance matrix does not. The expressions (p40) are obtained by defining a nonsingular matrix C as

$$C = \begin{bmatrix} I & -\Lambda_{XY}\Lambda_{YY}^{-1} \\ 0 & I\end{bmatrix} \tag{p41}$$

Noting that the determinant of C is 1, i.e., |C|=1, we have

$$\begin{aligned} |\Lambda| &= \left|C\Lambda C^T\right| = \left|\Lambda_{XX} - \Lambda_{XY}\Lambda_{YY}^{-1}\Lambda_{YX}\right|\left|\Lambda_{YY}\right| \\ \begin{bmatrix} x-m_X \\ y-m_Y\end{bmatrix}^T \Lambda^{-1} \begin{bmatrix} x-m_X \\ y-m_Y\end{bmatrix} &= \begin{bmatrix} x-m_X \\ y-m_Y\end{bmatrix}^T C^T\left(C\Lambda C^T\right)^{-1} C \begin{bmatrix} x-m_X \\ y-m_Y\end{bmatrix} \\ &= \begin{bmatrix} x-m_{X|y} \\ y-m_Y\end{bmatrix}^T \begin{bmatrix} \Lambda_{X|y} & 0 \\ 0 & \Lambda_{YY}\end{bmatrix}^{-1} \begin{bmatrix} x-m_{X|y} \\ y-m_Y\end{bmatrix} \\ &= \left[x-m_{X|y}\right]^T \Lambda_{X|y}^{-1}\left[x-m_{X|y}\right] + \left[y-m_Y\right]^T \Lambda_{YY}^{-1}\left[y-m_Y\right] \end{aligned} \tag{p42}$$

Thus, from (p38),

$$p_{XY}(x,y) = \frac{1}{\left(\sqrt{2\pi}\right)^n\sqrt{|\Lambda_{X|y}|}} e^{-\frac{1}{2}\left[x-m_{X|y}\right]^T \Lambda_{X|y}^{-1}\left[x-m_{X|y}\right]} p_Y(y) \tag{p43}$$

which indicates that

$$p_{X|y}(x|y) = \frac{1}{\left(\sqrt{2\pi}\right)^n\sqrt{|\Lambda_{X|y}|}} e^{-\frac{1}{2}\left[x-m_{X|y}\right]^T \Lambda_{X|y}^{-1}\left[x-m_{X|y}\right]} \tag{p44}$$

This shows that the conditional density function of X given $Y = y$ is also Gaussian with mean and covariance matrix given by (p40).

Random Process (Discrete Time Case)

Ensemble Averages

A random variable with the time as a parameter is called a random process, i.e., $\{X(k), \quad k=1,2,\ldots\}$ (see Fig. P.2). To completely specify the discrete time random process, we need to specify its complete joint probability density function

$$p(x(1),x(2),\ldots) \tag{p45}$$

which is not practical. Fortunately, in most case, it is sufficient to know the mean function

$$m_X(k)=E[X(k)], \qquad k=1,2,\ldots \tag{p46}$$

and the (auto-)covariance function

$$\Lambda_{XX}(j,k)=E\left[\left(X(j)-m_X(j)\right)\left(X(k)-m_X(k)\right)^T\right], \qquad j,k=1,2,\ldots \tag{p47}$$

The autocovariance function is a measure of the degree of association of the signal at time j with itself at time k. It is a simple matter to justify that the expected value operator may be identified with averaging across the set of records in Fig.P.2, or "*ensemble average*".

In the engineering literature, the autocorrelation function is sometimes defined to be

$$E[X(j)X(k)^T] \tag{p48}$$

If $X(k)$ has zero mean, the autocorrelation function is nothing but the autovariance function.

Stationary

In general the properties of a random process are time dependent but to simplify matters, we often assume that a sort of "steady state" has been reached in the sense that certain statistical properties are unchanged with respect to time,. For example, assume

i. $p_{X(j)}(x) = p_{X(k)}(x), \quad \forall j,k$ (implying that $m_X(k)$ and $\Lambda_{XX}(k,k)$ are constant)

ii. $p_{X(j)X(k)}(x_1, x_2)$ is a function of the time difference $k-j$ only and not both j and k, which implies that the autocovariance function $\Lambda_{XX}(j,k)$ is a function of the time difference $k-j$ only.

