Chapter 12

Multiple Continuous Random Variables

12.1 Introduction

In Chapter 7 we discussed multiple discrete random variables. We now proceed to parallel that discussion for *multiple continuous random variables.* We will consider in this chapter only the case of two random variables, also called *bivariate random variables*, with the extension to any number of continuous random variables to be presented in Chapter 14. In describing bivariate discrete random variables, we used the example of height and weight of a college student. Figure 7.1 displayed the probabilities of a student having a height in a given interval and a weight in a given interval. For example, the probability of having a height in the interval $[5'8'', 6']$ and a weight in the interval $[160, 190]$ lbs. is 0.14 as listed in Table 4.1 and as seen in Figure 7.1 for the values of $H = 70$ inches and $W = 175$ lbs. For physical measurements such as height and weight, however, we would expect to observe a continuum of values. As such, height and weight are more appropriately modeled by multiple *continuous* random variables. For example, we might have a population of college students, all of whose heights and weights lie in the intervals $60 \leq H \leq 80$ inches and $100 \leq W \leq 250$ lbs. Therefore, the *continuous* random variables (H, W) would take on values in the sample space

$$\mathcal{S}_{H,W} = \{(h, w) : 60 \leq h \leq 80, 100 \leq w \leq 250\}$$

which is a subset of the plane, i.e., R^2. We might wish to determine probabilities such as $P[61 \leq H \leq 67.5, 98.5 \leq W \leq 154]$, which cannot be found from Figure 7.1. In order to compute such a probability we will define a *joint PDF* for the continuous random variables H and W. It will be a two-dimensional function of h and w. In the case of a single random variable we needed to integrate to find the area under the PDF as the desired probability. Now integration of the joint PDF, which is a function of two variables, will produce the probability. However, we will now be determining the *volume* under the joint PDF. All our concepts for a single continuous random variable will extend to the case of two random variables. Computationally, however, we will encounter more difficulty since two-dimensional integrals, also known as double integrals, will need to be evaluated. Hence, the reader should be acquainted with double integrals and their evaluation using iterated integrals.

12.2 Summary

The concept of jointly distributed continuous random variables is introduced in Section 12.3. Given the joint PDF the probability of any event defined on the plane is given by (12.2). The standard bivariate Gaussian PDF is given by (12.3) and is plotted in Figure 12.9. The concept of constant PDF contours is also illustrated in Figure 12.9. The marginal PDF is found from the joint PDF using (12.4). The joint CDF is defined by (12.6) and is evaluated using (12.7). Its properties are listed in P12.1–P12.6. To obtain the joint PDF from the joint CDF we use (12.9). Independence of jointly distributed random variables is defined by (12.10) and can be verified by the factorization of either the PDF as in (12.11) or the CDF as in (12.12). Section 12.6 addresses the problem of determining the PDF of a function of two random variables—see (12.13), and that of determining the joint PDF of a function which maps two random variables into two new random variables. See (12.18) for a linear transformation and (12.22) for a nonlinear transformation. The general bivariate Gaussian PDF is defined in (12.24) and some useful properties are discussed in Section 12.7. In particular, Theorem 12.7.1 indicates that a linear transformation of a bivariate Gaussian random vector produces another bivariate Gaussian random vector, although with different means and covariances. Example 12.14 indicates how a bivariate Gaussian random vector may be transformed to one with independent components. Also, a formula for computation of the expected value of a function of two random variables is given as (12.28). Section 12.9 discusses prediction of a random variable from the observation of a second random variable while Section 12.10 summarizes the joint characteristic function and its properties. In particular, the use of (12.47) allows the determination of the PDF of the sum of two continuous and independent random variables. It is used to prove that two independent Gaussian random variables that are added together produce another Gaussian random variable in Example 12.15. Section 12.11 shows how to simulate on a computer a random vector with any desired mean vector and covariance matrix by using the Cholesky decomposition of the covariance matrix—see (12.53). If the desired random vector is bivariate Gaussian, then the procedure provides a general method for generating Gaussian random vectors on a computer. Finally, an application to optical character recognition is described in Section 12.12.

12.3 Jointly Distributed Random Variables

We consider two continuous random variables that will be denoted by X and Y. As alluded to in the introduction, they represent the functions that map an outcome s of an experiment to a point in the plane. Hence, we have that

$$\begin{bmatrix} X(s) \\ Y(s) \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

for all $s \in \mathcal{S}$. An example is shown in Figure 12.1 in which the outcome of a dart toss s, which is a point within a unit radius circular dartboard, is mapped into a point in the plane, which is within the unit circle. The random variables X and Y

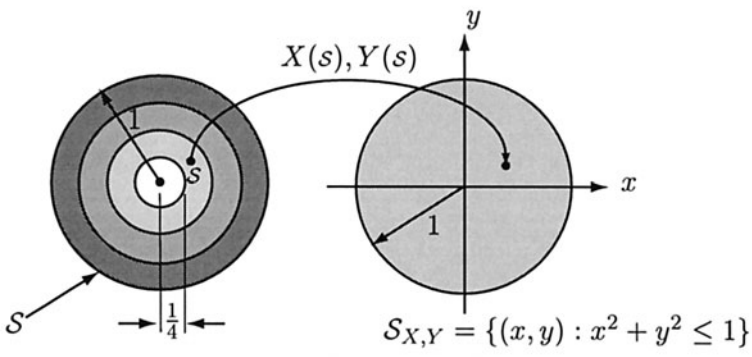


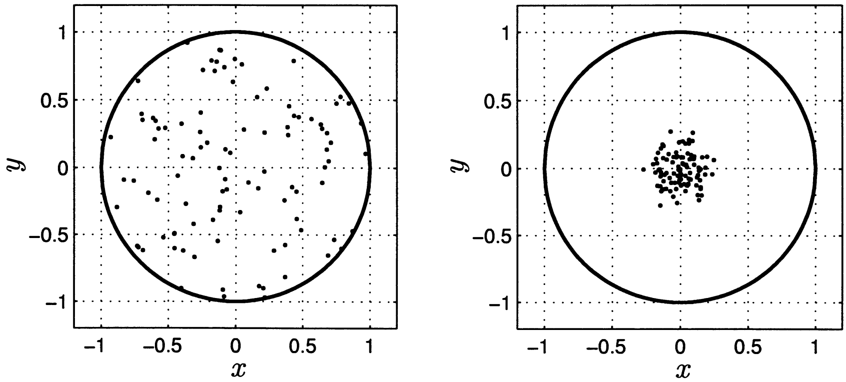
Figure 12.1: Mapping of the outcome of a thrown dart to the plane (example of jointly continuous random variables).

are said to be *jointly distributed continuous random variables*. As before, we will denote the random variables as (X, Y) or $[X\,Y]^T$, in either case referring to them as a random vector. Note that a different mapping would result if we chose to represent the point in $\mathcal{S}_{X,Y}$ in polar coordinates (r, θ). Then we would have

$$\mathcal{S}_{R,\Theta} = \{(r, \theta) : 0 \leq r \leq 1, 0 \leq \theta < 2\pi\}.$$

This is a different random vector but is of course related to (X, Y). Depending upon the shape of the mapped region in the plane, it may be more convenient to use either rectangular coordinates or polar coordinates for probability calculations (see also Problem 12.1).

