Chapter 11

Expected Values for Continuous Random Variables

11.1 Introduction

We now define the expectation of a continuous random variable. In doing so we parallel the discussion of expected values for discrete random variables given in Chapter 6. Based on the probability density function (PDF) description of a continuous random variable, the expected value is defined and its properties explored. The discussion is conceptually much the same as before, only the particular method of evaluating the expected value is different. Hence, we will concentrate on the manipulations required to obtain the expected value.

11.2 Summary

The expected value $E[X]$ for a continuous random variable is motivated from the analogous definition for a discrete random variable in Section 11.3. Its definition is given by (11.3). An analogy with the center of mass of a wedge is also described. For the expected value to exist we must have $E[|X|] < \infty$ or the expected value of the absolute value of the random variable must be finite. The expected values for the common continuous random variables are given in Section 11.4 with a summary given in Table 11.1. The expected value of a function of a continuous random variable can be easily found using (11.10), eliminating the need to find the PDF of the transformed random variable. The expectation is shown to be linear in Example 11.2. For a mixed random variable the expectation is computed using (11.11). The variance is defined by (11.12) with some examples given in Section 11.6. It has the same properties as for a discrete random variable, some of which are given in (11.13), and is a nonlinear operation. The moments of a continuous random variable are defined as $E[X^n]$ and can be found either by using a direct integral evaluation as in Example 11.6 or by using the characteristic function (11.18). The characteristic function is the Fourier transform of the PDF as given by (11.17). Central moments, which are the moments about the mean, are related to the moments by (11.15). The second central moment is just the variance. Although the probability of an event cannot in general be determined from the mean and variance, the Chebyshev inequality of (11.21) provides a formula for bounding the probability. The mean and variance can be estimated using (11.22) and (11.23). Finally, an application of mean estimation to test highly reliable software is described in Section 11.10. It is based on importance sampling, which provides a means of estimating small probabilities with a reasonable number of Monte Carlo trials.

11.3 Determining the Expected Value

The expected value for a discrete random variable X was defined in Chapter 6 to be

$$E[X] = \sum_i x_i p_X[x_i] \tag{11.1}$$

where $p_X[x_i]$ is the probability mass function (PMF) of X and the sum is over all i for which the PMF $p_X[x_i]$ is nonzero. In the case of a continuous random variable, the sample space $\mathcal{S}_X$ is not countable and hence (11.1) can no longer be used. For example, if $X \sim \mathcal{U}(0,1)$, then X can take on any value in the interval $(0,1)$, which consists of an uncountable number of values. We might expect that the average value is $E[X] = 1/2$ since the probability of X being in any *equal length* interval in $(0,1)$ is the same. To verify this conjecture we employ the same strategy used previously, that of approximating a uniform PDF by a uniform *PMF*, using a fine partitioning of the interval $(0,1)$. Letting

$$p_X[x_i] = \frac{1}{M} \qquad x_i = i\Delta x$$

for $i = 1, 2, \ldots, M$ and with $\Delta x = 1/M$, we have from (11.1)

$$\begin{aligned} E[X] &= \sum_{i=1}^{M} x_i p_X[x_i] = \sum_{i=1}^{M} (i\Delta x)\left(\frac{1}{M}\right) \\ &= \sum_{i=1}^{M} \frac{i}{M^2} = \frac{1}{M^2}\sum_{i=1}^{M} i. \end{aligned} \tag{11.2}$$

But $\sum_{i=1}^{M} i = (M/2)(M+1)$ so that

$$E[X] = \frac{\frac{M}{2}(M+1)}{M^2} = \frac{1}{2} + \frac{1}{2M}$$

and as $M \to \infty$ or the partition of $(0,1)$ becomes infinitely fine, we have $E[X] \to 1/2$, as expected. To extend these results to more general PDFs we first note from (11.2) that

$$\begin{aligned} E[X] &= \sum_{i=1}^{M} x_i P[x_i - \Delta x/2 \le X \le x_i + \Delta x/2] \\ &= \sum_{i=1}^{M} x_i \frac{P[x_i - \Delta x/2 \le X \le x_i + \Delta x/2]}{\Delta x} \Delta x. \end{aligned}$$

But

$$\frac{P[x_i - \Delta x/2 \le X \le x_i + \Delta x/2]}{\Delta x} = \frac{1/M}{\Delta x} = 1$$

and as $\Delta x \to 0$, this is the probability per unit length for all small intervals centered about x_i, which is the *PDF* evaluated at $x = x_i$. In this example, $p_X(x_i)$ does not depend on the interval center, which is x_i, so that the PDF is uniform or $p_X(x) = 1$ for $0 < x < 1$. Thus, as $\Delta x \to 0$

$$E[X] \to \sum_{i=1}^{M} x_i p_X(x_i) \Delta x$$

and this becomes the integral

$$E[X] = \int_0^1 x p_X(x) dx$$

where $p_X(x) = 1$ for $0 < x < 1$ and is zero otherwise. To confirm that this integral produces a result consistent with our earlier value of $E[X] = 1/2$, we have

$$\begin{aligned} E[X] &= \int_0^1 x p_X(x) dx \\ &= \int_0^1 x \cdot 1 dx = \frac{1}{2}x^2 \Big|_0^1 = \frac{1}{2}. \end{aligned}$$

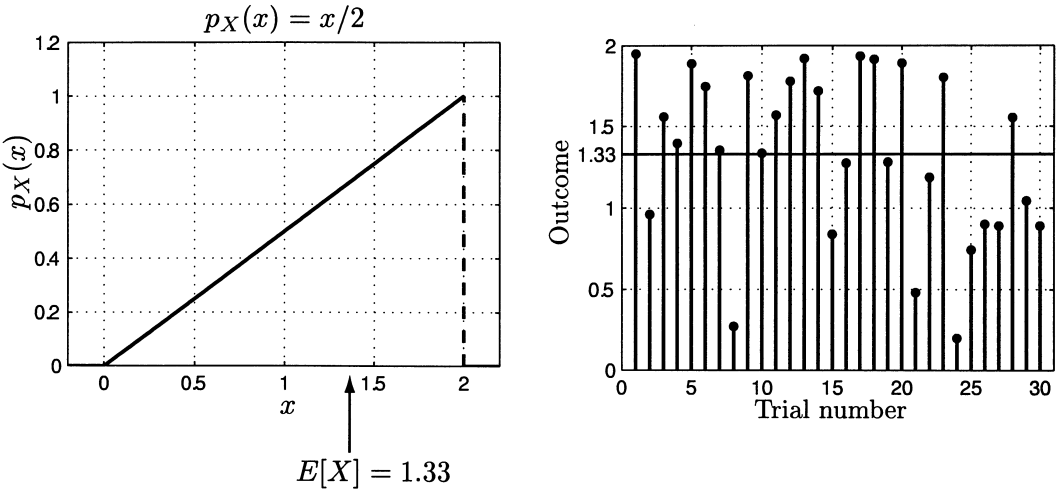
In general, the expected value for a continuous random variable X is defined as

$$E[X] = \int_{-\infty}^{\infty} x p_X(x) dx \tag{11.3}$$

where $p_X(x)$ is the PDF of X. Another example follows.

Example 11.1 – Expected value for random variable with a nonuniform PDF

Consider the computation of the expected value for the PDF shown in Figure 11.1a. From the PDF and some typical outcomes shown in Figure 11.1b the expected value



(a) PDF

(b) Typical outcomes and expected value of 1.33

Figure 11.1: Example of nonuniform PDF and its mean.

should be between 1 and 2. Using (11.3) we have

$$\begin{aligned} E[X] &= \int_0^2 x\left(\frac{1}{2}x\right)dx \\ &= \frac{1}{6}x^3\Big|_0^2 = \frac{4}{3} \end{aligned}$$

which appears to be reasonable.

$\diamond$

As an analogy to the expected value we can revisit our Jarlsberg cheese first described in Section 10.3, and which is shown in Figure 11.2. The integral

$$\mathrm{CM} = \int_0^2 x m(x) dx \tag{11.4}$$

is the *center of mass*, assuming that the total mass or $\int_0^2 m(x)dx$, is one. Here, $m(x)$ is the linear mass density or mass per unit length. The center of mass is the point at which one could balance the cheese on the point of a pencil. Recall that the linear mass density is $m(x) = x/2$ for which CM $= 4/3$ from Example 11.1. To show that CM is the balance point we first note that $\int_0^2 m(x)dx = 1$ so that we can

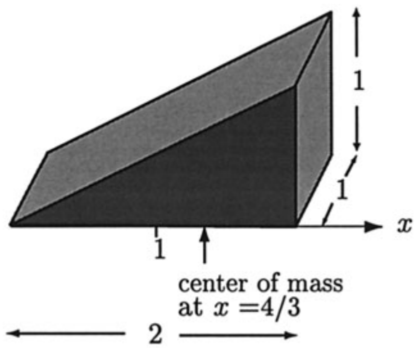


Figure 11.2: Center of mass (CM) analogy to average value.

write (11.4) as

$$\begin{aligned}
\int_0^2 xm(x)dx - \text{CM} &= 0 \\
\int_0^2 xm(x)dx - \text{CM}\int_0^2 m(x)dx &= 0 \\
\underbrace{\int_0^2}_{\text{sum}} \underbrace{(x-\text{CM})}_{\text{moment arm}} \underbrace{m(x)dx}_{\text{mass}} &= 0.
\end{aligned}$$

Since the "sum" of the mass times moment arms is zero, the cheese is balanced at $x = \text{CM} = 4/3$.

