CHAPTER 1

Review: Probability, Random Processes, and Linear Systems

1.1. Probability

In this section, we briefly review some necessary concepts of probability that will be used throughout this text.

1.1.1. Discrete-valued random variables. A discrete-valued random variable takes values on a discrete, finite set $\mathcal{S}$. For example, a single roll of a six-sided die takes values $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. The set need not take numerical values; for instance, the outcome of a coin flip might be $\mathcal{S} = \{\text{Heads}, \text{Tails}\}$.

The probabilities of each outcome in $\mathcal{S}$ are expressed in a *probability mass function* (pmf). For a discrete-valued random variable x, we will write the pmf as $p(x)$.

Example 1.1. *For a fair die, with $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, every possible outcome has the same probability. Thus, the pmf is given by*

$$p(x) = \begin{cases} \frac{1}{6}, & x \in \mathcal{S}, \\ 0, & x \notin \mathcal{S}. \end{cases} \tag{1.1}$$

We will make use of the following properties of the pmf:

(1) For all $x \in \mathcal{S}$, $p(x) \geq 0$, that is, probability is never negative.
(2) Let $\mathcal{R}$ be a subset of $\mathcal{S}$. Then the probability that an event in $\mathcal{R}$ occurs is $\sum_{x \in \mathcal{R}} p(x)$. (This is equivalent to saying that the individual outcomes in $\mathcal{S}$ are mutually exclusive.)
(3) $\sum_{x \in \mathcal{S}} p(x) = 1$, that is, the total probability is 1. (Combined with property 2, this means that some event in $\mathcal{S}$ must happen with probability 1.)

Let $g(x)$ represent some function of the random variable x. Then the *expected value* of $g(x)$, written $E[g(x)]$, is defined as

$$E[g(x)] = \sum_{x \in \mathcal{S}} g(x)p(x). \tag{1.2}$$

We will make use of the following properties of expected value.

(1) $E[g_1(x) + g_2(x)] = E[g_1(x)] + E[g_2(x)]$.
(2) If a is a deterministic (i.e., known, non-random) constant, then $E[ag(x)] = aE[g(x)]$, and $E[a] = a$.

The *mean* and *variance* are two important special cases of expectation. The mean, written μ, is given by

$$\begin{aligned} \mu &= E[x] && (1.3)\\ &= \sum_{x \in \mathcal{S}} xp(x). && (1.4) \end{aligned}$$

The variance, written either $\mathrm{Var}[x]$ or σ^2, is given by

$$\begin{aligned} \mathrm{Var}[x] &= E[(x-\mu)^2] && (1.5)\\ &= \sum_{x \in \mathcal{S}} (x-\mu)^2 p(x). && (1.6) \end{aligned}$$

There is an alternative way to calculate $\mathrm{Var}[x]$, making use of the properties of expectation. Starting with (1.5), we have

$$\begin{aligned} E[(x-\mu)^2] &= E[x^2 - 2\mu x + \mu^2] && (1.7)\\ &= E[x^2] - E[2\mu x] + E[\mu^2] && (1.8)\\ &= E[x^2] - 2\mu E[x] + \mu^2 && (1.9)\\ &= E[x^2] - \mu^2 && (1.10)\\ &= E[x^2] - E[x]^2, && (1.11) \end{aligned}$$

where (1.8) follows from the first property of expectation, (1.9) follows from the second property, and the remainder follows from the fact that $\mu = E[x]$, by definition.

Examples ...

1.1.2. Joint and conditional probability.

1.1.3. Continuous-valued random variables. A continuous-valued random variable takes values from the entire set of real numbers $\mathbb{R}$. For example, the temperature tomorrow at noon in downtown Toronto is a continuous-valued random variable.

We will normally use the probability density function (pdf) to describe

Probability density function; expected value; mean and variance; examples.

1.1.4. The Gaussian distribution. Definition; properties (e.g., even function).

A Gaussian random variable x with with mean μ and variance σ^2 has a probability density function given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right). \tag{1.12}$$

Integrals over this pdf may be expressed in terms of the *error function complementary*, $\mathrm{erfc}(\cdot)$, which is defined as

$$\mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_{t=z}^{\infty} \exp(-t^2) dt. \tag{1.13}$$