Typical outcomes of the random variables are shown in Figure 12.2 as points in $\mathcal{S}_{X,Y}$ for two different players. In Figure 12.2a 100 outcomes for a novice dart player are shown while those for a champion dart player are displayed in Figure 12.2b. We might be interested in the probability that $\sqrt{X^2 + Y^2} \leq 1/4$, which is the event that a bullseye is attained. Now our event of interest is a two-dimensional region as opposed to a one-dimensional interval for a single continuous random variable. In



(a) Novice

(b) Champion

Figure 12.2: Typical outcomes for novice and champion dart player.

the case of the novice dart player the dart is equally likely to land anywhere in the unit circle and hence the probability is

$$\begin{aligned} P[\text{bullseye}] &= \frac{\text{Area of bullseye}}{\text{Total area of dartboard}} \\ &= \frac{\pi(1/4)^2}{\pi(1)^2} = \frac{1}{16}. \end{aligned}$$

However, for a champion dart player we see from Figure 12.2b that the probability of a bullseye is much higher. How should we compute this probability? For the novice dart player we can interpret the probability calculation geometrically as shown in Figure 12.3 as the *volume* of the inner cylinder since

$$\begin{aligned} P[\text{bullseye}] &= \pi(1/4)^2 \times \frac{1}{\pi} \\ &= \underbrace{\text{Area of bullseye}}_{\text{Area of event}} \times \underbrace{\frac{1}{\pi}}_{\text{Height}} \quad . \end{aligned}$$

If we define a function

$$p_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} & x^2+y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{12.1}$$

then this volume is also given by

$$P[A] = \iint_A p_{X,Y}(x,y)dxdy \tag{12.2}$$

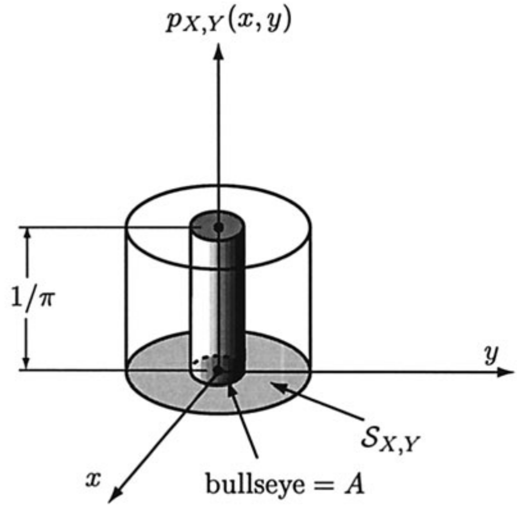


Figure 12.3: Geometric interpretation of bullseye probability calculation for novice dart thrower.

since then

$$\begin{aligned} P[A] &= \iint_{\{(x,y):x^2+y^2\leq(1/4)^2\}} \frac{1}{\pi} dx\, dy \\ &= \frac{1}{\pi} \iint_{\{(x,y):x^2+y^2\leq(1/4)^2\}} dx\, dy \\ &= \frac{1}{\pi} \times \text{Area of } A = \frac{1}{\pi}\pi\left(\frac{1}{4}\right)^2 = \frac{1}{16}. \end{aligned}$$

In analogy with the definition of the PDF for a single random variable X, we define $p_{X,Y}(x,y)$ as the *joint PDF* of X and Y. For this example, it is given by (12.1) and is used to evaluate the probability that (X,Y) lies in a given region A by (12.2). The region A can be any subset of the plane. Note that in using (12.2) we are determining the *volume* under $p_{X,Y}$, hence the need for a double integral. Another example follows.

Example 12.1 – Pyramid-like joint PDF

A joint PDF is given by

$$p_{X,Y}(x,y) = \begin{cases} 4(1-|2x-1|)(1-|2y-1|) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

We wish to first verify that the PDF integrates to one. Then, we consider the evaluation of $P[1/4 \leq X \leq 3/4, 1/4 \leq Y \leq 3/4]$. A three-dimensional plot of the PDF is shown in Figure 12.4 and appears pyramid-like. Since it is often difficult to visualize the PDF in 3-D, it is helpful to plot the contours of the PDF as shown

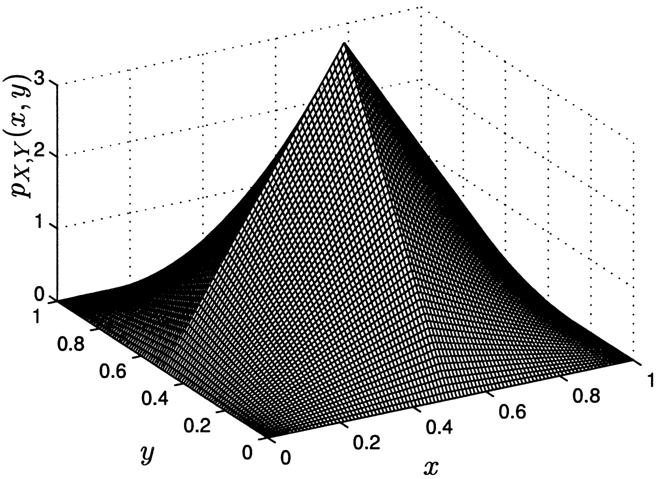


Figure 12.4: Three-dimensional plot of joint PDF.

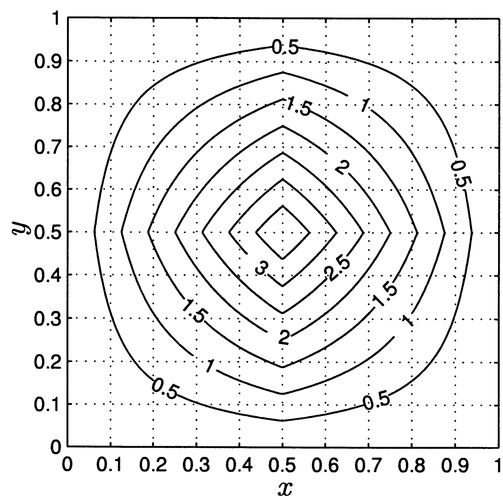


Figure 12.5: Contour plot of joint PDF.

in Figure 12.5. As seen in the contour plot (also called a topographical map) the innermost contour consists of all values of (x, y) for which $p_{X,Y}(x, y) = 3.5$. This contour is obtained by slicing the solid shown in Figure 12.4 with a plane parallel to the x-y plane and at a height of 3.5 and similarly for the other contours. These contours are called *contours of constant PDF.*

To verify that $p_{X,Y}$ is indeed a valid joint PDF, we need to show that the volume under the PDF is equal to one. Since the sample space is $\mathcal{S}_{X,Y} = \{(x, y) : 0 \leq x \leq$

$1, 0 \leq y \leq 1\}$ we have that

$$\begin{aligned}
P[\mathcal{S}_{X,Y}] &= \int_0^1 \int_0^1 4(1-|2x-1|)(1-|2y-1|)dx\,dy \\
&= \int_0^1 2(1-|2x-1|)dx \int_0^1 2(1-|2y-1|)dy.
\end{aligned}$$

The two definite integrals are seen to be identical and hence we need only evaluate one of these. But each integral is the area under the function shown in Figure 12.6a which is easily found to be 1. Hence, $P[\mathcal{S}_{X,Y}] = 1 \cdot 1 = 1$, verifying that $p_{X,Y}$ is a

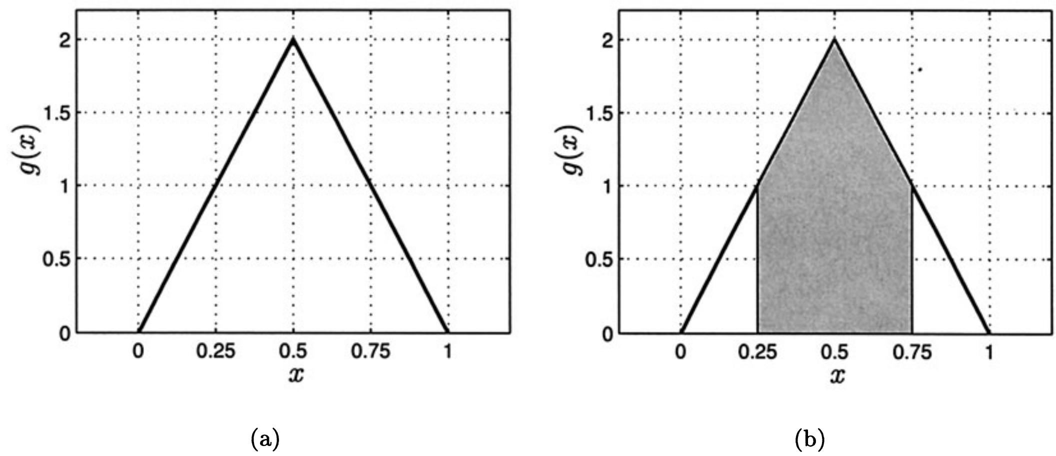


Figure 12.6: Plot of function $g(x) = 2(1-|2x-1|)$.

valid PDF. Next to find $P[1/4 \leq X \leq 3/4, 1/4 \leq Y \leq 3/4]$ we use (12.2) to yield

$$P[A] = \int_{1/4}^{3/4} \int_{1/4}^{3/4} 4(1-|2x-1|)(1-|2y-1|)dx\,dy.$$

By the same argument as before we have

$$P[A] = \left[\int_{1/4}^{3/4} 2(1-|2x-1|)dx\right]^2$$

and referring to Figure 12.6b, we have that each unshaded triangle has an area of $(1/2)(1/4)(1) = 1/8$ and so

$$P[A] = \left[1 - \frac{1}{8} - \frac{1}{8}\right]^2 = \left(\frac{6}{8}\right)^2 = \frac{9}{16}.$$

In summary, a joint PDF has the expected properties of being a nonnegative two-dimensional function that integrates to one over R^2.