A random process is said to be *weakly stationary* or simply stationary if the conditions i and ii hold. If similar conditions hold for all higher order joint probabilities, then, the process is said to be *completely stationary* or stationary in the strict sense. Fig. P.3 is an attempt to give a visual interpretation of the stationary concepts.

If a process is stationary, since the autocovariance function $\Lambda_{XX}(j,k)$ is a function of the time difference $k-j$ only, i.e., $\Lambda_{XX}(j,k) = \Lambda_{XX}(k-j)$, we can use the usual notation $R_{XX}(\tau)$ to represent the autocovariance function $\Lambda_{XX}(k-j)$, i.e., $R_{XX}(\tau) = \Lambda_{XX}(\tau)$, where $\tau = k-j$ is the "lag". It is easy to verify that $R_{XX}(\tau)$ is an even function having a typical shape in Fig. P.4.

The assumption of stationary (though an approximation) has important consequences, since it is the key to being able to replace the "ensemble average" across a set of records with the "time average" along a single record introduced below.

For a bivariate random processes, $\{X(k), \quad k=1,2,\ldots\}$ and $\{Y(k), \quad k=1,2,\ldots\}$, the cross-covariance (crosscorrelation) function is defined to be

$$\Lambda_{XY}(j,k) = E\left[\left(X(j)-m_X(j)\right)\left(Y(k)-m_Y(k)\right)^T\right], \qquad j,k=1,2,\ldots \tag{p49}$$

which is a measure of the degree of association of signal X at time j with signal Y at time k. As in the above, if the processes are stationary, the cross-covariance is a function of the time difference $k-j$ only and is denoted by $R_{XY}(\tau)$. A short list of important properties are

$$\begin{aligned} &\text{i.} && R_{XX}(\tau) = R_{XX}(-\tau) \\ &\text{ii.} && \left|R_{XX}(\tau)\right| \le R_{XX}(0) \quad (=\sigma_X^2) \\ &\text{iii.} && R_{XY}(\tau) = R_{YX}(-\tau) \\ &\text{iv.} && \left|R_{XY}(\tau)\right|^2 \le R_{XX}(0)R_{YY}(0) \end{aligned} \tag{p50}$$

Time Averages

The definitions above have all involved the underlying probability distributions (or equivalently, ensemble averaging). Often we find that we shall have only a single record from which to make estimates of the averages we have defined. This raises the question as to whether averages along a record, i.e., time average, might be used in place of the ensemble average. It seems self evident that "stationary" is a prerequisite if this is to be feasible. When the time average agree with the ensemble average, a random process is *ergodic*. Thus for a stationary and ergodic random process, the ensemble mean (p46) can be calculated by

$$m_X = \overline{x(k)} = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{j=-N}^{N} x(j) \tag{p51}$$

where the upper bar denotes the time average. Similarly, the cross covariance (correlation function) (p49) is

$$\begin{aligned} R_{XY}(\tau) &= E\left[\left(X(k)-m_X\right)\left(Y(k+\tau)-m_Y\right)^T\right] \\ &= \overline{\left(x(k)-m_X\right)\left(y(k+\tau)-m_Y\right)^T} \\ &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{j=-N}^{N} \left(x(j)-m_X\right)\left(y(j+\tau)-m_Y\right)^T \end{aligned} \tag{p52}$$

with $Y = X$ for the autocovariance function.

Spectral Density Functions

The Fourier transform of an auto-covariance function is the variational (auto) *spectral density function* $S_{XX}(\omega)$, i.e.,

$$S_{XX}(\omega) = \sum_{\tau=-\infty}^{\infty} R_{XX}(\tau) e^{-j\omega\tau} \qquad \text{and} \qquad R_{XX}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{XX}(\omega) e^{j\omega\tau} d\omega \tag{p53}$$

The Fourier transform of an autocorrelation function is called the (auto) *spectral density function.* (p53) is called variational because of the way that the auto-covariance is defined. It follows from (p53) that

$$\sigma_X^2 = R_{XX}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{XX}(\omega) d\omega \tag{p54}$$

which implies that the variance of X is equal to the area under the spectral density curve, or the spectral density function $S_{XX}(\omega)$ describes the decomposition of the average power of X(t) over frequency.

