Operations on a Single Random Variable

In our study of random variables, we use the probability density function (PDF), the cumulative distribution function (CDF), or the probability mass function (PMF) to provide a complete statistical description of the random variable. From these functions, we could, in theory, determine just about anything we might want to know about the random variable. In many cases, it is of interest to distill this information down to a few parameters which describe some of the important features of the random variable. For example, we saw in Chapter 3 that the Gaussian random variable is described by two parameters, which were referred to as the mean and variance. In this chapter, we look at these parameters as well as several others that describe various characteristics of random variables. We see that these parameters can be viewed as the results of performing various operations on a random variable.

4.1 Expected Value of a Random Variable

To begin, we introduce the idea of an average or expected value of a random variable. This is perhaps the single most important characteristic of a random variable and also is a concept that is very familiar to most students. After taking a test, one of the most common questions a student will ask after they see their grade is "What was the average?" On the other hand, how often does a student ask "What was the PDF of the exam scores?" While the answer to the second question would provide the student with more information about how the class performed, the student may not want all that information. Just knowing the average may be sufficient to tell the student how he/she performed relative to the rest of the class.

Definition 4.1: The *expected value* of a random variable X which has a PDF, $f_X(x)$, is

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx . \tag{4.1}$$

The terms *average*, *mean*, *expectation*, and *first moment* are all alternative names for the concept of expected value and will be used interchangeably throughout the text. Furthermore, an overbar is often used to denote expected value so that the symbol $\bar{X}$ is to be interpreted as meaning the same thing as $E[X]$. Another commonly used notation is to write $\mu_X = E[X]$.

For discrete random variables, the PDF can be written in terms of the PMF,

$$f_X(x) = \sum_k P_X(x_k)\delta(x-x_k). \tag{4.2}$$

In that case, using the properties of delta functions, the definition of expected values for discrete random variables reduces to

$$E[X] = \sum_k x_k P_X(x_k). \tag{4.3}$$

Hence, the expected value of a discrete random variable is simply a weighted average of the values that the random variable can take on, weighted by the probability mass of each value. Naturally, the expected value of a random variable only exists if the integral in Equation (4.1) or the series in Equation (4.3) converges. One can dream up many random variables for which the integral or series does not converge and hence their expected values don't exist (or less formally, their expected value is infinite). To gain some physical insight into this concept of expected value, we may think of $f_X(x)$ as a mass distribution of an object along the x axis, then Equation (4.1) calculates the centroid or center of gravity of the mass.

■ Example 4.1:

Consider a random variable that has an exponential PDF given by $f_X(x) = \frac{1}{b}\exp\left(-\frac{x}{b}\right)u(x)$.
Its expected value is calculated as follows:

$$E[X] = \int_0^\infty \frac{x}{b}\exp\left(-\frac{x}{b}\right)dx = b\int_0^\infty y\exp(-y)dy = b.$$

The last equality in the series is obtained by using integration by parts once. It is seen from this example that the parameter b which appears in this exponential distribution is in fact the mean (or expected value) of the random variable.

■

■ Example 4.2:

Next, consider a Poisson random variable whose PMF is given by
$P_X(k) = \alpha^k e^{-\alpha}/k!$, $k = 0, 1, 2, \ldots$. Its expected value is found in a similar manner.

$$E[X] = \sum_{k=0}^\infty k\frac{\alpha^k e^{-\alpha}}{k!} = e^{-\alpha}\sum_{k=1}^\infty \frac{\alpha^k}{(k-1)!} = \alpha e^{-\alpha}\sum_{k=1}^\infty \frac{\alpha^{k-1}}{(k-1)!} = \alpha e^{-\alpha}\sum_{m=0}^\infty \frac{\alpha^m}{m!} = \alpha e^{-\alpha}e^\alpha = \alpha.$$

Once again, we see that the parameter α in the Poisson distribution is equal to the mean.

■

■ Example 4.3:

In the last two examples, we saw random variables whose PDF or PMF was described by a single parameter which in both cases turned out to be the mean. We work one more example here to show that this does not always have to be the case. Consider a Rayleigh random variable with PDF

$$f_X(x) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)u(x).$$

The mean is calculated as follows:

$$E[X] = \int_0^\infty \frac{x^2}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)dx = \sqrt{2}\sigma\int_0^\infty y^{1/2}\exp(-y)dy = \sqrt{2}\sigma\Gamma(3/2) = \sqrt{\frac{\pi}{2}}\sigma.$$

The last equality is obtained using the fact that $\Gamma(3/2) = \sqrt{\pi}/2$. Alternatively (for those students not familiar with the properties of Γ functions), one could obtain this result using integration by parts once on the original integral (setting $u = x$ and $dv = (x/\sigma^2)\exp(-x^2/(2\sigma^2))$). In this case, neither the parameter σ nor σ^2 is equal to the expected value of the random variable. However, since the mean is proportional to σ, we could, if we wanted to, rewrite the Rayleigh PDF in terms of its expected value, μ_X, as follows:

$$f_X(x) = \frac{\pi x}{2\mu_X^2}\exp\left(-\frac{\pi x^2}{4\mu_X^2}\right)u(x) .$$

■

4.2 Expected Values of Functions of Random Variables

The concept of expectation can be applied to the functions of random variables as well as to the random variable itself. This will allow us to define many other parameters that describe various aspects of a random variable.

Definition 4.2: Given a random variable X with PDF $f_X(x)$, the expected value of a function, $g(X)$, of that random variable is given by

$$E[g(X)] = \int g(x)f_X(x)dx. \tag{4.4}$$

For a discrete random variable, this definition reduces to

$$E[g(X)] = \sum_k g(x_k)P_X(x_k). \tag{4.5}$$

To start with, we demonstrate one extremely useful property of expectations in the following theorem.

Theorem 4.1: For any constants a and b,

$$E[aX + b] = aE[X] + b. \qquad (4.6)$$

Furthermore, for any function $g(x)$ which can be written as a sum of several other functions (i.e., $g(x) = g_1(x) + g_2(x) + \ldots + g_N(x)$),

$$E\left[\sum_{k=1}^{N} g_k(X)\right] = \sum_{k=1}^{N} E[g_k(X)] \qquad (4.7)$$

In other words, expectation is a linear operation and the expectation operator can be exchanged (in order) with any other linear operation.

Proof: The proof follows directly from the linearity of the integration operator.

$$E[aX + b] = \int_{-\infty}^{\infty} (ax + b) f_X(x)\,\mathrm{d}x$$

$$= a\int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x + b = aE[X] + b. \qquad (4.8)$$

The second part of the theorem is proved in an identical manner:

$$E\left[\sum_{k=1}^{N} g_k(X)\right] = \int_{-\infty}^{\infty}\left[\sum_{k=1}^{N} g_k(x)\right] f_X(x)\,\mathrm{d}x = \sum_{k=1}^{N} \int_{-\infty}^{\infty} g_k(x) f_X(x)\,\mathrm{d}x$$

$$= \sum_{k=1}^{N} E[g_k(X)] \ . \ \square \qquad (4.9)$$

Different functional forms of $g(X)$ lead to various different parameters which describe the random variable and are known by special names. A few of the more common ones are listed in Table 4.1. In the following sections, selected parameters will be studied in more detail.

Table 4.1: Expected values of various functions of random variables

Name	Function of X	Expected Value, Notation
Mean, average, expected value, expectation, first moment	$g(x) = x$	$\mu_X = \bar{X} = E[X]$
nth moment	$g(x) = x^n$	$\overline{X^n} = E[X^n]$
nth central moment	$g(x) = (x - \mu_X)^n$	$\overline{(x - \mu_X)^n} = E[(X - \mu_X)^n]$
Variance	$g(x) = (x - \mu_X)^2$	$\sigma_X^2 = E[(X - \mu_X)^2]$
Coefficient of skewness	$g(x) = \left(\dfrac{x - \mu_X}{\sigma_X}\right)^3$	$c_s = E\left[\left(\dfrac{X - \mu_X}{\sigma_X}\right)^3\right]$
Coefficient of kurtosis	$g(x) = \left(\dfrac{x - \mu_X}{\sigma_X}\right)^4$	$c_k = E\left[\left(\dfrac{X - \mu_X}{\sigma_X}\right)^4\right]$
Characteristic function	$g(x) = e^{j\omega x}$	$\Phi_X(\omega) = E[e^{j\omega X}]$
Moment-generating function	$g(x) = e^{sx}$	$M_X(s) = E[e^{sX}]$
Probability-generating function	$g(x) = z^x$	$H_X(z) = E[z^X]$

4.3 Moments

Definition 4.3: The nth *moment* of a random variable X is defined as

$$E[X^n] = \int x^n f_X(x)\mathrm{d}x. \tag{4.10}$$

For a discrete random variable, this definition reduces to

$$E[X^n] = \sum_k x_k^n P_X(x_k). \tag{4.11}$$

The zeroth moment is simply the area under the PDF and must be one for any random variable. The most commonly used moments are the first and second moments. The first moment is what we previously referred to as the mean, while the second moment is the mean-squared value. For some random variables, the second moment might be a more meaningful characterization than the first. For example, suppose X is a sample of a noise waveform. We might expect that the distribution of the noise is symmetric about zero (i.e., just as likely to be positive as negative)

and hence the first moment will be zero. So if we are told that X has a zero mean, all this says is merely that the noise does not have a bias. However, the second moment of the random noise sample is in some sense a measure of the strength of the noise. In fact, in Chapter 10, we will associate the second moment of a noise process with the power in the process. Thus, specifying the second moment can give us some useful physical insight into the noise process.

■ Example 4.4:

Consider a discrete random variable that has a binomial distribution. Its probability mass function is given by

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \ k = 0, 1, 2, ..., n.$$

The first moment is calculated as follows:

$$E[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{kn!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np \sum_{m=0}^{n-1} \binom{n-1}{m} p^m (1-p)^{n-1-m}.$$

In this last expression, the summand is a valid PMF (i.e., that of a binomial random variable with parameters p and $n-1$) and therfore must sum to unity. Thus, $E[X] = np$. To calculate the second moment, we employ a helpful little trick. Note that we can write $k^2 = k(k-1) + k$. Then

$$E[X^2] = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}.$$

The second sum is simply the first moment, which has already been calculated. The first sum is evaluated in a manner similar to that used to calculate the mean.

$$\sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k}$$

$$= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} = n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} (1-p)^{n-k}$$

$$= n(n-1)p^2 \sum_{m=0}^{n-2} \binom{n-2}{m} p^m (1-p)^{n-2-m} = n(n-1)p^2$$

Putting these two results together gives

$$E[X^2] = n(n-1)p^2 + np = n^2p^2 + np(1-p).$$

■

■ Example 4.5:

Consider a random variable with a uniform PDF given as

$$f_X(x) = \begin{cases} 1/a, & 0 \le x \le a, \\ 0, & \text{otherwise.} \end{cases}$$

The mean is given by

$$E[X] = \int_0^a \frac{x}{a} dx = \frac{x^2}{2a}\Big|_0^a = \frac{a}{2},$$

while the second moment is

$$E[X^2] = \int_0^a \frac{x^2}{a} dx = \frac{x^3}{3a}\Big|_0^a = \frac{a^2}{3}.$$

In fact, it is not hard to see that in general, the *n*th moment of this uniform random variable is given by

$$E[X^n] = \int_0^a \frac{x^n}{a} dx = \frac{x^{n+1}}{(n+1)a}\Big|_0^a = \frac{a^n}{n+1}.$$

■

4.4 Central Moments

Consider a random variable Y which could be expressed as the sum, $Y = a + X$ of a deterministic (i.e., not random) part a and a random part X. Furthermore, suppose that the random part tends to be very small compared to the fixed part. That is, the random variable Y tends to take small fluctuations about a constant value, a. Such might be the case in a situation where there is a fixed signal corrupted by noise. In this case, we might write $Y^n = (a + X)^n \approx a^n$. As such, the *n*th moment of Y would be dominated by the fixed part. That is, it is difficult to characterize the randomness in Y by looking at the moments. To overcome this, we can use the concept of central moments.

Definition 4.4: The *n*th *central moment* of a random variable X is defined as

$$E[(X - \mu_X)^n] = \int (x - \mu_X)^n f_X(x) dx. \tag{4.12}$$

In the above equation, μ_X is the mean (first moment) of the random variable. For discrete random variables, this definition reduces to

$$E[(X - \mu_X)^n] = \sum_k (x_k - \mu_X)^k P_X(x_k). \tag{4.13}$$

With central moments, the mean is subtracted from the variable before the moment is taken in order to remove the bias in the higher moments due to the mean. Note that, like regular moments, the zeroth central moment is $E[(X-\mu_X)^0] = E[1] = 1$. Furthermore, the first central moment is $E[X-\mu_X] = E[X] - \mu_X = \mu_X - \mu_X = 0$. Hence, the lowest central moment of any real interest is the second central moment. This central moment is given a special name, the *variance*, and we quite often use the notation σ_X^2 to represent the variance of the random variable X. Note that

$$\sigma_X^2 = E[(X-\mu_X)^2] = E[X^2 - 2\mu_X X + \mu_X^2] = E[X^2] - 2\mu_X E[X] + \mu_X^2$$

$$= E[X^2] - \mu_X^2. \tag{4.14}$$

In many cases, the best way to calculate the variance of a random variable is to calculate the first two moments and then form the second moment minus the first moment squared.

■ Example 4.6:

For the binomial random variable in Example 4.4, recall that the mean was $E[X] = np$ and the second moment was $E[X^2] = n^2 p^2 + np(1-p)$. Therefore, the variance is given by $\sigma_X^2 = np(1-p)$. Similarly, for the uniform random variable in Example 4.5, $E[X] = a/2$, $E[X^2] = a^2/3$, and therfore $\sigma_X^2 = a^2/3 - a^2/4 = a^2/12$. Note that if the moments have not previously been calculated, it may be just as easy to compute the variance directly. In the case of the uniform random variable, once the mean has been calculated, the variance can be found as

$$\sigma_X^2 = \int_0^a (x-a/2)^2 \frac{1}{a} dx = \int_{-\frac{a}{2}}^{\frac{a}{2}} \frac{x^2}{a} dx = \frac{x^3}{3a}\Big|_{-a/2}^{a/2} = \frac{a^2}{12}.$$

■

Another common quantity related to the second central moment of a random variable is the *standard deviation,* which is defined as the square root of the variance,

$$\sigma_X = \sqrt{E[(X-\mu_X)^2]}. \tag{4.15}$$

Both the variance and the standard deviation serve as a measure of the width of the PDF of a random variable. Some of the higher-order central moments also have special names, although they are much less frequently used. The third central moment is known as the *skewness* and is a measure of the symmetry of the PDF about the mean. The fourth central moment is called the *kurtosis* and is a measure of the peakedness of a random variable near the mean. Note that not all random variables have finite moments and/or central moments. We give an example of this later for the Cauchy random variable. Some quantities related to these higher-order central moments are given in the following definition.

Definition 4.5: The *coefficient of skewness* is

$$c_\mathrm{s} = \frac{E[(X - \mu_X)^3]}{\sigma_X^3}. \tag{4.16}$$

This is a dimensionless quantity that is positive if the random variable has a PDF skewed to the right and negative if skewed to the left. The *coefficient of kurtosis* is also dimensionless and is given as

$$c_\mathrm{k} = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}. \tag{4.17}$$

The more the PDF is concentrated near its mean, the larger the coefficient of kurtosis. In other words, a random variable with a large coefficient of kurtosis will have a large peak near the mean.

■ Example 4.7:

An exponential random variable has a PDF given by

$$f_X(x) = b\exp(-bx)u(x).$$

The mean value of this random variable is $\mu_X = 1/b$. The nth central moment is given by

$$E[(X - \mu_X)^n] = \int_0^\infty (x - 1/b)^n b\exp(-bx)\mathrm{d}x$$

$$= \frac{b}{e}\int_{-1/b}^\infty y^n \exp(-by)\mathrm{d}y$$

$$= \frac{1}{b^n}\sum_{m=0}^{n} \frac{n!}{m!}(-1)^m.$$

In the preceding expression, it is understood that $0! = 1$. As expected, it is easily verified from the above expression that the 0th central moment is 1 and the first central moment is 0. Beyond these, the second central moment is $\sigma_X^2 = 1/b^2$, the third central moment is $E[(X - 1/b)^3] = -2/b^3$, and the fourth central moment is $E[(X - 1/b)^4] = 9/b^4$. The coefficients of skewness and kurtosis are given by

$$c_\mathrm{s} = \frac{E[(X - \mu_X)^3]}{\sigma_X^3} = \frac{-2/b^3}{1/b^3} = -2,$$

$$c_\mathrm{k} = \frac{E[(X - \mu_X)^4]}{\sigma_X^4} = \frac{9/b^4}{1/b^4} = 9.$$

The fact that the coefficient of skewness is negative shows that the exponential PDF is skewed to the left of its mean.

