Biomolecular Feedback Systems

Domitilla Del Vecchio
MIT

Richard M. Murray
Caltech

DRAFT v0.4a, December 26, 2010

Appendix C
Random Procesess

This appendix provides a summary of random processes in continuous time with continuous and discrete states. Some of the material in this section is drawn from the AM08 supplement on Optimization-Based Control [53].

C.1 Random Variables

Random variables and processes are defined in terms of an underlying *probability space* that captures the nature of the stochastic system we wish to study. A probability space has three elements:

- a *sample space* Ω that represents the set of all possible outcomes;
- a set of *events* $\mathcal{F}$ the captures combinations of elementary outcomes that are of interest; and
- a *probability measure* $\mathcal{P}$ that describes the likelihood of a given event occurring.

Ω can be any set, either with a finite, countable or infinite number of elements. The event space $\mathcal{F}$ consists of subsets of Ω. There are some mathematical limits on the properties of the sets in $\mathcal{F}$, but these are not critical for our purposes here. The probability measure $\mathcal{P}$ is a mapping from $\mathcal{P} : \mathcal{F} \to [0,1]$ that assigns a probability to each event. It must satisfy the property that given any two disjoint sets $A, B \subset \mathcal{F}$, $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B)$. The term *probability distribution* is also to describe a probability measure.

With these definitions, we can model many different stochastic phenomena. Given a probability space, we can choose samples $\omega \in \Omega$ and identify each sample with a collection of events chosen from $\mathcal{F}$. These events should correspond to phenomena of interest and the probability measure $\mathcal{P}$ should capture the likelihood of that event occurring in the system that we are modeling. This definition of a probability space is very general and allows us to consider a number of situations as special cases.

Need more details on $\omega, \mathcal{F}$?

A *random variable* X is a function $X : \Omega \to S$ that gives a value in S, called the state space, for any sample $\omega \in \Omega$. Given a subset $A \subset S$, we can write the

probability that $X \in A$ as

$$P(X \in A) = P(\omega \in \Omega : X(\omega) \in A).$$

We will often find it convenient to omit ω when working random variables and hence we write $X \in S$ rather than the more correct $X(\omega) \in S$.

A *discrete random variable* X is a variable that can take on any value from a discrete set S with some probability for each element of the set. We model a discrete random variable by its *probability mass function* $p_X(s)$, which gives the probability that the random variable X takes on the specific value $s \in S$:

$$p_X(s) = \text{probability that } X \text{ takes on the value } s \in S.$$

The sum of the probabilities over the entire set of states must be unity, and so we have that

$$\sum_{s \in S} p_X(s) = 1.$$

If A is a subset of S, then we can write $P(X \in A)$ for the probability that X will take on some value in the set A. It follows from our definition that

$$P(X \in A) = \sum_{s \in A} p(s).$$

Definition C.1 (Bernoulli distribution)**.** The Bernoulli distribution is used to model a random variable that takes the value 1 with probability p and 0 with probability $1 - p$:

$$P(X = 1) = p, \qquad P(X = 0) = 1 - p.$$

Alternatively, it can be written in terms of its probability mass function

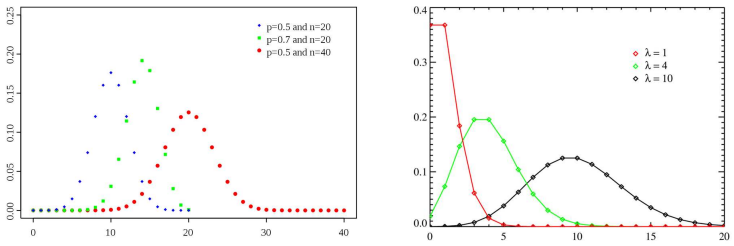
$$p(s) = \begin{cases} p & s = 1 \\ 1 - p & s = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Bernoulli distributions are used to model independent experiments with binary outcomes, such as flipping a coin.

Definition C.2 (Binomial distribution)**.** The *binomial distribution* models the probability of successful trials in n experiments, given that a single experiment has probability of success p. If we let K_n be a random variable that indicates the number of success in n trials, then the binomial distribution is given by

$$p_{K_n}(k) = P(K_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for $k = 1, \ldots, n$. The probability mass function is shown in Figure C.1a.



(a) Binomial distribution (b) Poisson distribution

Figure C.1: Probability mass functions for common discrete distributions.

Definition C.3 (Poisson distribution)**.** The *Poisson distribution* is used to describe the probability that a given number of events will occur in a fixed interval of time t. The Poisson distribution is defined as

$$p_{N_t}(k) = P(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \tag{C.1}$$

where N_t is the number of events that occur in a period t and λ is a real number parameterizing the distribution. This distribution can be considered as a model for a counting process, where we assume that the average rate of occurrences in a period t is given by λt and λ represents the rate of the counting process. Figure C.1b shows the form of the distribution for different values of k and λt.

A *continuous (real-valued) random variable* X is a variable that can take on any value in the set of real numbers $\mathbb{R}$. We can model the random variable X according to its *probability distribution* P:

$$P(x_l \leq X \leq x_u) = \text{probability that } x \text{ takes on a value in the range } x_l, x_u.$$

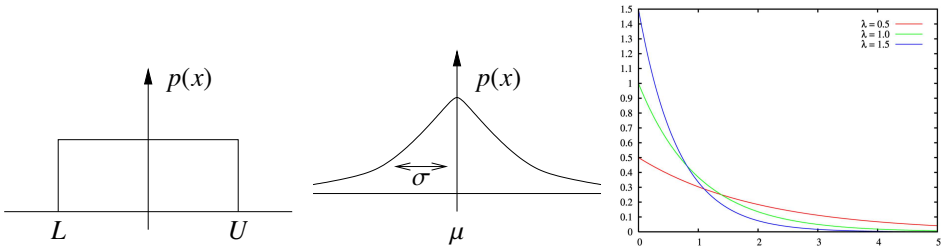
More generally, we write $P(A)$ as the probability that an event A will occur (e.g., $A = \{x_l \leq X \leq x_u\}$). It follows from the definition that if X is a random variable in the range $[L, U]$ then $P(L \leq X \leq U) = 1$. Similarly, if $Y \in [L, U]$ then $P(L \leq X \leq Y) = 1 - P(Y \leq X \leq U)$.