The function $\mathrm{erfc}(\cdot)$ has the following mathematical interpretation: if t is a Gaussian random variable with mean $\mu = 0$ and variance $\sigma^2 = 1/2$, then $\mathrm{erfc}(z) = \Pr(|t| > z)$. Furthermore, due to the symmetry of the Gaussian pdf about the mean, we illustrate in Figure X that

$$\Pr(t > z) = \Pr(t < z) = \frac{1}{2}\mathrm{erfc}(z). \tag{1.14}$$

Using a change of variables, $\mathrm{erfc}(\cdot)$ may be used to calculate an arbitrary Gaussian integral. For instance, for the random variable x with pdf $f(x)$ in (1.12), suppose we want to calculate the probability $\Pr(x > z)$. This probability can be expressed as

$$\Pr(x > z) = \int_{x=z}^{\infty} f(x) dx \tag{1.15}$$

$$= \int_{x=z}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx. \tag{1.16}$$

Now we make the substitution

$$t = \frac{x-\mu}{\sqrt{2\sigma^2}}. \tag{1.17}$$

To perform a change of variables in an integral, we need to replace both x and dx with the equivalent functions of t. Solving for x, we have that

$$x = \sqrt{2\sigma^2}t + \mu, \tag{1.18}$$

so taking the first derivative of x with respect to t, dx is given by

$$dx = \sqrt{2\sigma^2}dt. \tag{1.19}$$

Substituting (1.18)-(1.19) into (1.16), we get

$$\begin{aligned}
\Pr(x > z) &= \int_{x=z}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx && (1.20)\\
&= \int_{\sqrt{2\sigma^2}t+\mu=z}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-t^2\right) \sqrt{2\sigma^2}dt && (1.21)\\
&= \int_{t=(z-\mu)/\sqrt{2\sigma^2}}^{\infty} \frac{1}{\sqrt{\pi}} \exp\left(-t^2\right) dt && (1.22)\\
&= \frac{1}{2}\mathrm{erfc}\left(\frac{z-\mu}{\sqrt{2\sigma^2}}\right). && (1.23)
\end{aligned}$$

1.2. Discrete-Time Random Processes

There are many ways to define a random process, but for our purposes, the following is sufficient:

- A random process is a function of time $X(t)$, so that for each fixed time t^*, $X(t^*)$ is a random variable.

As a result, we can write the probability density function (pdf) of the random process at any given time. For example, $f_{X(t^*)}(x)$ represents the pdf of the random process at time t^*. Joint probability density functions measure the joint probability of the process at k different times; these are called *kth order statistics* of the random process. For example, for $k = 2$ and times t_1 and t_2, we can write the second order statistics as $f_{X(t_1),X(t_2)}(x_1, x_2)$.

1.2.1. Definition, Mean, and Variance. It's easy to imagine a random process in discrete time, as merely a sequence of random variables, one for each time interval. For instance, consider the following two random processes defined at integer times $t \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$:

EXAMPLE 1.2. *At each time* $t \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, *a fair coin is flipped. If the coin shows heads after the flip at time* t, *then* $X(t) = 1$; *otherwise,* $X(t) = -1$. *Thus, for any integer* t^*, *we can write*

$$f_{X(t^*)}(x) = \begin{cases} 0.5, & x = +1; \\ 0.5, & x = -1; \\ 0 & \text{otherwise.} \end{cases}$$

Since, at each fixed time t, the random process is a random variable, we can calculate the mean and variance of the process at each fixed time as usual for random variables. Thus, for the process as a whole, the mean and variance for a random process are calculated as functions of time. For instance, for the process in Example 1.2, the mean of this process is given by

$$\begin{aligned} \mu(t) &= \sum_{x \in \{+1,-1\}} x f_{X(t)}(x) \\ &= (+1)(0.5) + (-1)(0.5) \\ &= 0 \end{aligned}$$

for all t. The variance of the process is given by

$$\begin{aligned} \sigma^2(t) &= \sum_{x \in \{+1,-1\}} (x - \mu(t))^2 f_{X(t)}(x) \\ &= (+1-0)^2(0.5) + (-1-0)^2(0.5) \\ &= 1 \end{aligned}$$

for all t.

As an alternative, the following more compicated example has mean and variance that are non-trivial functions of time:

EXAMPLE 1.3. *Let* $X(0) = 0$. *For each* $t \in \{1, 2, \ldots\}$, *a fair coin is flipped. If the coin shows heads after the flip at time* t, *then* $X(t) = X(t-1) + 1$; *otherwise,* $X(t) = X(t-1)$.