■

■ Example 4.8:

Next consider a Laplace (two-sided exponential) random variable with a PDF given by

$$f_X(x) = \frac{b}{2}\exp(-b|x|).$$

Since this PDF is symmetric about zero, its mean is zero and therefore in this case the central moments are simply the moments,

$$E[(X - \mu_X)^n] = E[X^n] = \int_{-\infty}^{\infty} \frac{bx^n}{2}\exp(-b|x|)dx.$$

Since the two-sided exponential is an even function and x^n is an odd function for any odd n, the integrand is then an odd function for any odd n. The integral of any odd function over an interval symmetric about zero is equal to zero, and hence all odd moments of the Laplace random variable are zero. The even moments can be calculated individually:

$$\sigma_X^2 = E[X^2] = \int_{-\infty}^{\infty} \frac{bx^2}{2}\exp(-b|x|)dx = \int_0^{\infty} bx^2\exp(-bx)dx = \frac{2}{b^2},$$

$$E[(X - \mu_X)^4] = E[X^4] = \int_{-\infty}^{\infty} \frac{bx^4}{2}\exp(-b|x|)dx = \int_0^{\infty} bx^4\exp(-bx)dx = \frac{24}{b^4}.$$

The coefficient of skewness is zero (since the third central moment is zero) and the coefficient of kurtosis is

$$c_k = \frac{E[(X - \mu_X)^4]}{\sigma_X^4} = \frac{24/b^4}{4/b^4} = 6.$$

Note that the Laplace distribution has a sharp peak near its mean as evidenced by a large coefficient of kurtois. The fact that the coefficient of skewness is zero is consistent with the fact that the distribution is symmetric about its mean.

■

■ Example 4.9:
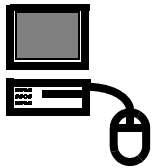
It is often the case that the PDF of random variables of practical interest may be too complicated to allow us to compute various moments and other important parameters of the distribution in an analytic fashion. In those cases, we can use a computer to calculate the needed quantities numerically. Take for example a Rician random variable whose PDF is given by

$$f_X(x) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2 + a^2}{2\sigma^2}\right)I_0\left(\frac{ax}{\sigma^2}\right)u(x).$$

Suppose we wanted to know the mean of this random variable. This requires us to evaluate

$$\mu_X = \int_0^{\infty} \frac{x^2}{\sigma^2}\exp\left(-\frac{x^2 + a^2}{2\sigma^2}\right)I_0\left(\frac{ax}{\sigma^2}\right)dx.$$

Note that the parameter σ^2 which shows up in the Rician PDF is *not* the variance of the Rician random variable. While analytical evaluation of this integral looks formidable, given numerical values for a and σ^2, this integral can be evaluated (at least approximately) using standard numerical integration techniques. In order to use the numerical integration routines built into MATLAB, we must first write a function which evaluates the integrand. For evaluating the mean of the Rician PDF, this can be accomplished as follows (see Appendix D, Equation (D.52) for an analytic expression for the mean):

```
function pdf=Rician_pdf(x,a,sigma)
% Evaluate the integrand needed for calculating the mean of a
% Rician random variable with parameters a and sigma.
pdf=(x./sigma).^2.*exp(-(x.^2+a^2)/(2*sigma^2));
pdf=pdf.*besseli(0,a*x/sigma^2);
```

Once this function is defined, the MATLAB function `quad8` can be called upon to perform the numerical integration. For example, if $a = 2$ and $\sigma = 3$, the mean could be calculated as follows:

```
a=2; sigma=3;                        % set parameters
limit1=0; limit2=20;                 % set limits of integration.
mean=quad8('Rician_pdf',limit1,limit2,[],[],a,sigma);
```

Executing this code produced an answer of $\mu_X = 4.1665$. Note that in order to calculate the mean, the upper limit of the integral should be infinite. However using `limit2=Inf` in the above code would have led MATLAB to produce a result of `NaN` ("not a number"). Instead, we must use an upper limit sufficiently large that for the purposes of evaluating the integral it is essentially infinite. This can be done by observing the integrand and seeing at what point the integrand dies off. The reader is encouraged to execute the code in this example using different values of the upper integration limit to see how large the upper limit must be to produce accurate results.

∎

4.5 Conditional Expected Values

Another important concept involving expectation is that of conditional expected value. As specified in Definition 4.6, the conditional expected value of a random variable is a weighted average of the values the random variable can take on, weighted by the conditional PDF of the random variable.

Definition 4.6: The expected value of a random variable X, conditioned on some event A is

$$E[X|A] = \int x f_{X|A}(x)\mathrm{d}x . \qquad (4.18)$$

For a discrete random variable, this definition reduces to

$$E[X|A] = \sum_k x_k P_{X|A}(x_k).$$

(4.19)

Similarly, the conditional expectation of a function, $g(\cdot)$, of a random variable, conditioned on the event A is

$$E[g(X)|A] = \int g(x) f_{X|A}(x) dx \text{ or } E[g(X)|A] = \sum_k g(x_k) P_{X|A}(x_k),$$

(4.20)

depending on whether the random variable is continuous or discrete.

Conditional expected values are computed in the same manner as regular expected values with the PDF or PMF replaced by a conditional PDF or conditional PMF.

■ Example 4.10:

Consider a Gaussian random variable of the form

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Suppose the event A is the event that the random variable X is positive, $A = \{X > 0\}$. Then

$$f_{X|A}(x) = \frac{f_X(x)}{\Pr(X > 0)} u(x) = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right) u(x).$$

The conditional expected value of X given that $X > 0$ is then

$$E[X|X > 0] = \int x f_{X|X>0}(x) dx = \sqrt{\frac{2}{\pi}} \int_0^\infty x \exp\left(-\frac{x^2}{2}\right) dx = -\sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right)\Big|_0^\infty = \sqrt{\frac{2}{\pi}}.$$

■

4.6 Transformations of Random Variables

Consider a random variable X with a PDF and CDF given by $f_X(x)$ and $F_X(x)$, respectively. Define a new random variable Y such that $Y = g(X)$ for some function $g(\cdot)$. What is the PDF, $f_Y(y)$, (or CDF) of the new random variable? This problem is often encountered in the study of systems where the PDF for the input random variable X is known and the PDF for the output random variable Y needs to be determined. In such a case, we say that the input random variable has undergone a transformation.

4.6.1 Monotonically Increasing Functions

To begin our exploration of transformations of random variables, let us assume that the function is continuous, one-to-one, and monotonically increasing. A typical function of this
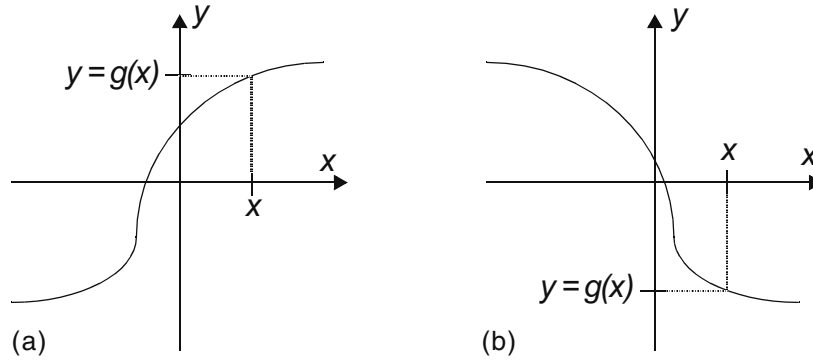
Figure 4.1

(a) A monotonic increasing function and (b) a monotonic decreasing function.

form is illustrated in Figure 4.1a. This assumption will be lifted later when more general functions are considered, but for now this simpler case applies. Under these assumptions, the inverse function, $X = g^{-1}(Y)$, exists and is well behaved. In order to obtain the PDF of Y, we first calculate the CDF. Recall that $F_Y(y) = \Pr(Y \le y)$. Since there is a one-to-one relationship between values of Y and their corresponding values of X, this CDF can be written in terms of X according to

$$F_Y(y) = \Pr(g(X) \le y) = \Pr(X \le g^{-1}(y)) = F_X(g^{-1}(y)). \tag{4.21}$$

Note, this can also be written as

$$F_X(x) = F_Y(g(x)). \tag{4.22}$$

Differentiating Equation (4.21) with respect to y produces

$$f_Y(y) = f_X(g^{-1}(y))\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} = f_X(x)\frac{\mathrm{d}x}{\mathrm{d}y}\bigg|_{x = g^{-1}(y)}, \tag{4.23}$$

while differentiating Equation (4.22) with respect to x gives

$$f_X(x) = f_Y(g(x))\frac{\mathrm{d}y}{\mathrm{d}x} \Rightarrow f_Y(y) = \frac{f_X(x)}{\dfrac{\mathrm{d}y}{\mathrm{d}x}}\bigg|_{x = g^{-1}(y)}. \tag{4.24}$$

Either Equation (4.23) or (4.24) can be used (whichever is more convenient) to compute the PDF of the new random variable.

■ **Example 4.11:**

Suppose X is a Gaussian random variable with mean, μ, and variance, σ^2. A new random variable is formed according to $Y = aX + b$, where $a > 0$ (so that the transformation is monotonically increasing). Since $dy/dx = a$, then applying Equation (4.24) produces

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

Furthermore, plugging in the Gaussian PDF of X results in

$$f_Y(y) = \frac{1}{a\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi(a\sigma)^2}} \exp\left(-\frac{(y - (b + a\mu))^2}{2(a\sigma)^2}\right).$$

Note that the PDF of Y still has a Gaussian form. In this example, the transformation did not change the form of the PDF; it merely changed the mean and variance. ■

■ **Example 4.12:**

Let X be an exponential random variable with $f_X(x) = 2e^{-2x}u(x)$ and let the transformation be $Y = X^3$. Then $dy/dx = 3x^2$ and

$$f_Y(y) = \left.\frac{f_X(x)}{3x^2}\right|_{x = \sqrt[3]{y}} = \frac{2}{3}y^{-2/3}\exp(-2y^{1/3})u(y).$$

■

■ **Example 4.13:**

Suppose a phase angle Θ is uniformly distributed over $(-\pi/2, \pi/2)$ and the transformation is $Y = \sin(\Theta)$. Note that in general $y = \sin(\theta)$ is not a monotonic transformation, but under the restriction $-\pi/2 < \theta < \pi/2$, this transformation is indeed monotonically increasing. Also note that with this transformation the resulting random variable, Y, must take on values in the range $(-1, 1)$. Therefore, whatever PDF is obtained for Y, it must be understood that the PDF is zero outside $(-1, 1)$. Applying Equation (4.24) gives

$$f_Y(y) = \left.\frac{f_\Theta(\theta)}{\cos(\theta)}\right|_{\theta = \sin^{-1}(y)} = \frac{1}{\pi\cos(\sin^{-1}(y))} = \frac{1}{\pi\sqrt{1-y^2}}, \quad -1 < y < 1.$$

This is known as an *arcsine distribution*.

■

4.6.2 Monotonically Decreasing Functions

If the transformation is monotonically decreasing rather than increasing, a simple modification to the previous derivations can lead to a similar result. First, note that for monotonic decreasing functions, the event $\{Y \le y\}$ is equivalent to the event $X \ge g^{-1}(y)$, giving us

$$F_Y(y) = \Pr(Y \le y) = \Pr(X \ge g^{-1}(y)) = 1 - F_X(g^{-1}(y)). \tag{4.25}$$

Differentiating with respect to y gives

$$f_Y(y) = -f_X(x)\frac{dx}{dy}\bigg|_{x = g^{-1}(y)}. \tag{4.26}$$

Similarly, writing $F_Y(g(x)) = 1 - F_X(x)$ and differentiating with respect to x results in

$$f_Y(y) = -\frac{f_X(x)}{\dfrac{dy}{dx}}\bigg|_{x = g^{-1}(y)}. \tag{4.27}$$

Equations (4.23), (4.24), (4.26), and (4.27) can be consolidated into the following compact form:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|\bigg|_{x = g^{-1}(y)} = \frac{f_X(x)}{\left|\dfrac{dy}{dx}\right|}\bigg|_{x = g^{-1}(y)}, \tag{4.28}$$

where now the sign differences have been accounted for by the absolute value operation. This equation is valid for any monotonic function, either monotonic increasing or monotonic decreasing.

4.6.3 Nonmonotonic Functions

Finally, we consider a general function which is not necessarily monotonic. Figure 4.2 illustrates one such example. In this case, we cannot associate the event $\{Y \le y\}$ with events
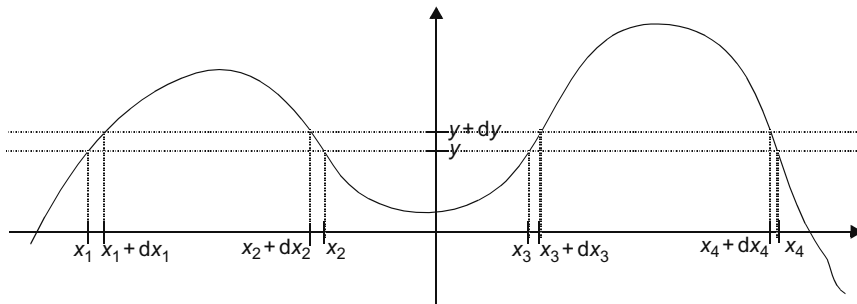
Figure 4.2
A nonmonotonic function; the inverse function may have multiple roots.

of the form $\{X \le g^{-1}(y)\}$ or $\{X \ge g^{-1}(y)\}$ because the transformation is not monotonic. To avoid this problem, we calculate the PDF of Y directly, rather than first finding the CDF. Consider an event of the form $\{y \le Y < y + dy\}$ for an infinitesimal dy. The probability of this event is $\Pr(y \le Y < y + dy) = f_Y(y)dy$. In order to relate the PDF of Y to the PDF of X, we relate the event $\{y \le Y < y + dy\}$ to events involving the random variable X. Because the transformation is not monotonic, there may be several values of x which map to the same value of y. These are the roots of the equation $x = g^{-1}(y)$. Let us refer to these roots as $x_1, x_2, \ldots, x_N$. Furthermore, let X^+ be the subset of these roots at which the function $g(x)$ has a positive slope and similarly let X^- be the remaining roots for which the slope of the function is negative. Then

$$\{y \le Y < y + dy\} = \left[\bigcup_{i:x_i \in X^+} \{x_i \le X < x_i + dx_i\} \right] \cup \left[\bigcup_{i:x_i \in X^-} \{x_i + dx_i < X \le x_i\} \right]. \quad (4.29)$$

Since each of the events on the right-hand side is mutually exclusive, the probability of the union is simply the sum of the probabilities so that

$$f_Y(y)dy = \sum_{x_i \in X^+} \Pr(x_i \le X < x_i + dx_i) + \sum_{x_i \in X^-} \Pr(x_i + dx_i < X \le x_i)$$

$$= \sum_{x_i \in X^+} f_X(x_i)dx_i + \sum_{x_i \in X^-} f_X(x_i)(-dx_i) \quad (4.30)$$

Again, invoking absolute value signs to circumvent the need to have two separate sums and dividing by dy, the following result is obtained:

$$f_Y(y) = \sum_{x_i} f_X(x) \left| \frac{dx}{dy} \right| \bigg|_{x_i = g^{-1}(y)}. \quad (4.31)$$

When it is more convenient, the equivalent expression

$$f_Y(y) = \sum_{x_i} \frac{f_X(x)}{\left| \frac{dy}{dx} \right|} \bigg|_{x_i = g^{-1}(y)}, \quad (4.32)$$

can also be used. The following theorem summarizes the general formula for transformations of random variables.

Theorem 4.2: Given a random variable X with known PDF, $f_X(x)$, and a transformation $Y = g(X)$. The PDF of Y is

$$f_Y(y) = \sum_{x_i} f_X(x)\left|\frac{dx}{dy}\right|\Bigg|_{x_i = g^{-1}(y)} = \sum_{x_i} \frac{f_X(x)}{\left|\frac{dy}{dx}\right|}\Bigg|_{x_i = g^{-1}(y)}, \tag{4.33}$$

where the x_i are the roots of the equation $y = g(x)$. The proof precedes the theorem.