We characterize a random variable in terms of the *probability density function* (pdf) $p(x)$. The density function is defined so that its integral over an interval gives the probability that the random variable takes its value in that interval:

$$P(x_l \leq X \leq x_u) = \int_{x_l}^{x_u} p(x)dx. \tag{C.2}$$

It is also possible to compute $p(x)$ given the distribution P as long as the distribution is suitably smooth:

$$p(x) = \left.\frac{\partial P(x_l \leq x \leq x_u)}{\partial x_u}\right|_{\substack{x_l \text{ fixed}, \\ x_u = x,}} \qquad x > x_l.$$



(a) Uniform distribution (b) Gaussian distribution (c) Exponentialdistribution

Figure C.2: Probability density function (pdf) for uniform, Gaussian and exponential distributions.

We will sometimes write $p_X(x)$ when we wish to make explicit that the pdf is associated with the random variable X. Note that we use capital letters to refer to a random variable and lower case letters to refer to a specific value.

Definition C.4 (Uniform distribution)**.** The *uniform distribution* on an interval $[L, U]$ assigns equal probability to any number in the interval. Its pdf is given by

$$p(x) = \frac{1}{U - L}. \tag{C.3}$$

The uniform distribution is illustrated in Figure C.2a.

Definition C.5 (Gaussian distribution)**.** The *Gaussian distribution* (also called a *normal distribution*) has a pdf of the form

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{C.4}$$

The parameter μ is called the *mean* of the distribution and σ is called the *standard deviation* of the distribution. Figure C.2b shows a graphical representation a Gaussian pdf.

Definition C.6 (Exponential distribution)**.** The exponential distribution is defined for positive numbers and has a pdf of the form

$$p(x) = \lambda e^{-\lambda x}, \qquad x > 0$$

where λ is a parameter defining the distribution. A plot of the pdf for an exponential distribution is shown in Figure C.2c. The exponential distribution can be shown to describe the amount of time between two events in a Poisson process.

We now define a number of properties of collections of random variables. We focus on the continuous random variable case, but unless noted otherwise these

concepts can all be defined similarly for discrete random variables (using the probability mass function in place of the probability density function).

If two random variables are related, we can talk about their *joint probability distribution*: $P_{X,Y}(A, B)$ is the probability that both event A occurs for X and B occurs for Y. This is sometimes written as $P(A \cap B)$, where we abuse notation by implicitly assuming that A is associated with X and B with Y. For continuous random variables, the joint probability distribution can be characterized in terms of a *joint probability density function*

$$P(x_l \le X \le x_u, y_l \le Y \le y_u) = \int_{y_l}^{y_u} \int_{x_l}^{x_u} p(x,y)\, dx dy. \tag{C.5}$$

The joint pdf thus describes the relationship between X and Y, and for sufficiently smooth distributions we have

$$p(x,y) = \left. \frac{\partial^2 P(x_l \le X \le x_u, y_l \le Y \le y_u)}{\partial x_u \partial y_u} \right|_{\substack{x_l, y_l \text{ fixed}, \\ x_u = x,\, y_u = y,}} \quad \begin{array}{l} x > x_l, \\ y > y_l. \end{array}$$

We say that X and Y are *independent* if $p(x,y) = p(x)p(y)$, which implies that $P_{X,Y}(A,B) = P_X(A)P_Y(B)$ for events A associated with X and B associated with Y. Equivalently, $P(A \cap B) = P(A)P(B)$ if A and B are independent.

The *conditional probability* for an event A given that an event B has occurred, written as $P(A \mid B)$, is given by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{C.6}$$

If the events A and B are independent, then $P(A \mid B) = P(A)$. Note that the individual, joint and conditional probability distributions are all different, so we should really write $P_{X,Y}(A \cap B)$, $P_{X|Y}(A \mid B)$ and $P_Y(B)$.

If X is dependent on Y then Y is also dependent on X. *Bayes' theorem* relates the conditional and individual probabilities:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}, \qquad P(B) \neq 0. \tag{C.7}$$

Bayes' theorem gives the conditional probability of event A on event B given the inverse relationship (B given A). It can be used in situations in which we wish to evaluate a hypothesis H given data D when we have some model for how likely the data is given the hypothesis, along with the unconditioned probabilities for both the hypothesis and the data.

The analog of the probability density function for conditional probability is the *conditional probability density function* $p(x \mid y)$

$$p(x \mid y) = \begin{cases} \dfrac{p(x,y)}{p(y)} & 0 < p(y) < \infty \\ 0 & \text{otherwise.} \end{cases} \tag{C.8}$$

It follows that

$$p(x,y) = p(x \mid y)p(y) \tag{C.9}$$

and

$$\begin{aligned} P(x_l \le X \le x_u \mid y) &:= P(x_l \le X \le x_u \mid Y = y) \\ &= \int_{x_l}^{x_u} p(x \mid y)dx = \frac{\int_{x_l}^{x_u} p(x,y)dx}{p(y)}. \end{aligned} \tag{C.10}$$

If X and Y are independent than $p(x \mid y) = p(x)$ and $p(y \mid x) = p(y)$. Note that $p(x,y)$ and $p(x \mid y)$ are different density functions, though they are related through equation (C.9). If X and Y are related with joint probability density function $p(x,y)$ and conditional probability density function $p(x \mid y)$ then

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

Example C.1 (Conditional probability for sum)**.** Consider three random variables X, Y and Z related by the expression

$$Z = X + Y.$$

In other words, the value of the random variable Z is given by choosing values from two random variables X and Y and adding them. We assume that X and Y are independent Gaussian random variables with mean μ_1 and μ_2 and standard deviation $\sigma = 1$ (the same for both variables).