By the same argument the expected value can also be found by solving

$$\int_{-\infty}^{\infty} (x - E[X])p_X(x)dx = 0 \tag{11.5}$$

for $E[X]$. If, however, the PDF is symmetric about some point $x = a$, which is to say that $p_X(a+u) = p_X(a-u)$ for $-\infty < u < \infty$, then (see Problem 11.2)

$$\int_{-\infty}^{\infty} (x-a)p_X(x)dx = 0 \tag{11.6}$$

and therefore $E[X] = a$. Such was the case for $X \sim \mathcal{U}(0,1)$, whose PDF is symmetric about $a = 1/2$. Another example is the Gaussian PDF which is symmetric about $a = \mu$ as seen in Figures 10.8a and 10.8c. Hence, $E[X] = \mu$ for a Gaussian random variable (see also the next section for a direct derivation). In summary, if the PDF is symmetric about a point, then that point is $E[X]$. However, the PDF need not be symmetric about any point as in Example 11.1.

Not all PDFs have expected values.

Before computing the expected value of a random variable using (11.3) we must make sure that it exists (see similar discussion in Section 6.4 for discrete random variables). Not all integrals of the form $\int_{-\infty}^{\infty} x p_X(x)dx$ exist, even if $\int_{-\infty}^{\infty} p_X(x)dx = 1$. For example, if

$$p_X(x) = \begin{cases} \dfrac{1}{2x^{3/2}} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

then

$$\int_1^{\infty} \frac{1}{2} x^{-3/2} dx = -\frac{1}{\sqrt{x}}\bigg|_1^{\infty} = 1$$

but

$$\int_1^{\infty} x \frac{1}{2} x^{-3/2} dx = \sqrt{x}\big|_1^{\infty} \to \infty.$$

A more subtle and somewhat surprising example is the Cauchy PDF. Recall that it is given by

$$p_X(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty.$$

Since the PDF is symmetric about $x = 0$, we would expect that $E[X] = 0$. However, if we are careful about our definition of expected value by correctly interpreting the region of integration in a limiting sense, we would have

$$E[X] = \lim_{L\to-\infty} \int_L^0 x p_X(x)dx + \lim_{U\to\infty} \int_0^U x p_X(x)dx.$$

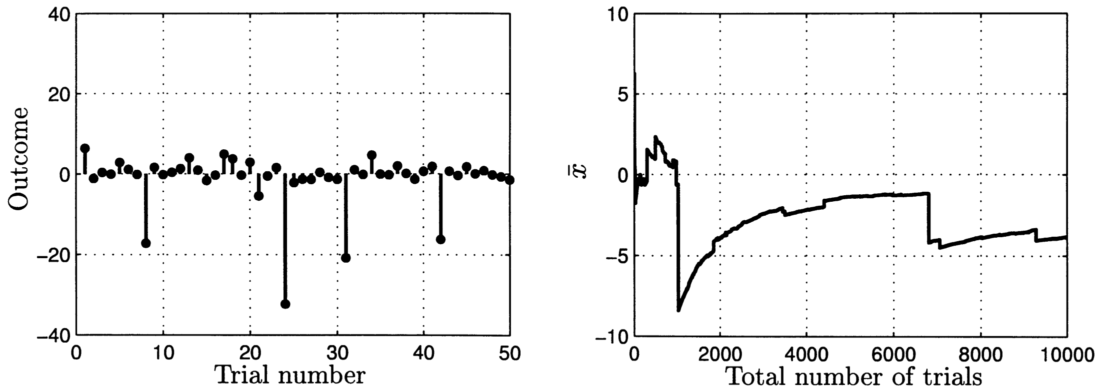
But for a Cauchy PDF

$$\begin{aligned} E[X] &= \lim_{L\to-\infty} \int_L^0 x \frac{1}{\pi(1+x^2)} dx + \lim_{U\to\infty} \int_0^U x \frac{1}{\pi(1+x^2)} dx \\ &= \lim_{L\to-\infty} \frac{1}{2\pi} \ln(1+x^2)\bigg|_L^0 + \lim_{U\to\infty} \frac{1}{2\pi} \ln(1+x^2)\bigg|_0^U \\ &= \lim_{L\to-\infty} -\frac{1}{2\pi} \ln(1+L^2) + \lim_{U\to\infty} \frac{1}{2\pi} \ln(1+U^2) \\ &= -\infty + \infty = ? \end{aligned}$$

Hence, if the limits are taken independently, then the result is indeterminate. To make the expected value useful in practice the independent choice of limits (and not $L = U$) is necessary. The indeterminancy can be avoided, however, if we require "absolute convergence" or

$$\int_{-\infty}^{\infty} |x| p_X(x)dx < \infty.$$

Hence, $E[X]$ is defined to exist if $E[|X|] < \infty$. This surprising result can be "verified" by a computer simulation, the results of which are shown in Figure 11.3. In



(a) First 50 outcomes

(b) Sample mean

Figure 11.3: Illustration of nonexistence of Cauchy PDF mean.

Figure 11.3a the first 50 outcomes of a total of 10,000 are shown. Because of the slow decay of the "tails" of the PDF or since the PDF decays only as $1/x^2$, very large outcomes are possible. As seen in Figure 11.3b the sample mean does not converge to zero as might be expected because of these infrequent but very large outcomes (see also Problem 12.41). See also Problem 11.3 on the simulation used in this example and Problems 11.4 and 11.9 on how to make the sample mean converge by truncating the PDF.

Finally, as for discrete random variables the expected value is the best guess of the outcome of the random variable. By "best" we mean that the use of $b = E[X]$ as our estimator. This estimator minimizes the mean square error, which is defined as $\text{mse} = E[(X - b)^2]$ (see Problem 11.5).

11.4 Expected Values for Important PDFs

We now determine the expected values for the important PDFs described in Chapter 10. Of course, the Cauchy PDF is omitted.

11.4.1 Uniform

If $X \sim \mathcal{U}(a, b)$, then it is easy to prove that $E[X] = (a + b)/2$ or the mean lies at the midpoint of the interval (see Problem 11.8).

11.4.2 Exponential

If $X \sim \exp(\lambda)$, then

$$
\begin{aligned}
E[X] &= \int_0^\infty x\lambda \exp(-\lambda x)dx \\
&= \left[-x\exp(-\lambda x) - \frac{1}{\lambda}\exp(-\lambda x)\right]\Bigg|_0^\infty = \frac{1}{\lambda}.
\end{aligned} \tag{11.7}
$$

Recall that the exponential PDF spreads out as λ decreases (see Figure 10.6) and hence so does the mean.

11.4.3 Gaussian or Normal

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then since the PDF is symmetric about the point $x = \mu$, we know that $E[X] = \mu$. A direct appeal to the definition of the expected value yields

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx \\
&= \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx \\
&\quad + \int_{-\infty}^{\infty} \mu \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx.
\end{aligned}
$$

Letting $u = x - \mu$ in the first integral we have

$$
E[X] = \underbrace{\int_{-\infty}^{\infty} u \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}u^2\right] du}_{0} + \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx}_{=1} = \mu.
$$

The first integral is zero since the integrand is an odd function ($g(-u) = -g(u)$, see also Problem 11.6) and the second integral is one since it is the total area under the Gaussian PDF.

11.4.4 Laplacian

The Laplacian PDF is given by

$$
p_X(x) = \frac{1}{\sqrt{2\sigma^2}} \exp\left[-\sqrt{\frac{2}{\sigma^2}}|x|\right] \qquad -\infty < x < \infty \tag{11.8}
$$

and since it is symmetric about $x = 0$ (and the expected value *exists* – needed to avoid the situation of the Cauchy PDF), we must have $E[X] = 0$.