■ Example 4.14:

Suppose X is a Gaussian random variable with zero mean and variance σ^2 together with a quadratic transformation, $Y = X^2$. For any positive value of y, $y = x^2$ has two real roots, namely $x = \pm\sqrt{y}$ (for negative values of y there are no real roots). Application of Equation (4.33) gives

$$f_Y(y) = \left[\frac{f_X(+\sqrt{y})}{2|+\sqrt{y}|} + \frac{f_X(-\sqrt{y})}{2|-\sqrt{y}|}\right]u(y) = \frac{f_X(+\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}u(y).$$

For a zero-mean Gaussian PDF, $f_X(x)$ is an even function so that $f_X(+\sqrt{y}) = f_X(-\sqrt{y})$. Therefore,

$$f_Y(y) = \frac{1}{\sqrt{y}}f_X(\sqrt{y})u(y) = \sqrt{\frac{1}{2\pi y\sigma^2}}\exp\left(-\frac{y}{2\sigma^2}\right)u(y).$$

Hence, Y is a Gamma random variable.

■

■ Example 4.15:

Suppose the same Gaussian random variable from the previous example is passed through a half-wave rectifier which is described by the input-output relationship

$$y = g(x) = \begin{cases} x, & x \geq 0, \\ 0, & x \leq 0. \end{cases}$$

For $x > 0$, $dy/dx = 1$ so that $f_Y(y) = f_X(y)$. However, when $x < 0$, $dy/dx = 0$, which will create a problem if we try to insert this directly into Equation (4.33). To treat this case, we note that the event $X < 0$ is equivalent to the event $Y = 0$; therefore, $\Pr(Y=0) = \Pr(X<0)$. Since the input Gaussian PDF is symmetric about zero, $\Pr(X<0) = 1/2$. Basically, the random variable Y is a mixed random variable. It has a continuous part over the region $y > 0$ and a discrete part at $y = 0$. Using a delta function, we can write the PDF of Y as

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{y^2}{2\sigma^2}\right)u(y) + \frac{1}{2}\delta(y).$$

■

Example 4.15 illustrates how to deal with transformations that are flat over some interval of nonzero length. In general, suppose the transformation $y = g(x)$ is such that $g(x) = y_0$ for any x in the interval $x_1 \le x \le x_2$. Then the PDF of Y will include a discrete component (*a delta function*) of height $\Pr(Y=y_0) = \Pr(x_1 \le x \le x_2)$ at the point $y = y_0$. One often encounters transformations that have several different flat regions. One such "staircase" function is shown in Figure 4.3. Hence, a random variable X that may be continuous will be converted into a discrete random variable. The classical example of this is analog-to-digital conversion of signals. Suppose the transformation is of a general staircase form,

$$y = \begin{cases} y_0, & x < x_1, \\ y_i, & x_i \le x < x_{i+1}, \quad i = 1, 2, ..., N-1, \\ y_N, & x \ge x_N. \end{cases} \tag{4.34}$$

Then Y will be a discrete random variable whose PMF is

$$P(Y=y_i) = \begin{cases} \Pr(X < x_1), & i = 0, \\ \Pr(x_i \le x < x_{i+1}), & i = 1, 2, ..., N-1, \\ \Pr(X \ge x_N), & i = N. \end{cases} \tag{4.35}$$
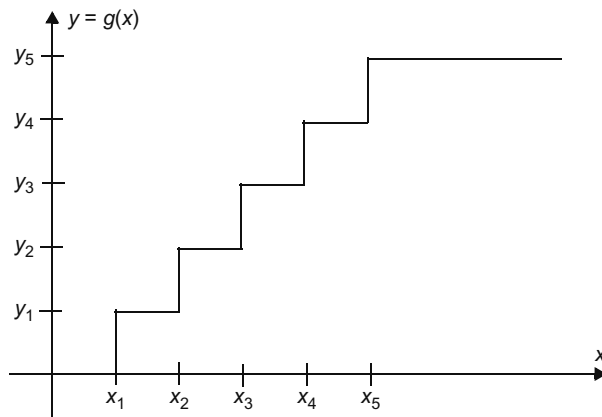
Figure 4.3
A staircase (quantizer) transformation: a continuous random variable will be converted into a discrete random variable.

■ **Example 4.16:**

Suppose X is an exponential random variable with a PDF $f_X(x) = \exp(-x)u(x)$ and we form a new random variable Y by rounding X down to the nearest integer. That is,
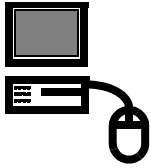
$$Y = g(X) = \text{floor}(X) = k, \ k \leq X < k+1.$$

Then, the PMF of Y is

$$P(Y=k) = \Pr(k \leq X < k+1) = \int_k^{k+1} e^{-x}dx = e^{-k} - e^{-(k+1)} = e^{-k}(1-1/e), \ k = 0, 1, 2, \dots.$$

Hence, quantization of an exponential random variable produces a geometric random variable.

■

■ **Example 4.17:**

Let X be a random variable uniformly distributed over $(-a/2, a/2)$. Accordingly, its PDF is of the form

$$f_X(x) = \frac{1}{a}\left(u\left(x+\frac{a}{2}\right) - u\left(x-\frac{a}{2}\right)\right).$$

A new random variable is to be formed according to the square law transformation $Y = X^2$. Applying the theory developed in this section you should be able to demonstrate that the new random variable has a PDF given by

$$f_Y(y) = \frac{1}{a\sqrt{y}}(u(y) - u(y - a^2/4)).$$

Using MATLAB, we create a large number of samples of the uniform random variable, pass these samples through the square law transformation, and then construct a histogram of the resulting probability densities. The MATLAB code to do so follows and the results of running this code are shown in Figure 4.4.

```
clear
N=10000;
a=5; ymax=a^2/4;                    % set parameters.
x=a*(rand(1,N)-0.5);                % generate uniform RVs
y=x.^2;                             % square law transformation
bw=0.25;                            % bin width
bins=[bw/2:bw:ymax];                % histogram bins.
[yvals,xvals]=hist(y,bins);         % compute histogram values
pdf_est=yvals/(N*bw);               % convert to prob. densities
bar(xvals,pdf_est)                  % plot histogram

% Compare true PDF with histogram.
y=[0.01:0.01:ymax];
pdf=1./(a*sqrt(y));
hold on
plot(y,pdf)
xlabel('y'); ylabel('f_Y(y)')
hold off
```

(Continued)
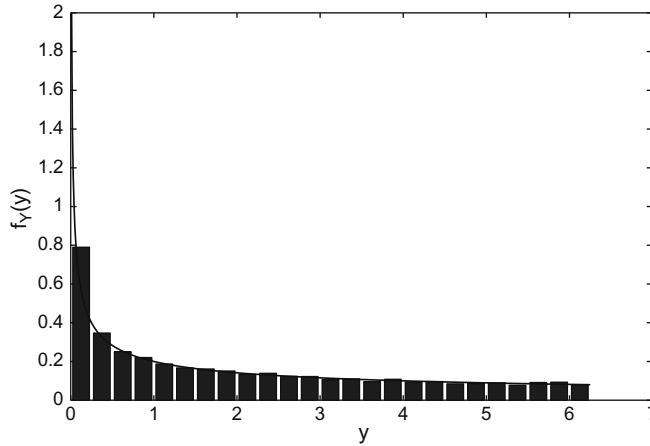
Figure 4.4
Comparison of estimated and true PDF for Example 4.17. (For color version of this figure, the reader is refered to the web version of this chapter).

∎

4.7. Characteristic Functions

In this section, we introduce the concept of a characteristic function. The characteristic function of a random variable is closely related to the Fourier transform of the PDF of that random variable. Thus, the characteristic function provides a sort of "frequency domain" representation of a random variable, although in this context there is no connection between our frequency variable ω and any physical frequency. In studies of deterministic signals, it was found that the use of Fourier transforms greatly simplified many problems, especially those involving convolutions. We will see in future chapters the need for performing convolution operations on PDFs of random variables, and hence frequency domain tools will become quite useful. Furthermore, we will find that characteristic functions have many other uses. For example, the characteristic function is quite useful for finding moments of a random variable. In addition to the characteristic function, two other related functions, namely, the moment-generating function (analogous to the Laplace transform) and the probability-generating function (analogous to the z-transform), will also be studied in the following sections.

Definition 4.7: The *characteristic function* of a random variable, X, is given by

$$\Phi_X(\omega) = E[e^{j\omega X}] = \int_{-\infty}^{\infty} e^{j\omega x} f_X(x)\mathrm{d}x. \tag{4.36}$$

Note the similarity between this integral and the Fourier transform. In most of the electrical engineering literature, the Fourier transform of the function $f_X(x)$ would be $\Phi(-\omega)$. Given this relationship between the PDF and the characteristic function, it should be clear that one can get the PDF of a random variable from its characteristic function through an inverse Fourier transform operation:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\omega x}\Phi_X(\omega)\,d\omega. \tag{4.37}$$

The characteristic functions associated with various random variables can be easily found using tables of commonly used Fourier transforms, but one must be careful since the Fourier integral used in Equation (4.36) may be different from the definition used to generate common tables of Fourier transforms. In addition, various properties of Fourier transforms can also be used to help calculate characteristic functions as shown in the following example.

■ **Example 4.18:**

An exponential random variable has a PDF given by $f_X(x) = \exp(-x)u(x)$. Its characteristic function is found to be

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} e^{j\omega x}f_X(x)\,dx = \int_0^{\infty} e^{j\omega x}e^{-x}\,dx = -\left.\frac{e^{-(1-j\omega)x}}{1-j\omega}\right|_0^{\infty} = \frac{1}{1-j\omega}.$$

This result assumes that ω is a real quantity. Now suppose another random variable Y has a PDF given by $f_Y(y) = a\exp(-ay)u(y)$. Note that $f_Y(y) = af_X(ay)$, thus using the scaling property of Fourier transforms, the characteristic function associated with the random variable Y is given by

$$\Phi_Y(\omega) = a\frac{1}{|a|}\Phi_X\left(\frac{\omega}{a}\right) = \frac{1}{1-j\omega/a} = \frac{a}{a-j\omega},$$

assuming a is a positive constant (which it must be for Y to have a valid PDF). Finally, suppose that Z has a PDF given by $f_Z(z) = a\exp(-a(z-b))u(z-b)$. Since $f_Z(z) = f_Y(z-b)$, the shifting property of Fourier transforms can be used to help find the characteristic function associated with the random variable Z:

$$\Phi_Z(\omega) = \Phi_Y(\omega)e^{-j\omega b} = \frac{ae^{-j\omega b}}{a-j\omega}.$$

■

The next example demonstrates that the characteristic function can also be computed for discrete random variables. In Section 4.8, the probability-generating function will be introduced which is preferred by some when dealing with discrete random variables.

■ Example 4.19:

A binomial random variable has a PDF which can be expressed as

$$f_X(x) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta(x-k).$$

Its characteristic function is computed as follows:

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} e^{j\omega x} \left(\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta(x-k) \right) dx = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \int_{-\infty}^{\infty} \delta(x-k) e^{j\omega x} dx$$

$$= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{j\omega k} = \sum_{k=0}^{n} \binom{n}{k} (pe^{j\omega})^k (1-p)^{n-k} = (1-p+pe^{j\omega})^n.$$

■

Since the Gaussian random variable plays such an important role in so many studies, we derive its characteristic function in Example 4.20. We recommend that the student commit the result of this example to memory. The techniques used to arrive at this result are also important and should be carefully understood.

■ Example 4.20:

For a standard normal random variable, the characteristic function can be found as follows:

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{j\omega x} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x^2 - 2j\omega x)}{2} \right) dx.$$

To evaluate this integral, we complete the square in the exponent.

$$\Phi_X(\omega) = \exp\left(-\frac{\omega^2}{2} \right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x^2 - 2j\omega x - \omega^2)}{2} \right) dx = \exp\left(-\frac{\omega^2}{2} \right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-j\omega)^2}{2} \right) dx.$$

The integrand in the above expression looks like the properly normalized PDF of a Gaussian random variable, and since the integral is over all values of x, the integral must be unity. However, close examination of the integrand reveals that the "mean" of this Gaussian integrand is complex. It is left to the student to rigorously verify that this integral still evaluates to unity even though the integrand is not truly a Gaussian PDF (since it is a complex function and hence not a PDF at all). The resulting characteristic function is then

$$\Phi_X(\omega) = \exp\left(-\frac{\omega^2}{2} \right).$$

For a Gaussian random variable whose mean is not zero or whose standard deviation is not unity (or both), the shifting and scaling properties of Fourier transforms can be used to show that

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \leftrightarrow \Phi_X(\omega) = \exp\left(j\mu\omega - \frac{\omega^2 \sigma^2}{2} \right).$$

■

Theorem 4.3: For any random variable whose characteristic function is differentiable at $\omega = 0$,

$$E[X] = -j\frac{\mathrm{d}}{\mathrm{d}\omega}\Phi_X(\omega)\bigg|_{\omega = 0}. \tag{4.38}$$

Proof: The proof follows directly from the fact that the expectation and differentiation operations are both linear and consequently the order of these operations can be exchanged.

$$\frac{\mathrm{d}}{\mathrm{d}\omega}\Phi_X(\omega) = \frac{\mathrm{d}}{\mathrm{d}\omega}(E[e^{j\omega X}]) = E\left[\frac{\mathrm{d}}{\mathrm{d}\omega}(e^{j\omega X})\right] = E[jXe^{j\omega X}] = jE[Xe^{j\omega X}].$$

Multiplying both sides by $-j$ and evaluating at $\omega = 0$ produces the desired result. □

Theorem 4.3 demonstrates a very powerful use of the characteristic function. Once the characteristic function of a random variable has been found, it is generally a very straightforward thing to produce the mean of the random variable. Furthermore, by taking the kth derivative of the characteristic function and evaluating at $\omega = 0$, an expression proportional to the kth moment of the random variable is produced. In particular,

$$E[X^k] = (-j)^k\frac{\mathrm{d}^k}{\mathrm{d}\omega^k}\Phi_X(\omega)\bigg|_{\omega = 0}. \tag{4.39}$$

Hence, the characteristic function represents a convenient tool to easily determine the moments of a random variable.

■ Example 4.21:

Consider the exponential random variable of Example 4.18 where $f_Y(y) = a\exp(-ay)u(y)$. The characteristic function was found to be

$$\Phi_Y(\omega) = \frac{a}{a - j\omega}.$$

The derivative of the characteristic function is

$$\frac{\mathrm{d}}{\mathrm{d}\omega}\Phi_Y(\omega) = \frac{ja}{(a - j\omega)^2},$$

and thus the first moment of Y is

$$E[Y] = -j\frac{\mathrm{d}}{\mathrm{d}\omega}\Phi_Y(\omega)\bigg|_{\omega = 0} = \frac{a}{(a - j\omega)^2}\bigg|_{\omega = 0} = \frac{1}{a}.$$

(Continued)

For this example, it is not difficult to show that the kth derivative of the characteristic function is

$$\frac{d^k}{d\omega^k}\Phi_Y(\omega) = \frac{j^k k! a}{(a-j\omega)^{k+1}},$$

and from this, the kth moment of the random variable is found to be

$$E[Y^k] = (-j)^k \frac{d^k}{d\omega^k}\Phi_Y(\omega)\bigg|_{\omega=0} = \frac{k!a}{(a-j\omega)^{k+1}}\bigg|_{\omega=0} = \frac{k!}{a^k}.$$

∎

For random variables that have a more complicated characteristic function, evaluating the kth derivative in general may not be an easy task. However, Equation (4.39) only calls for the kth derivative evaluated at a single point ($\omega = 0$), which can be extracted from the Taylor series expansion of the characteristic function. To see this, note that from Taylor's theorem, the characteristic function can be expanded in a power series as

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{d^k}{d\omega^k}\Phi_X(\omega)\bigg|_{\omega=0}\right)\omega^k. \tag{4.40}$$

If one can obtain a power series expansion of the characteristic function, then the required derivatives are proportional to the coefficients of the power series. Specifically, suppose an expansion of the form

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} \phi_k \omega^k, \tag{4.41}$$

is obtained. Then the derivatives of the characteristic function are given by

$$\frac{d^k}{d\omega^k}\Phi_X(\omega)\bigg|_{\omega=0} = k!\,\phi_k. \tag{4.42}$$

The moments of the random variable are then given by

$$E[X^k] = (-j)^k k!\,\phi_k. \tag{4.43}$$

This procedure is illustrated using a Gaussian random variable in the next example.