Clearly the random variable Z is not independent of X (or Y) since if we know the values of X then it provides information about the likely value of Z. To see this, we compute the joint probability between Z and X. Let

$$A = \{x_l \le x \le x_u\}, \qquad B = \{z_l \le z \le z_u\}.$$

The joint probability of both events A and B occurring is given by

$$\begin{aligned} P_{X,Z}(A \cap B) &= P(x_l \le x \le x_u, z_l \le x + y \le z_u) \\ &= P(x_l \le x \le x_u, z_l - x \le y \le z_u - x). \end{aligned}$$

We can compute this probability by using the probability density functions for X and Y:

$$\begin{aligned} P(A \cap B) &= \int_{x_l}^{x_u} \Big(\int_{z_l - x}^{z_u - x} p_Y(y)dy \Big) p_X(x)dx \\ &= \int_{x_l}^{x_u} \int_{z_l}^{z_u} p_Y(z-x)p_X(x)dzdx =: \int_{z_l}^{z_u} \int_{x_l}^{x_u} p_{Z,X}(z,x)dxdz. \end{aligned}$$

Using Gaussians for X and Y we have

$$\begin{aligned} p_{Z,X}(z,x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-x-\mu_Y)^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu_X)^2} \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}((z-x-\mu_Y)^2 + (x-\mu_X)^2)}. \end{aligned}$$

A similar expression holds for $p_{Z,Y}$. ∇

Given a random variable X, we can define various standard measures of the distribution. The *expectation* or *mean* of a random variable is defined as

$$\mathbb{E}\{X\} = \langle X \rangle = \int_{-\infty}^{\infty} x\, p(x)\, dx,$$

and the *mean square* of a random variable is

$$\mathbb{E}\{X^2\} = \langle X^2 \rangle = \int_{-\infty}^{\infty} x^2\, p(x)\, dx.$$

If we let μ represent the expectation (or mean) of X then we define the *variance* of X as

$$\mathbb{E}\{(X-\mu)^2\} = \langle (X - \langle X \rangle)^2 \rangle = \int_{-\infty}^{\infty} (x-\mu)^2\, p(x)\, dx.$$

We will often write the variance as σ^2. As the notation indicates, if we have a Gaussian random variable with mean μ and (stationary) standard deviation σ, then the expectation and variance as computed above return μ and σ^2.

Example C.2 (Exponential distribution)**.** The exponential distribution has mean and variance given by

$$\mu = \frac{1}{\lambda}, \qquad \sigma^2 = \frac{1}{\lambda^2}.$$

The exponential distribution can be shown to describe the amount of time between two events in a Poisson process. ∇

Several useful properties follow from the definitions.

Proposition C.1 (Properties of random variables)**.**

1. *If X is a random variable with mean μ and variance σ^2, then αX is random variable with mean αX and variance $\alpha^2\sigma^2$.*

2. *If X and Y are two random variables, then $\mathbb{E}\{\alpha X + \beta Y\} = \alpha\mathbb{E}\{X\} + \beta\mathbb{E}\{Y\}$.*

3. *If X and Y are Gaussian random variables with means μ_X, μ_Y and variances σ_X^2, σ_Y^2,*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}, \qquad p(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2},$$

then $X+Y$ is a Gaussian random variable with mean $\mu_Z = \mu_X + \mu_Y$ and variance $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$,

$$p(x+y) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{1}{2}\left(\frac{x+y-\mu_Z}{\sigma_Z}\right)^2}.$$

Proof. The first property follows from the definition of mean and variance:

$$\mathbb{E}\{\alpha X\} = \int_{-\infty}^{\infty} \alpha x\, p(x)\, dx = \alpha \int_{-\infty}^{\infty} \alpha x\, p(x)\, dx = \alpha \mathbb{E}\{X\}$$
$$\mathbb{E}\{(\alpha X)^2\} = \int_{-\infty}^{\infty} (\alpha x)^2\, p(x)\, dx = \alpha^2 \int_{-\infty}^{\infty} x^2\, p(x)\, dx = \alpha^2 \mathbb{E}\{X^2\}.$$

The second property follows similarly, remembering that we must take the expectation using the joint distribution (since we are evaluating a function of two random variables):

$$\begin{aligned}\mathbb{E}\{\alpha X + \beta Y\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\alpha x + \beta y)\, p_{X,Y}(x,y)\, dxdy \\ &= \alpha \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, p_{X,Y}(x,y)\, dxdy + \beta \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, p_{X,Y}(x,y)\, dxdy \\ &= \alpha \int_{-\infty}^{\infty} x\, p_X(x)\, dx + \beta \int_{-\infty}^{\infty} y\, p_Y(y)\, dy = \alpha \mathbb{E}\{X\} + \beta \mathbb{E}\{Y\}.\end{aligned}$$

The third item is left as an exercise. □

C.2 Continuous-State Random Processes

A *random process* is a collection of time-indexed random variables. Formally, we consider a random process X to be a joint mapping of sample and a time to a state: $X : \Omega \times \mathcal{T} \to S$, where $\mathcal{T}$ is an appropriate time set. We view this mapping as a generalized random variable: a sample corresponds to choosing an entire function of time. Of course, we can always fix the time and interpret $X(\omega, t)$ as a regular random variable, with $X(\omega, t')$ representing a different random variable if $t \neq t'$. Our description of random processes will consist of describing how the random variable at a time t relates to the value of the random variable at an earlier time s.

To build up some intuition about random processes, we will begin with the discrete time case, where the calculations are a bit more straightforward, and then proceed to the continuous time case.

A *discrete-time random process* is a stochastic system characterized by the *evolution* of a sequence of random variables $X[k]$, where k is an integer. As an example, consider a discrete-time linear system with dynamics

$$X[k+1] = AX[k] + BU[k] + FW[k], \qquad Y[k] \quad = CX[k] + V[k]. \tag{C.11}$$

As in AM08, $X \in \mathbb{R}^n$ represents the state of the system, $U \in \mathbb{R}^p$ is the vector of inputs and $Y \in \mathbb{R}^q$ is the vector of outputs. The (possibly vector-valued) signal W represents disturbances to the process dynamics and V represents noise in the measurements. To try to fix the basic ideas, we will take $u = 0$, $n = 1$ (single state) and $F = 1$ for now.

We wish to describe the evolution of the dynamics when the disturbances and noise are not given as deterministic signals, but rather are chosen from some probability distribution. Thus we will let $W[k]$ be a collection of random variables where the values at each instant k are chosen from a probability distribution with pdf $p_{W,k}$. As the notation indicates, the distributions might depend on the time instant k, although the most common case is to have a *stationary* distribution in which the distributions are independent of k (defined more formally below).