◊

For the previous example the double integral was easily evaluated since

1. The integrand $p_{X,Y}(x,y)$ was separable (we will see shortly that this property will hold when the random variables are independent).

2. The integration region in the x-y plane was rectangular.

More generally this will not be the case. Consider, for example, the computation of $P[Y \leq X]$. We need to integrate $p_{X,Y}$ over the shaded region shown in Figure 12.7. To do so we first integrate in the y direction for a fixed x, shown as the darkly

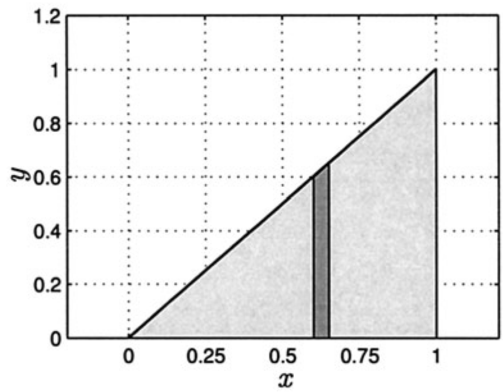


Figure 12.7: Integration region to determine $P[Y \leq X]$.

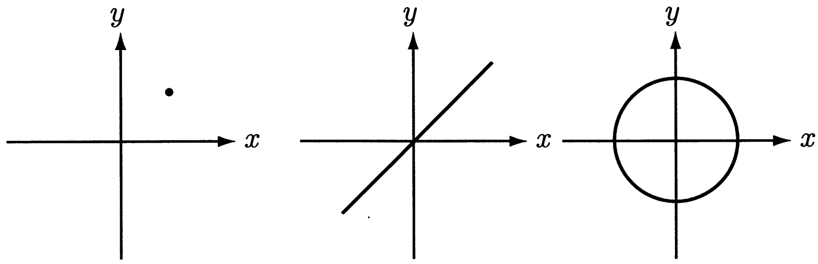
shaded region. Since $0 \leq y \leq x$ for a fixed x, we have the limits of 0 to x for the integration over y and the limits of 0 to 1 for the final integration over x. This results in

$$\begin{aligned} P[Y \leq X] &= \int_0^1 \int_0^x p_{X,Y}(x,y)dy\,dx \\ &= \int_0^1 \int_0^x 4(1-|2x-1|)(1-|2y-1|)dy\,dx. \end{aligned}$$

Although the integration can be carried out, it is tedious. In this illustration the joint PDF is separable but the integration region is not rectangular.

⚠ **Zero probability events are more complex in two dimensions.**

Recall that for a single continuous random variable the probability of X attaining any value is zero. This is because the area under the PDF is zero for any zero length *interval.* Similarly, for jointly continuous random variables X and Y the probability of any event defined on the x-y plane will be zero *if the region of the event in the plane has zero area.* Then, the volume under the joint PDF will be zero. Some examples of these zero probability events are shown in Figure 12.8.



(a) Point (b) Line (c) Curve

Figure 12.8: Examples of zero probability events for jointly distributed continuous random variables X and Y. All regions in the x-y plane have zero area.

An important joint PDF is the *standard bivariate Gaussian or normal PDF*, which is defined as

$$p_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right] \qquad \begin{array}{c} -\infty < x < \infty \\ -\infty < y < \infty \end{array} \tag{12.3}$$

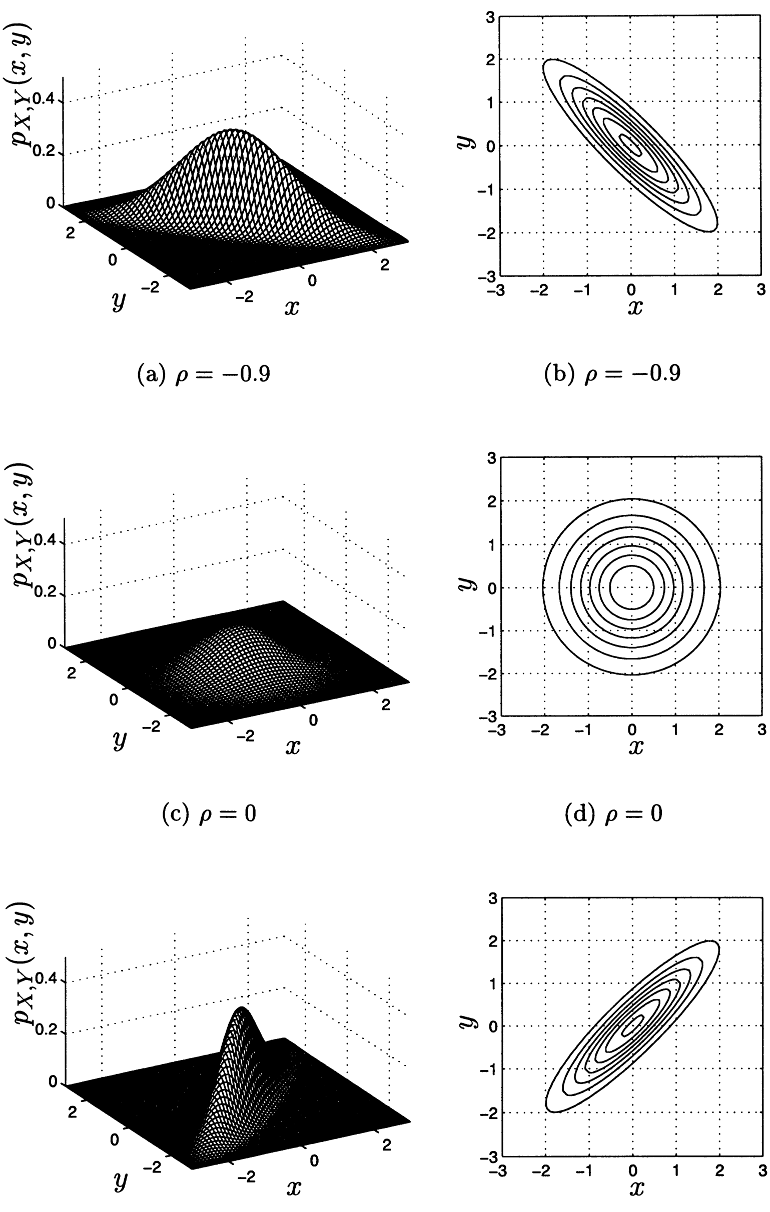
where ρ is a parameter that takes on values $-1 < \rho < 1$. (The use of the term *standard* is because as is shown later the means of X and Y are 0 and the variances are 1.) The joint PDF is shown in Figure 12.9 for various values of ρ. We will see shortly that ρ is actually the correlation coefficient $\rho_{X,Y}$ first introduced in Section 7.9. The contours of constant PDF shown in Figures 12.9b,d,f are given by the values of (x, y) for which

$$x^2 - 2\rho xy + y^2 = r^2$$

where r is a constant. This is because for these values of (x, y) the joint PDF takes on the fixed value

$$p_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}r^2\right].$$

If $\rho = 0$, these contours are circular as seen in Figure 12.9d and otherwise they are elliptical. Note that our use of r^2, which implies that $x^2 - 2\rho xy + y^2 > 0$, is valid



(a) $\rho = -0.9$ (b) $\rho = -0.9$

(c) $\rho = 0$ (d) $\rho = 0$

(e) $\rho = 0.9$ (f) $\rho = 0.9$

Figure 12.9: Three-dimensional and constant PDF contour plots of standard bivariate Gaussian PDF.

since in vector/matrix notation

$$x^2 - 2\rho xy + y^2 = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

which is a *quadratic form*. Because $-1 < \rho < 1$, the matrix is positive definite (its principal minors are all positive—see Appendix C) and hence the quadratic form is positive. We will frequently use the standard bivariate Gaussian PDF and its generalizations as examples to illustrate other concepts. This is because its mathematical tractability lends itself to easy algebraic manipulations.