■ Example 4.22:

Consider a Gaussian random variable with a mean of $\mu = 0$ and variance σ^2. Using the result of Example 4.20, the characteristic function is $\Phi_X(\omega) = \exp(-\omega^2\sigma^2/2)$. Using the well-known Taylor series expansion of the exponential function, the characteristic function is expressed as

$$\Phi_X(\omega) = \sum_{n=0}^{\infty} \frac{(-\omega^2\sigma^2/2)^n}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n\sigma^{2n}}{2^n n!}\omega^{2n}.$$

The coefficients of the general power series as expressed in Equation (4.41) are given by

$$\phi_k = \begin{cases} \dfrac{j^k(\sigma/\sqrt{2})^k}{(k/2)!}, & k \text{ even,} \\ 0, & k \text{ odd.} \end{cases}$$

Hence, the moments of the zero-mean Gaussian random variable are

$$E[X^k] = \begin{cases} \dfrac{k!}{(k/2)!}\left(\dfrac{\sigma}{\sqrt{2}}\right)^k, & k \text{ even,} \\ 0, & k \text{ odd.} \end{cases}$$

As expected, $E[X^0] = 1$, $E[X] = 0$ (since it was specified that $\mu = 0$), and $E[X^2] = \sigma^2$ (since in the case of zero-mean variables, the second moment and variance are one and the same). Now, we also see that $E[X^3] = 0$ (as are all odd moments), $E[X^4] = 3\sigma^4$, $E[X^6] = 15\sigma^6$, and so on. We can also conclude from this that for Gaussian random variables, the coefficient of skewness is $c_s = 0$ while the coefficient of kurtosis is $c_k = 3$. ■

In many cases of interest, the characteristic function has an exponential form. The Gaussian random variable is a typical example. In such cases, it is convenient to deal with the natural logarithm of the characteristic function.

Definition 4. 8: In general, we can write a series expansion of $\ln[\Phi_X(\omega)]$ as

$$\ln[\Phi_X(\omega)] = \sum_{n=1}^{\infty} \lambda_n \frac{(j\omega)^n}{n!}, \tag{4.44}$$

where the coefficients, λ_n, are called the *cumulants* and are given as

$$\lambda_n = \frac{d^n}{d(j\omega)^n}\{\ln[\Phi_X(\omega)]\}\Big|_{\omega=0}, \quad n = 1, 2, 3, \dots \tag{4.45}$$

The cumulants are related to the moments of the random variable. By taking the derivatives specified in Equation (4.45) we obtain

$$\lambda_1 = \mu_X, \tag{4.46}$$

$$\lambda_2 = E[X^2] - \mu_X^2 = \sigma_X^2, \tag{4.47}$$

$$\lambda_3 = E[X^3] - 3\mu_X E[X^2] + 2\mu_X^3 = E[(X - \mu_X)^3]. \tag{4.48}$$

Thus, λ_1 is the mean, λ_2 is the second central moment (or the variance), and λ_3 is the third central moment. However, higher-order cumulants are not as simply related to the central moments.

4.8. Probability-Generating Functions

In the world of signal analysis, we often use Fourier transforms to describe continuous time signals, but when we deal with discrete time signals, it is common to use a z-transform instead. In the same way, the characteristic function is a useful tool for working with continuous random variables, but when discrete random variables are concerned, it is often more convenient to use a device similar to the z-transform which is known as the probability-generating function.

> **Definition 4.9:** For a discrete random variable with a PMF, $P_X(k)$, defined on the nonnegative integers,[1] $k = 0, 1, 2, \ldots$, the *probability-generating function*, $H_X(z)$, is defined as
>
> $$H_X(z) = \sum_{k=0}^{\infty} P_X(k) z^k. \tag{4.49}$$
>
> Note the similarity between the probability-generating function and the unilateral z-transform of the PMF.

Since the PMF is seen as the coefficients of the Taylor series expansion of $H_X(z)$, it should be apparent that the PMF can be obtained from the probability-generating function through

$$P_X(k) = \frac{1}{k!} \frac{\mathrm{d}^k}{\mathrm{d}z^k} H_X(z) \bigg|_{z=0}. \tag{4.50}$$

---

[1] Note that this definition assumes that the discrete random variable, X, is defined on nonnegative integer values, k. One could also define a probability-generating function based on a bilateral z-transform which would allow for random variables which can take on negative integer values as well. However, since this is less common, we do not consider it further here.

The derivatives of the probability-generating function evaluated at zero return the PMF and not the moments as with the characteristic function. However, the moments of the random variable can be obtained from the derivatives of the probability-generating function at $z = 1$.

> **Theorem 4.4:** The mean of a discrete random variable can be found from its probability-generating function according to
>
> $$E[X] = \frac{\mathrm{d}}{\mathrm{d}z}H_X(z)\bigg|_{z=1}. \tag{4.51}$$
>
> Furthermore, the higher-order derivatives of the probability-generating function evaluated at $z = 1$ lead to quantities which are known as the *factorial moments*,
>
> $$h_k = \frac{\mathrm{d}^k}{\mathrm{d}z^k}H_X(z)\bigg|_{z=1} = E[X(X-1)(X-2)\ldots(X-k+1)]. \tag{4.52}$$
>
> **Proof:** The result follows directly from differentiating Equation (4.49). The details are left to the reader. $\square$

It is a little unfortunate that these derivatives do not produce the moments directly, but the moments can be calculated from the factorial moments. For example,

$$h_2 = E[X(X-1)] = E[X^2] - E[X] = E[X^2] - h_1,$$

$$\Rightarrow E[X^2] = h_2 + h_1. \tag{4.53}$$

Hence, the second moment is simply the sum of the first two factorial moments. Furthermore, if we were interested in the variance as well, we would obtain

$$\sigma_X^2 = E[X^2] - \mu_X^2 = h_2 + h_1 - h_1^2. \tag{4.54}$$

∎ Example 4.23:

Consider the binomial random variable of Example 4.4 whose PMF is

$$P_X(k) = \binom{n}{k}p^k(1-p)^{n-k}, \; k = 0, 1, 2, \ldots, n.$$

The corresponding probability-generating function is

$$H_X(z) = \sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}z^k = \sum_{k=0}^{n}\binom{n}{k}(pz)^k(1-p)^{n-k} = (1-p+pz)^n.$$

(Continued)

Evaluating the first few derivatives at $z = 1$ produces

$$h_1 = \frac{d}{dz}H_X(z)\Big|_{z=1} = np(1-p+pz)^{n-1}\big|_{z=1} = np,$$

$$h_2 = \frac{d^2}{dz^2}H_X(z)\Big|_{z=1} = n(n-1)p^2(1-p+pz)^{n-2}\big|_{z=1} = n(n-1)p^2.$$

From these factorial moments, we calculate the mean, second moment, and variance of a binomial random variable as

$$\mu_X = h_1 = np, \ E[X^2] = h_1 + h_2 = (np)^2 + np(1-p), \ \sigma_X^2 = h_2 + h_1 - h_1^2 = np(1-p).$$

∎

In order to gain an appreciation for the power of these "frequency domain" tools, compare the amount of work used to calculate the mean and variance of the binomial random variable using the probability-generating function in Example 4.23 with the direct method used in Example 4.4.

As was the case with the characteristic function, we can compute higher-order factorial moments without having to take many derivatives, by expanding the probability-generating function into a Taylor series. In this case, the Taylor series must be about the point $z = 1$.

$$H_X(z) = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{d^k}{dz^k}H_X(z)\Big|_{z=1}\right)(z-1)^k = \sum_{k=0}^{\infty} \frac{1}{k!}h_k(z-1)^k. \tag{4.55}$$

Once this series is obtained, one can easily identify all of the factorial moments. This is illustrated using a geometric random variable in Example 4.24.

∎ **Example 4.24:**

A geometric random variable has a PMF given by $P_X(k) = (1-p)p^k$, $k = 0, 1, 2, \ldots$ The probability-generating function is found to be

$$H_X(z) = \sum_{k=0}^{\infty} (1-p)(pz)^k = \frac{1-p}{1-pz}.$$

In order to facilitate forming a Taylor series expansion of this function about the point $z = 1$, it is written explicitly as a function of $z - 1$. From there, the power series expansion is fairly simple.

$$H_X(z) = \frac{1-p}{1-p-p(z-1)} = \frac{1}{1-\frac{p}{1-p}(z-1)} = \sum_{k=0}^{\infty} \left(\frac{p}{1-p}\right)^k(z-1)^k.$$

Comparing the coefficients of this series with the coefficients given in Equation (4.55) leads to immediate identification of the factorial moments,

$$h_k = \frac{k!p^k}{(1-p)^k}.$$

∎

4.9 Moment-Generating Functions

In many problems, the random quantities we are studying are often inherently nonnegative. Examples include the magnitude of a random signal, the time between arrivals of successive customers in a queueing system, or the number of points scored by your favorite football team. The resulting PDFs of these quantities are naturally one-sided. For such one-sided waveforms, it is common to use Laplace transforms as a frequency domain tool. The moment-generating function is the equivalent tool for studying random variables.

> **Definition 4.10:** The *moment-generating function*, $M_X(u)$, of a nonnegative[2] random variable, X, is
>
> $$M_X(u) = E[e^{uX}] = \int_0^\infty f_X(x)e^{ux}dx. \tag{4.56}$$
>
> Note the similarity between the moment-generating function and the Laplace transform of the PDF.

The PDF can in principle be retrieved from the moment-generating function through an operation similar to an inverse Laplace transform,

$$f_X(x) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} M_X(u)e^{-ux}du. \tag{4.57}$$

Because the sign in the exponential term in the integral in Equation (4.56) is the opposite of the traditional Laplace transform, the contour of integration (the so-called Bromwich contour) in the integral specified in Equation (4.57) must now be placed to the left of all poles of the moment-generating function. As with the characteristic function, the moments of the random variable can be found from the derivatives of the moment-generating function (hence, its name) according to

$$E[X^k] = \frac{d^k}{du^k}M_X(u)\bigg|_{u=0}. \tag{4.58}$$

---

[2] One may also define a moment-generating function for random variables which are not necessarily nonnegative. In that case, a two-sided Laplace transform would be appropriate. This would be identical to the characteristic function with the association $u = j\omega$.

It is also noted that if the moment-generating function is expanded in a power series of the form

$$M_X(u) = \sum_{k=0}^{\infty} m_k u^k , \qquad (4.59)$$

then the moments of the random variable are given by $E[X^k] = k! m_k$.

■ Example 4.25:

Consider an Erlang random variable with a PDF of the form

$$f_X(x) = \frac{x^{n-1} \exp(-x) u(x)}{(n-1)!}$$

The moment-generating function is calculated according to

$$M_X(u) = \int_0^{\infty} f_X(x) e^{ux} dx = \int_0^{\infty} \frac{x^{n-1} \exp(-(1-u)x)}{(n-1)!} dx .$$

To evaluate this function, we note that the integral looks like the Laplace transform of the function $x^{n-1}/(n-1)!$ evaluated at $s = 1-u$. Using standard tables of Laplace transforms (or using integration by parts several times) we get

$$M_X(u) = \frac{1}{(1-u)^n} .$$

The first two moments are then found as follows:

$$E[X] = \frac{d}{du}(1-u)^{-n}\bigg|_{u=0} = n(1-u)^{-(n+1)}\big|_{u=0} = n ,$$

$$E[X^2] = \frac{d^2}{du^2}(1-u)^{-n}\bigg|_{u=0} = n(n+1)(1-u)^{-(n+2)}\big|_{u=0} = n(n+1) .$$

From this, we could also infer that the variance is $\sigma_X^2 = n(n+1) - n^2 = n$. If we wanted a general expression for the kth moment, it is not hard to see that

$$E[X^k] = \frac{d^k}{du^k}(1-u)^{-n}\bigg|_{u=0} = n(n+1)\ldots(n+k-1) = \frac{(n+k-1)!}{(n-1)!} .$$

■

4.10 Evaluating Tail Probabilities

A common problem encountered in a variety of applications is the need to compute the probability that a random variable exceeds a threshold, $\Pr(X > x_o)$. Alternatively we might want to know, $\Pr(|X - \mu_X| > x_o)$. These quantities are referred to as tail probabilities. That is, we are asking, what is the probability that the random variable takes on a value that is in the tail of the distribution? While this can be found directly from the CDF of the random variable, quite often, the CDF may be difficult or even impossible to find. In those cases, one can always resort to numerical integration of the PDF. However, this involves a numerical integration over a

semi-infinite region, which in some cases may be problematic. Then, too, in some situations, we might not even have the PDF of the random variable, but rather the random variable may be described in some other fashion. For example, we may only know the mean, or the mean and variance, or the random variable may be described by one of the frequency domain functions discussed in the previous sections. Obviously, if we are only given partial information about the random variable, we would not expect to be able to perfectly evaluate the tail probabilities, but we can obtain bounds on these probabilities. In this section, we present several techniques for obtaining various bounds on tail probabilities based on different information about the random variable. We then conclude the section by showing how to exactly evaluate the tail probabilities directly from one of the frequency domain descriptions of the random variables.

Theorem 4.5 (Markov's inequality): Suppose that X is a nonnegative random variable (i.e., one whose PDF is nonzero only over the range $[0, \infty)$). Then,

$$\Pr(X \geq x_o) \leq \frac{E[X]}{x_o}. \tag{4.60}$$

Proof: For nonnegative random variables, the expected value is

$$E[X] = \int_0^\infty x f_X(x)\,dx = \int_0^{x_o} x f_X(x)\,dx + \int_{x_o}^\infty x f_X(x)\,dx \geq \int_{x_o}^\infty x f_X(x)\,dx \geq x_o \int_{x_o}^\infty f_X(x)\,dx . \tag{4.61}$$

Dividing both sides by x_o gives the desired result. □

Markov's inequality provides a bound on the tail probability. The bound requires only knowledge of the mean of the random variable. Because the bound uses such limited information, it has the potential of being very loose. In fact, if $x_o < E[X]$, then the Markov inequality states that $\Pr(X \geq x_o)$ is bounded by a number that is greater than 1. While this is true, in this case the Markov inequality gives us no useful information. Even in less extreme cases, the result can still be very loose as shown by the next example.

■ Example 4.26:

Suppose the average life span of a person was 78 years. The probability of a human living to be 110 years would then be bounded by

$$\Pr(X \geq 110) \leq \frac{78}{110} = 0.7091 .$$

Of course, we know that in fact very few people live to be 110 years old, and hence this bound is almost useless to us.

■

If we know more about the random variable than just its mean, we can obtain a more precise estimate of its tail probability. In Example 4.26, we know that the bound given by the Markov's inequality is ridiculously loose because we know something about the variability of the human life span. The next result allows us to use the variance as well as the mean of a random variable to form a different bound on the tail probability.

Theorem 4.6 (Chebyshev's inequality): Suppose that X is a random variable with mean, μ_X, and variance, σ_X^2. The probability that the random variable takes on a value that is removed from the mean by more than x_0 is given by

$$\Pr(|X - \mu_X| \geq x_0) \leq \frac{\sigma_X^2}{x_0^2}. \tag{4.62}$$

Proof: Chebyshev's inequality is a direct result of Markov's inequality. Note that the event $\{|X - \mu_X| \geq x_0\}$ is equivalent to the event $\{(X - \mu_X)^2 \geq x_0^2\}$. Applying Markov's inequality to the later event results in

$$\Pr((X - \mu_X)^2 \geq x_0^2) \leq \frac{E[(X - \mu_X)^2]}{x_0^2} = \frac{\sigma_X^2}{x_0^2}. \quad \square \tag{4.63}$$

Chebyshev's inequality gives a bound on the two-sided tail probability, whereas the Markov inequality applies to the one-sided tail probability. Also, the Chebyshev inequality can be applied to any random variable, not just those that are nonnegative.

■ **Example 4.27:**

Continuing the previous example, suppose that in addition to a mean of 78 years, the human life span had a standard deviation of 15 years. In this case

$$\Pr(X \geq 110) \leq \Pr(X \geq 110) + \Pr(X \leq 46) = \Pr(|X - 78| \geq 32) .$$

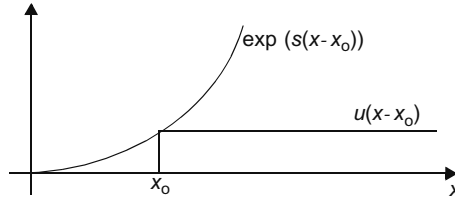Now the Chebyshev inequality can be applied to give

$$\Pr(|X - 78| \geq 32) \leq \left(\frac{15}{32}\right)^2 = 0.2197 .$$

While this result may still be quite loose, by using the extra piece of information provided by the variance, a better bound is obtained.

■

Theorem 4.7 (Chernoff bound): Suppose X is a random variable whose moment generating function is $M_X(s)$. Then

$$\Pr(X \geq x_0) \leq \min_{s \geq 0} \ e^{-s x_0} M_X(s). \tag{4.64}$$

Figure 4.5
The unit step function and an exponential upper bound.