In addition to stationarity, we will often also assume that distribution of values of W at time k is independent of the values of W at time l if $k \neq l$. In other words, $W[k]$ and $W[l]$ are two separate random variables that are independent of each other. We say that the corresponding random process is *uncorrelated* (also defined more formally below). As a consequence of our independence assumption, we have that

$$\mathbb{E}\{W[k]W[l]\} = \mathbb{E}\{W^2[k]\}\delta(k-l) = \begin{cases} \mathbb{E}\{W^2[k]\} & k = l \\ 0 & k \neq l. \end{cases}$$

In the case that $W[k]$ is a Gaussian with mean zero and (stationary) standard deviation σ, then $\mathbb{E}\{W[k]W[l]\} = \sigma^2\,\delta(k-l)$.

We next wish to describe the evolution of the state x in equation (C.11) in the case when W is a random variable. In order to do this, we describe the state x as a sequence of random variables $X[k]$, $k = 1, \cdots, N$. Looking back at equation (C.11), we see that even if $W[k]$ is an uncorrelated sequence of random variables, then the states $X[k]$ are not uncorrelated since

$$X[k+1] = AX[k] + FW[k],$$

and hence the probability distribution for X at time $k+1$ depends on the value of X at time k (as well as the value of W at time k), similar to the situation in Example C.1.

Since each $X[k]$ is a random variable, we can define the mean and variance as $\mu[k]$ and $\sigma^2[k]$ using the previous definitions at each time k:

$$\mu[k] := \mathbb{E}\{X[k]\} = \int_{-\infty}^{\infty} x\, p(x,k)\, dx,$$
$$\sigma^2[k] := \mathbb{E}\{(X[k]-\mu[k])^2\} = \int_{-\infty}^{\infty} (x-\mu[k])^2\, p(x,k)\, dx.$$

To capture the relationship between the current state and the future state, we define the *correlation function* for a random process as

$$\rho(k_1,k_2) := \mathbb{E}\{X[k_1]X[k_2]\} = \int_{-\infty}^{\infty} x_1 x_2\, p(x_1,x_2;k_1,k_2)\, dx_1 dx_2$$

The function $p(x_i,x_j;k_1,k_2)$ is the *joint probability density function*, which depends on the times k_1 and k_2. A process is *stationary* if $p(x,k+d) = p(x,d)$ for all k, $p(x_i,x_j;k_1+d,k_2+d) = p(x_i,x_j;k_1,k_2)$, etc. In this case we can write $p(x_i,x_j;d)$ for the joint probability distribution. We will almost always restrict to this case. Similarly, we will write $p(k_1,k_2)$ as $p(d) = p(k,k+d)$.

We can compute the correlation function by explicitly computing the joint pdf (see Example C.1) or by directly computing the expectation. Suppose that we take a random process of the form (C.11) with $x[0] = 0$ and W having zero mean and standard deviation σ. The correlation function is given by

$$\begin{aligned}\mathbb{E}\{X[k_1]X[k_2]\} &= E\Big\{\Big(\sum_{i=0}^{k_1-1} A^{k_1-i}BW[i]\Big)\Big(\sum_{j=0}^{k_2-1} A^{k_2-j}BW[j]\Big)\Big\} \\ &= E\Big\{\sum_{i=0}^{k_1-1}\sum_{j=0}^{k_2-1} A^{k_1-i}BW[i]W[j]BA^{k_2-j}\Big\}.\end{aligned}$$

We can now use the linearity of the expectation operator to pull this inside the summations:

$$\begin{aligned}\mathbb{E}\{X[k_1]X[k_2]\} &= \sum_{i=0}^{k_1-1}\sum_{j=0}^{k_2-1} A^{k_1-i}B\mathbb{E}\{W[i]W[j]\}BA^{k_2-j} \\ &= \sum_{i=0}^{k_1-1}\sum_{j=0}^{k_2-1} A^{k_1-i}B\sigma^2\delta(i-j)BA^{k_2-j} \\ &= \sum_{i=0}^{k_1-1} A^{k_1-i}B\sigma^2 BA^{k_2-i}.\end{aligned}$$

Note that the correlation function depends on k_1 and k_2.

We can see the dependence of the correlation function on the time more clearly by letting $d = k_2 - k_1$ and writing

$$\rho(k, k+d) = \mathbb{E}\{X[k]X[k+d]\} = \sum_{i=0}^{k_1-1} A^{k-i} B\sigma^2 B A^{d+k-i}$$
$$= \sum_{j=1}^{k} A^j B\sigma^2 B A^{j+d} = \Big(\sum_{j=1}^{k} A^j B\sigma^2 B A^j\Big) A^d.$$

In particular, if the discrete time system is stable then $|A| < 1$ and the correlation function decays as we take points that are further departed in time (d large). Furthermore, if we let $k \to \infty$ (i.e., look at the steady state solution) then the correlation function only depends on d (assuming the sum converges) and hence the steady state random process is stationary.

In our derivation so far, we have assumed that $X[k+1]$ only depends on the value of the state at time k (this was implicit in our use of equation (C.11) and the assumption that $W[k]$ is independent of X). This particular assumption is known as the *Markov property* for a random process: a Markovian process is one in which the distribution of possible values of the state at time k depends only on the values of the state at the prior time and not earlier. Written more formally, we say that a discrete random process is Markovian if

$$p_{X,k}(x \mid X[k-1], X[k-2], \ldots, X[0]) = p_{X,k}(x \mid X[k-1]).$$

Markov processes are roughly equivalent to state space dynamical systems, where the future evolution of the system can be completely characterized in terms of the current value of the state (and not it history of values prior to that).

We now consider the case where our time index is no longer discrete, but instead varies continuously. A fully rigorous derivation requires careful use of measure theory and is beyond the scope of this text, so we focus here on the concepts that will be useful for modeling and analysis of important physical properties.

A *continuous-time random process* is a stochastic system characterized by the evolution of a random variable $X(t)$, $t \in [0, T]$. We are interested in understanding how the (random) state of the system is related at separate times. The process is defined in terms of the "correlation" of $X(t_1)$ with $X(t_2)$. We assume, as above, that the process is described by continuous random variables, but the discrete state case (with time still modeled as a real variable) can be handled in a similar fashion.