11.4.5 Gamma

If $X \sim \Gamma(\alpha, \lambda)$, then from (10.10)

$$E[X] = \int_0^\infty x \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x) dx.$$

To evaluate this integral we attempt to modify the integrand so that it becomes the PDF of a $\Gamma(\alpha', \lambda')$ random variable. Then, we can immediately equate the integral to one. Using this strategy

$$\begin{aligned}
E[X] &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha \exp(-\lambda x) dx \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \\
&= \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} \qquad \text{(integrand is } \Gamma(\alpha+1, \lambda) \text{ PDF)} \\
&= \frac{\alpha \Gamma(\alpha)}{\lambda \Gamma(\alpha)} \qquad \text{(using Property 10.3)} \\
&= \frac{\alpha}{\lambda}.
\end{aligned}$$

11.4.6 Rayleigh

It can be shown that $E[X] = \sqrt{(\pi\sigma^2)/2}$ (see Problem 11.16).

The reader should indicate on Figures 10.6–10.10, 10.12, and 10.13 where the mean occurs.

11.5 Expected Value for a Function of a Random Variable

If $Y = g(X)$, where X is a continuous random variable, then assuming that Y is also a continuous random variable with PDF $p_Y(y)$, we have by the definition of expected value of a continuous random variable

$$E[Y] = \int_{-\infty}^\infty y p_Y(y) dy. \tag{11.9}$$

Even if Y is a mixed random variable, its expected value is still given by (11.9), although in this case $p_Y(y)$ will contain impulses. Such would be the case if for example, $Y = \max(0, X)$ for X taking on values $-\infty < x < \infty$ (see Section 10.8). As in the case of a discrete random variable, it is not necessary to use (11.9) directly, which requires us to first determine $p_Y(y)$ from $p_X(x)$. Instead, we can use for $Y = g(X)$ the formula

$$E[g(X)] = \int_{-\infty}^\infty g(x) p_X(x) dx. \tag{11.10}$$

A partial proof of this formula is given in Appendix 11A. Some examples of its use follows.

Example 11.2 – Expectation of linear (affine) function
If $Y = aX + b$, then since $g(x) = ax + b$, we have from (11.10) that

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty} (ax+b)p_X(x)dx \\ &= a\int_{-\infty}^{\infty} xp_X(x)dx + b\int_{-\infty}^{\infty} p_X(x)dx \\ &= aE[X] + b \end{aligned}$$

or equivalently

$$E[aX+b] = aE[X] + b.$$

It indicates how to easily change the expectation or mean of a random variable. For example, to increase the mean value by b just replace X by $X + b$. More generally, it is easily shown that

$$E[a_1g_1(X) + a_2g_2(X)] = a_1E[g_1(X)] + a_2E[g_2(X)].$$

This says that the expectation operator is *linear*.

◇

Example 11.3 – Power of $\mathcal{N}(0,1)$ random variable
If $X \sim \mathcal{N}(0,1)$ and $Y = X^2$, consider $E[Y] = E[X^2]$. The quantity $E[X^2]$ is the average squared value of X and can be interpreted physically as a *power*. If X is a voltage across a 1 ohm resistor, then X^2 is the power and therefore $E[X^2]$ is the *average power*. Now according to (11.10)

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \\ &= 2\int_{0}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \qquad \text{(integrand is symmetric about } x = 0\text{).} \end{aligned}$$

To evaluate this integral we use integration by parts ($\int UdV = UV - \int VdU$, see also Problem 11.7) with $U = x$, $dU = dx$, $dV = (1/\sqrt{2\pi})x\exp[-(1/2)x^2]dx$ and therefore $V = -(1/\sqrt{2\pi})\exp[-(1/2)x^2]$ to yield

$$\begin{aligned} E[X^2] &= 2\left[-x\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)\Big|_0^{\infty} - \int_0^{\infty} -\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)dx\right] \\ &= 0 + 1 = 1. \end{aligned}$$

The first term is zero since

$$\lim_{x\to\infty} x\exp\left(-\frac{1}{2}x^2\right) = \lim_{x\to\infty}\frac{x}{\exp\left(\frac{1}{2}x^2\right)} = \lim_{x\to\infty}\frac{1}{x\exp\left(\frac{1}{2}x^2\right)} = 0$$

using L'Hospital's rule and the second term is evaluated using

$$\int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx = \frac{1}{2} \qquad \text{(Why?)}.$$

$\diamondsuit$

Example 11.4 – Expected value of indicator random variable

An indicator function indicates whether a point is in a given set. For example, if the set is $A = [3, 4]$, then the indicator function is defined as

$$I_A(x) = \begin{cases} 1 & 3 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

and is shown in Figure 11.4. The subscript on I refers to the set of interest. The

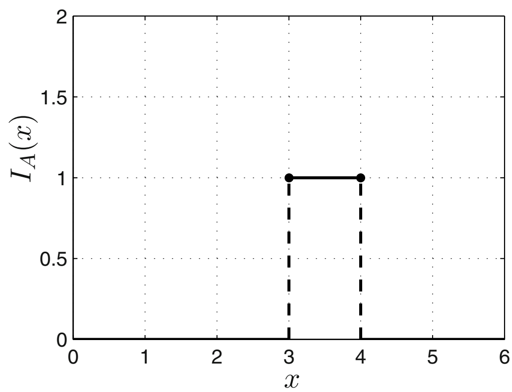


Figure 11.4: Example of indicator function for set $A = [3, 4]$.

indicator function may be thought of as a generalization of the unit step function since if $u(x) = 1$ for $x \ge 0$ and zero otherwise, we have that

$$I_{[0,\infty)}(x) = u(x).$$

Now if X is a random variable, then $I_A(X)$ is a transformed random variable that takes on values 1 and 0, depending upon whether the outcome of the experiment lies within the set A or not, respectively. (It is actually a Bernoulli random variable.) On the average, however, it has a value *between* 0 and 1, which from (11.10) is

$$\begin{aligned} E[I_A(X)] &= \int_{-\infty}^{\infty} I_A(x) p_X(x) dx \\ &= \int_{\{x:x\in A\}} 1 \cdot p_X(x) dx \qquad \text{(definition)} \\ &= \int_{\{x:x\in A\}} p_X(x) dx \\ &= P[A]. \end{aligned}$$

Therefore, *the expected value of the indicator random variable is the probability of the set or event.* As an example of its utility, consider the estimation of $P[3 \leq X \leq 4]$. But this is just $E[I_A(X)]$ when $I_A(x)$ is given in Figure 11.4. To estimate the expected value of a transformed random variable we first generate the outcomes of X, say $x_1, x_2, \ldots, x_M$, then transform each one to the new random variable producing for $i = 1, 2, \ldots, M$

$$I_A(x_i) = \begin{cases} 1 & 3 \leq x_i \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

and finally compute the sample mean for our estimate using

$$\widehat{E[I_A(X)]} = \frac{1}{M} \sum_{i=1}^{M} I_A(x_i).$$

However, since $P[A] = E[I_A(X)]$, we have as our estimate of the probability

$$\widehat{P[A]} = \frac{1}{M} \sum_{i=1}^{M} I_A(x_i).$$

But this is just what we have been using all along, since $\sum_{i=1}^{M} I_A(x_i)$ counts all the outcomes for which $3 \leq x \leq 4$. Thus, *the indicator function provides a means to connect the expected value with the probability.* This is a very useful for later theoretical work in probability.

◇

Lastly, if the random variable is a mixed one with PDF

$$p_X(x) = p_c(x) + \sum_{i=1}^{\infty} p_i \delta(x - x_i)$$

where $p_c(x)$ is the continuous part of the PDF, then the expected value becomes

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \left(p_c(x) + \sum_{i=1}^{\infty} p_i \delta(x - x_i) \right) dx \\ &= \int_{-\infty}^{\infty} x p_c(x) dx + \int_{-\infty}^{\infty} x \sum_{i=1}^{\infty} p_i \delta(x - x_i) dx \\ &= \int_{-\infty}^{\infty} x p_c(x) dx + \sum_{i=1}^{\infty} p_i \int_{-\infty}^{\infty} x \delta(x - x_i) dx \\ &= \int_{-\infty}^{\infty} x p_c(x) dx + \sum_{i=1}^{\infty} x_i p_i \qquad (11.11) \end{aligned}$$

since $\int_{-\infty}^{\infty} g(x)\delta(x - x_i)dx = g(x_i)$ for $g(x)$ a function continuous at $x = x_i$. This is known as the *sifting* property of a Dirac delta function (see Appendix D). A

	Values	PDF	$E[X]$	$\text{var}(X)$	$\phi_X(\omega)$
Uniform	$a<x<b$	$\frac{1}{b-a}$	$\frac{1}{2}(a+b)$	$\frac{(b-a)^2}{12}$	$\frac{\exp(j\omega b)-\exp(j\omega a)}{j\omega(b-a)}$
Exponential	$x \geq 0$	$\lambda\exp(-\lambda x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-j\omega}$
Gaussian	$-\infty<x<\infty$	$\frac{\exp[-(1/(2\sigma^2))(x-\mu)^2]}{\sqrt{2\pi\sigma^2}}$	μ	σ^2	$\exp[j\omega\mu-\sigma^2\omega^2/2]$
Laplacian	$-\infty<x<\infty$	$\frac{1}{\sqrt{2\sigma^2}}\exp(-\sqrt{2/\sigma^2}\lvert x\rvert)$	0	σ^2	$\frac{2/\sigma^2}{\omega^2+2/\sigma^2}$
Gamma	$x \geq 0$	$\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\lambda x)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\frac{1}{(1-j\omega/\lambda)^\alpha}$
Rayleigh	$x \geq 0$	$\frac{x}{\sigma^2}\exp[-x^2/(2\sigma^2)]$	$\sqrt{\frac{\pi\sigma^2}{2}}$	$(2-\pi/2)\sigma^2$	[Johnson et al 1994]

Table 11.1: Properties of continuous random variables.

summary of the means for the important PDFs is given in Table 11.1. Lastly, note that the expected value of a random variable can also be determined from the CDF as shown in Problem 11.28.