The *cross spectral density* $S_{XY}(\omega)$ is similarly defined as

$$S_{XY}(\omega) = \sum_{\tau=-\infty}^{\infty} R_{XY}(\tau) e^{-j\omega\tau} \qquad \text{and} \qquad R_{XY}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{XY}(\omega) e^{j\omega\tau} d\omega \tag{p55}$$

The cross spectral density $S_{XY}(\omega)$ is in general complex and may be expressed by $\left|S_{XY}(\omega)\right| e^{j\arg(S_{XY}(\omega))}$, in which $\left|S_{XY}(\omega)\right|$ is the cross amplitude spectral density and indicates whether frequency components in one time series are associated with large or small amplitudes at the same frequency in the other series. $\arg\left(S_{XY}(\omega)\right)$ is the phase

spectral density and indicates whether frequency components in one series lag or lead the components at the same frequency in the other series. Some properties of spectra are as follows

$$\begin{aligned} &\text{i.} \quad S_{XY}(\omega) = S_{YX}{}^{*}(-\omega) \\ &\text{ii.} \quad \left|S_{XY}(\omega)\right|^2 \le S_{XX}(\omega) S_{YY}(\omega) \end{aligned} \tag{p56}$$

where star * stands for the complex conjugate operation. The following figure lists some of the common auto-correlation functions and associated power spectral densities.

Linear Time-Invariant System

Consider a single-input/single-output (SISO) linear time-invariant (LTI) system given by

$$Y(z) = G(z)U(z) \;\; or \;\; y(k) = \sum_{i=-\infty}^{\infty} g(i)u(k-i) \tag{p57}$$

where u is the input and y is the output. If u is a random process, y is also a random process. Assuming that u is zero mean and that the ergodic hypothesis holds, the cross-covariance function for u and y and the autocovariance function for y can be obtained as follows.

$$R_{uy}(l) = \overline{u(k)\sum_{i=-\infty}^{\infty} u(k+l-i)g(i)} = \sum_{i=-\infty}^{\infty} \overline{u(k)u(k+l-i)}g(i) = \sum_{i=-\infty}^{\infty} R_{uu}(l-i)g(i) \tag{p58}$$

$$R_{yy}(l) = \overline{y(k)y(k+l)} = \overline{y(k)\sum_{i=-\infty}^{\infty} u(k+l-i)g(i)} = \sum_{i=-\infty}^{\infty} \overline{y(k)u(k+l-i)}g(i) = \sum_{i=-\infty}^{\infty} R_{yu}(l-i)g(i) \tag{p59}$$

Eq. (p58) implies that the cross-covariance function, R_{uy}, is the output of the linear system (p57) excited by the autocovariance function of u, R_{uu}. Eq. (p59) implies that the autocovariance function of y, R_{yy}, is the output of the linear system (p57) excited by the cross-covariance function, R_{yu}, as illustrated by the following figure.

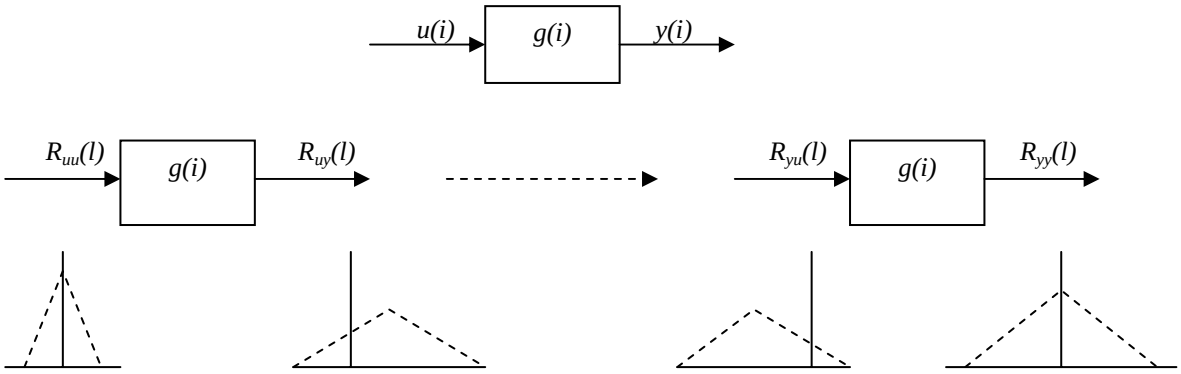

Fig. P-3 Generation of R_{uy} and R_{yy} from R_{uu}

From (p58) and (p59), the cross-spectral density function of the output and the input of the linear system (p57) and the spectral density function of the output are related to the spectral density function of the input by

$$\begin{aligned} S_{uy}(\omega) &= G(e^{j\omega})S_{uu}(\omega) \\ S_{yy}(\omega) &= G(e^{j\omega})G(e^{-j\omega})S_{uu}(\omega) = \left|G(e^{j\omega})\right|^2 S_{uu}(\omega) \end{aligned} \tag{p60}$$