We call $X(t) \in \mathbb{R}^n$ the *state* of the random process at time t. For the case $n > 1$, we have a vector of random processes:

$$X(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{pmatrix}$$

We can characterize the state in terms of a (vector-valued) time-varying pdf,

$$P(x_l \le X_i(t) \le x_u) = \int_{x_l}^{x_u} p_{X_i}(x;t)dx.$$

Note that the state of a random process is not enough to determine the next state (otherwise it would be a deterministic process). We typically omit indexing of the individual states unless the meaning is not clear from context.

We can characterize the dynamics of a random process by its statistical characteristics, written in terms of joint probability density functions:

$$P(x_{1l} \le X_i(t_1) \le x_{1u}, x_{2l} \le X_j(t_2) \le x_{2u}) = \int_{x_{2l}}^{x_{2u}} \int_{x_{1l}}^{x_{1u}} p_{X_i,Y_i}(x_1, x_2; t_1, t_2)\, dx_1 dx_2$$

The function $p(x_i, x_j; t_1, t_2)$ is called a *joint probability density function* and depends both on the individual states that are being compared and the time instants over which they are compared. Note that if $i = j$, then p_{X_i,X_i} describes how X_i at time t_1 is related to X_i at time t_2.

In general, the distributions used to describe a random process depend on the specific time or times that we evaluate the random variables. However, in some cases the relationship only depends on the difference in time and not the absolute times (similar to the notion of time invariance in deterministic systems, as described in AM08). A process is *stationary* if $p(x, t+\tau) = p(x, t)$ for all τ, $p(x_i, x_j; t_1 + \tau, t_2 + \tau) = p(x_i, x_j; t_1, t_2)$, etc. In this case we can write $p(x_i, x_j; \tau)$ for the joint probability distribution. Stationary distributions roughly correspond to the steady state properties of a random process and we will often restrict our attention to this case.

In looking at biomolecular systems, we are going to be interested in random processes in which the changes in the state occur when a random event occurs (such as a molecular reaction or binding event). In this case, it is natural to describe the state of the system in terms of a set of times $t_0 < t_1 < t_2 < \cdots < t_n$ and $X(t_i)$ is the random variable that corresponds to the possible states of the system at time t_i. Note that time time instants do not have to be uniformly spaced and most often (for biomolecular systems) they will not be. All of the definitions above carry through, and the process can now be described by a probability distribution of the form

$$P\Big(X(t_i) \in [x_i, x_i + dx_i], i = 1, \ldots, n\Big) = \int \ldots \int p(x_n, x_{n-1}, \ldots, x_0; t_n, t_{n-1}, \ldots, t_0) dx_n\, dx_{n-1}\, dx_1,$$

where dx_i are taken as infinitesimal quantities.

An important class of stochastic systems is those for which the next state of the system depends only on the current state of the system and not the history of the

process. Suppose that

$$P\Big(X(t_n) \in [x_n, x_n + dx_n] \mid X(t_i) \in [x_i, x_i + dx_i], i = 1, \ldots, n-1\Big) \\ = P\Big(X(t_n) \in [x_n, x_n + dx_n] \mid X(t_{n-1}) \in [x_{n-1}, x_{n-1} + dx_{n-1}]\Big). \quad \text{(C.12)}$$

That is, the probability of being in a given state at time t_n depends *only* on the state that we were in at the previous time instant t_{n-1} and not the entire history of states prior to t_{n-1}. A stochastic process that satisfies this property is called a *Markov process*.

In practice we do not usually specify random processes via the joint probability distribution $p(x_i, x_j; t_1, t_2)$ but instead describe them in terms of a *propogater function*. Let $X(t)$ be a Markov process and define the Markov propogater as

$$\Xi(dt; x, t) = X(t + dt) - X(t), \text{ given } X(t) = x.$$

The propogater function describes how the random variable at time t is related to the random variable at time $t + dt$. Since both $X(t + dt)$ and $X(t)$ are random variables, $\Xi(dt; x, t)$ is also a random variable and hence it can be described by its density function, which we denote as $\Pi(\xi, x; dt, t)$:

$$P(x \le X(t + dt) \le x + \xi) = \int_x^{x+\xi} \Pi(dx, x; dt, t)\, dx.$$

The previous definitions for mean, variance and correlation can be extended to the continuous time, vector-valued case by indexing the individual states:

$$E\{X(t)\} = \begin{pmatrix} E\{X_1(t)\} \\ \vdots \\ E\{X_n(t)\} \end{pmatrix} =: \mu(t)$$

$$E\{(X(t) - \mu(t))(X(t) - \mu(t))^T\} = \begin{pmatrix} E\{X_1(t)X_1(t)\} & \ldots & E\{X_1(t)X_n(t)\} \\ & \ddots & \vdots \\ & & E\{X_n(t)X_n(t)\} \end{pmatrix} =: \Sigma(t)$$

$$E\{X(t)X^T(s)\} = \begin{pmatrix} E\{X_1(t)X_1(s)\} & \ldots & E\{X_1(t)X_n(s)\} \\ & \ddots & \vdots \\ & & E\{X_n(t)X_n(s)\} \end{pmatrix} =: R(t, s)$$

Note that the random variables and their statistical properties are all indexed by the time t (and s). The matrix $R(t, s)$ is called the *correlation matrix* for $X(t) \in \mathbb{R}^n$. If $t = s$ then $R(t, t)$ describes how the elements of x are correlated at time t (with each other) and in the case that the processes have zero mean, $R(t, t) = \Sigma(t)$. The elements on the diagonal of $\Sigma(t)$ are the variances of the corresponding scalar variables. A random process is uncorrelated if $R(t, s) = 0$ for all $t \neq s$. This implies that $X(t)$ and $X(s)$ are independent random events and is equivalent to $p_{X,Y}(x, y) = p_X(x)p_Y(y)$.

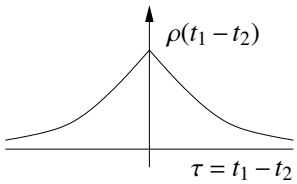


Figure C.3: Correlation function for a first-order Markov process.