11.6 Variance and Moments of a Continuous Random Variable

The variance of a continuous random variable, as for a discrete random variable, measures the average squared deviation from the mean. It is defined as $\text{var}(X) = E[(X - E[X])^2]$ (exactly the same as for a discrete random variable). To evaluate the variance we use (11.10) to yield

$$\text{var}(X) = \int_{-\infty}^{\infty} (x - E[X])^2 p_X(x)dx. \tag{11.12}$$

As an example, consider a $\mathcal{N}(\mu, \sigma^2)$ random variable. In Figure 10.9 we saw that the width of the PDF increases as σ^2 increases. This is because the parameter σ^2 is actually the variance, as we now show. Using (11.12) and the definition of a Gaussian PDF

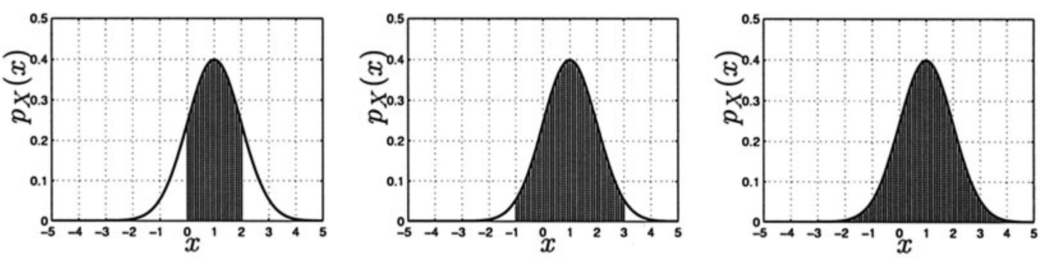
$$\begin{aligned}
\text{var}(X) &= \int_{-\infty}^{\infty} (x - E[X])^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx \\
&= \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx \qquad \text{(recall that } E[X] = \mu).
\end{aligned}$$

Letting $u = (x - \mu)/\sigma$ produces (recall that $\sigma = \sqrt{\sigma^2} > 0$)

$$\begin{aligned}
\operatorname{var}(X) &= \int_{-\infty}^{\infty} \sigma^2 u^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}u^2\right] \sigma du \\
&= \sigma^2 \underbrace{\int_{-\infty}^{\infty} u^2 \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}u^2\right] du}_{=1} \qquad \text{(see Example 11.3)} \\
&= \sigma^2.
\end{aligned}$$

Hence, we now know that a $\mathcal{N}(\mu, \sigma^2)$ random variable has a mean of μ and a variance of σ^2.

It is common to refer to the square-root of the variance as the *standard deviation.* For a $\mathcal{N}(\mu, \sigma^2)$ random variable it is given by σ. The standard deviation indicates how closely outcomes tend to cluster about the mean. (See Problem 11.29 for an alternative interpretation.) Again if the random variable is $\mathcal{N}(\mu, \sigma^2)$, then 68.2% of the outcomes will be within the interval $[\mu - \sigma, \mu + \sigma]$, 95.5% will be within $[\mu - 2\sigma, \mu + 2\sigma]$, and 99.8% will be within $[\mu - 3\sigma, \mu + 3\sigma]$. This is illustrated in Figure 11.5. Of course, other PDFs will have concentrations that are different for $E[X] \pm k\sqrt{\operatorname{var}(X)}$. Another example follows.



(a) 68.2% for 1 standard deviation

(b) 95.5% for 2 standard deviations

(c) 99.8% for 3 standard deviations

Figure 11.5: Percentage of outcomes of $\mathcal{N}(1, 1)$ random variable that are within $k = 1, 2$, and 3 standard deviations from the mean. Shaded regions denote area within interval $\mu - k\sigma \leq x \leq \mu + k\sigma$.

Example 11.5 – Variance of a uniform random variable

If $X \sim \mathcal{U}(a, b)$, then

$$\begin{aligned}
\operatorname{var}(X) &= \int_{-\infty}^{\infty} (x - E[X])^2 p_X(x) dx \\
&= \int_a^b \left(x - \frac{1}{2}(a+b)\right)^2 \frac{1}{b-a} dx
\end{aligned}$$

and letting $u = x - (a+b)/2$, we have

$$\begin{aligned}
\operatorname{var}(X) &= \frac{1}{b-a}\int_{-(b-a)/2}^{(b-a)/2} u^2 du \\
&= \frac{1}{b-a}\frac{1}{3}u^3\bigg|_{-(b-a)/2}^{(b-a)/2} \\
&= \frac{(b-a)^2}{12}.
\end{aligned}$$

◊

A summary of the variances for the important PDFs is given in Table 11.1. The variance of a continuous random variable enjoys the same properties as for a discrete random variable. Recall that an alternate form for variance computation is

$$\operatorname{var}(X) = E[X^2] - E^2[X]$$

and if c is a constant then

$$\begin{aligned}
\operatorname{var}(c) &= 0 \\
\operatorname{var}(X+c) &= \operatorname{var}(X) \\
\operatorname{var}(cX) &= c^2\operatorname{var}(X).
\end{aligned} \tag{11.13}$$

Also, the variance is a nonlinear type of operation in that

$$\operatorname{var}(g_1(X) + g_2(X)) \neq \operatorname{var}(g_1(X)) + \operatorname{var}(g_2(X))$$

(see Problem 11.32). Recall from the discussions for a discrete random variable that $E[X]$ and $E[X^2]$ are termed the *first and second moments*, respectively. In general, $E[X^n]$ is termed the nth moment and it is defined to exist if $E[|X|^n] < \infty$. If it is known that $E[X^s]$ exists, then it can be shown that $E[X^r]$ exists for $r < s$ (see Problem 6.23). This also says that *if* $E[X^r]$ *is known not to exist, then* $E[X^s]$ *cannot exist for* $s > r$. An example is the Cauchy PDF for which we saw that $E[X]$ does not exist and therefore all the higher order moments do not exist. In particular, the Cauchy PDF does not have a second-order moment and therefore its variance does not exist. We next give an example of the computation of all the moments of a PDF.

Example 11.6 – Moments of an exponential random variable

Using (11.10) we have for $X \sim \exp(\lambda)$ that

$$E[X^n] = \int_0^\infty x^n \lambda \exp(-\lambda x) dx.$$

To evalute this we first show how the nth moment can be written recursively in terms of the $(n-1)$st moment. Since we know that $E[X] = 1/\lambda$, we can then determine all the moments using the recursion. We can begin to evaluate the integral using integration by parts. This will yield the recursive formula for the moments. Letting $U = x^n$ and $dV = \lambda \exp(-\lambda x)dx$ so that $dU = nx^{n-1}dx$ and $V = -\exp(-\lambda x)$, we have

$$\begin{aligned} E[X^n] &= -x^n \exp(-\lambda x)|_0^\infty - \int_0^\infty -\exp(-\lambda x) n x^{n-1} dx \\ &= 0 + n \int_0^\infty x^{n-1} \exp(-\lambda x) dx \\ &= \frac{n}{\lambda} \int_0^\infty x^{n-1} \lambda \exp(-\lambda x) dx \\ &= \frac{n}{\lambda} E[X^{n-1}]. \end{aligned}$$

Hence, the nth moment can be written in term of the $(n-1)$st moment. Since we know that $E[X] = 1/\lambda$, we have upon using the recursion that

$$\begin{aligned} E[X^2] &= \frac{2}{\lambda} E[X] = \frac{2}{\lambda}\frac{1}{\lambda} = \frac{2}{\lambda^2} \\ E[X^3] &= \frac{3}{\lambda} E[X^2] = \frac{3}{\lambda}\frac{2}{\lambda^2} = \frac{3 \cdot 2}{\lambda^3} \\ &\text{etc.} \end{aligned}$$

and in general

$$E[X^n] = \frac{n!}{\lambda^n}. \tag{11.14}$$

The variance can be found to be $\mathrm{var}(X) = 1/\lambda^2$ using these results.