12.4 Marginal PDFs and the Joint CDF

The marginal PDF $p_X(x)$ of jointly distributed continuous random variables X and Y is the usual PDF which yields the probability of $a \leq X \leq b$ when integrated over the interval $[a,b]$. To determine $p_X(x)$ if we are given the joint PDF $p_{X,Y}(x,y)$, we consider the event

$$A = \{(x,y) : a \leq x \leq b, -\infty < y < \infty\}$$

whose probability must be the same as

$$A_X = \{x : a \leq x \leq b\}.$$

Thus, using (12.2)

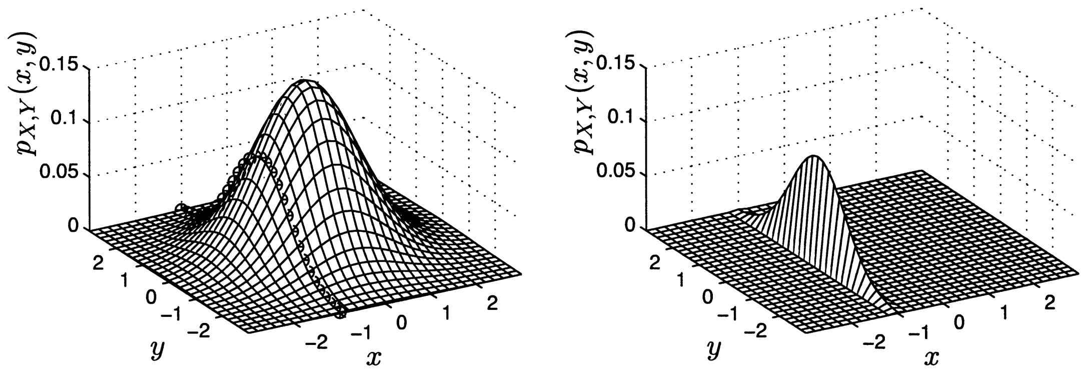
$$\begin{aligned}
P[a \leq X \leq b] &= P[A_X] = P[A] \\
&= \iint_A p_{X,Y}(x,y)dx\,dy \\
&= \int_{-\infty}^{\infty} \int_a^b p_{X,Y}(x,y)dx\,dy \\
&= \int_a^b \underbrace{\int_{-\infty}^{\infty} p_{X,Y}(x,y)dy}_{p_X(x)}\,dx.
\end{aligned}$$

Clearly then, we must have that

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)dy \tag{12.4}$$

as the marginal PDF for X. This operation is shown in Figure 12.10. In effect, we "sum" the probabilites of all the y values associated with the desired x, much the same as summing along a row to determine the marginal PMF $p_X[x_i]$ from the joint PMF $p_{X,Y}[x_i, y_j]$. The marginal PDF can also be viewed as the limit as $\Delta x \to 0$ of

$$\begin{aligned}
p_X(x_0) &= \frac{P[x_0 - \Delta x/2 \leq X \leq x_0 + \Delta x/2, -\infty < Y < \infty]}{\Delta x} \\
&= \frac{\int_{x_0-\Delta x/2}^{x_0+\Delta x/2} \int_{-\infty}^{\infty} p_{X,Y}(x,y)dy\,dx}{\Delta x}
\end{aligned}$$



(a) Curve is $p_{X,Y}(-1, y)$ (b) Area under curve is $p_X(-1)$

Figure 12.10: Obtaining the marginal PDF of X from the joint PDF of (X, Y).

for a small Δx. An example follows.

Example 12.2 – Marginal PDFs for Standard Bivariate Gaussian PDF

From (12.3) and (12.4) we have that

$$p_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right] dy. \qquad (12.5)$$

To carry out the integration we convert the integrand to one we recognize, i.e., a Gaussian, for which the integral over $(-\infty, \infty)$ is known. The trick here is to "complete the square" in y as follows:

$$\begin{aligned} Q &= y^2 - 2\rho xy + x^2 \\ &= y^2 - 2\rho xy + \rho^2 x^2 + x^2 - \rho^2 x^2 \\ &= (y - \rho x)^2 + (1 - \rho^2)x^2. \end{aligned}$$

Substituting into (12.5) produces

$$\begin{aligned} p_X(x) &= \exp(-(1/2)x^2) \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(y - \rho x)^2\right] dy \\ &= \frac{1}{\sqrt{2\pi}} \exp(-(1/2)x^2) \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y - \mu)^2)\right] dy}_{=1} \end{aligned}$$

where $\mu = \rho x$ and $\sigma^2 = 1 - \rho^2$, so that we have

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

or $X \sim \mathcal{N}(0,1)$. Hence, the marginal PDF for X is a standard Gaussian PDF. By reversing the roles of X and Y, we will also find that $Y \sim \mathcal{N}(0,1)$. Note that since the marginal PDFs are standard Gaussian PDFs, the corresponding bivariate Gaussian PDF is also referred to as a standard one.

◇

In the previous example we saw that the marginals could be found from the joint PDF. However, in general the reverse process is not possible—given the marginal PDFs we cannot determine the joint PDF. For example, knowing that $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ does not allow us to determine ρ, which characterizes the joint PDF. Furthermore, the marginal PDFs are the same for any ρ in the interval $(-1,1)$. This is just a restatement of the conclusion that we arrived at for joint and marginal PMFs. In that case there were many possible two-dimensional sets of numbers, i.e., specified by a joint PMF, that could sum to the same one-dimensional set, i.e., specified by a marginal PMF.

We next define the joint CDF for continuous random variables (X,Y). It is given by

$$F_{X,Y}(x,y) = P[X \le x, Y \le y]. \tag{12.6}$$

From (12.2) it is evaluated using

$$F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} p_{X,Y}(t,u)\,dt\,du. \tag{12.7}$$

Some examples follow.

Example 12.3 – Joint CDF for an exponential joint PDF

If (X,Y) have the joint PDF

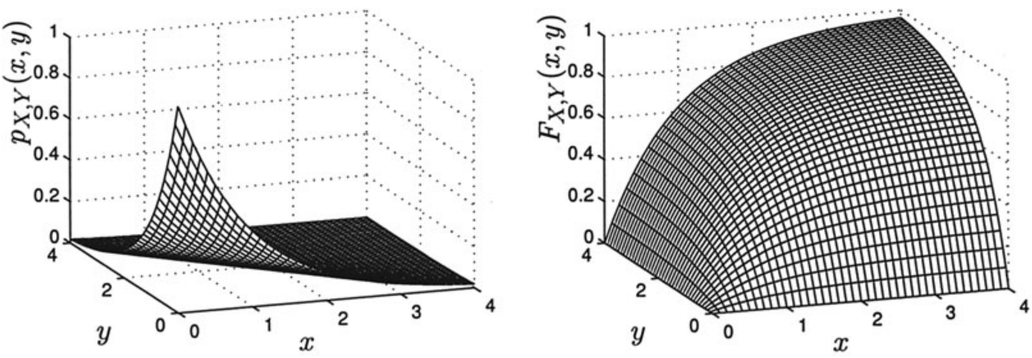
$$p_{X,Y}(x,y) = \begin{cases} \exp[-(x+y)] & x \ge 0, y \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

then for $x \ge 0$, $y \ge 0$

$$\begin{aligned} F_{X,Y}(x,y) &= \int_0^y \int_0^x \exp[-(t+u)]\,dt\,du \\ &= \int_0^y \exp(-u) \underbrace{\int_0^x \exp(-t)\,dt}_{1-\exp(-x)}\,du \\ &= \int_0^y [1-\exp(-x)]\exp(-u)\,du \\ &= [1-\exp(-x)] \int_0^y \exp(-u)\,du \end{aligned}$$

so that

$$F_{X,Y}(x,y) = \begin{cases} [1-\exp(-x)][1-\exp(-y)] & x \ge 0, y \ge 0 \\ 0 & \text{otherwise.} \end{cases} \tag{12.8}$$



(a) PDF (b) CDF

Figure 12.11: Joint exponential PDF and CDF.

The joint CDF is shown in Figure 12.11 along with the joint PDF. Once the joint CDF is obtained the probability for any rectangular region is easily found.