For any t, *it is clear that* $X(t)$ *is the number of heads in the previous* t *trials. Such random variables are represented by the binomial distribution* [**1**]. *Thus,*

$$f_{X(t)}(x) = \binom{t}{x} \frac{1}{2^t}.$$

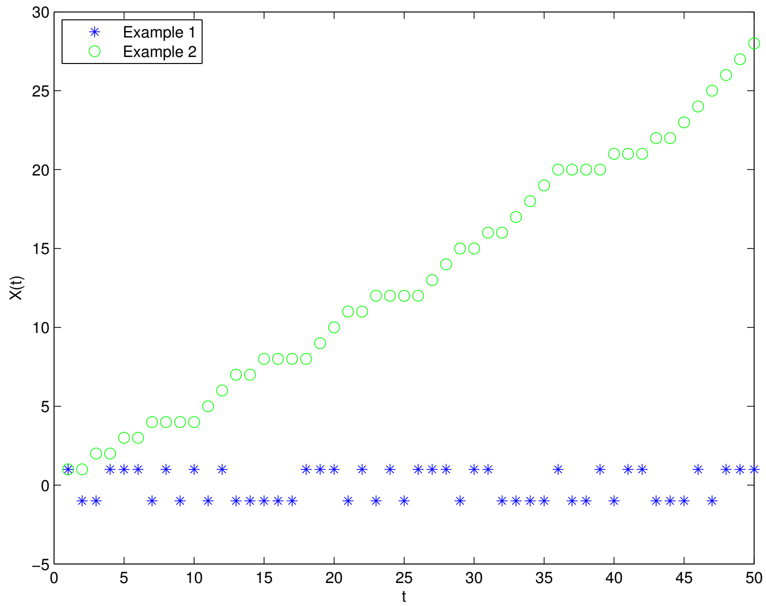

FIGURE 1.1. Illustration of the discrete-time random processes from Examples 1.2 and 1.3.

The mean of this random process is given by

$$\mu(t) = \frac{t}{2},$$

and the variance is given by

$$\sigma^2(t) = \frac{t}{4}.$$

The reader is asked to prove these values in the exercises.

Instances of the random processes from Examples 1.2 and 1.3 are given in Figure 1.1.

1.2.2. Autocorrelation. Suppose you wanted a measure of correlation between two random variables, X_1 and X_2, with the same mean $\mu = 0$ and the same variance $\sigma^2 > 0$. As a candidate for this measure, consider

$$R = E[X_1 X_2]. \tag{1.24}$$

If the random variables are independent (i.e., uncorrelated), then since $E[X_1X_2] = E[X_1]E[X_2]$ for independent random variables, we would have

$$R = E[X_1]E[X_2] = \mu^2 = 0,$$

bearing in mind that each of the random variables are zero mean. On the other hand, if the two random variables are completely correlated (i.e., $X_1 = X_2$), we would have

$$R = E[X_1 X_2] = E[X_1^2] = \sigma^2.$$

Further, if they were completely anticorrelated (i.e., $X_1 = -X_2$), it is easy to see that $R = -\sigma^2$.

This measure of correlation also has the following nice property:

THEOREM 1.1. *Given the above definitions,* $|R| \leq \sigma^2$.

Proof: Start with $E[(X_1 + X_2)^2]$. We can write:

$$\begin{aligned} E[(X_1 + X_2)^2] &= E[X_1^2 + 2X_1X_2 + X_2^2] \\ &= E[X_1^2] + 2E[X_1X_2] + E[X_2^2] \\ &= \sigma^2 + 2R + \sigma^2 \\ &= 2\sigma^2 + 2R. \end{aligned}$$

Since $(X_1 + X_2)^2 \geq 0$ for all X_1 and X_2, it is true that $E[(X_1 + X_2)^2] \geq 0$. Thus, $2\sigma^2 + 2R \geq 0$, so $R \geq -\sigma^2$. Repeating the same procedure but starting with $E[(X_1 - X_2)^2]$, we have that $R \leq \sigma^2$, and the theorem follows. ■

Since $R = 0$ when X_1 and X_2 are independent, $R = \sigma^2$ (the maximum possible value) when they are completely correlated, and $R = -\sigma^2$ (the minimum possible value) when they are completely anticorrelated, R is a good candidate for a correlation measure. The magnitude of R indicates the degree of correlation between X_1 and X_2, while the sign indicates whether the variables are correlated or anticorrelated. Properties of this correlation measure when the variances are unequal, or when the means are nonzero, are considered in the exercises.

We apply this correlation measure to different time instants of the same random process, which we refer to as the *autocorrelation*. In particular, let $X(t)$ be a discrete-time random process defined on $t \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Then the autocorrelation between $X(t_1)$ and $X(t_2)$ is defined as

$$R(t_1, t_2) = E[X(t_1)X(t_2)]. \tag{1.25}$$

Note the similarity with (1.24), since $X(t)$ is merely a random variable for each time t. For the same reason, $R(t_1, t_2)$ has all the same properties as R.