The ***ordinary coherence function*** between the input $u(t)$ and the output $y(t)$ is defined as

$$\gamma_{uy}{}^2(\omega) = \frac{\left|S_{uy}(\omega)\right|^2}{S_{uu}(\omega)S_{yy}(\omega)} \tag{p61}$$

From the inequility (p50) for cross spectra, it follows that

$$0 \le \gamma_{uy}{}^2(\omega) \le 1 \tag{p62}$$

If $u(t)$ and $y(t)$ are completely unrelated, $\gamma_{uy}{}^2(\omega) = 0$, while if $y(t)$ and $u(t)$ are linearly related (e.g. the LTI system (p57)), then $\gamma_{uy}{}^2(\omega) = 1$.

An instantaneous value of a purely random process (***white process***) is not correlated with its value at other time instances. Namely, the autocovariance function of a ***white process*** is

$$R_{xx}(0) = \sigma_{xx}^2 \qquad \text{and} \qquad R_{xx}(l) = 0 \quad \text{for} \quad l \neq 0 \tag{p63}$$

If a zero mean white process is not stationary, its auto-covariance function is given by

$$E\left[x(k)x(j)\right] = Q(k)\delta_{kj} \quad , \quad \delta_{kj} = 1 \text{ for } k = j \text{ and } = 0 \text{ for } k \neq j \tag{p64}$$

From Eqs. (p58) and (p59), the output of a linear system is correlated (or colored) even when the input is white. Consider a discrete transfer function

$$G(z) = \frac{b_n z^n + b_{n-1}z^{n-1} + \cdots + b_0}{z^n + a_{n-1}z^{n-1} + \cdots + a_0} \tag{p65}$$

If the input u is white with a variance of 1, from Eqs. (p60), the spectral density of the output, y, is

$$S_{yy}(\omega) = \left|G\left(e^{j\omega}\right)\right|^2 = G(z)G(z^{-1})\,|_{z=e^{j\omega}} \tag{p66}$$

We now consider a discrete time system described in the state space. Let w be a zero mean white random vector with the covariance $E[w(k)w(k)^T] = W(k)$, and consider

$$x(k+1) = A(k)x(k) + B_w(k)w(k) \tag{p67}$$

The initial state is random, and its mean and covariance are given by

$$E[x(0)] = m_{x0} \quad and \quad E\left[(x(0)-m_{x0})(x(0)-m_{x0})^T\right] = X_0 \tag{p68}$$

Assume that x(0) and w(k) are not correlated. Under these assumptions, the mean and covariance of x(k) are obtained as follows. By taking the expectation of Eq. (p67), we obtain the equation for the mean

$$m_x(k+1) = A(k)m_x(k), \quad m_x(0) = m_{x0} \tag{p69}$$

where we have noted E[w(k)] = 0. Subtraction of (p69) from (p67) yields

$$x(k+1) - m_x(k+1) = A(k)[x(k) - m_x(k)] + B_w(k)w(k) \tag{p70}$$

Multiplying the left and right hand sides of (p70) by their transposed quantities from right and taking expectation, we obtain

$$X(k+1) = A(k)X(k)A^T(k) + B_w(k)W(k)B_w^T(k), \quad X(0) = X_0 \tag{p71}$$

where $X(k) = E\left[(x(k)-m_x(k))(x(k)-m_x(k))^T\right]$ represents the covariance of transient state $x(k)$, and we have set $E[(x(k)-m_x)w^T(k)] = 0$ because x(k) depends only on x(0), w(0), ... , w(k-2) and w(k-1), which are all uncorrelated with w(k).