◇

In the next section we will see how to use characteristic functions to simplify the complicated integration process required for moment evaluation.

Lastly, it is sometimes important to be able to compute moments *about some point*. For example, the variance is the second moment about the point $E[X]$. In general, the nth *central moment* about the point $E[X]$ is defined as $E[(X - E[X])^n]$. The relationship between the moments and the central moments is of interest. For $n = 2$ the central moment is related to the moments by the usual formula $E[(X - E[X])^2] = E[X^2] - E^2[X]$. More generally, this relationship is found using the binomial theorem as follows.

$$\begin{aligned} E[(X - E[X])^n] &= E\left[\sum_{k=0}^{n} \binom{n}{k} X^k (-E[X])^{n-k}\right] \\ &= \sum_{k=0}^{n} \binom{n}{k} E[X^k](-E[X])^{n-k} \quad \text{(linearity of expectation operator)} \end{aligned}$$

or finally we have that

$$E[(X - E[X])^n] = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} (E[X])^{n-k} E[X^k]. \tag{11.15}$$

11.7 Characteristic Functions

As first introduced for discrete random variables, the characteristic function is a valuable tool for the calculation of moments. It is defined as

$$\phi_X(\omega) = E[\exp(j\omega X)] \tag{11.16}$$

and always exists (even though the moments of a PDF may not). For a continuous random variable it is evaluated using (11.10) for the real and imaginary parts of $E[\exp(j\omega X)]$, which are $E[\cos(\omega X)]$ and $E[\sin(\omega X)]$. This results in

$$\phi_X(\omega) = \int_{-\infty}^{\infty} \exp(j\omega x) p_X(x) dx$$

or in more familiar form as

$$\phi_X(\omega) = \int_{-\infty}^{\infty} p_X(x) \exp(j\omega x) dx. \tag{11.17}$$

The characteristic function is seen to be the Fourier transform of the PDF, although with a $+j$ in the definition as opposed to the more common $-j$. Once the characteristic function has been found, the moments are given as

$$E[X^n] = \frac{1}{j^n} \left. \frac{d^n \phi_X(\omega)}{d\omega^n} \right|_{\omega=0}. \tag{11.18}$$

An example follows.

Example 11.7 – Moments of the exponential PDF

Using the definition of the exponential PDF (see (10.5)) we have

$$\begin{aligned} \phi_X(\omega) &= \int_0^{\infty} \lambda \exp(-\lambda x) \exp(j\omega x) dx \\ &= \int_0^{\infty} \lambda \exp[-(\lambda - j\omega)x] dx \\ &= \lambda \left. \frac{\exp[-(\lambda - j\omega)x]}{-(\lambda - j\omega)} \right|_0^{\infty} \\ &= -\frac{\lambda}{\lambda - j\omega} \left(\exp[-(\lambda - j\omega)\infty] - 1 \right). \end{aligned}$$

But $\exp[-(\lambda - j\omega)x] \to 0$ as $x \to \infty$ since $\lambda > 0$ and hence we have

$$\phi_X(\omega) = \frac{\lambda}{\lambda - j\omega}. \tag{11.19}$$

To find the moments using (11.18) we need to differentiate the characteristic function n times. Proceeding to do so

$$\begin{aligned}
\frac{d\phi_X(\omega)}{d\omega} &= \frac{d}{d\omega}\lambda(\lambda - j\omega)^{-1} \\
&= \lambda(-1)(\lambda - j\omega)^{-2}(-j) \\
\frac{d^2\phi_X(\omega)}{d\omega^2} &= \lambda(-1)(-2)(\lambda - j\omega)^{-3}(-j)^2 \\
&\vdots \\
\frac{d^n\phi_X(\omega)}{d\omega^n} &= \lambda(-1)(-2)\dots(-n)(\lambda - j\omega)^{-n-1}(-j)^n \\
&= \lambda j^n n!(\lambda - j\omega)^{-n-1}
\end{aligned}$$

and therefore

$$\begin{aligned}
E[X^n] &= \frac{1}{j^n}\left.\frac{d^n\phi_X(\omega)}{d\omega^n}\right|_{\omega=0} \\
&= \lambda n!\left.(\lambda - j\omega)^{-n-1}\right|_{\omega=0} \\
&= \frac{n!}{\lambda^n}
\end{aligned}$$

which agrees with our earlier results (see (11.14)).

◊

⚠ Moment formula only valid if moments exist

Just because a PDF has a characteristic function, and all do, does not mean that (11.18) can be applied. For example, the Cauchy PDF has the characteristic function (see Problem 11.40)

$$\phi_X(\omega) = \exp(-|\omega|)$$

(although the derivative does not exist at $\omega = 0$). However, as we have already seen, the mean does not exist and hence all higher order moments also do not exist. Thus, no moments exist at all for the Cauchy PDF.

⚠

The characteristic function has nearly the same properties as for a discrete random variable, namely

1. The characteristic function always exists.

2. The PDF can be recovered from the characteristic function by the inverse Fourier transform, which in this case is

$$p_X(x) = \int_{-\infty}^{\infty} \phi_X(\omega) \exp(-j\omega x) \frac{d\omega}{2\pi}. \tag{11.20}$$

3. Convergence of a sequence of characteristic functions $\phi_X^{(n)}(\omega)$ for $n = 1, 2, \ldots$ to a given characteristic function $\phi(\omega)$ guarantees that the corresponding sequence of PDFs $p_X^{(n)}(x)$ for $n = 1, 2, \ldots$ converges to $p(x)$, where from (11.20)

$$p(x) = \int_{-\infty}^{\infty} \phi(\omega) \exp(-j\omega x) \frac{d\omega}{2\pi}.$$

(See Problem 11.42 for an example.) This property is also essential for proving the central limit theorem described in Chapter 15.

A slight difference from the characteristic function of a discrete random variable is that now $\phi_X(\omega)$ is *not* periodic in ω. It does, however, have the usual properties of the continuous-time Fourier transform [Jackson 1991]. A summary of the characteristic functions for the important PDFs is given in Table 11.1.

11.8 Probability, Moments, and the Chebyshev Inequality

The mean and variance of a random variable indicate the average value and variability of the outcomes of a repeated experiment. As such, they summarize important information about the PDF. However, they are not sufficient to determine probabilities of events. For example, the PDFs

$$\begin{aligned} p_X(x) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) && \text{(Gaussian)} \\ p_X(x) &= \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right) && \text{(Laplacian)} \end{aligned}$$

both have $E[X] = 0$ (due to symmetry about $x = 0$) and $\text{var}(X) = 1$. Yet, the probability of a given interval can be very different. Although the relationship between the mean and variance, and the probability of an event is not a direct one, we can still obtain some information about the probabilities based on the mean and variance. In particular, it is possible to *bound* the probability or to be able to assert that

$$P[|X - E[X]| > \gamma] \leq B$$

where B is a number less than one. This is especially useful if we only wish to make sure the probability is below a certain value, without explicitly having to find the probability. For example, if the probability of a speech signal of mean 0 and variance 1 exceeding a given magnitude γ (see Section 10.10) is to be no more than 1%, then we would be satisfied if we could determine a γ so that

$$P[|X - E[X]| > \gamma] \leq 0.01.$$

We now show that the probability for the event $|X - E[X]| > \gamma$ can be bounded if we know the mean and variance. Computation of the probability is not required and therefore *the PDF does not need to be known.* Estimating the mean and variance is much easier than the entire PDF (see Section 11.9). The inequality to be developed is called the *Chebyshev inequality.* Using the definition of the variance we have

$$\begin{aligned}
\operatorname{var}(X) &= \int_{-\infty}^{\infty} (x - E[X])^2 p_X(x)dx \\
&= \int_{\{x:|x-E[X]|>\gamma\}} (x - E[X])^2 p_X(x)dx + \int_{\{x:|x-E[X]|\leq\gamma\}} (x - E[X])^2 p_X(x)dx \\
&\geq \int_{\{x:|x-E[X]|>\gamma\}} (x - E[X])^2 p_X(x)dx \qquad \text{(omitted integral is nonnegative)} \\
&\geq \int_{\{x:|x-E[X]|>\gamma\}} \gamma^2 p_X(x)dx \qquad \text{(since for each } x,\ |x - E[X]| > \gamma) \\
&= \gamma^2 \int_{\{x:|x-E[X]|>\gamma\}} p_X(x)dx \\
&= \gamma^2 P[|X - E[X]| > \gamma]
\end{aligned}$$

so that we have the Chebyshev inequality

$$P[|X - E[X]| > \gamma] \leq \frac{\operatorname{var}(X)}{\gamma^2}. \tag{11.21}$$

Hence, the probability that a random variable deviates from its mean by more than γ (in either direction) is less than or equal to $\operatorname{var}(X)/\gamma^2$. This agrees with our intuition in that the probability of an outcome departing from the mean must become smaller as the width of the PDF decreases or equivalently as the variance decreases. An example follows.