1.2.3. Stationary random processes. A *stationary* discrete-time random process is a process for which the statistics do not change with time. Formally, a process is stationary if and only if

$$f_{X(t_1),X(t_2),\ldots,X(t_k)}(x_1, x_2, \ldots, x_k) = f_{X(t_1+\tau),X(t_2+\tau),\ldots,X(t_k+\tau)}(x_1, x_2, \ldots, x_k) \tag{1.26}$$

for all $k \in \{1, 2, \ldots\}$ and all $\tau \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. This does *not* imply that the process $X(t)$ is constant with respect to time, only that the statistical variation of the process is the same, regardless of when you examine the process. The process in Example 1.2 is stationary; intuitively, this is because we keep flipping the same unchanging coin, and recording the outcome in the same way at all t.

We now examine the effects of stationarity on the mean, variance, and autocorrelation of a discrete-time random process $X(t)$. The mean $\mu(t)$ is calculated as follows:

$$\begin{aligned} \mu(t) &= \int_x x f_{X(t)}(x) \mathrm{d}x \\ &= \int_x x f_{X(t+\tau)}(x) \mathrm{d}x \\ &= \mu(t+\tau), \end{aligned}$$

where the second line follows from the fact that $f_{X(t)} = f_{X(t+\tau)}$ for all $\tau \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Thus, $\mu(t) = \mu(t+\tau)$ for all τ, so $\mu(t)$ must be a constant with respect to t. Using a similar line of reasoning, we can show that $\sigma^2(t)$ is a constant with respect to t. Thus, for stationary random processes, we will write $\mu(t) = \mu$ and $\sigma^2(t) = \sigma^2$ for all t.

For the autocorrelation, we can write

$$\begin{aligned} R(t_1, t_2) &= E[X(t_1)X(t_2)] \\ &= \int_{x_1}\int_{x_2} x_1 x_2 f_{X(t_1),X(t_2)}(x_1, x_2) \mathrm{d}x_2 \mathrm{d}x_1 && (1.27) \\ &= \int_{x_1}\int_{x_2} x_1 x_2 f_{X(t_1+\tau),X(t_2+\tau)}(x_1, x_2) \mathrm{d}x_2 \mathrm{d}x_1. && (1.28) \end{aligned}$$

Let $\tau = \tau' - t_1$. Substituting back into (1.28), we have

$$\begin{aligned} R(t_1, t_2) &= \int_{x_1} \int_{x_2} x_1 x_2 f_{X(t_1+\tau'-t_1), X(t_2+\tau'-t_1)}(x_1, x_2) \mathrm{d}x_2 \mathrm{d}x_1 \\ &= \int_{x_1} \int_{x_2} x_1 x_2 f_{X(\tau'), X(t_2-t_1+\tau')}(x_1, x_2) \mathrm{d}x_2 \mathrm{d}x_1. \end{aligned} \tag{1.29}$$

However, in (1.29), since $X(t)$ is stationary, $f_{X(\tau'), X(t_2-t_1+\tau')}(x_1, x_2)$ does not change for any value of τ'. Thus, setting $\tau' = 0$, we can write

$$R(t_1, t_2) = \int_{x_1} \int_{x_2} x_1 x_2 f_{X(0), X(t_2-t_1)}(x_1, x_2) \mathrm{d}x_2 \mathrm{d}x_1,$$

which is not dependent on the exact values of t_1 or t_2, but only on the difference $t_2 - t_1$. As a result, we can redefine the autocorrelation function for stationary random processes as $R(t_2 - t_1)$; further, reusing τ to represent this difference, we will usually write $R(\tau)$, where

$$R(\tau) = E[X(t)X(t + \tau)]$$

for all t.

The properties that $\mu(t) = \mu$, $\sigma^2(t) = \sigma^2$, and $R(t_1, t_2) = R(t_2 - t_1)$ apply only to the first and second order statistics of the process $X(t)$. In order to verify whether a process is stationary, it is necessary to prove the condition (1.26) for every order of statistics. In general this is a difficult task. However, in some circumstances, only first and second order statistics are required. In this case, we define a *wide-sense stationary* (WSS) process as any process which satisfies the first and second order conditions of $\mu(t) = \mu$, $\sigma^2(t) = \sigma^2$, and $R(t_1, t_2) = R(t_2 - t_1)$. We have shown that all stationary processes are WSS, but it should seem clear that a WSS process is not necessarily stationary.