If a random process is stationary, then it can be shown that $R(t+\tau, s+\tau) = R(t,s)$ and it follows that the correlation matrix depends only on $t-s$. In this case we will often write $R(t,s) = R(s-t)$ or simple $R(\tau)$ where τ is the correlation time. The correlation matrix in this case is simply $R(0)$.

In the case where X is also scalar random process, the correlation matrix is also a scalar and we will write $\rho(\tau)$, which we refer to as the (scalar) correlation function. Furthermore, for stationary scalar random processes, the correlation function depends only on the absolute value of the correlation function, so $\rho(\tau) = \rho(-\tau) = \rho(|\tau|)$. This property also holds for the diagonal entries of the correlation matrix since $R_{ii}(s,t) = R_{ii}(t,s)$ from the definition.

Definition C.7 (Ornstein-Uhlenbeck process)**.** Consider a scalar random process defined by a Gaussian pdf with $\mu = 0$,

$$p(x,t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}},$$

and a correlation function given by

$$\rho(t_1, t_2) = \frac{Q}{2\omega_0} e^{-\omega_0 |t_2 - t_1|}.$$

The correlation function is illustrated in Figure C.3. This process is known as an *Ornstein-Uhlenbeck process* and it is a stationary process.

Note on terminology. The terminology and notation for covariance and correlation varies between disciplines. The term covariance is often used to refer to both the relationship between different variables X and Y and the relationship between a single variable at different times, $X(t)$ and $X(s)$. The term "cross-covariance" is used to refer to the covariance between two random vectors X and Y, to distinguish this from the covariance of the elements of X with each other. The term "cross-correlation" is sometimes also used. Finally, the term "correlation coefficient" refers to the normalized correlation $\bar{\rho}(t,s) = \mathbb{E}\{X(t)X(s)\}/\mathbb{E}\{X(t)X(t)\}$..

`MATLAB` has a number of functions to implement covariance and correlation, which mostly match the terminology here:

- `cov(X)` - this returns the variance of the vector `X` that represents samples of a given random variable or the covariance of the columns of a matrix X where the rows represent observations.
- `cov(X, Y)` - equivalent to `cov([X(:), Y(:)])`. Computes the covariance between the columns of X and Y, where the rows are observations.
- `xcorr(X, Y)` - the "cross-correlation" between two random sequences. If these sequences came from a random process, this is correlation function $\rho(t)$.
- `xcov(X, Y)` - this returns the "cross-covariance", which `MATLAB` defines as the "mean-removed cross-correlation".

The `MATLAB` help pages give the exact formulas used for each, so the main point here is to be careful to make sure you know what you really want.

We will also make use of a special type of random process referred to as "white noise". A *white noise process* $X(t)$ satisfies $E\{X(t)\} = 0$ and $R(t,s) = W\delta(s-t)$, where $\delta(\tau)$ is the impulse function and W is called the *noise intensity*. White noise is an idealized process, similar to the impulse function or Heaviside (step) function in deterministic systems. In particular, we note that $\rho(0) = E\{X^2(t)\} = \infty$, so the covariance is infinite and we never see this signal in practice. However, like the step function, it is very useful for characterizing the responds of a linear system, as described in the following proposition. It can be shown that the integral of a white noise process is a Wiener process, and so often white noise is described as the derivative of a Wiener process.

C.3 Discrete-State Random Processes

There are a number of specialized discrete random processes that are relevant for biochemical systems. In this section we give a brief introduction to these processes.

A *birth-death* process is one in which the states of the process represent integer-value counts of different species populations and the transitions between states are restricted to either incrementing (birth) or decrementing (death) a given species. This type of model is often used to represent chemical reactions such as the production and degradation of proteins.

Example C.3 (Protein production)**.** ∇

A more general type of discrete random process is a *Markov chain*. In a Markov chain, evolution of the discrete states occurs by execution of allowable transitions between two states. Each transition has a specified probability, which is used to determine whether a system will transition from its current state into a different state (corresponding to an allowable transition). An important property, called the

Markov property, is that the transition probability only depends on the value of the current state, not the previous values of the state.

We define a Markov chain by giving the set of transition probabilities

$$q_{ij}(t,\tau) = P(X(t+\tau) = s_j | X(t) = s_i),$$

where $s_i, s_j \in S$, t is the current time and τ is the time interval over which we are interested. If $q_{ij}(t,\tau) \neq 0$ for some $\tau \neq 0$ then we say that the transition is allowable at time t. If q_{ij} is independent of t then we say that the process is *stationary* and we omit the argument t. In the special case that we are only interested in a fixed τ (i.e., we are using a discrete-time model) then we omit this argument as well.

It is generally difficult to describe the probability of being in a particular state in a Markov process at a given time. Instead, we often resort to describing the steady state distributions, assuming that they exist. For a stationary Markov chain, we can look at the equilibrium distributions, which are those distributions π that satisfy

$$\pi_i = q_{ij}(\tau)\pi_j, \quad \text{for all } i, j.$$

Example C.4 (Protein expression)**.** ∇

C.4 Input/Output Linear Stochastic Systems

We now consider the problem of how to compute the response of a linear system to a random process. We assume we have a linear system described in state space as

$$\dot{X} = AX + FW, \qquad Y = CX \tag{C.13}$$

Given an "input" W, which is itself a random process with mean $\mu(t)$, variance $\sigma^2(t)$ and correlation $\rho(t, t+\tau)$, what is the description of the random process Y?

Let W be a white noise process, with zero mean and noise intensity Q:

$$\rho(\tau) = Q\delta(\tau).$$

We can write the output of the system in terms of the convolution integral

$$Y(t) = \int_0^t h(t-\tau)W(\tau)\,d\tau,$$

where $h(t-\tau)$ is the impulse response for the system

$$h(t-\tau) = Ce^{A(t-\tau)}B + D\delta(t-\tau).$$

We now compute the statistics of the output, starting with the mean:

$$\begin{aligned}\mathbb{E}\{Y(t)\} &= E\{\int_0^t h(t-\eta)W(\eta)\,d\eta\} \\ &= \int_0^t h(t-\eta)E\{W(\eta)\}\,d\eta = 0.\end{aligned}$$

Note here that we have relied on the linearity of the convolution integral to pull the expectation inside the integral.