Example 11.8 – Bounds for different PDFs

Assuming $E[X] = 0$ and $\operatorname{var}(X) = 1$, we have from (11.21)

$$P[|X| > \gamma] \leq \frac{1}{\gamma^2}.$$

If $\gamma = 3$, then we have that $P[|X| > 3] \leq 1/9 \approx 0.11$. This is a rather "loose" bound in that if $X \sim \mathcal{N}(0,1)$, then the actual value of this probability is $P[|X| >$

3] = 2Q(3) = 0.0027. Hence, the actual probability is indeed less than or equal to the bound of 0.11, but quite a bit less. In the case of a Laplacian random variable with mean 0 and variance 1, the bound is the same but the actual value is now

$$\begin{aligned}
P[|X| > 3] &= \int_{-\infty}^{-3} \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right) dx + \int_{3}^{\infty} \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right) dx \\
&= 2\int_{3}^{\infty} \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}x\right) dx \qquad \text{(PDF is symmetric about } x = 0) \\
&= -\exp\left(-\sqrt{2}x\right)\Big|_{3}^{\infty} \\
&= \exp\left(-3\sqrt{2}\right) = 0.0144.
\end{aligned}$$

Once again the bound is seen to be correct but provides a gross overestimation of the probability. A graph of the Chebyshev bound as well as the actual probabilities of $P[|X| > \gamma]$ versus γ is shown in Figure 11.6. The reader may also wish to consider

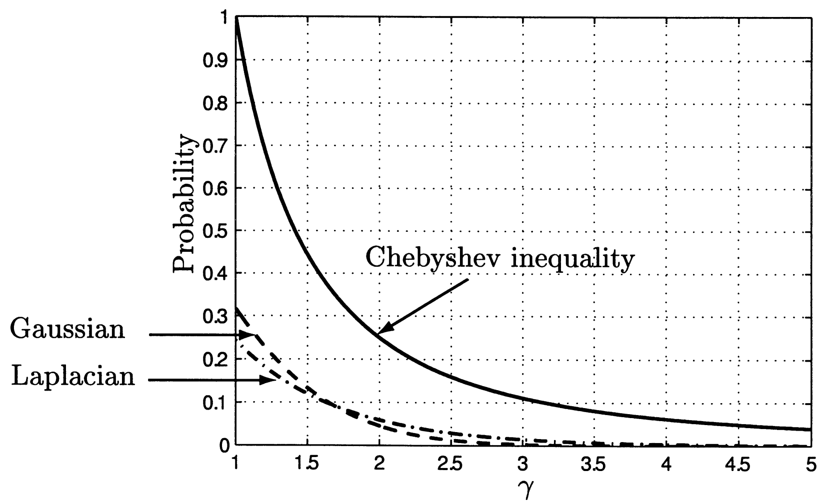


Figure 11.6: Probabilities $P[|X| > \gamma]$ for Gaussian and Laplacian random variables with zero mean and unity variance compared to Chebyshev inequality.

what would happen if we used the Chebyshev inequality to bound $P[|X| > 0.5]$ if $X \sim \mathcal{N}(0, 1)$.

◇

11.9 Estimating the Mean and Variance

The mean and variance of a continuous random variable are estimated in exactly the same way as for a discrete random variable (see Section 6.8). Assuming that we have the M outcomes $\{x_1, x_2, \ldots, x_M\}$ of a random variable X the mean estimate is

$$\widehat{E[X]} = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{11.22}$$

and the variance estimate is

$$\begin{aligned}
\widehat{\operatorname{var}(X)} &= \widehat{E[X^2]} - \left(\widehat{E[X]}\right)^2 \\
&= \frac{1}{M}\sum_{i=1}^{M} x_i^2 - \left(\frac{1}{M}\sum_{i=1}^{M} x_i\right)^2 .
\end{aligned} \tag{11.23}$$

An example of the use of (11.22) was given in Example 2.6 for a $\mathcal{N}(0,1)$ random variable. Some practice with the estimation of the mean and variance is provided in Problem 11.46.

11.10 Real-World Example – Critical Software Testing Using Importance Sampling

Computer software is a critical component of nearly every device used today. The failure of such software can range from being an annoyance, as in the outage of a cellular telephone, to being a catastrophe, as in the breakdown of the control system for a nuclear power plant. Testing of software is of course a prerequisite for reliable operation, but some events, although potentially catastrophic, will (hopefully) occur only rarely. Therefore, the question naturally arises as to how to test software that is designed to only fail once every 10^7 hours (≈ 1400 years). In other words, although a theoretical analysis might predict such a low failure rate, there is no way to test the software by running it and waiting for a failure. A technique that is often used in other fields to test a system is to "stress" the system to induce more frequent failures, say by a factor of 10^5, then estimate the probability of failure per hour, and finally readjust the probability for the increased stress factor. An analogous approach can be used for highly reliable software if we can induce a higher failure rate and then readjust our failure probability estimate by the increased factor. A proposed method to do this is to stress the software to cause the probability of a failure to increase [Hecht and Hecht 2000]. Conceivably we could do this by inputting data to the software that is suspected to cause failures but at a much higher rate than is normally encountered in practice. This means that if T is the time to failure, then we would like to replace the PDF of T so that $P[T > \gamma]$ increases by a significant factor. Then, after estimating this probability by exercising the software we could adjust the estimate back to the original unstressed value. This probabilitic approach is called *importance sampling* [Rubinstein 1981].

As an example of the use of importance sampling, assume that X is a continuous random variable and we wish to estimate $P[X > \gamma]$. As usual, we could generate realizations of X, count the number that exceed γ, and then divide this by the total number of realizations. But what if the probability sought is 10^{-7}? Then we would need about 10^9 realizations to do this. As a specific example, suppose that $X \sim \mathcal{N}(0,1)$, although in practice we would not have knowledge of the PDF at our disposal, and that we wish to estimate $P[X > 5]$ based on observed realization values. The true probability is known to be $Q(5) = 2.86 \times 10^{-7}$. The importance sampling approach first recognizes that the desired probability is given by

$$\mathcal{I} = \int_5^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx$$

and is equivalent to

$$\mathcal{I} = \int_5^\infty \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)}{p_{X'}(x)} p_{X'}(x) dx$$

where $p_{X'}(x)$ is a more suitable PDF. By "more suitable" we mean that its probability of $X' > 5$ is larger, and therefore, generating realizations based on it will produce more occurrences of the desired event. One possibility is $X' \sim \exp(1)$ or $p_{X'}(x) = \exp(-x)u(x)$ for which $P[X > 5] = \exp(-5) = 0.0067$. Using this new PDF we have the desired probability

$$\mathcal{I} = \int_5^\infty \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)}{\exp(-x)} \exp(-x) dx$$

or using the indicator function, this can be written as

$$\mathcal{I} = \int_0^\infty \underbrace{I_{(5,\infty)}(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2 + x\right)}_{g(x)} p_{X'}(x) dx.$$

Now the desired probability can be interpreted as $E[g(X')]$, where $X' \sim \exp(1)$. To estimate it using a Monte Carlo computer simulation we first generate M realizations of an $\exp(1)$ random variable and then use as our estimate

$$\begin{aligned} \hat{\mathcal{I}} &= \frac{1}{M} \sum_{i=1}^M g(x_i) \\ &= \frac{1}{M} \sum_{i=1}^M I_{(5,\infty)}(x_i) \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2 + x_i\right)}_{\substack{\text{weight with value} \ll 1 \\ \text{for } x_i \gg 5}}. \end{aligned} \tag{11.24}$$

The advantage of the importance sampling approach is that the realizations whose values exceed 5, which are the ones contributing to the sum, are *much more probable*. In fact, as we have noted $P[X' > 5] = 0.0067$ and therefore with $N = 10,000$

realizations we would expect about 67 realizations to contribute to the sum. Contrast this with a $\mathcal{N}(0,1)$ random variable for which we would expect $NQ(5) = (10^4)(2.86 \times 10^{-7}) \approx 0$ realizations to exceed 5. The new PDF $p_{X'}$ is called the *importance function* and hence the generation of realizations from this PDF, which is also called *sampling from the PDF*, is termed *importance sampling.* As seen from (11.24), its success requires a weighting factor that downweights the counting of threshold exceedances.

In software testing the portions of software that are critical to the operation of the overall system would be exercised more often than in normal operation, thus effectively replacing the operational PDF or p_X by the importance function PDF or $p_{X'}$. The ratio of these two would be needed as seen in (11.24) to adjust the weight for each incidence of a failure. *This ratio would also need to be estimated in practice.* In this way a good estimate of the probability of failure could be obtained by exercising the software a reasonable number of times with different inputs. Otherwise, the critical software might not exhibit a failure a sufficient number of times to estimate its probability.