◊

Example 12.4 – Probability from CDF for exponential random variables

Consider the rectangular region $A = \{(x, y) : 1 < x \leq 2, 2 < y \leq 3\}$. Then referring

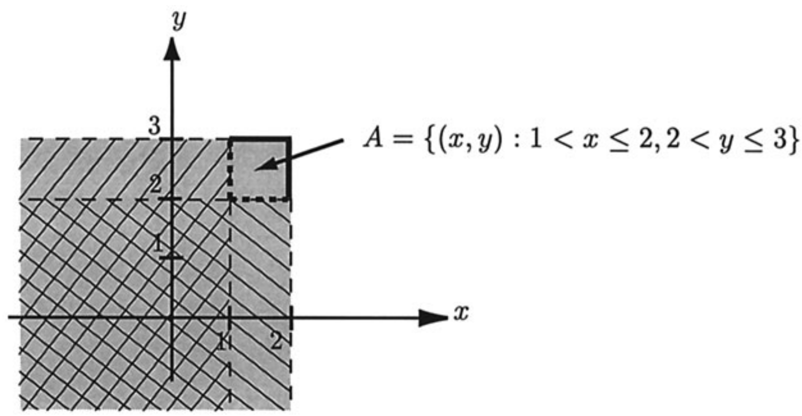


Figure 12.12: Evaluation of probability of rectangular region A using joint CDF.

to Figure 12.12 we determine the probability of A by determining the probability of the shaded region, then subtracting out the probability of each cross-hatched region (one running from south-east to north-west and the other running from south-west to north-east), and finally adding back in the probability of the double cross-hatched region (which has been subtracted out twice). This results in

$$\begin{aligned} P[A] &= P[-\infty < X \leq 2, -\infty < Y \leq 3] - P[-\infty < X \leq 2, -\infty < Y \leq 2] \\ &\quad - P[-\infty < X \leq 1, -\infty < Y \leq 3] + P[-\infty < X \leq 1, -\infty < Y \leq 2] \\ &= F_{X,Y}[2,3] - F_{X,Y}[2,2] - F_{X,Y}[1,3] + F_{X,Y}[1,2]. \end{aligned}$$

For the joint CDF given by (12.8) this becomes

$$\begin{aligned} P[A] &= [1-\exp(-2)][1-\exp(-3)] - [1-\exp(-2)]^2 \\ &\quad -[1-\exp(-1)][1-\exp(-3)] + [1-\exp(-1)][1-\exp(-2)]. \end{aligned}$$

Upon simplication we have the result

$$P[A] = [\exp(-1) - \exp(-2)][\exp(-2) - \exp(-3)]$$

which can also be verified by a direct evaluation as

$$P[A] = \int_2^3 \int_1^2 \exp[-(x+y)] dx\, dy.$$

We see that the advantage here is that no integration is required. However, the event A must be a rectangular region.

◊

The joint PDF can be recovered from the joint CDF by partial differentiation as

$$p_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \tag{12.9}$$

which is the two-dimensional version of the fundamental theorem of calculus. As an example we continue the previous one.

Example 12.5 – Obtaining the joint PDF from the joint CDF for exponential random variables

Continuing with the previous example we have from (12.8) that

$$p_{X,Y}(x,y) = \begin{cases} \frac{\partial^2 [1-\exp(-x)][1-\exp(-y)]}{\partial x \partial y} & x \geq 0, y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

For $x > 0, y > 0$

$$\begin{aligned} p_{X,Y}(x,y) &= \frac{\partial}{\partial x} \frac{\partial [1-\exp(-x)][1-\exp(-y)]}{\partial y} \\ &= \frac{\partial [1-\exp(-x)]}{\partial x} \frac{\partial [1-\exp(-y)]}{\partial y} \\ &= \exp(-x)\exp(-y) = \exp[-(x+y)]. \end{aligned}$$

◊

Finally, the properties of the joint CDF are for the most part identical to those for the CDF (see Section 7.4 for the properties of the joint CDF for discrete random variables). They are (see Figure 12.11b for an illustration):

P12.1 $F_{X,Y}(-\infty, -\infty) = 0$

P12.2 $F_{X,Y}(+\infty, +\infty) = 1$

P12.3 $F_{X,Y}(x, \infty) = F_X(x)$

P12.4 $F_{X,Y}(\infty, y) = F_Y(y)$

P12.5 $F_{X,Y}(x, y)$ is monotonically increasing, which means that if $x_2 \geq x_1$ and $y_2 \geq y_1$, then $F_{X,Y}(x_2, y_2) \geq F_{X,Y}(x_1, y_1)$.

P12.6 $F_{X,Y}(x, y)$ is continuous with no jumps (assuming that X and Y are jointly continuous random variables). This property is different from the case of jointly discrete random variables.

12.5 Independence of Multiple Random Variables

The definition of independence of two continuous random variables is the same as for discrete random variables. Two continuous random variables X and Y are defined to be independent if for *all* events $A \in R$ and $B \in R$

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B]. \tag{12.10}$$

Using the definition of conditional probability this is equivalent to

$$\begin{aligned} P[Y \in B | X \in A] &= \frac{P[X \in A, Y \in B]}{P[X \in A]} \\ &= P[Y \in B] \end{aligned}$$

and similarly $P[X \in A | Y \in B] = P[X \in A]$. It can be shown that X and Y are independent if and only if the joint PDF factors as (see Problem 12.20)

$$p_{X,Y}(x, y) = p_X(x) p_Y(y). \tag{12.11}$$

Alternatively, X and Y are independent if and only if (see Problem 12.21)

$$F_{X,Y}(x, y) = F_X(x) F_Y(y). \tag{12.12}$$

An example follows.

Example 12.6 – Independence of exponential random variables

From Example 12.3 we have for the joint PDF

$$p_{X,Y}(x, y) = \begin{cases} \exp[-(x+y)] & x \geq 0, y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Recalling that the unit step function $u(x)$ is defined as $u(x) = 1$ for $x \geq 0$ and $u(x) = 0$ for $x < 0$, we have

$$p_{X,Y}(x,y) = \exp[-(x+y)]u(x)u(y)$$

since $u(x)u(y) = 1$ if and only if $u(x) = 1$ *and* $u(y) = 1$, which will be true for $x \geq 0, y \geq 0$. Hence, we have

$$p_{X,Y}(x,y) = \underbrace{\exp(-x)u(x)}_{p_X(x)}\underbrace{\exp(-y)u(y)}_{p_Y(y)}.$$

To assert independence we need only factor $p_{X,Y}(x,y)$ as $g(x)h(y)$, where g and h are nonnegative functions. However, to assert that $g(x)$ is actually $p_X(x)$ and $h(y)$ is actually $p_Y(y)$, each function, g and h, must integrate to one. For example, we could have factored $p_{X,Y}(x,y)$ into $(1/2)\exp(-x)u(x)$ and $2\exp(-y)u(y)$, but then we could not claim that $p_X(x) = (1/2)\exp(-x)u(x)$ since it does not integrate to one. Note also that the joint CDF given in Example 12.3 is also factorable as given in (12.8) and in general, factorization of the CDF is also necessary and sufficient to assert independence.

◇

Assessing independence – careful with domain of PDF

The joint PDF given by

$$p_{X,Y}(x,y) = \begin{cases} 2\exp[-(x+y)] & x \geq 0, y \geq 0, \text{ and } y < x \\ 0 & \text{otherwise} \end{cases}$$

is *not* factorable, although it is very similar to our previous example. The reason is that the region in the x-y plane where $p_{X,Y}(x,y) \neq 0$ cannot be written as $u(x)u(y)$ or for that matter as any $g(x)h(y)$. See Figure 12.13.

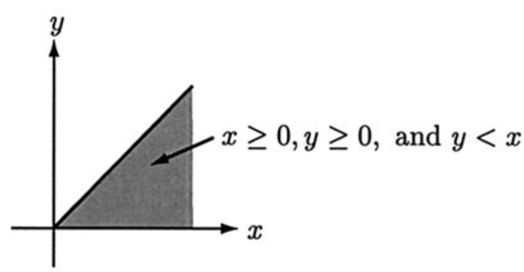


Figure 12.13: Nonfactorable region in x-y plane.

Example 12.7 – Standard bivariate Gaussian PDF

From (12.3) we see that $p_{X,Y}(x,y)$ is only factorable if $\rho = 0$. From Figure 12.9d this corresponds to the case of circular PDF contours. Specifically, for $\rho = 0$, we have

$$\begin{aligned} p_{X,Y}(x,y) &= \frac{1}{2\pi}\exp\left[-\frac{1}{2}(x^2+y^2)\right] \qquad -\infty < x < \infty, -\infty < y < \infty \\ &= \underbrace{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}x^2\right]}_{p_X(x)}\underbrace{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}y^2\right]}_{p_Y(y)}. \end{aligned}$$

Hence, we observe that if $\rho = 0$, then X and Y are independent. Furthermore, each marginal PDF is a standard Gaussian (normal) PDF, but as shown in Example 12.2 this holds regardless of the value of ρ.

$\diamond$

Finally, note that if we can assume that X and Y are independent, then knowledge of $p_X(x)$ and $p_Y(y)$ is sufficient to determine the joint PDF according to (12.11). In practice, the independence assumption greatly simplifies the problem of joint PDF estimation as we need only to estimate the two one-dimensional PDFs $p_X(x)$ and $p_Y(y)$.