Proof: First, note that

$$\Pr(X \geq x_o) = \int_{x_o}^{\infty} f_X(x)\mathrm{d}x = \int_{-\infty}^{\infty} f_X(x)u(x - x_o)\mathrm{d}x. \tag{4.65}$$

Next, upper bound the unit step function in the above integrand by an exponential function of the form $u(x - x_o) \leq \exp(s(x - x_o))$. This bound is illustrated in Figure 4.5. Note that the bound is valid for any real $s \geq 0$. The tail probability is then upper bounded by

$$\Pr(X \geq x_o) \leq \mathrm{e}^{-sx_o} \int_{-\infty}^{\infty} f_X(x)\mathrm{e}^{sx}\mathrm{d}x = \mathrm{e}^{-sx_o}M_X(s). \tag{4.66}$$

Since this bound is valid for any $s \geq 0$, it can be tightened by finding the value of s that minimizes the right-hand side. In this expression, a two-sided Laplace transform must be used to obtain the moment-generating function if the random variable is not nonnegative (see footnote 2 associated with Definition 4.10). □

■ Example 4.28:

Consider a standard normal random variable whose moment-generating function is given by $M_X(u) = \exp(u^2/2)$ (see the result of Example 4.20, where the characteristic function is found and replace ω with $-ju$). The tail probability, $\Pr(X \geq x_o)$, in this case is simply the Q-function, $Q(x_o)$. According to (4.66), this tail probability can be bounded by

$$Q(x_o) \leq \exp\left(-ux_o + \frac{u^2}{2}\right)$$

for any $u \geq 0$. Minimizing with respect to u, we get

$$\frac{\mathrm{d}}{\mathrm{d}u}\exp\left(-ux_o + \frac{u^2}{2}\right) = (-x_o + u)\exp\left(-ux_o + \frac{u^2}{2}\right) = 0 \Rightarrow u = x_o.$$

(Continued)

Hence, the Chernoff bound on the tail probability for a standard normal random variable is

$$Q(x_o) \leq \exp\left(-\frac{x_o^2}{2}\right).$$

The result of this example provides a convenient upper bound on the Q-function. ∎

Evaluating the Chernoff bound requires knowledge of the moment-generating function of the random variable. This information is sufficient to calculate the tail probability exactly since, in theory, one can obtain the PDF from the moment-generating function, and from there the exact tail probability can be obtained. However, in cases where the moment-generating function is of a complicated analytical form determining the PDF may be exceedingly difficult. Indeed, in some cases, it may not be possible to express the tail probability in closed form (like with the Gaussian random variable). In these cases, the Chernoff bound will often provide an analytically tractable expression which can give a crude bound on the tail probability. If a precise expression for the tail probability is required, the result of Theorem 4.8 will show how this can be obtained directly from the moment-generating function (without having to explicitly find the PDF).

Theorem 4.8: For a random variable, X, with a moment-generating function, $M_X(u)$, an exact expression for the tail probability, $\Pr(X \geq x_o)$, is given by

$$\Pr(X \geq x_o) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} \frac{M_X(u)}{u} e^{-ux_o} du, \qquad (4.67)$$

where the contour of integration is to the right of the origin, but to the left of all singularities of the moment-generating function in the right half plane.

Proof: The right tail probability is given in general by

$$\Pr(X \geq x_o) = \int_{x_o}^{\infty} f_X(x) dx. \qquad (4.68)$$

Then, replace the PDF in this integral, with an inverse transform of the moment-generating function as specified in Equation (4.57).

$$\Pr(X \geq x_o) = \int_{x_o}^{\infty} \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} M_X(u)e^{-ux} du\, dx = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} M_X(u)\int_{x_o}^{\infty} e^{-ux} dx\, du. \quad (4.69)$$

Evaluating the inner integral results in

$$\Pr(X \geq x_0) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{M_X(u)}{u} e^{-ux_0} du, \text{ for } Re[u] > 0. \;\; \square \qquad (4.70)$$

The integral specified in Equation (4.67) can be evaluated numerically or, when convenient to do so, it can also be evaluated by computing the appropriate residues.[3] For $x_0 > 0$ the contour of integration can be closed to the right. The resulting closed contour will encompass all the singularities of the moment-generating function in the right half plane. According to Cauchy's residue theorem, the value of the integral will then be $-2\pi j$ times the sum of the residues of all the singularities encompassed by the contour. Hence,

$$\Pr(X \geq x_0) = \sum_{\text{right half plane}} \text{Residues}\left\{ \frac{M_X(u)}{u} e^{-ux_0} \right\}. \qquad (4.71)$$

If a precise evaluation of the tail probability is not necessary, several approximations to the integral in Equation (4.67) are available. Perhaps the simplest and most useful is known as the *saddle point approximation*. To develop the saddle point approximation, define $\psi(u) = \ln(M_X(u))$ and

$$\lambda(u) = \ln\left(\frac{M_X(u)}{u} e^{-ux_0} \right) = \psi(u) - ux_0 - \ln(u). \qquad (4.72)$$

Furthermore, consider a Taylor series expansion of the function $\lambda(u)$ about some point $u = u_0$,

$$\lambda(u) = \lambda(u_0) + \lambda'(u_0)(u - u_0) + \frac{1}{2}\lambda''(u_0)(u - u_0)^2 + \dots. \qquad (4.73)$$

In particular, if u_0 is chosen so that $\lambda'(u_o) = 0$, then near the point $u = u_0$, the integrand in Equation (4.67) behaves approximately like

$$\frac{M_X(u)}{u} e^{-ux_0} = e^{\lambda(u)} \approx \exp(\lambda(u_0)) \exp\left(\frac{1}{2}\lambda''(u_0)(u - u_0)^2 \right). \qquad (4.74)$$

In general, the point $u = u_0$ will be a minima of the integrand as viewed along the real axis. This follows from the fact that the integrand is a concave function of u. A useful property of complex (analytic) functions tells us that if the function has a minima at some point u_0 as u

---

[3] The remainder of this section assumes the student is familiar with the concepts of contour integration and residue calculus. For those students not familiar with these topics, the remainder of this section can be skipped without any loss in continuity.

passes through it in one direction, the function will also have a maxima as u passes through the same point in the orthogonal direction. Such a point is called a "saddle point" since the shape of the function resembles a saddle near that point. If the contour of integration is selected to pass through the saddle point, the integrand will reach a local maximum at the saddle point. As just seen in Equation (4.74), the integrand also has a Gaussian behavior at and around the saddle point. Hence, using the approximation of (4.74) and running the contour of integration through the saddle point so that $u = u_o + j\omega$ along the integration contour, the tail probability is approximated by

$$
\begin{aligned}
\Pr(X \geq x_o) &\approx \frac{\exp(\lambda(u_o))}{2\pi j} \int_{u_o - j\infty}^{u_o + j\infty} \exp\left(\frac{1}{2}\lambda''(u_o)(u - u_o)^2\right) du \\
&= \frac{\exp(\lambda(u_o))}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\lambda''(u_o)\omega^2\right) d\omega \\
&= \frac{\exp(\lambda(u_o))}{\sqrt{2\pi\lambda''(u_o)}} \\
&= \frac{M_X(u_o)\exp(-u_o x_o)}{u_o\sqrt{2\pi\lambda''(u_o)}} .
\end{aligned}
\tag{4.75}
$$

The third step is accomplished using the normalization integral for Gaussian PDFs.

The saddle point approximation is usually quite accurate provided that $x_o \gg E[X]$. That is, the farther we go out into the tail of the distribution, the better the approximation. If it is required to calculate $\Pr(X \geq x_o)$ for $x_o < E[X]$, it is usually better to calculate the left tail probability in which case the saddle point approximation is

$$
\Pr(X \leq x_o) \approx -\frac{M_X(u_o)\exp(-u_o x_o)}{u_o\sqrt{2\pi\lambda''(u_o)}} ,
\tag{4.76}
$$

where in this case, the saddle point, u_o, must be negative.

■ **Example 4.29:**

In this example, we form the saddle point approximation to the Q-function which is the right tail probability for a standard normal random variable. The corresponding moment-generating function is $M_X(u) = \exp(u^2/2)$. To find the saddle point, we note that

$$
\lambda(u) = \frac{u^2}{2} - \ln(u) - ux_o .
$$

We will need the first two derivatives of this function:

$$\lambda'(u) = u - \frac{1}{u} - x_o, \quad \lambda''(u) = 1 + \frac{1}{u^2}.$$

The saddle point is the solution to $\lambda'(u_o) = 0$. This results in a quadratic equation whose roots are

$$u_o = \frac{x_o \pm \sqrt{x_o^2 + 4}}{2}.$$

When calculating the right tail probability, the saddle point must be to the right of the imaginary axis, hence the positive root must be used:

$$u_o = \frac{x_o + \sqrt{x_o^2 + 4}}{2}.$$

The saddle point approximation then becomes

$$Q(x_o) \approx \frac{M_x(u_o)\exp(-u_o x_o)}{u_o\sqrt{2\pi\lambda''(u_o)}} = \frac{\exp\left(\frac{u_o^2}{2} - u_o x_o\right)}{\sqrt{2\pi(1 + u_o^2)}}.$$

The exact value of the Q-function and the saddle point approximation are compared in Figure 4.6. As long as x_o is not close to zero, this approximation is quite accurate.
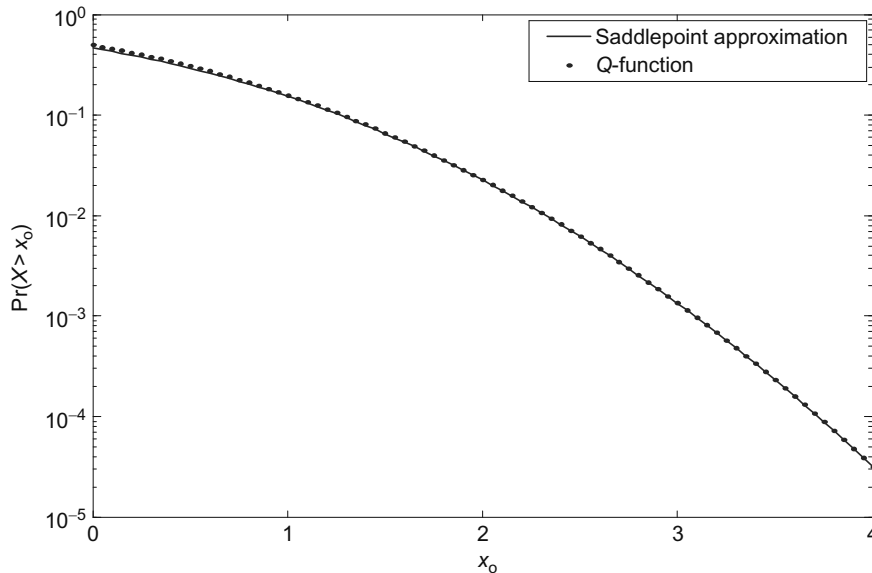
Figure 4.6
The Q-function and its saddle point approximation.

4.11   Engineering Application—Scalar Quantization

In many applications, it is convenient to convert a signal which is analog in nature to a digital one. This is typically done in three steps. First, the signal is sampled, which converts the signal from continuous time to discrete time. Then samples of the signal are quantized. This second action converts the signal from one with a continuous amplitude to one whose amplitude can only take on discrete values. Once the signal is converted to discrete time/discrete amplitude, the signal can then easily be represented by a sequence of bits. In this third step, each discrete amplitude level is represented by a binary codeword. While the first step (sampling) and the third step (encoding) are invertible, the second step (quantization) is not. That is, we can perfectly recover the analog signal from its discrete time samples (provided the samples are taken at a rate above the Nyquist rate), and the discrete amplitude levels can easily be recovered from the codewords which represent them (provided a lossless source code is used). However, the act of quantization causes distortion of the signal which cannot be undone. Given the discrete amplitude of a sample, it is not possible to determine the exact value of the original (continuous amplitude) sample. For this reason, careful attention is paid to the quantization process in order to minimize the amount of distortion.

In order to determine efficient ways to quantize signals, we must first quantify this concept of signal distortion. Suppose a signal is sampled and we focus attention on one of those samples. Let the random variable X represent the value of that sample, which in general will draw from a continuous sample space. Now suppose that sample is quantized (using some quantization function $q(x)$) to form a new (discrete) random variable $Y = q(X)$. The difference between the original sample value and its quantized value, $X - q(X)$, is the error caused by the quantizer, or the quantizer noise. It is common to measure signal distortion as the mean-squared quantizer error,

$$d = E[(X - q(X))^2] = \int_{-\infty}^{\infty} (x - q(x))^2 f_X(x)\,\mathrm{d}x. \tag{4.77}$$

We will see in Chapter 10 that the mean-squared value of a signal has the physical interpretation of the signal's power. Thus, the quantity d can be interpreted as the quantization noise power. Often the fidelity of a quantized signal is measured in terms of the ratio of the original signal power, $E[X^2]$, to the quantization noise power. This is referred to as the signal-to-quantization-noise power ratio (SQNR)

$$\mathrm{SQNR} = \frac{E[X^2]}{E[(X - q(X))^2]}. \tag{4.78}$$

The goal of the quantizer design is to choose a quantization function which minimizes the distortion, d. Normally, the quantizer maps the sample space of X into one of $M = 2^n$ levels. Then each quantization level can be represented by a unique n-bit codeword. We refer to this

as an *n*-bit quantizer. As indicated in Equation (4.77), the expected value is with respect to the PDF of X. Hence, the function $q(x)$ that minimizes the distortion will depend on the distribution of X.

To start with consider a random variable X which is uniformly distributed over the interval $(-a/2, a/2)$. Since the sample X is equally likely to fall anywhere in the region, it would make sense for the quantizer to divide that region into M equally spaced subintervals of width $\Delta = a/M$. For each subinterval, the quantization level (i.e., the value of $q(x)$ for that subinterval) should be chosen as the midpoint of the subinterval. This is referred to as a uniform quantizer. A 3-bit uniform quantizer is illustrated in Figure 4.7. For example, if $X \in (0, a/8)$, then $q(X) = a/16$. To measure the distortion for this signal together with the uniform quantizer, condition on the event that the signal falls within one of the quantization intervals, and then use the theorem of total probability:

$$d = E[(X - q(X))^2] = \sum_{k=1}^{\infty} E[(X - q(X))^2 | X \in X_k] \Pr(X \in X_k), \qquad (4.79)$$

where X_k refers to the *k*th quantization interval. Consider, for example, $X_5 = (0, a/8)$, so that

$$E[(X - q(X))^2 | X \in X_5] = E[(X - a/16)^2 | X \in (0, a/8)]. \qquad (4.80)$$
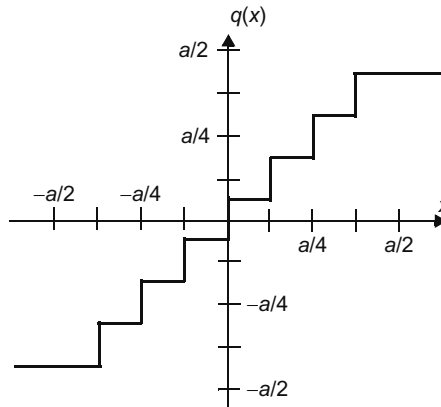
Figure 4.7
A 3-bit uniform quantizer on the interval (-*a*/2,*a*/2).

To calculate this conditional expected value requires the conditional PDF of X. From Equation (3.41) this is

$$f_X(x|X \in (0, a/8)) = \frac{f_X(x)}{\Pr(X \in (0, a/8))} = \frac{1/a}{1/8} = \frac{8}{a}, \; x \in (0, a/8). \tag{4.81}$$

Not surprisingly, conditioned on $X \in (0, a/8)$, X is uniformly distributed over $(0, a/8)$. The conditional expected value is then

$$E[(X - a/16)^2 | X \in (0, a/8)] = \frac{8}{a} \int_0^{a/8} (x - a/16)^2 dx = \frac{a^2}{768}. \tag{4.82}$$

Due to the symmetry of the uniform distribution, this conditional distortion is the same regardless of what quantization interval the signal falls in. Hence, Equation (4.82) is also the unconditional distortion. Note that the power of the original signal is

$$E[X^2] = \frac{1}{a} \int_{-a/2}^{a/2} x^2 dx = \frac{a^2}{12}. \tag{4.83}$$

The resulting SQNR is then

$$\text{SQNR} = \frac{E[X^2]}{d} = \frac{a^2/12}{a^2/768} = 64 = 18.06 \, \text{dB}. \tag{4.84}$$

The preceding result for the three-bit uniform quantizer can be generalized to any uniform quantizer. In general, for an n-bit uniform quantizer, there will be $M = 2^n$ quantization intervals of width $\Delta = a/M$. Consider the quantization interval $(0, \Delta)$ and suppose the quantization level for that interval is chosen to be the midpoint, $\Delta/2$. Then the distortion for that interval (and hence the distortion for the quantizer) is

$$E[(X - q(X))^2 | X \in (0, \Delta)] = \frac{1}{\Delta} \int_0^{\Delta} (x - \Delta/2)^2 dx = \Delta^2/12. \tag{4.85}$$

The SQNR for an n-bit uniform quantizer with a uniformly distributed input is

$$\text{SQNR} = \frac{E[X^2]}{d} = \frac{a^2/12}{\Delta^2/12} = M^2 = 2^{2n} \text{ or SQNR (dB)} = 2n \log_{10}(2) = 6.02n \, \text{dB}. \tag{4.86}$$

This is the so-called 6 dB rule whereby the SQNR is increased by approximately 6 dB for each extra bit added to the quantizer. For example, in wireline digital telephony, 8-bit quantization is used which would result in an SQNR of approximately 48 dB.