We can compute the covariance of the output by computing the correlation $\rho(\tau)$ and setting $\sigma^2 = \rho(0)$. The correlation function for y is

$$\begin{aligned}\rho_Y(t,s) = E\{Y(t)Y(s)\} &= E\{\int_0^t h(t-\eta)W(\eta)\,d\eta \cdot \int_0^s h(s-\xi)W(\xi)\,d\xi\} \\ &= E\{\int_0^t \int_0^s h(t-\eta)W(\eta)W(\xi)h(s-\xi)\,d\eta d\xi\}\end{aligned}$$

Once again linearity allows us to exchange expectation and integration

$$\begin{aligned}\rho_Y(t,s) &= \int_0^t \int_0^s h(t-\eta)E\{W(\eta)W(\xi)\}h(s-\xi)\,d\eta d\xi \\ &= \int_0^t \int_0^s h(t-\eta)Q\delta(\eta-\xi)h(s-\xi)\,d\eta d\xi \\ &= \int_0^t h(t-\eta)Qh(s-\eta)\,d\eta\end{aligned}$$

Now let $\tau = s - t$ and write

$$\begin{aligned}\rho_Y(\tau) = \rho_Y(t,t+\tau) &= \int_0^t h(t-\eta)Qh(t+\tau-\eta)\,d\eta \\ &= \int_0^t h(\xi)Qh(\xi+\tau)\,d\xi \qquad (\text{setting } \xi = t-\eta)\end{aligned}$$

Finally, we let $t \to \infty$ (steady state)

$$\lim_{t\to\infty} \rho_Y(t,t+\tau) = \bar{\rho}_Y(\tau) = \int_0^\infty h(\xi)Qh(\xi+\tau)d\xi \tag{C.14}$$

If this integral exists, then we can compute the second order statistics for the output Y.

We can provide a more explicit formula for the correlation function ρ in terms of the matrices A, F and C by expanding equation (C.14). We will consider the general case where $W \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ and use the correlation matrix $R(t,s)$ instead of the correlation function $\rho(t,s)$. Define the *state transition matrix* $\Phi(t,t_0) = e^{A(t-t_0)}$ so that the solution of system (C.13) is given by

$$x(t) = \Phi(t,t_0)x(t_0) + \int_{t_0}^t \Phi(t,\lambda)Fw(\lambda)d\lambda$$

Proposition C.2 (Stochastic response to white noise)**.** *Let $E\{X(t_0)X^T(t_0)\} = P(t_0)$ and W be white noise with $E\{W(\lambda)W^T(\xi)\} = R_W\delta(\lambda-\xi)$. Then the correlation matrix for X is given by*

$$R_X(t,s) = P(t)\Phi^T(s,t)$$

where $P(t)$ satisfies the linear matrix differential equation

$$\dot{P}(t) = AP + PA^T + FR_W F, \qquad P(0) \quad = P_0.$$

Proof. Using the definition of the correlation matrix, we have

$$\begin{aligned} E\{X(t)X^T(s)\} &= E\Big\{\Phi(t,0)X(0)X^T(0)\Phi^T(t,0) + \text{cross terms} \\ &\qquad + \int_0^t \Phi(t,\xi)FW(\xi)\,d\xi \int_0^s W^t(\lambda)F^T\Phi(s,\lambda)\,d\lambda\Big\} \\ &= \Phi(t,0)E\{X(0)X^T(0)\}\Phi(s,0) \\ &\qquad + \int_0^t \int_0^s \Phi(t,\xi)FE\{W(\xi)W^T(\lambda)\}F^T\Phi(s,\lambda)\,d\xi\,d\lambda \\ &= \Phi(t,0)P(0)\phi^T(s,0) + \int_0^t \Phi(t,\lambda)FR_W(\lambda)F^T\Phi(s,\lambda)\,d\lambda. \end{aligned}$$

Now use the fact that $\Phi(s,0) = \Phi(s,t)\Phi(t,0)$ (and similar relations) to obtain

$$R_X(t,s) = P(t)\Phi^T(s,t)$$

where

$$P(t) = \Phi(t,0)P(0)\Phi^T(t,0) + \int_0^T \Phi(t,\lambda)FR_W F^T(\lambda)\Phi^T(t,\lambda)d\lambda$$

Finally, differentiate to obtain

$$\dot{P}(t) = AP + PA^T + FR_W F, \qquad P(0) \quad = P_0$$

(see Friedland for details). □

The correlation matrix for the output Y can be computed using the fact that $Y = CX$ and hence $R_Y = C^T R_X C$. We will often be interested in the steady state properties of the output, which are given by the following proposition.

Proposition C.3 (Steady state response to white noise)**.** *For a time-invariant linear system driven by white noise, the correlation matrices for the state and output converge in steady state to*

$$R_X(\tau) = R_X(t,t+\tau) = Pe^{A^T\tau}, \qquad R_Y(\tau) = CR_X(\tau)C^T$$

where P satisfies the algebraic equation

$$AP + PA^T + FR_W F^T = 0 \qquad P > 0. \tag{C.15}$$

Equation (C.15) is called the *Lyapunov equation* and can be solved in MATLAB using the function `lyap`.

Example C.5 (First-order system)**.** Consider a scalar linear process

$$\dot{X} = -aX + W, \qquad Y = cX,$$

where W is a white, Gaussian random process with noise intensity σ^2. Using the results of Proposition C.2, the correlation function for X is given by

$$R_X(t, t+\tau) = p(t)e^{-a\tau}$$

where $p(t) > 0$ satisfies

$$p(t) = -2ap + \sigma^2.$$

We can solve explicitly for $p(t)$ since it is a (non-homogeneous) linear differential equation:

$$p(t) = e^{-2at}p(0) + (1 - e^{-2at})\frac{\sigma^2}{2a}.$$

Finally, making use of the fact that $Y = cX$ we have

$$\rho(t, t+\tau) = c^2(e^{-2at}p(0) + (1 - e^{-2at})\frac{\sigma^2}{2a})e^{-a\tau}.$$

In steady state, the correlation function for the output becomes

$$\rho(\tau) = \frac{c^2\sigma^2}{2a}e^{-a\tau}.$$

Note correlation function has the same form as the Ornstein-Uhlenbeck process in Example C.7 (with $Q = c^2\sigma^2$). ∇

As in the case of deterministic linear systems, we can analyze a stochastic linear system either in the state space or the frequency domain. The frequency domain approach provides a very rich set of tools for modeling and analysis of interconnected systems, relying on the frequency response and transfer functions to represent the flow of signals around the system.