12.6 Transformations

We will consider two types of transformations. The first one maps two continuous random variables into a single continuous random variable as $Z = g(X,Y)$, and the second one maps two continuous random variables into two new continuous random variables as $W = g(X,Y)$ and $Z = h(X,Y)$. The first type of transformation $Z = g(X,Y)$ is now discussed. The approach is to find the CDF of Z and then differentiate it to obtain the PDF. The CDF of Z is given as

$$\begin{aligned} F_Z(z) &= P[Z \le z] && \text{(definition of CDF)} \\ &= P[g(X,Y) \le z] && \text{(definition of } Z\text{)} \\ &= \iint_{\{(x,y):g(x,y)\le z\}} p_{X,Y}(x,y)\,dx\,dy && \text{(from (12.2)).} \end{aligned} \tag{12.13}$$

We see that it is necessary to integrate the joint PDF over the region in the plane where $g(x,y) \le z$. Depending upon the form of g, this may be a simple task or unfortunately a very complicated one. A simple example follows. It is the continuous version of (7.22), which yields the PMF for the sum of two independent discrete random variables.

Example 12.8 – Sum of independent $\mathcal{U}(0,1)$ random variables

In Section 2.3 we inquired as to the distribution of the outcomes of an experiment that added U_1, a number chosen at random from 0 to 1, to U_2, another number chosen at random from 0 to 1. A histogram of the outcomes of a computer simulation indicated that there is a higher probability of the sum being near 1, as opposed to being near 0 or 2. We now know that $U_1 \sim \mathcal{U}(0,1)$, $U_2 \sim \mathcal{U}(0,1)$. Also, in the experiment of Section 2.3 the two numbers were chosen independently of each other. Hence, we can determine the probabilities of the sum random variable if we first find the CDF of $X = U_1 + U_2$, where U_1 and U_2 are independent, and then differentiate it to find the PDF of X. We will use (12.13) and replace x, y, z, and $g(x,y)$ by u_1, u_2, x, and $g(u_1,u_2)$, respectively. Then

$$F_X(x) = \iint_{\{(u_1,u_2):u_1+u_2\leq x\}} p_{U_1,U_2}(u_1,u_2)du_1\,du_2.$$

To determine the possible values of X, we note that both U_1 and U_2 take on values in $(0,1)$ and so $0 < X < 2$. In evaluating the CDF we need two different intervals for x as shown in Figure 12.14. Since U_1 and U_2 are independent, we have $p_{U_1,U_2} = p_{U_1}p_{U_2}$

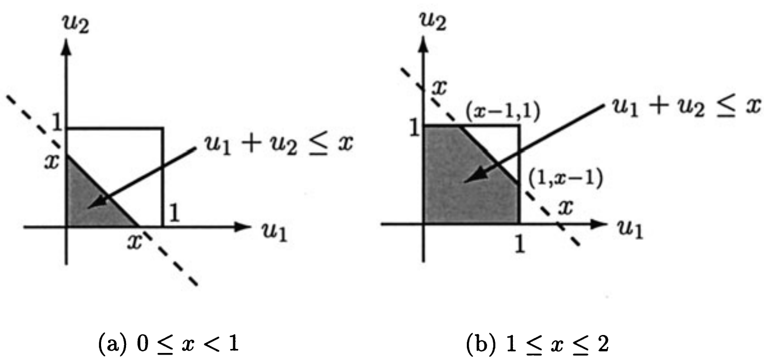


Figure 12.14: Shaded areas are regions of integration used to find CDF.

and therefore $p_{U_1,U_2}(u_1,u_2) = 1$ for $0 < u_1 < 1$ and $0 < u_2 < 1$, which results in

$$F_X(x) = \iint_{\{(u_1,u_2):u_1+u_2\leq x\}} 1\,du_1\,du_2 = \text{shaded area in Figure 12.14.}$$

Hence, the CDF is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2}x^2 & 0 \leq x < 1 \\ 1 - \frac{1}{2}(2-x)^2 & 1 \leq x \leq 2 \\ 1 & x > 2. \end{cases}$$

and the PDF is finally

$$\begin{aligned} p_X(x) &= \frac{dF_X(x)}{dx} \\ &= \begin{cases} 0 & x < 0 \\ x & 0 \le x < 1 \\ 2 - x & 1 \le x \le 2 \\ 0 & x > 2. \end{cases} \end{aligned}$$

This PDF is shown in Figure 12.15. This is in agreement with our computer results

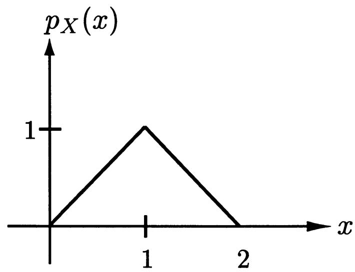


Figure 12.15: PDF for the sum of two independent $\mathcal{U}(0, 1)$ random variables.

shown in Figure 2.2. The highest probability is at $x = 1$, which concurs with our computer generated results of Section 2.3. Also, note that $p_X(x) = p_{U_1}(x) \star p_{U_2}(x)$, where $\star$ denotes integral *convolution* (see Problem 12.28).

◇

More generally, we can derive a useful formula for the PDF of the sum of two independent continuous random variables. According to (12.13), we first need to determine the region in the plane for which $x+y \le z$. This inequality can be written as $y \le z - x$, where z is to be regarded for the present as a *constant*. To integrate $p_{X,Y}(x, y)$ over this region, which is shown in Figure 12.16 as the shaded region, we can use an iterated integral. Thus,

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} p_{X,Y}(x, y) dy\, dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} p_X(x) p_Y(y) dy\, dx & \text{(independence)} \\ &= \int_{-\infty}^{\infty} p_X(x) \int_{-\infty}^{z-x} p_Y(y) dy\, dx \\ &= \int_{-\infty}^{\infty} p_X(x) F_Y(z - x) dx & \text{(definition of CDF).} \end{aligned}$$

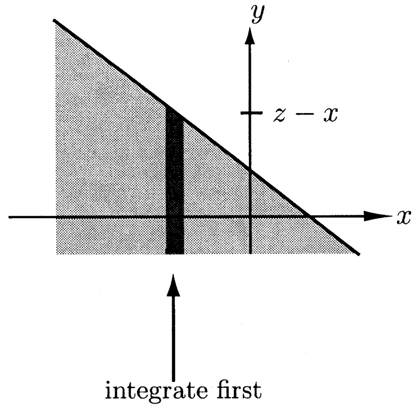


Figure 12.16: Iterated integral evaluation – shaded region is $y \leq z - x$. Integrate first in y direction for a fixed x and then integrate over $-\infty < x < \infty$.

If we now differentiate the CDF, we have

$$\begin{aligned} p_Z(z) &= \frac{d}{dz}\int_{-\infty}^{\infty} p_X(x)F_Y(z-x)dx \\ &= \int_{-\infty}^{\infty} p_X(x)\frac{d}{dz}F_Y(z-x)dx && \text{(assume interchange is valid)} \\ &= \int_{-\infty}^{\infty} p_X(x)\left.\frac{d}{du}F_Y(u)\right|_{u=z-x}\frac{du}{dz}dx && \text{(chain rule with } u = z-x\text{)} \end{aligned}$$

so that finally we have our formula

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(x)p_Y(z-x)dx. \tag{12.14}$$

This is the analogous result to (7.22). It is recognized as a convolution integral, which we can express more succinctly as $p_Z = p_X \star p_Y$, and thus may be more easily evaluated by using characteristic functions. The latter approach is explored in Section 12.10.

A second approach to obtaining the PDF of $g(X,Y)$ is to let $W = X$, $Z = g(X,Y)$, find the joint PDF of W and Z, i.e., $p_{W,Z}(w,z)$, and finally integrate out W to yield the desired PDF for Z. This method was encountered previously in Chapter 7, where it was used for discrete random variables, and was termed the method of auxiliary random variables. To implement it now requires us to determine the joint PDF of two new random variables that result from having transformed two random variables. This is the second type of transformation we were interested in. Hence, we now consider the more general transformation

$$\begin{aligned} W &= g(X,Y) \\ Z &= h(X,Y). \end{aligned}$$

The final result will be a formula relating the joint PDF of (W, Z) to that of the given joint PDF of (X, Y). It will be a generalization of the single random variable transformation formula

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \tag{12.15}$$

for $Y = g(X)$.