The previous results assumed that the input to the quantizer followed a uniform probability distribution. This is rarely the case. Speech signals, for example, are commonly modeled using a Laplace (two-sided exponential) distribution. For such signals, small sample values are much more frequent than larger values. In such a situation, it would make sense to use finer resolution (i.e., narrower quantization intervals) for the most frequent smaller amplitudes in order to keep the distortion minimized in those regions, at the cost of more distortion in the less frequent larger amplitude regions.

Given that an *n*-bit quantizer is to be used, the design of an optimum quantizer involves two separate problems. First, the ranges for each quantization interval must be specified; then the quantization level for each interval must be chosen. The following theorem specifies how each of these two tasks should be accomplished.

Theorem 4.9: A random variable X with PDF $f_X(x)$ is to be quantized with an M-level quantizer that produces a discrete random variable Y according to

$$y = q(x) = y_i, \text{ for } x_{i-1} < x < x_i, \ i = 1, 2, ..., M, \qquad (4.87)$$

where it is assumed that the lower limit of the first quantization interval is $x_0 = -\infty$ and the upper limit of the last quantization interval is $x_M = \infty$. The (mean-squared) distortion is minimized by choosing the quantization intervals (i.e., the x_i) and the quantization levels (i.e., the y_i) according to

(i) $y_i = E[X | x_{i-1} < X < x_i], \ i = 1, 2, ..., M$, (the conditional mean criterion), (4.88)

(ii) $x_i = \dfrac{y_i + y_{i+1}}{2}, \ i = 1, 2, ..., M-1$, \qquad (the midpoint criterion). \qquad (4.89)

These two criteria provide a system of $2M - 1$ equations with which to solve for the $2M - 1$ quantities $(x_1, x_2, ..., x_{M-1}, y_1, y_2, ..., y_M)$ which specify the optimum M-level quantizer.

Proof: The distortion is given by

$$d = \sum_{i=1}^{M} \int_{x_{i-1}}^{x_i} (x - y_i)^2 f_X(x) dx. \qquad (4.90)$$

To minimize d with respect to y_j,

$$\frac{\partial d}{\partial y_j} = -2 \int_{x_{i-1}}^{x_i} (x - y_i) f_X(x) dx = 0. \qquad (4.91)$$

Solving for y_j in this equation establishes the conditional mean criterion. Similarly differentiating with respect to x_j gives

$$\frac{\partial d}{\partial x_j} = (x_j - y_j)^2 f_X(x_j) - (x_j - y_{j+1})^2 f_X(x_j) = 0 \ . \tag{4.92}$$

Solving for x_j produces the midpoint criterion. □

■ Example 4.30:

Using the criteria set forth in Theorem 4.9, an ideal nonuniform 2-bit quantizer will be designed for a signal whose samples have a Laplace distribution, $f_X(x) = (1/2)\exp(-|x|)$. A 2-bit quantizer will have four quantization levels, $\{y_1, y_2, y_3, y_4\}$, and four corresponding quantization intervals which can be specified by three boundary points, $\{x_1, x_2, x_3\}$. The generic form of the 2-bit quantizer is illustrated in Figure 4.8. Due to the symmetry of the Laplace distribution, it seems reasonable to expect that the quantizer should have a negative symmetry about the y-axis. That is, $x_1 = -x_3$, $x_2 = 0$, $y_1 = -y_4$, and $y_2 = -y_3$. Hence, it is sufficient to determine just three unknowns, such as $\{x_3, y_3, y_4\}$. The rest can be inferred from the symmetry. Application of the conditional mean criterion and the midpoint criterion leads to the following set of three equations:

$$y_3 = E[X|0 < X < x_3] = \frac{\int_0^{x_3} \frac{x}{2}\exp(-x)dx}{\int_0^{x_3} \frac{1}{2}\exp(-x)dx} = 1 - \frac{x_3}{e^{x_3} - 1},$$

$$y_4 = E[X|X > x_3] = \frac{\int_{x_3}^{\infty} \frac{x}{2}\exp(-x)dx}{\int_{x_3}^{\infty} \frac{1}{2}\exp(-x)dx} = x_3 + 1,$$

$$x_3 = \frac{y_3 + y_4}{2} \ .$$
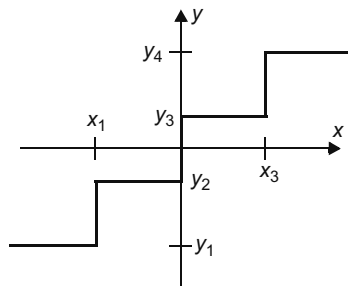
Figure 4.8
A 2-bit quantizer.

Plugging the expressions for y_3 and y_4 into the last equation results in a single equation to solve for the variable x_3. Unfortunately, the equation is transcendental and must be solved numerically. Doing so results in the solution $\{x_3, y_3, y_4\} = \{1.594, 0.594, 2.594\}$. The (mean-squared) distortion of this 2-bit quantizer is given by

$$d = 2\int_0^{x_3} (x - y_3)^2 f_X(x)\mathrm{d}x + 2\int_{x_3}^{\infty} (x - y_4)^2 f_X(x)\mathrm{d}x = 0.3524.$$

Note that the power in the original (unquantized) signal is $E[X^2] = 2$ so that the SQNR of this quantizer is

$$\mathrm{SQNR} = \frac{E[X^2]}{d} = \frac{2}{0.3524} = 5.675 = 7.54 \text{ dB}.$$

∎

■ Example 4.31:
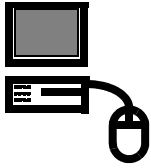
In this example, we generalize the results of the last example for an arbitrary number of quantization levels. When the number of quantization levels get large, the number of equations to solve becomes too difficult to do by hand, so we use MATLAB to help us with this task. Again, we assume that the random variable X follows a Laplace distribution given by $f_X(x) = (1/2)\exp(-|x|)$. Because of the symmetry of this distribution, we again take advantage of the fact that the optimum quantizer will be symmetric. We design a quantizer with M levels for positive X (and hence M levels for negative X as well, for a total of $2M$ levels). The M quantization levels are at $y_1, y_2, ..., y_M$ and the quantization bin edges are at $x_0 = 0, x_1, x_2, ..., x_{N-1}, x_N = \infty$. We compute the optimum quantizer in an iterative fashion. We start by arbitrarily setting the quantization bins in a uniform fashion. We choose to start with $x_i = 2i/M$. We then iterate between computing new quantization levels according to Equation (4.88) and new quantization bin edges according to Equation (4.89). After going back and forth between these two equations several times, the results converge toward a final optimum quantizer. For the Laplace distribution, the conditional mean criterion results in

$$y_i = \frac{(x_{i-1} + 1)\exp(-x_{i-1}) - (x_i + 1)\exp(-x_i)}{\exp(-x_{i-1}) - \exp(-x_i)}.$$

At each iteration stage (after computing new quantization levels), we also compute the the SQNR. By observing the SQNR at each stage, we can verify that this iterative process is in fact improving the quantizer design at each iteration. For this example, the SQNR is computed according to (see Exercise 4.82).

$$SQNR = \frac{1}{1 - \dfrac{1}{2}\displaystyle\sum_{i=1}^{M} p_i y_i^2}, \quad p_i = \exp(-x_{i-1}) - \exp(-x_i).$$

(Continued)

The MATLAB code we used to implement this process is included below. Figure 4.9 shows the results of running this code for the case of $M = 8$ (16 level, 4-bit quantizer).

```
M=8;
x=[0 2*[1:M-1]/M];                        % Initialize quantization bin edges
iterations=50;
for k=1:iterations
    % Update quantization levels
    x1=x(1:M-1);
    x2=x(2:M);
    y=(x1+1).*exp(-x1)-(x2+1).*exp(-x2);
    y=y./(exp(-x1)-exp(-x2));
    y=[y x(length(x))+1];
    % Calculate SQNR
    p=exp(-x1)-exp(-x2);
    p=[p exp(-x(length(x)))];
    SQNR(k)=1/(1-(y.^2)*p'/2);
    % Update quantization bin edges
    y1=y(1:M-1);
    y2=y(2:M);
    x=[0 (y1+y2)/2];
end
plot(10*log10(SQNR),'o')
```
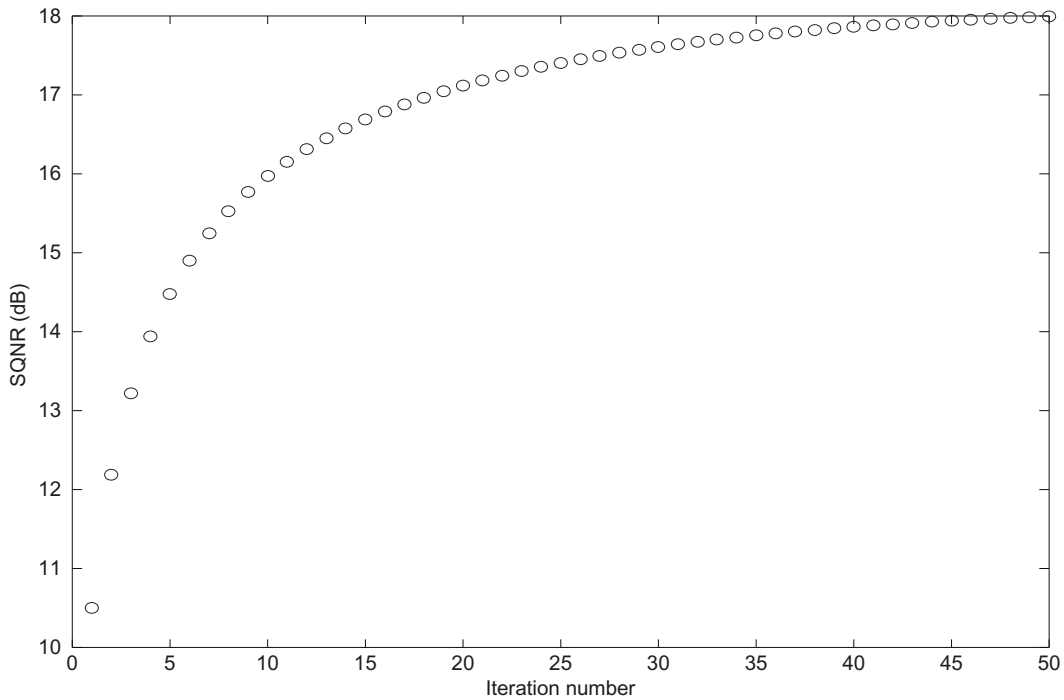

Figure 4.9

SQNR measurements for the iterative quantizer design in Example 4.31. (For color version of this figure, the reader is refered to the web version of this chapter).

∎

4.12 Engineering Application—Entropy and Source Coding

The concept of information is something we hear about frequently. After all, we supposedly live in the "information age" with the internet is often referred to as the "information superhighway." But, what is information? In this section, we give a quantitative definition of information and show how this concept is used in the world of digital communications.

To motivate the forthcoming definition, imagine a situation where we had a genie who could tell us about certain future events. Suppose that this genie told us that on July 15th of next year, the high temperature in the state of Texas would be above $90°F$. Anyone familiar with the weather trends in Texas would know that our genie has actually given us very little information. Since we know that the temperature in July in Texas is above $90°F$ with probability approaching one, the statement made by the genie does not tell us anything new. Next, suppose that the genie tells us that on July 15th of next year, the high temperature in the state of Texas would be below $80°F$. Since this event is improbable, in this case the genie would be giving us a great deal of information.

To define a numerical quantity which we will call information, we note from the previous discussion that

* Information should be a function of various events. Observing (or being told) that some event A occurs (or will occur) provides a certain amount of information, $I(A)$.
* The amount of information associated with an event should be inversely related to the probability of the event. Observing highly probable events provides very little information, while observing very unlikely events gives a large amount of information.

At this point, there are many definitions which could satisfy the two previous bullet items. We include one more observation that will limit the possibilities:

* If it is observed that two events, A and B, have occurred and if those two events are independent, then we expect that the information $I(A \cap B) = I(A) + I(B)$.

Since we observed that information should be a function of the probability of an event, the last bullet item requires us to define information as a function of probability that satisfies

$$I(p_A p_B) = I(p_A) + I(p_B). \tag{4.93}$$

Since a logarithmic function satisfies this property, we obtain the following definition.

> **Definition 4.11:** If some event A occurs with probability p_A, then observing the event A provides an amount of *information* given by
>
> $$I(A) = -\log(p_A). \tag{4.94}$$

The units associated with this measure of information depend on the base of the logarithm used. If base 2 logs are used, then the unit of information is a "bit"; if natural logs are used, the unit of information is the "nat."

Note that with this definition, an event which is sure to happen ($p_A = 1$) provides $I(A) = 0$ bits of information. This makes sense since if we know the event must happen, observing that it does happen provides us with no information.

Next, suppose we conduct some experiment which has a finite number of outcomes. The random variable X will be used to map those outcomes into the set of integers, $0, 1, 2, \ldots, n-1$. How much information do we obtain when we observe the outcome of the experiment? Since information is a function of the probability of each outcome, the amount of information is random and depends on which outcome occurs. We can, however, talk about the average information associated with the observation of the experiment.

Definition 4.12: Suppose a discrete random variable X takes on the values $0, 1, 2, \ldots, n-1$ with probabilities $p_0, p_1, \ldots, p_{n-1}$. The *average information* or *(Shannon) entropy* associated with observing a realization of X is

$$H(X) = \sum_{k=0}^{n-1} \Pr(X=k)I(X=k) = \sum_{k=0}^{n-1} p_k \log\left(\frac{1}{p_k}\right). \tag{4.95}$$

Entropy provides a numerical measure of how much randomness or uncertainty there is in a random variable. In the context of a digital communication system, the random variable might represent the output of a data source. For example, suppose a binary source outputs the letters $X = 0$ and $X = 1$ with probabilities p and $1-p$, respectively. The entropy associated with each letter the source outputs is

$$H(X) = p\log\left(\frac{1}{p}\right) + (1-p)\log\left(\frac{1}{1-p}\right) = \mathcal{H}(p). \tag{4.96}$$

The function $\mathcal{H}(p)$ is known as the binary entropy function and is plotted in Figure 4.10. Note that this function has a maximum value of 1 bit when $p = 1/2$. Consequently, to maximize the information content of a binary source, the source symbols should be equally likely.

Next, suppose a digital source described by a discrete random variable X periodically outputs symbols. Then the information rate of the source is given by $H(X)$ bits/source symbol. Furthermore, suppose we wish to represent the source symbols with binary codewords such that the resulting binary representation of the source outputs uses r bits/symbol. A fundamental result of source coding, which we will not attempt to prove here, is that if we desire the source code to be lossless (that is, we can always recover the original source
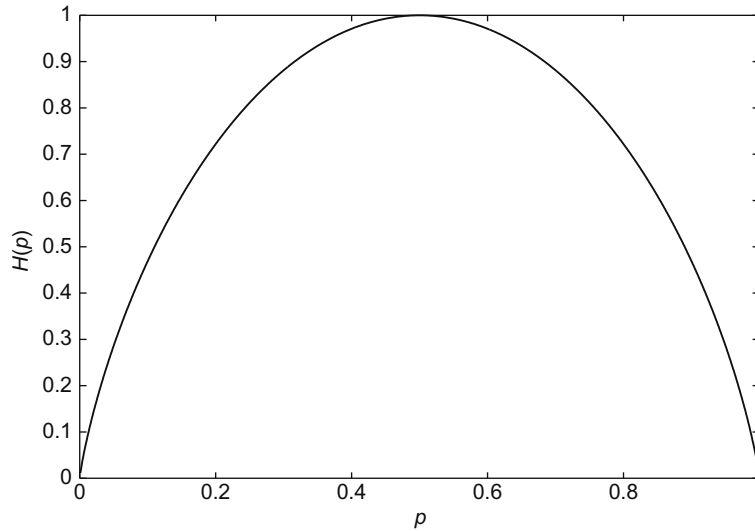
Figure 4.10
The binary entropy function.

symbols from their binary representation), then the source code rate must satisfy $r \geq H(X)$. In other words, the entropy of a source provides a lower bound on the average number of bits that are needed to represent each source output.