Throughout this book, we normally consider discrete-time random processes. In this case, it is important to remember that $t_1, t_2 \in \mathbb{Z}$,

1.2.4. Power spectral density. For a wide-sense stationary random process, the *power spectral density* (PSD) of that process is the Fourier transform of the autocorrelation function:

$$S_x(j\omega) = \mathcal{F}[R_x(\tau)] = \int_{\tau=-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau. \tag{1.30}$$

Properties of PSD:

(1) **Variance.**

$$\mathrm{Var}(x[k]) = R_x(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(j\omega)d\omega. \tag{1.31}$$

(2) **Positive and real.** $S_x(j\omega)$ is positive and real for all ω.

1.3. Linear time-invariant systems

1.3.1. Review of linear time-invariant systems. A linear time-invariant (LTI) system has the following two properties:

(1) **Linear.** If input $x_1(t)$ produces output $y_1(t)$, and input $x_2(t)$ produces output $y_2(t)$, then for any constants a and b, input $ax_1(t)+bx_2(t)$ produces output $ay_1(t) + by_2(t)$.

(2) **Time invariant.** If input $x(t)$ produces output $y(t)$, then for any τ, input $x(t + \tau)$ produces output $y(t + \tau)$.

An LTI system is completely characterized by its *impulse response* $h(t)$. That is, $h(t)$ is the system output if the system input is $\delta(t)$. Given $h(t)$ and an arbitrary input $x(t)$, the output $y(t)$ of an LTI system is given by

$$\begin{aligned} y(t) &= x(t) \star h(t) && (1.32)\\ &= \int_{\tau=-\infty}^{\infty} x(\tau)h(t-\tau)d\tau. && (1.33) \end{aligned}$$

Furthermore, the following relationship holds in the Fourier domain:

$$\mathcal{F}[y(t)] = \mathcal{F}[x(t)]\mathcal{F}[h(t)]. \tag{1.34}$$

Discrete time ... example ...

For further details, the reader is directed to [**4**].

1.3.2. LTI and random processes. Apply a linear filter with frequency-domain transfer function $H(j\omega)$ to a wide-sense stationary random process with PSD $S_x(j\omega)$. The output is a random process with PSD $S_w(j\omega)$, where

$$S_w(j\omega) = S_x(j\omega)|H(j\omega)|^2. \tag{1.35}$$

1.4. Problems

(1) For the random process in Example 1.3, show that $\mu(t) = t/2$, and $\sigma^2(t) = t/4$. Is this process stationary? Explain.

(2) Suppose X_1 and X_2 are zero-mean random variables with variances σ_1^2 and σ_2^2, respectively. For the correlation measure R defined in (1.24), show that
$$|R| \leq \frac{\sigma_1^2 + \sigma_2^2}{2}.$$

(3) Suppose X_1 and X_2 have the same *nonzero* mean μ, and the same variance σ^2. For the correlation measure R defined in (1.24), show that $|R| \leq \sigma^2 + \mu^2$.

(4) Give an example of a discrete-time random process for which $\mu(t) = \mu$ and $\sigma^2(t) = \sigma^2$ for all t, but there exist t_1 and t_2 such that $R(t_1, t_2) \neq R(t_2 - t_1)$.

(5) Calculate $\mu(t)$ and $R(t_1, t_2)$ for the continuous time random process given in Example 1.2. Is this process stationary? Explain.

(6) Let $X(t) = X \sin(2\pi t)$, where X is a random variable corresponding to the result of a single fair coin flip: $X = 1$ if the coin is heads, and $X = -1$ is the coin is tails. Does $X(t)$ satisfy the definition of a continuous-time random process? If so, calculate $f_{X(t)}(x)$; if not, explain why not.

1.5. Laboratory Exercise: Probability and Random Processes

In this laboratory exercise, you will investigate the properties of discrete-valued random variables and random processes.

1.5.1. Generating arbitrary random variables. Let x be a discrete-valued random variable, taking values on 1, 2, . . . , 6, with probability mass function $p(x)$.

- MATLAB provides a routine, `rand`, which generates uniformly distributed random variables on the range from 0 to 1. Given $p(x)$, propose a way to generate instances of x, with probabilities $p(x)$, from `rand`.
- Write a MATLAB function, called `xrand`, implementing the method you describe. The routine takes a 1×6 vector, where the first element of the vector is $p(1)$, the second is $p(2)$, and so on. The routine returns a value on 1, 2, . . . , 6 at random according to the probabilities $p(x)$.

Discussion of empirical distributions.

Given a distribution, write a function to calculate the mean and variance, both empirically and theoretically.

Consider the following Gaussian random process: ... Plot the autocorrelation, both empirically and