Given a random process $X(t)$, we can look at the frequency content of the properties of the response. In particular, if we let $\rho(\tau)$ be the correlation function for a (scalar) random process, then we define the *power spectral density function* as the Fourier transform of ρ:

$$S(\omega) = \int_{-\infty}^{\infty} \rho(\tau)e^{-j\omega\tau}\,d\tau, \qquad \rho(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{j\omega\tau}\,d\tau.$$

The power spectral density provides an indication of how quickly the values of a random process can change through the frequency content: if there is high frequency content in the power spectral density, the values of the random variable can change quickly in time.

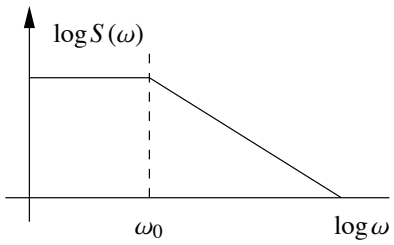


Figure C.4: Spectral power density for a first-order Markov process.

.

Example C.6 (First-order Markov process)**.** To illustrate the use of these measures, consider a first-order Markov process as defined in Example C.7. The correlation function is

$$\rho(\tau) = \frac{Q}{2\omega_0} e^{-\omega_0(\tau)}.$$

The power spectral density becomes

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} \frac{Q}{2\omega_0} e^{-\omega|\tau|} e^{-j\omega\tau}\, d\tau \\ &= \int_{-\infty}^{0} \frac{Q}{2\omega_0} e^{(\omega - j\omega)\tau}\, d\tau + \int_{0}^{\infty} \frac{Q}{2\omega_0} e^{(-\omega - j\omega)\tau}\, d\tau = \frac{Q}{\omega^2 + \omega_0^2}. \end{aligned}$$

We see that the power spectral density is similar to a transfer function and we can plot $S(\omega)$ as a function of ω in a manner similar to a Bode plot, as shown in Figure C.4. Note that although $S(\omega)$ has a form similar to a transfer function, it is a real-valued function and is not defined for complex s. ∇

Using the power spectral density, we can more formally define "white noise": a *white noise process* is a zero-mean, random process with power spectral density $S(\omega) = W$ = constant for all ω. If $X(t) \in \mathbb{R}^n$ (a random vector), then $W \in \mathbb{R}^{n\times n}$. We see that a random process is white if all frequencies are equally represented in its power spectral density; this spectral property is the reason for the terminology "white". The following proposition verifies that this formal definition agrees with our previous (time domain) definition.

Proposition C.4. *For a white noise process,*

$$\rho(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{j\omega\tau}\, d\tau = W\delta(\tau),$$

where $\delta(\tau)$ is the unit impulse function.

Proof. If $\tau \neq 0$ then

$$\rho(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(\cos(\omega\tau) + j\sin(\omega\tau)\, d\tau = 0$$

If $\tau = 0$ then $\rho(\tau) = \infty$. Can show that

$$\rho(0) = \lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} \int_{-\infty}^{\infty} (\cdots)\, d\omega d\tau = W\delta(0)$$

□

Given a linear system

$$\dot{X} = AX + FW, \qquad Y = CX,$$

with W given by white noise, we can compute the spectral density function corresponding to the output Y. We start by computing the Fourier transform of the steady state correlation function (C.14):

$$\begin{aligned} S_Y(\omega) &= \int_{-\infty}^{\infty} \left[\int_0^{\infty} h(\xi) Q h(\xi + \tau) d\xi \right] e^{-j\omega\tau}\, d\tau \\ &= \int_0^{\infty} h(\xi) Q \left[\int_{-\infty}^{\infty} h(\xi + \tau) e^{-j\omega\tau}\, d\tau \right] d\xi \\ &= \int_0^{\infty} h(\xi) Q \left[\int_0^{\infty} h(\lambda) e^{-j\omega(\lambda - \xi)}\, d\lambda \right] d\xi \\ &= \int_0^{\infty} h(\xi) e^{j\omega\xi}\, d\xi \cdot QH(j\omega) = H(-j\omega) Q_u H(j\omega) \end{aligned}$$

This is then the (steady state) response of a linear system to white noise.

As with transfer functions, one of the advantages of computations in the frequency domain is that the composition of two linear systems can be represented by multiplication. In the case of the power spectral density, if we pass white noise through a system with transfer function $H_1(s)$ followed by transfer function $H_2(s)$, the resulting power spectral density of the output is given by

$$S_Y(\omega) = H_1(-j\omega) H_2(-j\omega) Q_u H_2(j\omega) H_1(j\omega).$$

As stated earlier, white noise is an idealized signal that is not seen in practice. One of the ways to produced more realistic models of noise and disturbances it to apply a filter to white noise that matches a measured power spectral density function. Thus, we wish to find a covariance W and filter $H(s)$ such that we match the statistics $S(\omega)$ of a measured noise or disturbance signal. In other words, given $S(\omega)$, find $W > 0$ and $H(s)$ such that $S(\omega) = H(-j\omega) W H(j\omega)$. This problem is know as the *spectral factorization problem*.

Figure C.5 summarizes the relationship between the time and frequency domains.

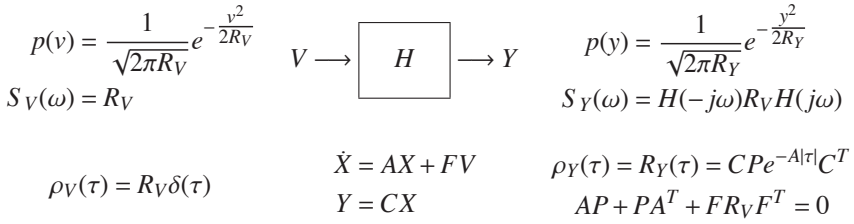


Figure C.5: Summary of steady state stochastic response.