■ Example 4.32:

Consider a source that outputs symbols from a four-letter alphabet. That is, suppose $X \in \{a, b, c, d\}$. Let $p_a = 1/2$, $p_b = 1/4$, and $p_c = p_d = 1/8$ be the probability of each of the four source symbols. Given this source distribution, the source has an entropy of

$$H(X) = \frac{1}{2}\log(2) + \frac{1}{4}\log(4) + 2*\frac{1}{8}\log(8) = 1.75 \text{ bits/source symbol}.$$

Table 4.2 shows several different possible binary representations of this source. This first code is the simplest and most obvious representation. Since there are four letters, we can always assign a unique 2-bit codeword to represent each source letter. This results in a code rate of $r_1 = 2$ bits/symbol which is indeed greater than the entropy of the source. The second code uses variable length codewords. The average codeword length is

$$r_2 = \frac{1}{2}*1 + \frac{1}{4}*2 + \frac{1}{4}*3 = 1.75 \text{ bits/symbol}.$$

Therefore, this code produces the most efficient representation of any lossless source coded since the code rate is equal to the source entropy. Note that Code 3 from Table 4.2 produces a code rate of

(Continued)

$$r_3 = \frac{3}{4}*1 + \frac{1}{4}*2 = 1.25 \text{ bits/symbol} ,$$

which is lower than the entropy, but this code is not lossless. This can easily be seen by noting that the source sequences "*d*" and "*a,b*" both lead to the same encoded sequence "01."

Table 4.2: Three possible codes for a four-letter source

	Source Letters			
	a	*b*	*c*	*d*
Code 1	00	01	10	11
Code 2	0	10	110	111
Code 3	0	1	10	01

Exercises

Section 4.1: Expected Values of a Random Variable

4.1 Find the mean of the random variables described by each of the following probability density functions:

(a) $f_W(w) = \dfrac{1}{b-a}$, $a < w < b$;   (c) $f_Y(y) = y(1-y)$, $0 < y < 1$;

(b) $f_X(x) = 3x^2$, $0 < x < 1$;   (d) $f_Z(z) = \dfrac{2}{\pi}\dfrac{1}{1+z^2}$, $z > 0$.

Note: Each of the PDFs described above is zero outside the intervals indicated.

4.2 Find the mean of the random variables described by each of the following probability mass functions:

(a) $P_K(k) = \dfrac{1}{k_0+1}$, $k = 0, 1, 2, ..., k_0$;

(b) $P_M(m) = \dfrac{2m}{m_0(m_0+1)}$, $m = 0, 1, 2, ..., m_0$;

(c) $P_N(n) = \left(\dfrac{1}{2}\right)^n$, $n = 1, 2, 3, ...$;

(d) $P_Q(q) = \left(\dfrac{q-1}{q_0-1}\right)\left(\dfrac{1}{2}\right)^q$, $q = q_0, q_0+1, q_0+2, ...$.

Note: Each of the PMFs described above is zero outside the intervals indicated.

4.3 Find the mean of the random variables described by each of the following cumulative distribution functions:

(a) $F_W(w) = \begin{cases} 0, & w < 0, \\ w/10, & 0 < w < 10 \\ 1, & w > 10; \end{cases}$ (c) $F_Y(y) = \begin{cases} 0, & y < 0, \\ \dfrac{1}{4}y^2, & 0 \le y \le 2, \\ 1, & y > 2; \end{cases}$

(b) $F_X(x) = [1 - \exp(-2x)]u(x)$;   (d) $F_Z(z) = [1 - \exp(-z^2)]u(z)$.

4.4 In each of the following cases, find the value of the parameter a which causes the indicated random variable to have a mean value of 10.

(a) PDF: $f_W(w) = \begin{cases} \dfrac{2}{a}, & \dfrac{a}{2} < w < a, \\ 0, & \text{otherwise} \end{cases}$

(b) PMF: $P_M(m) = \left(\dfrac{1}{2}\right)^{m-a+1}, \quad m = a, a+1, a+2, \ldots$

(c) CDF: $F_X(x) = \begin{cases} 0, & x < 0, \\ ax^2, & 0 \le x \le 10, \\ \dfrac{1-a}{10}x + 2a - 1, & 10 \le x \le 20, \\ 1, & x > 20. \end{cases}$

4.5 Suppose a random variable X has a PDF which is nonzero only on the interval $[0, \infty)$. That is, the random variable cannot take on negative values. Prove that

$$E[X] = \int_0^\infty [1 - F_X(x)]\,dx.$$

Section 4.2: Expected Values of Functions of a Random Variable

4.6 Two players compete against each other in a game of chance where Player A wins with probability 1/3 and Player B wins with probability 2/3. Every time Player A loses he must pay Player B \$1, while every time Player B loses he must pay Player A \$3. Each time the two play the game, the results are independent of any other game. If the two players repeat the game 10 times, what is the expected value of Player A's winnings?

4.7 The current flowing through a 75 Ω resistor is modelled as a Gaussian random variable with parameters, $m = 0A$ and $\sigma = 15\,\text{mA}$. Find the average value of the power consumed in the resistor.

4.8 The received voltage in a 75 Ω antenna of a wireless communication system is modeled as a Rayleigh random variable, $f_V(v) = \dfrac{v}{\sigma^2}\exp\left(-\dfrac{v^2}{2\sigma^2}\right)u(v)$. What does the value of the parameter σ need to be for the received power to be $10\,\mu\text{W}$?

4.9 Suppose X is a Gaussian random variable with a mean of μ and a variance of σ^2 (i.e., $X \sim N(\mu, \sigma^2)$). Find an expression for $E[|X|]$.

4.10 Prove *Jensen's inequality,* which states that for any convex function $g(x)$ and any random variable X,
$$E[g(X)] \ge g(E[X]) .$$

Section 4.3: Moments

4.11 Find an expression for the m th moment of an Erlang random variable whose PDF is given by $f_X(x) = \dfrac{b^n}{(n-1)!} x^{n-1} e^{-bx} u(x)$ for some positive integer n and positive constant b.

4.12 Find an expression for the even moments of a Rayleigh random variable. That is, find $E[Y^{2m}]$ for any positive integer m if the random variable, Y, has a PDF given by
$$f_Y(y) = \frac{y}{\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right) u(y).$$

4.13 Find the first three moments of a geometric random variable whose PMF is $P_N(n) = (1-p)p^n$, $n = 0, 1, 2, \dots$.

4.14 Find the first three moments of a Poisson random variable whose PMF is
$$P_M(m) = \frac{\alpha^m e^{-\alpha}}{m!}, \ m = 0, 1, 2, \dots.$$

4.15 For the Rayleigh random variable described in Exercise 4.12, find a relationship between the n th moment, $E[Y^n]$, and the n th moment of a standard normal random variable.

4.16 Suppose X is a random variable whose n th moment is g_n, $n = 1, 2, 3\dots$ In terms of the g_n, find an expression for the m th moment of the random variable $Y = aX + b$ for constants a and b.

4.17 Suppose X is a random variable whose n th moment is g_n, $n = 1, 2, 3\dots$. In terms of the g_n, find an expression for $E[e^X]$.

Section 4.4: Central Moments

4.18 Calculate the mean value, second moment, and variance of each of the following random variables:

(a) Binomial, $P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \ k = 0, 1, 2, \ldots, n$;

(b) Poisson, $P_X(k) = \dfrac{\alpha^k}{k!} e^{-\alpha}, \ k = 0, 1, 2, \ldots$;

(c) Laplace, $f_X(x) = \dfrac{1}{2b} \exp\left(-\dfrac{|x|}{b}\right)$;

(d) Gamma, $f_X(x) = \dfrac{(x/b)^{c-1} \exp\left(-\dfrac{x}{b}\right)}{b\Gamma(c)} u(x)$;

4.19 For a Gaussian random variable, derive expressions for the coefficient of skewness and the coefficient of kurtosis in terms of the mean and variance, μ and σ^2.

4.20 Prove that all odd central moments of a Gaussian random variable are equal to zero. Furthermore, develop an expression for all even central moments of a Gaussian random variable.

4.21 Show that the variance of a Cauchy random variable is undefined (infinite).

4.22 Let c_n be the nth central moment of a random variable and μ_n be its nth moment. Find a relationship between c_n and $\mu_k, \ k = 0, 1, 2, \ldots, n$.

4.23 Let X be a random variable with $E[X] = 1$ and $\text{var}(X) = 4$. Find the following:

(a) $E[2X - 4]$;

(b) $E[X^2]$;

(c) $E[(2X - 4)^2]$.

4.24 A random variable X has a uniform distribution over the interval $(-a/2, a/2)$ for some positive constant a.

(a) Find the coefficient of skewness for X;

(b) Find the coefficient of kurtosis for X;

(c) Compare the results of (a) and (b) with the same quantities for a standard normal random variable.

4.25 Suppose Θ is a random variable uniformly distributed over the interval $[0, 2\pi)$.

    (a)   Find the PDF of $Y = \sin(\Theta)$.

    (b)   Find the PDF of $Z = \cos(\Theta)$.

    (c)   Find the PDF of $W = \tan(\Theta)$.

4.26 A random variable has a CDF given by

$$F_X(x) = \begin{cases} 0, & x < 0, \\ x^2, & 0 \le x \le 1, \\ 1, & x > 1 \end{cases}$$

    (a)   Find the mean of X;

    (b)   Find the variance of X;

    (c)   Find the coefficient of skewness of X;

    (d)   Find the coefficient of kurtosis of X.

4.27 Find the variance and coefficient of skewness for a geometric random variable whose PMF is $P_N(n) = (1-p)p^n$, $n = 0, 1, 2, \ldots$. *Hint:* You may want to use the results of Exercise 4.13.

4.28 Find the variance and coefficient of skewness for a Poisson random variable whose PMF is $P_M(m) = \dfrac{\alpha^m e^{-\alpha}}{m!}$, $m = 0, 1, 2, \ldots$. *Hint:* You may want to use the results of Exercise 4.14.

Section 4.5: Conditional Expected Values

4.29 Show that the concept of total probability can be extended to expected values. That is, if $\{A_i\}$, $i = 1, 2, 3, \ldots, n$ is a set of mutually exclusive and exhaustive events, then

$$E[X] = \sum_{k=1}^{n} E[X|A_k]Pr(A_k).$$

4.30 An exponential random variable has a PDF given by $f_X(x) = \exp(-x)u(x)$.

    (a)   Find the mean and variance of X.

    (b)   Find the conditional mean and the conditional variance given that $X > 1$.

4.31 A uniform random variable has a PDF given by $f_X(x) = u(x) - u(x-1)$.

(a) Find the mean and variance of X.

(b) Find the conditional mean and the conditional variance given that $\frac{1}{2} < X < \frac{3}{4}$.

4.32 Consider a Gaussian random variable, X, with mean μ and variance σ^2.

(a) Find $E[X|X > \mu + \sigma]$.

(b) Find $E[X||X - \mu| < \sigma]$.

4.33 A professor is studying the performance of various student groups in his class. Let the random variable X represent a students final score in the class and define the following conditioning events:

- F = {student is a freshman}
- So = {student is a sophomore}
- J = {student is a junior}
- Se = {student is a senior}
- M = {student is a male}
- F = {student is a female}

(a) Suppose the average score among males is $E[X|M] = 73.2$ and the average score among females is $E[X|F] = 75.8$. If the overall class average score is $E[X] = 74.6$, what percentage of the class is female?

(b) Suppose the conditional average scores by class are $E[X|F] = 65.8$, $E[X|So] = 71.2$, $E[X|J] = 75.4$, $E[X|Se] = 79.1$. If the overall class average score is $E[X] = 72.4$, what can we say about the percentage of freshmen, sophomore, juniors and seniors in the class?

(c) Given the class statistics in part (b). If it is known that there are 10 freshmen, 12 sophomores, and 9 juniors in the class, how many seniors are in the class?

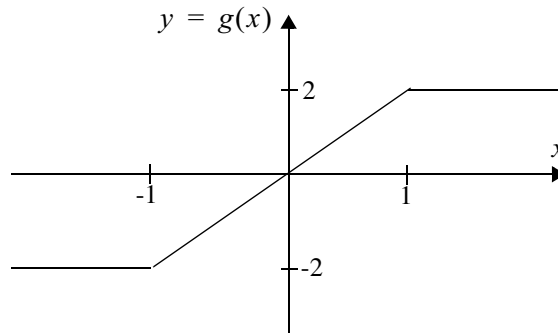## Section 4.6: Transformations of Random Variables

4.34 Suppose X is uniformly distributed over $(-a, a)$, where a is some positive constant. Find the PDF of $Y = X^2$.

4.35 Suppose X is a random variable with an exponential PDF of the form $f_X(x) = 2e^{-2x}u(x)$. A new random variable is created according to the transformation $Y = 1 - X$.

(a) Find the domain for X and Y.

(b) Find $f_Y(y)$.

4.36 Let X be a standard normal random variable (i.e., $X \sim N(0, 1)$). Find the PDF of $Y = |X|$.

4.37 Repeat Exercise 4.36 if the transformation is

$$Y = \begin{cases} X, & X > 0, \\ 0, & X \leq 0. \end{cases}$$

4.38 Suppose a random variable, X, has a Gaussian PDF with zero mean and variance σ_X^2. The random variable is transformed by the device whose input–output relationship is shown in the accompanying figure. Find and sketch the PDF of the transformed random variable, Y.

4.39 Let X be a Gaussian random variable with zero mean and arbitrary variance, σ^2. Given the transformation $Y = X^3$, find $f_Y(y)$.

4.40 A real number between 0 and 100 is randomly selected according to a uniform distribution and rounded off to the nearest integer. For example, 36.5001 is rounded off to 37; $\sqrt{3}$ is rounded off to 2; and 69.49 is rounded off to 69. Define a random variable to be $X = $ (number selected) – (nearest integer).
(a)   What is the domain of this random variable?
(b)   Determine the PDF for X.
(c)   Determine the mean square value of X.

4.41 A Gaussian random variable with zero mean and variance σ_X^2 is applied to a device that has only two possible outputs, 0 or 1. The output 0 occurs when the input is negative, and the output 1 occurs when the input is nonnegative.
(a)   What is the probability mass function of the output?
(b)   Rework the problem when $\mu_X = 1/2$ and $\sigma_X^2 = 1$.

4.42 Let X be a Cauchy random variable whose PDF is given by

$$f_X(x) = \frac{b/\pi}{b^2 + x^2}$$

Find the PDF of $Y = 1/X$.

4.43 Let X be a Chi-square random variable with a PDF given by

$$f_X(x) = \frac{x^{c-1}\exp(-x/2)u(x)}{2^c \Gamma(c)},$$

where $c = n/2$ for any positive integer n. Find the PDF of $Y = \sqrt{X}$.

4.44 Suppose a random variable has some PDF given by $f_X(x)$. Find a function $g(x)$ such that $Y = g(X)$ is a uniform random variable over the interval $(0, 1)$. Next, suppose that X is a uniform random variable. Find a function $g(x)$ such that $Y = g(X)$ has some specified PDF, $f_Y(y)$.

4.45 Suppose X is uniformly distributed over $(0, 1)$. Using the results of Exercise 4.44, find transformations $Y = g(X)$ to produce random variables with the following distributions:
   (a)   Exponential,
   (b)   Rayleigh,
   (c)   Cauchy,
   (d)   Geometric,
   (e)   Poisson.

4.46 Suppose X is a binomial random variable with parameters n and p. That is, the PMF of X is given by

$$P_X(k) = \binom{n}{k}p^k(1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

Find the PMF of a new random variable generated through the transformation, $Y = n - X$.

4.47 Suppose X is a Gaussian random variable with mean μ_X and variance σ_X^2. Suppose we form a new random variable according to $Y = aX + b$ for constants a and b.

   (a)   Prove that Y is also Gaussian for any $a \neq 0$.

   (b)   What values for the constants a and b will lead to the new random variable Y having zero mean and unit variance?

   (c)   What values for the constants a and b will lead to the new random variable Y having a mean of μ_Y and a variance of σ_Y^2?

4.48 The amplitude, A, of a received signal in a wireless communication system is often modelled as a Rayleigh random variable whose PDF is of the form, $f_A(a) = \dfrac{a}{\sigma^2} \exp\left(-\dfrac{a^2}{2\sigma^2}\right) u(a)$.

The (normalized) power of the same signal is given by $P = A^2$. Find the PDF of P.