To understand what is involved, consider as an example the transformation

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} U_1 \\ (U_1 + U_2)/2 \end{bmatrix} \tag{12.16}$$

where $U_1 \sim \mathcal{U}(0,1)$, $U_2 \sim \mathcal{U}(0,1)$, and U_1 and U_2 are independent. In Figure 2.13 we plotted realizations of $[U_1\, U_2]^T$ and $[X_1\, X_2]^T$. Note that the original joint PDF p_{U_1,U_2} is nonzero on the unit square while the transformed PDF is nonzero on a parallelogram. In either case the PDFs appear to be uniformly distributed. Similar observations about the region for which the PDF of the transformed random variable is nonzero were made in the one-dimensional case for $Y = g(X)$, where $X \sim \mathcal{U}(0,1)$, in Figure 10.22. In general, a linear transformation will change the support area of the joint PDF, which is the region in the plane where the PDF is nonzero. In Figure 2.13 it is seen that the area of the square is 1 while that for the parallelogram is 1/2. It can furthermore be shown that if we have the linear transformation (see Problem 12.29)

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\mathbf{G}} \begin{bmatrix} X \\ Y \end{bmatrix} \tag{12.17}$$

then

$$\begin{aligned} \frac{\text{Area in } w\text{-}z \text{ plane}}{\text{Area in } x\text{-}y \text{ plane}} &= |\det(\mathbf{G})| \\ &= |ad - bc|. \end{aligned}$$

It is always assumed that $\mathbf{G}$ is invertible so that $\det(\mathbf{G}) \neq 0$. In the previous example of (12.16) for which in our new notation we have $W = X$ and $Z = (X+Y)/2$, the linear transformation matrix is

$$\mathbf{G} = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

and it is seen that $|\det(\mathbf{G})| = 1/2$. Thus, the PDF support region is decreased by a factor of 2. We therefore expect the joint PDF of $[X\ (X+Y)/2]^T$ to be uniform with a height of 2 (as opposed to a height of 1 for the original joint PDF). Hence, the transformed PDF should have a factor of $1/|\det(\mathbf{G})|$ to make it integrate to one.

This amplification factor, which is $1/|\det(\mathbf{G})| = |\det(\mathbf{G}^{-1})|$ must be included in the expression for the transformed joint PDF. Also, we have that $[x\,y]^T = \mathbf{G}^{-1}[w\,z]^T$. Hence, it should not be surprising that for the linear transformation of (12.17) we have the formula for the transformed joint PDF

$$p_{W,Z}(w,z) = p_{X,Y}\left(\mathbf{G}^{-1}\begin{bmatrix} w \\ z \end{bmatrix}\right)\left|\det(\mathbf{G}^{-1})\right|. \tag{12.18}$$

An example follows.

Example 12.9 – Linear transformation for standard bivariate Gaussian PDF

Assume that (X, Y) has the PDF of (12.3) and consider the linear transformation

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \underbrace{\begin{bmatrix} \sigma_W & 0 \\ 0 & \sigma_Z \end{bmatrix}}_{\mathbf{G}}\begin{bmatrix} X \\ Y \end{bmatrix}.$$

Then,

$$\mathbf{G}^{-1} = \begin{bmatrix} 1/\sigma_W & 0 \\ 0 & 1/\sigma_Z \end{bmatrix}$$

and

$$\begin{aligned} \mathbf{G}^{-1}\begin{bmatrix} w \\ z \end{bmatrix} &= \begin{bmatrix} w/\sigma_W \\ z/\sigma_Z \end{bmatrix} \\ \det(\mathbf{G}^{-1}) &= \frac{1}{\sigma_W\sigma_Z} \end{aligned}$$

so that from (12.3) and (12.18)

$$\begin{aligned} & p_{W,Z}(w,z) \\ = {} & \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left((w/\sigma_W)^2 - 2\rho wz/(\sigma_W\sigma_Z) + (z/\sigma_Z)^2\right)\right]\frac{1}{\sigma_W\sigma_Z} \\ = {} & \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_W^2\sigma_Z^2}} \\ & \cdot\exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{w}{\sigma_W}\right)^2 - 2\rho\left(\frac{w}{\sigma_W}\right)\left(\frac{z}{\sigma_Z}\right) + \left(\frac{z}{\sigma_Z}\right)^2\right)\right]. \end{aligned} \tag{12.19}$$

Note that since $-\infty < x < \infty$, $-\infty < y < \infty$, we have that the region of support for $p_{W,Z}$ is $-\infty < w < \infty$, $-\infty < z < \infty$. Also, the joint PDF can be written in vector/matrix form as (see Problem 12.31)

$$p_{W,Z}(w,z) = \frac{1}{2\pi\det^{1/2}(\mathbf{C})}\exp\left(-\frac{1}{2}\begin{bmatrix} w \\ z \end{bmatrix}^T\mathbf{C}^{-1}\begin{bmatrix} w \\ z \end{bmatrix}\right) \tag{12.20}$$

where

$$\mathbf{C} = \begin{bmatrix} \sigma_W^2 & \rho\sigma_W\sigma_Z \\ \rho\sigma_Z\sigma_W & \sigma_Z^2 \end{bmatrix}. \tag{12.21}$$

The matrix $\mathbf{C}$ will be shown later to be the *covariance matrix* of W and Z (see Section 9.5 for the definition of the covariance matrix, which is also valid for continuous random variables).

◇

For nonlinear transformations a result similar to (12.18) is obtained. This is because a two-dimensional nonlinear function can be linearized about a point by replacing the usual tangent or derivative approximation for a one-dimensional function by a tangent plane approximation (see Problem 12.32). Hence, if the transformation is given by

$$\begin{aligned} W &= g(X,Y) \\ Z &= h(X,Y) \end{aligned}$$

then a given point in the w-z plane is obtained via $w = g(x,y)$, $z = h(x,y)$. Assume that the latter set of equations has a single solution for all (w,z), say

$$\begin{aligned} x &= g^{-1}(w,z) \\ y &= h^{-1}(w,z). \end{aligned}$$

Then it can be shown that

$$p_{W,Z}(w,z) = p_{X,Y}(\underbrace{g^{-1}(w,z)}_{x}, \underbrace{h^{-1}(w,z)}_{y}) \left| \det \left(\frac{\partial(x,y)}{\partial(w,z)} \right) \right| \tag{12.22}$$

where

$$\frac{\partial(x,y)}{\partial(w,z)} = \begin{bmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial z} \end{bmatrix} \tag{12.23}$$

is called the Jacobian matrix of the inverse transformation from $[w\, z]^T$ to $[x\, y]^T$ and is sometimes referred to as $\mathbf{J}^{-1}$. It represents the compensation for the amplification/reduction of the areas due to the transformation. For a linear transformation $\mathbf{G}$ it is given by $\mathbf{J} = \mathbf{G}$ (see also (12.15) for a single random variable). We now illustrate the use of this formula.

Example 12.10 – Affine transformation for standard bivariate Gaussian PDF

Let (X,Y) have a standard bivariate Gaussian PDF and consider the affine transformation

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \begin{bmatrix} \sigma_W & 0 \\ 0 & \sigma_Z \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} \mu_W \\ \mu_Z \end{bmatrix}.$$

Then using (12.22) we first solve for (x, y) as

$$
\begin{aligned}
x &= \frac{w-\mu_W}{\sigma_W} \\
y &= \frac{z-\mu_Z}{\sigma_Z}.
\end{aligned}
$$

The inverse Jacobian matrix becomes

$$
\frac{\partial(x,y)}{\partial(w,z)} = \begin{bmatrix} 1/\sigma_W & 0 \\ 0 & 1/\sigma_Z \end{bmatrix}
$$

and therefore, since

$$
p_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right]
$$

we have from (12.22)

$$
p_{W,Z}(w,z) = \frac{1}{2\pi\sqrt{1-\rho^2}}
$$

$$
\cdot \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{w-\mu_W}{\sigma_W}\right)^2 - 2\rho\left(\frac{w-\mu_W}{\sigma_W}\right)\left(\frac{z-\mu_Z}{\sigma_Z}\right) + \left(\frac{z-\mu_Z}{\sigma_Z}\right)^2\right)\right]\frac{1}{\sigma_W\sigma_Z}
$$

or finally

$$
p_{W,Z}(w,z) = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_W^2\sigma_Z^2}}
$$

$$
\cdot \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{w-\mu_W}{\sigma_W}\right)^2 - 2\rho\left(\frac{w-\mu_W}{\sigma_W}\right)\left(\frac{z-\mu_Z}{\sigma_Z}\right) + \left(\frac{z-\mu_Z}{\sigma_Z}\right)^2\right)\right]. \tag{12.24}
$$