As a numerical example, if $X' \sim \exp(1)$, we can generate realizations using the inverse probability transformation method (see Section 10.9) via $X' = -\ln(1-U)$, where $U \sim \mathcal{U}(0,1)$. A MATLAB computer program to estimate $\mathcal{I}$ is given below.

```
rand('state',0) % sets random number generator to
                % initial value
M=10000;gamma=5;% change M for different estimates
u=rand(M,1);    % generates M U(0,1) realizations
x=-log(1-u);    % generates M exp(1) realizations
k=0;
for i=1:M       % computes estimate of P[X>gamma]
   if x(i)>gamma
      k=k+1;
      y(k,1)=(1/sqrt(2*pi))*exp(-0.5*x(i)^2+x(i)); % computes weights
                                                   % for estimate
    end
end
Qest=sum(y)/M  % final estimate of P[X>gamma]
```

The results are summarized in Table 11.2 for different values of M, along with the true value of $Q(5)$. Also shown are the number of times γ was exceeded. Without the use of importance sampling the number of exceedances would be expected to be $MQ(5) \approx 0$ in all cases.

M	Estimated $P[X > 5]$	True $P[X > 5]$	Exceedances
10^3	1.11×10^{-7}	2.86×10^{-7}	4
10^4	2.96×10^{-7}	2.86×10^{-7}	66
10^5	2.51×10^{-7}	2.86×10^{-7}	630
10^6	2.87×10^{-7}	2.86×10^{-7}	6751

Table 11.2: Importance sampling approach to estimation of small probabilities.

References

Hecht, M., H. Hecht, "Use of Importance Sampling and Related Techniques to Measure Very High Reliability Software," 2000 IEEE Aerospace Conference Proc., Vol. 4, pp. 533–546.

Jackson, L.B., *Signals, Systems, and Transforms*, Addison-Wesley, Reading, MA, 1991.

Johnson, N.L., S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions, Vol. 1*, see pp. 456–459 for moments, John Wiley & Sons, New York, 1994.

Parzen, E., *Modern Probability Theory and its Applications*, John Wiley & Sons, New York, 1960.

Rubinstein, R.Y., *Simulation and the Monte Carlo Method*, John Wiley & Sons, New York, 1981.

Problems

11.1 (⌣̈) **(f)** The block shown in Figure 11.7 has a mass of 1 kg. Find the center of mass for the block, which is the point along the x-axis where the block could be balanced (in practice the point would also be situated in the depth direction at 1/2).

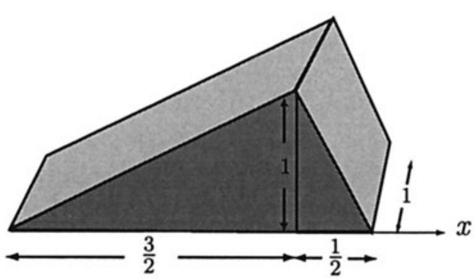


Figure 11.7: Block for Problem 11.1.

11.2 (t) Prove that if the PDF is symmetric about a point $x = a$, which is to say that it satisfies $p_X(a+u) = p_X(a-u)$ for all $-\infty < u < \infty$, then the mean will be a. Hint: Write the integral $\int_{-\infty}^{\infty} x p_X(x)dx$ as $\int_{-\infty}^{a} x p_X(x)dx + \int_{a}^{\infty} x p_X(x)dx$ and then let $u = x - a$ in the first integral and $u = a - x$ in the second integral.

11.3 (c) Generate and plot 50 realizations of a Cauchy random variable. Do so by using the inverse probability integral transformation method. You should be able to show that $X = \tan(\pi(U - 1/2))$, where $U \sim \mathcal{U}(0,1)$, will generate the Cauchy realizations.

11.4 (c) In this problem we show via a computer simulation that the mean of a *truncated Cauchy* PDF exists and is equal to zero. A truncated Cauchy random variable is one in which the realizations of a Cauchy PDF are set to $x = x_{\max}$ if $x > x_{\max}$ and $x = -x_{\max}$ if $x < -x_{\max}$. Generate realizations of this random variable with $x_{\max} = 50$ and plot the sample mean versus the number of realizations. What does the sample mean converge to?

11.5 (t) Prove that the best prediction of the outcome of a continuous random variable is its mean. Best is to be interpreted as the value that minimizes the mean square error $\text{mse}(b) = E[(X - b)^2]$.

11.6 (t) An even function is one for which $g(-x) = g(x)$, as for example $\cos(x)$. An odd function is one for which $g(-x) = -g(x)$, as for example $\sin(x)$. First prove that $\int_{-\infty}^{\infty} g(x)dx = 2\int_{0}^{\infty} g(x)dx$ if $g(x)$ is even and that $\int_{-\infty}^{\infty} g(x)dx = 0$ if $g(x)$ is odd. Next, prove that if $p_X(x)$ is even, then $E[X] = 0$ and also that $\int_{0}^{\infty} p_X(x)dx = 1/2$.

11.7 (f) Many integrals encountered in probability can be evaluated using *integration by parts.* This useful formula is

$$\int U dV = UV - \int V dU$$

where U and V are functions of x. As an example, if we wish to evaluate $\int x \exp(ax)dx$, we let $U = x$ and $dV = \exp(ax)dx$. The function U is easily differentiated to yield $dU = dx$ and the differential dV is easily integrated to yield $V = (1/a)\exp(ax)$. Continue the derivation to determine the integral of the function $x \exp(ax)$.

11.8 (f) Find the mean for a uniform PDF. Do so by first using the definition and then rederive it using the results of Problem 11.2.

11.9 (t) Consider a continuous random variable that can take on values $x_{\min} \leq x \leq x_{\max}$. Prove that the expected value of this random variable must satisfy $x_{\min} \leq E[X] \leq x_{\max}$. Hint: Use the fact that if $M_1 \leq g(x) \leq M_2$, then $M_1 a \leq \int_a^b g(x)dx \leq M_2 b$.

11.10 (⌣) **(w)** The signal-to-noise ratio (SNR) of a random variable quantifies the accuracy of a measurement of a physical quantity. It is defined as $E^2[X]/\text{var}(X)$ and is seen to increase as the mean, which represents the true value, increases and also as the variance, which represents the power of the measurement error, i.e., $X - E[X]$, decreases. For example, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\text{SNR} = \mu^2/\sigma^2$. Determine the SNR if the measurement is $X = A + U$, where A is the true value and U is the measurement error with $U \sim \mathcal{U}(-1/2, 1/2)$. For an SNR of 1000 what should A be?

11.11 (⌣) **(w)** A toaster oven has a failure time that has an exponential PDF. If the mean time to failure is 1000 hours, what is the probability that it will not fail for at least 2000 hours?

11.12 (w) A bus always arrives late. On the average it is 10 minutes late. If the lateness time is an exponential random variable, determine the probability that the bus will be less than 1 minute late.

11.13 (w) In Section 1.3 we described the amount of time an office worker spends on the phone in a 10-minute period. From Figure 1.5 what is the average amount of time he spends on the phone?

11.14 (⌣) **(f)** Determine the mean of a χ^2_N PDF. See Chapter 10 for the definition of this PDF.

11.15 (f) Determine the mean of an Erlang PDF using the definition of expected value. See Chapter 10 for the definition of this PDF.

11.16 (f) Determine the mean of a Rayleigh PDF using the definition of expected value. See Chapter 10 for the definition of this PDF.

11.17 (w) The *mode* of a PDF is the value of x for which the PDF is maximum. It can be thought of as the most probable value of a random variable (actually most probable small interval). Find the mode for a Gaussian PDF and a Rayleigh PDF. How do they relate to the mean?

11.18 (f) Indicate on the PDFs shown in Figures 10.7–10.13 the location of the mean value.

11.19 (⌣) **(w)** A dart is thrown at a circular dartboard. If the distance from the bullseye is a Rayleigh random variable with a mean value of 10, what is the probability that the dart will land within 1 unit of the bullseye?

11.20 (f) For the random variables described in Problems 2.8–2.11 what are the means? Note that the uniform random variable is $\mathcal{U}(0, 1)$ and the Gaussian random variable is $\mathcal{N}(0, 1)$.

11.21 (⌣) **(w)** In Problem 2.14 it was asked whether the mean of $\sqrt{U}$, where $U \sim \mathcal{U}(0,1)$, is equal to $\sqrt{\text{mean of } U}$. There we relied on a computer simulation to answer the question. Now prove or disprove this equivalence.