If w(k) is stationary, A and B_w do not depend on k, and A is asymptotically stable, then, $m_x(k)$ converges to zero and $X(k)$ converges to the stationary solution, X_{ss}. Eq. (p71) thus becomes a Lyapunov equation. Because of the whiteness of w(k), the autocorrelation (autocovariance) of x(k) at the steady state is

$$R_{ss}(l) = E[x(k)x^T(k+l)] = X_{ss}(A^T)^l \quad \text{and}$$
$$R_{ss}(-l) = E[x(k)x^T(k-l)] = E[x(k+l)x^T(k)] = A^l X_{ss} \tag{p72}$$

Example:

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

x(0) and w(k) are uncorrelated. Note that $m_x(k) = 0$ at the steady state. The steady state solution of Eq. (p71) satisfies

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

From Eq. (p72),

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

Thus, x(k) has a stronger correlation for increasing a. For stability, a must be less than 1.

Note: In discussing the random process, we did not specify the shape of the probability density function. Namely, **ideas such as whiteness and correlation are independent from the density function**. In most cases, we assume that the density (distribution) function is Gaussian. For example, we say that random processes are white and Gaussian. Recall that if the input of a linear system is Gaussian, the output is also Gaussian.

Random Process (Continuous Time)

Various ideas and quantities introduced for discrete time random processes are similarly defined for continuous time random processes. We will summarize them below.

Let x(t) be a random process. For a sample process, the time average of x(t) is

$$\overline{x(t)} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)dt \tag{p51C}$$

x(t) is stationary in the strict sense if

$$\Pr\{x(t_1) \le x_1, \ldots, x(t_i) \le x_i, \ldots\} = \Pr\{x(t_1 + l) \le x_1, \ldots, x(t_i + l) \le x_i, \ldots\} \quad \text{(p72)}$$

for all t_i's, x_i's and l. It is stationary in the weak (wide) sense if its mean does not depend on t and its auto-covariance function depends only on the time difference.

For a stationary and ergodic process, its mean and auto-covariance function can be computed by (p51C) and

$$R_{xx}(\tau) = E[x(t) - m_x)(x(t+\tau) - m_x)] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} (x(t) - m_x)(x(t+\tau) - m_x) dt \quad \text{(p52C)}$$

For two random processes, x and y, the cross-covariance function satisfies

$$R_{xy}(\tau) = R_{yx}(-\tau) \quad \text{(p50C)}$$

The (variational) (auto) spectral density function is

$$S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j\omega\tau} d\tau \quad \text{and} \quad R_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{xx}(\omega) e^{j\omega\tau} d\omega \quad \text{(p53C)}$$

If u is a random process and y is the output of a linear system

$$Y(s) = G(s)U(s) \quad or \quad y(t) = \int_{-\infty}^{\infty} g(\eta) u(t-\eta) d\eta \quad \text{(p57C)}$$

the cross-covariance function for u and y and the auto-covariance function for y are

$$R_{uy}(\tau) = u(t) \int_{-\infty}^{\infty} u(t+\tau-\eta) g(\eta) d\eta = \int_{-\infty}^{\infty} u(t) u(t+\tau+\eta) g(\eta) d\eta = \int_{-\infty}^{\infty} R_{uu}(\tau-\eta) g(\eta) d\eta \quad \text{(p58C)}$$

$$R_{yy}(\tau) = \int_{-\infty}^{\infty} R_{yu}(\tau-\eta) g(\eta) d\eta \quad \text{(p59C)}$$

The cross-spectral density function of the output and the input of the linear system (p57C) and the spectral density function of the output are related to the spectral density function of the input by

$$\begin{aligned} S_{uy}(\omega) &= G(j\omega) S_{uu}(\omega) \\ S_{yy}(\omega) &= G(j\omega) G(-j\omega) S_{uu}(\omega) = |G(j\omega)|^2 S_{uu}(\omega) \end{aligned} \quad \text{(p60C)}$$

The auto-covariance function of a purely random process (***white process***) is

$$R_{xx}(\tau) = \sigma_{xx}^2 \delta(\tau) \quad \text{(P63C)}$$

where δ is a delta function. If a zero mean white process is not stationary, its auto-covariance function is

$$E[x(t)x(t+\tau)] = Q(t)\delta(\tau) \quad \text{(p64C)}$$

If the input u(t) of the transfer function

$$G(s) = \frac{b_m s^m + \ldots + b_0}{s^n + a_{n-1} s^{n-1} + \ldots + a_0}, \quad m < n \quad \text{(p65C)}$$

is white and has an unity variance, the (variational) spectral density function of the output is

$$S_{yy}(\omega) = \left(\frac{B(s)B(-s)}{A(s)A(-s)} \right)_{s=j\omega} = |G(j\omega)|^2 \quad \text{(p66C)}$$