4.49 Suppose X is an exponential random variable with PDF, $f_X(x) = \exp(-x)u(x)$. Find a transformation, $Y = g(X)$ so that the new random variable Y has a Cauchy PDF given by $f_Y(y) = \dfrac{1/\pi}{1+x^2}$. *Hint:* Use the results of Exercise 4.44.

Section 4.7: Characteristic Functions

4.50 A random variable X has a characteristic function, $\phi_X(\omega)$. Write the characteristic function of $Y = aX + b$ in terms of $\phi_X(\omega)$ and the constants a and b.

4.51 Prove that the characteristic function of any random variable must satisfy the following properties.

(a) $\phi_X^*(\omega) = \phi_X(-\omega)$.

(b) $\phi_X(0) = 1$.

(c) For real ω, $|\phi_X(\omega)| \le 1$.

(d) If the PDF is symmetric about the origin (i.e, an even function), then $\phi_X(\omega)$ is real.

(e) $\phi_X(\omega)$ cannot be purely imaginary.

4.52 Suppose X is an integer-valued random variable. Show that in this case, $\phi_X(2\pi n) = 1$ for any integer, n. Likewise, prove the reverse is also true. That is, show that if $\phi_X(2\pi n) = 1$ for any integer, n, the random variable X must be integer-valued.

4.53 For a Laplace random variable whose PDF is given by $f_X(x) = \dfrac{1}{2b} \exp\left(-\dfrac{|x|}{b}\right)$, find the following:
(a) the characteristic function, $\phi_X(\omega)$,
(b) the Taylor series expansion of $\phi_X(\omega)$,
(c) a general expression for the kth moment of X.

4.54 An Erlang Random variable has a PDF of the form $f_X(x) = \dfrac{b^n}{(n-1)!} x^{n-1} e^{-bx} u(x)$.
(a) Find the characteristic function, $\phi_X(\omega)$.
(b) Find the Taylor series expansion of $\phi_X(\omega)$.
(c) Find a general expression for the kth moment of X.

4.55 A Cauchy random variable has a PDF $f_X(x) = \dfrac{b/\pi}{b^2 + x^2}$.

   (a) Find the characteristic function, $\phi_X(\omega)$.

   (b) Show that the derivatives $\dfrac{d^k}{d\omega^k}(\phi_X(\omega))$ do not exist at $\omega = 0$. What does this mean?

4.56 A Poisson random variable has a PMF of the form $P_X(k) = \dfrac{\alpha^k}{k!}\exp(-\alpha)$, $k = 0, 1, 2, \ldots$.

   (a) Find the characteristic function, $\phi_X(\omega)$.

   (b) Find the first three nonzero terms in the Taylor series expansion of $\ln[\phi_X(\omega)]$.

   (c) Use the results of part (b) to find the mean, variance, and skewness of the Poisson random variable.

4.57 A certain random variable has a characteristic function given by

$$\phi_X(\omega) = \frac{\sin^3(\pi\omega)}{(\pi\omega)^3}.$$

Find the PDF of this random variable.

4.58 Which of the following functions could be the characteristic function of a random variable? See Appendix E, Section 5 for definitions of these functions.

   (a) $\phi_a(\omega) = \text{rect}(\omega)$.

   (b) $\phi_b(\omega) = \text{tri}(\omega)$.

   (c) $\phi_c(\omega) = \text{sinc}(\omega)$.

   (d) $\phi_d(\omega) = \text{sinc}^2(\omega)$.

Section 4.8: Probability-Generating Functions

4.59 Derive a formula expressing the variance of a random variable in terms of its factorial moments.

4.60 Derive a relationship between the kth factorial moment for a nonnegative, integer-valued random variable and the coefficients of the Taylor series expansion of its probability-generating function, $H_X(z)$, about the point $z = 1$.

4.61 For a Poisson random variable whose PMF is given by $P_X(k) = \dfrac{\alpha^k}{k!}e^{-\alpha}$, $k = 0, 1, 2, \ldots$, find the following:
   (a) the probability-generating function, $H_X(z)$,
   (b) the Taylor series expansion of $H_X(z)$ about the point $z = 1$,
   (c) a general expression for the kth factorial moment.

4.62 A certain random variable has a probability-generating function given by

$$H_X(z) = \frac{1}{n}\frac{1 - z^n}{1 - z}.$$

Find the PMF for this random variable.

4.63 Show that for any probability-generating function, $H(z)$, $H(1) = 1$.

4.64 Suppose $H_X(z)$ is the probability-generating function of some random variable X with PMF $P_X(k)$. In terms of $P_X(k)$, find the PMF of the random variable Y if its probability-generating function is given as in each of the following cases.

   (a) $H_Y(z) = H_X\!\left(\dfrac{z}{2}\right)$.

   (b) $H_Y(z) = H_X^2(z)$.

   (c) $H_Y(z) = z^m H_X(z)$.

   (d) $H_Y(z) = H_X(z^{-1})$.

Section 4.9: Moment-Generating Functions

4.65 Derive an expression for the moment-generating function of a Rayleigh random variable whose PDF is

$$f_X(x) = x\exp\!\left(-\frac{x^2}{2}\right)u(x).$$

4.66 Suppose X is a Rician random variable with a PDF given by

$$f_X(x) = x\exp\!\left(-\frac{x^2 + a^2}{2}\right)I_0(ax)u(x).$$

Derive an expression for $E[e^{uX^2}]$. Note that this is not quite the moment-generating function, but it can be used in a similar way.

4.67 A Gaussian mixture is a random variable whose PDF is a linear combination of two Gaussian PDFs,

$$f_X(x) = \frac{p}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) + \frac{1-p}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right).$$

(a)　Find the moment-generating function, $M_X(u)$, of the Gaussian mixture.

(b)　Use the moment-generating function to find the mean of the Gaussian mixture.

4.68 A random variable has a moment-generating function given by

$$M_X(u) = \frac{2}{(1-u)(2-u)}.$$

(a)　Find the PDF of the random variable.

(b)　Use the moment-generating function to find an expression for the kth moment of the random variable.

4.69 Consider a moment-generating function of the general form

$$M_X(u) = \frac{c}{au^2 + bu + c},$$

for constants a, b, and c. Find constraints that the constants a, b, and c must satisfy so that $M_X(u)$ is the MGF of a valid random variable.

4.70 Prove the following properties of moment-generating functions.

(a)　$M_X(0) = 1$.

(b)　For a nonnegative random variable X, and for real $u < 0$, $M_X(u) \le 1$.

4.71 Find the mean of the random variable whose moment-generating function is

$$M_X(u) = \frac{6}{(1-u)^2(2-u)(3-u)}.$$

Section 4.10: Evaluating Tail Probabilities

4.72 Prove that for a random variable X with mean μ_X,

$$\Pr(|X - \mu_X| > \varepsilon) < \frac{E[|X - \mu_X|^n]}{\varepsilon^n},$$

where n is any positive integer.

4.73 Suppose we are interested in finding the left tail probability for a random variable, X. That is, we want to find $\Pr(X \le x_o)$. Rederive an expression for the Chernoff bound for the left tail probability.

4.74 Suppose X is a Poisson random variable with PMF, $P_X(k) = \dfrac{\alpha^k}{k!}\exp(-\alpha)$, $k = 0, 1, 2, \ldots$. Find the Chernoff bound for the tail probability, $\Pr(X \ge n_o)$.

4.75 Suppose X is a Gamma random variable with PDF, $f_X(x) = \dfrac{(x/b)^{c-1}\exp(-x/b)}{b\Gamma(c)}u(x)$. Find the Chernoff bound for the tail probability, $\Pr(X > x_o)$.

4.76 Let X be an Erlang random variable with PDF, $f_X(x) = \dfrac{x^{n-1}e^{-x}u(x)}{(n-1)!}$. Derive a saddle point approximation for the left tail probability, $\Pr(X < x_o)$. Compare your result with the exact value for $0 \le x_o < E[X]$.

4.77 In Exercise 4.66, an expression was derived for $E[e^{uX^2}]$ for a Rician random variable. Use this function to obtain a saddle point approximation for the tail probability of a Rician random variable, $\Pr(X \ge x_o)$.

Hint: For one-sided random variables, $\Pr(X \ge x_o) = \Pr(X^2 \ge x_o^2)$.

4.78 The average number of cars per hour arriving at a toll booth is 57 while the standard deviation is 15.

(a) Use Markov's inequality to find an upper bound on the probability of having more than 200 cars arrive in an hour.

(b) Use Chebyshev's inequality to find an upper bound on the probability of having more than 200 cars arrive in an hour.

4.79 In a certain communication system, bits are encoded into blocks of length 128 bits. Error correction techniques are used such that the block will be decoded correctly as long as no more than 7 of the 128 bits are received in error. We observe that 2% of our blocks are decoded incorrectly. From Markov's inequality, what can we say about the average number of bit errors that occur in each 128 bit block?

4.80 A nonnegative random variable X has moments which are known to be $E[X] = 1$, $E[X^2] = 2$, $E[X^3] = 5$, $E[X^4] = 9$, $E[X^5] = 14$, $E[X^6] = 33$.

(a) Show that for any nonnegative random variable, $\Pr(X \ge x_0) \le \dfrac{E[X^n]}{x_o^n}$.

(b)   Using the result of part (a) and the values of the moments given, find the tightest bound on $\Pr(X \geq 2)$.

(c)   Using the result of part (a) and the values of the moments given, find the tightest bound on $\Pr(X \geq 3)$.

4.81 Since the Q-function represents the tail probability of a Gaussian random variable, we can use the various bounds on tail probabilities to produce bounds on the Q-function.

(a)   Use Markov's inequality to produce an upper bound on the Q-function. *Hint:* a Gaussian random variable has a two-sided PDF, and Markov's inequality requires the random variable to be one-sided. You will need to work with absolute values to resolve this issue.

(b)   Use Chebyshev's inequality to produce an upper bound on the Q-function.

(c)   Plot your results from parts (a) and (b) along with the bound obtained from the Chernoff bound from Example 4.28. In order to determine how tight (or loose) these various bounds are, also include on your plot the exact value of the Q-function.

Section 4.11: Scalar Quantization

4.82 Consider a quantizer that is designed to minimize the mean square quantization error. That is, the quantization levels, y_i, are chosen according to the conditional mean criterion and the bin edges, x_i, are chosen according to the midpoint criterion. Show that the distortion is given by

$$d = E[(X - q(X))^2] = E[X^2] - \sum_{i=1}^{M} p_i y_i^2 \, ,$$

and therefore the SQNR can be written as

$$\text{SQNR} = \frac{E[X^2]}{d} = \left(1 - \frac{\displaystyle\sum_{i=1}^{M} p_i y_i^2}{E[X^2]} \right)^{-1} .$$

4.83 Following the lead of Example 4.30, design an optimum 2-bit quantizer for a signal whose samples follow a triangular PDF,

$$f_X(x) = \begin{cases} 1 - |x|, & |x| \leq 1, \\ 0, & |x| \geq 1. \end{cases}$$

(a)   Find the four quantization levels, $\{y_1, y_2, y_3, y_4\}$, and the three boundary points, $\{x_1, x_2, x_3\}$.

(b)   Find the mean-squared distortion, d.

(c)   Find the signal-to-quantization-noise ratio in dB.

Section 4.12: Entropy and Source Coding

4.84 Suppose a source sends symbols from a three letter alphabet with $X \in \{a, b, c\}$ and $p_a = 1/2$, $p_b = 1/4$, $p_c = 1/4$ are the source symbol probabilities.

(a)   Determine the entropy of this source.

(b)   Give a source code that has an average codeword length that matches the entropy.

4.85 Suppose a random variable, X, has $N = 2^n$ equally likely outcomes. What is the entropy of X in bits?

4.86 Suppose a fair coin is flipped n times and the random variable Y counts the number of times heads occurs. What is the entropy of Y in bits? Compare your answer to that of Exercise 4.85 and explain any difference.

4.87 Consider an N-letter source with probabilities, p_i, $i = 1, 2, 3, \ldots, N$. The entropy of

the source is given by $H(\boldsymbol{p}) = -\sum_{i=1}^{N} p_i \log(p_i)$. Prove that the discrete distribution that

maximizes the entropy is a uniform distribution. *Hint:* You need to perform a constrained optimization since the source probabilities must form a valid probability mass function, and thus $p_1 + p_2 + \ldots + p_N = 1$.

4.88 Consider a geometric random variable, Z, whose PMF is $P_Z(k) = (1-p)p^k$,

$k = 0, 1, 2, \ldots$. Find the entropy of this random variable as a function of p.

Miscellaneous Exercises

4.89 Imagine that you are trapped in a circular room with three doors symmetrically placed around the perimeter. You are told by a mysterious voice that one door leads to the outside after a 2-h trip through a maze. However, the other two doors lead to mazes that terminate back in the room after a 2-h trip at which time you are unable to tell through

which door you exited or entered. What is the average time for escape to the outside? Can you guess the answer ahead of time? If not, can you provide a physical explanation for the answer you calculate?

4.90 A communication system sends data in the form of packets of fixed length. Noise in the communication channel may cause a packet to be received incorrectly. If this happens, then the packet is retransmitted. Let the probability that a packet is received incorrectly be q. Determine the average number of transmissions that are necessary before a packet is received correctly.

4.91 In Exercise 4.90 let the transmission time be T_t seconds for a packet. If the packet was received incorrectly, then a message is sent back to the transmitter that states that the message was received incorrectly. Let the time for sending such a message be T_i. Assume that if the packet is received correctly that we do not send an acknowledgement. What is the average time for a successful transmission?

4.92 Use the characteristic function (or the moment-generating function or the probability-generating function) to show that a Poisson PMF is the limit of a binomial PMF with n approaching infinity and p approaching zero in such a way that $np = \mu = \text{constant}$.

MATLAB Exercises

4.93 Let X be a random variable that is uniformly distributed over the interval $(0, 100)$. Form a new random variable Y by rounding X to the nearest integer. In MATLAB code, this could be represented by `Y=round(X)`. Finally, form the random roundoff error according to $Z = X - Y$.
   (a) Using analytical methods, find the PDF of Z as well as the mean-squared value, $E[Z^2]$.
   (b) Using MATLAB, create a histogram for the probability densities for the random variable Z. Compare with the PDF found analytically in part (a).

4.94 Suppose you have a random variable X with PDF, $f_X(x) = 2x(u(x) - u(x - 1))$ and that this random variable is transformed as $Y = 2 - X$. Calculate $f_Y(y)$. Repeat this problem using MATLAB. Compare the estimate of the PDF from MATLAB with the analytically determined PDF. Note that for this problem there is no function in MATLAB that provides a sequence of data samples that has the PDF specified in this problem for X. Thus you must find an appropriate way to transform a uniform random variable to produce the desired X. The results of Exercise 4.44 will be helpful here.

4.95 Use MATLAB to generate a large number of samples from a Gaussian distribution with mean $\mu = 20$ and variance $\sigma = 4$. *Hint:* the MATLAB command `sigma*randn(1,N)+mu` will create N such numbers with mean `mu` and standard deviation `sigma`. Let $x_1, x_2, ..., x_N$ represent the samples you generated. Compute each of the following "mean" values:

(a) Sample mean, $\hat{\mu}_{sm} = \dfrac{1}{N} \sum_{k=1}^{N} x_k$;

(b) Geometric mean, $\hat{\mu}_{gm} = \left(\prod_{k=1}^{N} x_k \right)^{1/N}$;

(c) Harmonic mean, $\hat{\mu}_{hm} = \left(\dfrac{1}{N} \sum_{k=1}^{N} \dfrac{1}{x_k} \right)^{-1}$;

(d) Quadratic mean (root mean square), $\hat{\mu}_{qm} = \sqrt{\dfrac{1}{N} \sum_{k=1}^{N} x_k^2}$.

Which of these "estimates" give a decent estimate of the true mean?

4.96 Write a MATLAB program to simulate the problem described in Exercise 4.89. Estimate the average time until escape. Do your MATLAB results agree with your analytically determined results in Exercise 4.89?

4.97 Copy a segment of text into MATLAB as a string (you choose the source of the text). Then write a MATLAB program to count the relative frequency of each character (ignore all characters that do not correspond to one of the 26 letters and do not make a distinction between upper and lower case). Using the results of your program, calculate the entropy of a source that outputs the 26 English characters with the probabilities you calculated.

4.98 Suppose a random variable has a PDF given by

$$f_X(x) = \begin{cases} 1+x, & -1 < x < 0, \\ 1-x, & 0 \le x < 1, \\ 0, & |x| \ge 1. \end{cases}$$

Following the procedure laid out in Example 4.31, write a MATLAB program to design an optimum 4-bit (16 level) quantizer for this random variable. Compute the SQNR in decibels of the quantizer you designed. How does this SQNR compare with that obtained in Example 4.31. Can you explain any differences?