11.22 (⌣) **(w)** A sinusoidal oscillator outputs a waveform $s(t) = \cos(2\pi F_0 t + \phi)$, where t indicates time, F_0 is the frequency in Hz, and ϕ is a phase angle that varies depending upon when the oscillator is turned on. If the phase is modeled as a random variable with $\phi \sim \mathcal{U}(0, 2\pi)$, determine the average value of $s(t)$ for a given $t = t_0$. Also, determine the average power, which is defined as $E[s^2(t)]$ for a given $t = t_0$. Does this make sense? Explain your results.

11.23 (f) Determine $E[X^2]$ for a $\mathcal{N}(\mu, \sigma^2)$ random variable.

11.24 (f) Determine $E[(2X+1)^2]$ for a $\mathcal{N}(\mu, \sigma^2)$ random variable.

11.25 (f) Determine the mean and variance for the indicator random variable $I_A(X)$ as a function of $P[A]$.

11.26 (⌣) **(w)** A half-wave rectifier passes a zero or positive voltage undisturbed but blocks any negative voltage by outputting a zero voltage. If a noise sample with PDF $\mathcal{N}(0, \sigma^2)$ is input to a half-wave rectifier, what is the average power at the output? Explain your result.

11.27 (⌣) **(w)** A mixed PDF is given as

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

What is $E[X^2]$ for this PDF? Can this PDF be interpreted physically? Hint: See Problem 11.26.

11.28 (t) In this problem we derive an alternative formula for the mean of a nonnegative random variable. A more general formula exists for random variables that can take on both positive and negative values [Parzen 1960]. If X can only take on values $x \geq 0$, then

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

First verify that this formula holds for $X \sim \exp(\lambda)$. To prove that the formula is true in general, we use integration by parts (see Problem 11.7) as follows.

$$\begin{aligned} E[X] &= \int_0^\infty (1 - F_X(x))\, dx \\ &= \int_0^\infty \underbrace{\int_x^\infty p_X(t)dt}_{U} \underbrace{dx}_{dV}. \end{aligned}$$

Finish the proof by using $\lim_{x\to\infty} x\int_x^\infty p_X(t)dt = 0$, which must be true if the expected value exists (see if this holds for $X \sim \exp(\lambda)$).

11.29 (t) The standard deviation σ of a Gaussian PDF can be interpreted as the distance from the mean at which the PDF curve goes through an inflection point. This means that at the points $x = \mu \pm \sigma$ the second derivative of $p_X(x)$ is zero. The curve then changes from being concave (shaped like a $\cap$) to being convex (shaped like a $\cup$). Show that the second derivative is zero at these points.

11.30 (⌣̈) (w) The office worker described in Section 1.3 will spend an average of 7 minutes on the phone in any 10-minute interval. However, the probability that he will spend *exactly* 7 minutes on the phone is zero since the length of this interval is zero. If we wish to assert that he will spend between T_{min} and T_{max} minutes on the phone 95% of the time, what should T_{min} and T_{max} be? Hint: There are multiple solutions – choose any convenient one.

11.31 (w) A group of students is found to weigh an average of 150 lbs. with a standard deviation of 30 lbs. If we assume a normal population (in the probabilistic sense!) of students, what is the range of weights for which approximately 99.8% of the students will lie? Hint: There are multiple solutions – choose any convenient one.

11.32 (w) Provide a counterexample to disprove that $\text{var}(g_1(X) + g_2(X)) = \text{var}(g_1(X)) + \text{var}(g_2(X))$ in general.

11.33 (w) The SNR of a random variable was defined in Problem 11.10. Determine the SNR for exponential random variable and explain why it doesn't increase as the mean increases. Compare your results to a $\mathcal{N}(\mu, \sigma^2)$ random variable and explain.

11.34 (f) Verify the mean and variance for a Laplacian random variable given in Table 11.1.

11.35 (⌣̈) (f) Determine $E[X^3]$ if $X \sim \mathcal{N}(\mu, \sigma^2)$. Next find the third *central* moment.

11.36 (f) An example of a Gaussian mixture PDF is

$$p_X(x) = \frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(x-1)^2\right] + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(x+1)^2\right].$$

Determine its mean and variance.

11.37 (t) Prove that if a PDF is symmetric about $x = 0$, then all its odd-order moments are zero.

11.38 (⌣) (f) For a Laplacian PDF with $\sigma^2 = 2$ determine all the moments. Hint: Let

$$\frac{1}{\omega^2+1} = \frac{1}{2j}\left(\frac{1}{\omega - j} - \frac{1}{\omega + j}\right).$$

11.39 (f) If $X \sim \mathcal{N}(0, \sigma^2)$, determine $E[X^2]$ using the characteristic function approach.

11.40 (t) To determine the characteristic function of a Cauchy random variable we must evaluate the integral

$$\int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)} \exp(j\omega x)dx.$$

A result from Fourier transform theory called the *duality theorem* asserts that the Fourier transform and inverse Fourier transform are nearly the same if we replace x by ω and ω by x. As an example, for a Laplacian PDF with $\sigma^2 = 2$ we have from Table 11.1 that

$$\int_{-\infty}^{\infty} p_X(x) \exp(j\omega x)dx = \int_{-\infty}^{\infty} \frac{1}{2} \exp(-|x|) \exp(j\omega x)dx = \frac{1}{1+\omega^2}.$$

The inverse Fourier transform relationship is therefore

$$\int_{-\infty}^{\infty} \frac{1}{1+\omega^2} \exp(-j\omega x)\frac{d\omega}{2\pi} = \frac{1}{2}\exp(-|x|).$$

Use the latter integral, with appropriate modifications (note that x and ω are just variables which we can redefine as desired), to obtain the characteristic function of a Cauchy random variable.

11.41 (f) If the characteristic function of a random variable is

$$\phi_X(\omega) = \left(\frac{\sin\omega}{\omega}\right)^2$$

find the PDF. Hint: Recall that when we convolve two functions together the Fourier transform of the new function is the product of the individual Fourier transforms. Also, see Table 11.1 for the characteristic function of a $\mathcal{U}(-1,1)$ random variable.

11.42 (⌣) (w) If $X^{(n)} \sim \mathcal{N}(\mu, 1/n)$, determine the PDF of the limiting random variable X as $n \to \infty$. Use characteristic functions to do so.

11.43 (f) Find the mean and variance of a χ^2_N random variable using the characteristic function.

11.44 (⌣̈) (f) The probability that a random variable deviates from its mean by an amount γ in either direction is to be less than or equal to $1/2$. What should γ be?

11.45 (f) Determine the probability that $|X| > \gamma$ if $X \sim \mathcal{U}[-a, a]$. Next compare these results to the Chebyshev bound for $a = 2$.

11.46 (⌣̈) (c) Estimate the mean and variance of a Rayleigh random variable with $\sigma^2 = 1$ using a computer simulation. Compare your estimated results to the theoretical values.

11.47 (c) Use the importance sampling method described in Section 11.10 to determine $Q(7)$. If you were to generate M realizations of a $\mathcal{N}(0, 1)$ random variable and count the number that exceed $\gamma = 7$ as is usually done to estimate a right-tail probability, what would M have to be (in terms of order of magnitude)?

Appendix 11A

Partial Proof of Expected Value of Function of Continuous Random Variable

For simplicity assume that $Y = g(X)$ is a continuous random variable with PDF $p_Y(y)$ (having no impulses). Also, assume that $y = g(x)$ is monotonically increasing so that it has a single solution to the equation $y = g(x)$ for all y as shown in Figure 11A.1. Then

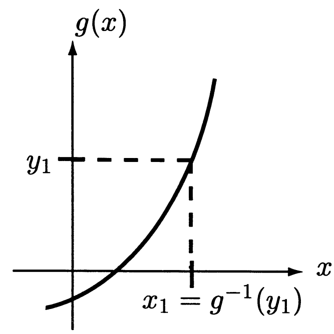


Figure 11A.1: Monotonically increasing function used to derive $E[g(X)]$.

$$
\begin{aligned}
E[Y] &= \int_{-\infty}^{\infty} y p_Y(y) dy \\
&= \int_{-\infty}^{\infty} y p_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| dy \qquad \text{(from (10.30)}.
\end{aligned}
$$

Next change variables from y to x using $x = g^{-1}(y)$. Since we have assumed that $g(x)$ is monotonically increasing, the limits for y of $\pm\infty$ also become $\pm\infty$ for x.

Then, since $x = g^{-1}(y)$, we have that $yp_X(g^{-1}(y))$ becomes $g(x)p_X(x)$ and

$$\begin{aligned}\left|\frac{dg^{-1}(y)}{dy}\right| dy &= \frac{dg^{-1}(y)}{dy} dy \qquad \begin{array}{l}(g \text{ is monotonically increasing,} \\ \text{implies } g^{-1} \text{ is monotonically increasing,} \\ \text{implies derivative is positive)}\end{array} \\ &= \frac{dx}{dy} dy = dx\end{aligned}$$

from which (11.10) follows. The more general result for nonmonotonic functions follows along these lines.