Consider

$$\frac{dx(t)}{dt} = A(t)x(t) + B_w w(t) \quad \text{(p67C)}$$

where $E[w(t)] = 0$, $E[w(t)w(t+\gamma)] = W(t)\delta(\gamma)$, $E[x(t_0)] = m_{x0}$, $E[x(t_0) - m_{m0})(x(t_0) - m_{x0})^T] = X_0$ and $E[(x(t_0) - m_{x0}) w(t)^T] = 0$. Then, the mean and covariance of x(t) are given by

$$\frac{dm_x(t)}{dt} = A(t)m_x(t), \quad m(t_0) = m_{x0} \quad \text{(p69C)}$$

$$\frac{dX(t)}{dt} = A(t)X(t) + X(t)A(t) + B_w(t)W(t)B_w(t), \quad X(t_0) = X_0 \qquad \text{(p70C)}$$

If W, A and B_w do not depend on t and A is asymptotically stable, $m_x(t)$ converges to zero and X(t) converges to the steady state solution, X_{ss}, of Eq. (p70C). X_{ss} is the positive (semi-)definite solution of the Lyapunov equation.

$$AX + XA^T = -B_w W B_w^T \qquad \text{(p73)}$$

The auto-covariance of x(t) is

$$R_{ss}(\tau) = E[x(t)x^T(t+\tau)] = X_{ss}e^{A\tau} \text{ and } R_{ss}(-\tau) = E[x(t)x^T(t-\tau)] = e^{A^T\tau}X_{ss} \qquad \text{(p72C)}$$

Derivation of Eqs. (p69C) and (p70C):

We derive these two equations in the following three steps: 1). obtain a discrete time approximation of Eq.(p67C) (sampling time = Δt); 2). apply Eqs. (p69) and (p71) to the approximate discrete time system; and 3). take the limit of Δt approaching to zero.

Step 1.

$$\frac{x((k+1)\Delta t)}{\Delta t} = Ax(k\Delta t) + B_w w(k)$$
$$\Rightarrow x(k+1) = (I + A\Delta t)x(k) + B_w \Delta t w(k)$$

where w(k) is the average of w(t) over $[k\Delta t, (k+1)\Delta t]$,

$$w(k) = \frac{1}{\Delta t}\int_{k\Delta t}^{(k+1)\Delta t} w(t)dt$$

The mean and covariance of w(k) are

$$E[w(k)] = \frac{1}{\Delta t}\int_{k\Delta t}^{(k+1)\Delta t} E[w(t)dt = 0$$

$$E[w(k)w^T(j)] = E[\frac{1}{\Delta t}\int_{k\Delta t}^{(k+1)\Delta t} w(t_1)dt_1 \frac{1}{\Delta t}\int_{j\Delta t}^{(j+1)\Delta t} w^T(t_2)dt_2] = \frac{1}{\Delta t^2}\int_{j\Delta t}^{(j+1)\Delta t}\int_{k\Delta t}^{(k+1)\Delta t} E[w(t_1)w^T(t_2)]dt_1 dt_2$$
$$= 0 \ \textit{for} \ j \neq k$$
$$= \frac{1}{\Delta t}W \ \textit{for} \ j = k$$

Step 2.

$$m_x((k+1)\Delta t = (I + A\Delta t)m_x(k\Delta t), \quad m_x(0) = x_0$$
$$X((k+1)\Delta t = (I + A\Delta t)X(k\Delta t)(I + A\Delta t)^T + B_w\Delta t\frac{W}{\Delta t}B_w^T, X(0) = X_0$$
$$\Rightarrow \frac{m_x((k+1)\Delta t) - m_x(k\Delta t)}{\Delta t} = Am_x(k\Delta t)$$
$$\frac{X((k+1)\Delta t) - X(k\Delta t)}{\Delta t} = AX(k\Delta t) + X(k\Delta t)A^T + AX(k\Delta t)A^T\Delta t + B_w W B_w^T$$

Step 3. Let $\Delta t \to 0$ in the above expressions to obtain (p69C) and (p70C).