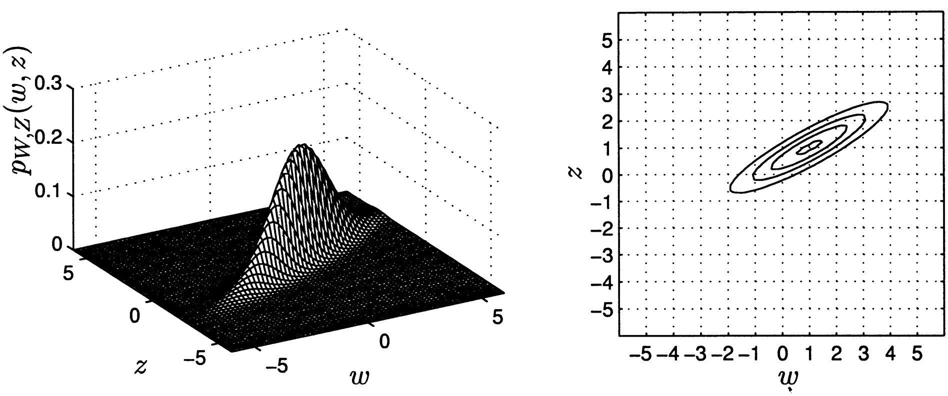
This is called the *bivariate Gaussian PDF*. If $\mu_W = \mu_Z = 0$ and $\sigma_W = \sigma_Z = 1$, then it reverts back to the usual *standard* bivariate Gaussian PDF. If $\mu_W = \mu_Z = 0$, we have the joint PDF in Example 12.9. An example of the PDF is shown in Figure 12.17.

◇

The bivariate Gaussian PDF can also be written more compactly in vector/matrix form as

$$
p_{W,Z}(w,z) = \frac{1}{2\pi\det^{1/2}(\mathbf{C})}\exp\left(-\frac{1}{2}\begin{bmatrix} w-\mu_W \\ z-\mu_Z \end{bmatrix}^T \mathbf{C}^{-1}\begin{bmatrix} w-\mu_W \\ z-\mu_Z \end{bmatrix}\right) \tag{12.25}
$$

where $\mathbf{C}$ is the covariance matrix given by (12.21). It can also be shown that the marginal PDFs are $W \sim \mathcal{N}(\mu_W, \sigma_W^2)$ and $Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2)$ (see Problem 12.36).



(a) Joint PDF (b) Contours of constant PDF

Figure 12.17: Example of bivariate Gaussian PDF with $\mu_W = 1, \mu_Z = 1$, $\sigma_W^2 = 3, \sigma_Z^2 = 1$, and $\rho = 0.9$.

Hence, *the marginal PDFs of the bivariate Gaussian PDF are obtained by inspection* (see Problem 12.37).

Example 12.11 – Transformation of independent Gaussian random variables to a Cauchy random variable

Let $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$, and X and Y be independent. Then consider the transformation $W = X, Z = Y/X$. To determine $\mathcal{S}_{W,Z}$ note that $w = x$ so that $-\infty < w < \infty$ and since $z = y/x$ with $-\infty < x < \infty, -\infty < y < \infty$, we have $-\infty < z < \infty$. Hence, $\mathcal{S}_{W,Z}$ is the entire plane. To find the joint PDF we first solve for (x,y) as $x = w$ and $y = xz = wz$. The inverse Jacobian matrix is

$$\frac{\partial(x,y)}{\partial(w,z)} = \begin{bmatrix} 1 & 0 \\ z & w \end{bmatrix}$$

so that $|\det(\partial(x,y)/\partial(w,z))| = |w|$. Using (12.22), we have

$$\begin{aligned} p_{W,Z}(w,z) &= \frac{1}{2\pi}\exp\left[-\frac{1}{2}(x^2+y^2)\right]\Bigg|_{x=w,y=wz} |w| \\ &= \frac{1}{2\pi}\exp\left[-\frac{1}{2}(w^2+w^2z^2)\right]|w| \\ &= \frac{1}{2\pi}\exp\left[-\frac{1}{2}(1+z^2)w^2\right]|w|. \end{aligned}$$

It is of interest to determine the marginal PDFs. Clearly, the marginal of $W = X$ is just the original PDF $\mathcal{N}(0,1)$. The marginal PDF for Z, which is the ratio of two independent $\mathcal{N}(0,1)$ random variables, is found from (12.4) as

$$\begin{aligned}
p_Z(z) &= \int_{-\infty}^{\infty} p_{W,Z}(w,z)dw \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi}\exp\left[-\frac{1}{2}(1+z^2)w^2\right]|w|dw \\
&= \frac{1}{\pi}\int_0^{\infty} w\exp\left[-\frac{1}{2}(1+z^2)w^2\right]dw \quad \text{(integrand is even function)} \\
&= \frac{1}{\pi}\left.\frac{\exp[-(1/2)(1+z^2)w^2]}{-(1+z^2)}\right|_0^{\infty} \\
&= \frac{1}{\pi(1+z^2)} \qquad -\infty < z < \infty
\end{aligned}$$

which is recognized as the Cauchy PDF. Hence, the PDF of Y/X, where X and Y are independent standard Gaussian random variables is Cauchy. We have implicitly used the method of *auxiliary random variables* to derive this result. Finally, the observation that the denominator of Y/X is a standard Gaussian random variable, with significant probability of being near zero, may help to explain why the outcomes of a Cauchy random variable are as large as they are. See Figure 11.3.

$\diamond$

The next example is of great importance in many fields of science and engineering.

Example 12.12 – Magnitude and angle of jointly Gaussian distributed random variables

Let $X \sim \mathcal{N}(0,\sigma^2)$, $Y \sim \mathcal{N}(0,\sigma^2)$, and X and Y be independent random variables. Then, it is desired to find the joint PDF when X and Y, considered as Cartesian coordinates, are converted to polar coordinates via

$$\begin{aligned}
R &= \sqrt{X^2+Y^2} \qquad R \geq 0 \\
\Theta &= \arctan\frac{Y}{X} \qquad 0 \leq \Theta < 2\pi.
\end{aligned} \tag{12.26}$$

It is common in many engineering disciplines, for example, in radar, sonar, and communications, to transmit a sinusoidal signal and to process the received signal by a digital computer. The received signal will be given by $s(t) = A\cos(2\pi F_0 t) + B\sin(2\pi F_0 t)$ for a transmit frequency of F_0 Hz. However, because the received signal is due to the sum of multiple reflections from an aircraft, as in the radar example, the values of A and B are generally not known. Consequently, they are modeled as continuous random variables with marginal PDFs $A \sim \mathcal{N}(0,\sigma^2)$, $B \sim \mathcal{N}(0,\sigma^2)$, and where A and B are independent. Since the received signal can equivalently be written in terms of a single sinusoid as (see Problem 12.42)

$$s(t) = \sqrt{A^2+B^2}\cos(2\pi F_0 t - \arctan(B/A))$$

the amplitude is a random variable as is the phase angle. Thus, the transformation of (12.26) is of interest in order to determine the joint PDF of the sinusoid's amplitude and phase. This motivates our interest in this particular transformation.

We first solve for (x, y) as $x = r\cos\theta$, $y = r\sin\theta$. Then using (12.22) and replacing w by r and z by θ we have the inverse Jacobian

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

and thus

$$\det\left(\frac{\partial(x,y)}{\partial(r,\theta)}\right) = r \geq 0.$$

Since

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}(x^2+y^2)\right]$$

we have upon using (12.22)

$$\begin{aligned} p_{R,\Theta}(r,\theta) &= \frac{1}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}r^2\right]r \qquad r \geq 0,\ 0 \leq \theta < 2\pi \\ &= \underbrace{\frac{r}{\sigma^2}\exp\left[-\frac{1}{2\sigma^2}r^2\right]}_{p_R(r)}\underbrace{\frac{1}{2\pi}}_{p_\Theta(\theta)} \qquad r \geq 0,\ 0 \leq \theta < 2\pi. \end{aligned}$$

Here we see that R has a Rayleigh PDF with parameter σ^2, Θ has a uniform PDF, and R and Θ are independent random variables.

◊

12.7 Expected Values

The expected value of two jointly distributed continuous random variables X and Y, or equivalently the random vector $[X\,Y]^T$, is defined as the vector of the expected values. That is to say

$$E_{X,Y}\left[\begin{bmatrix} X \\ Y \end{bmatrix}\right] = \begin{bmatrix} E_X[X] \\ E_Y[Y] \end{bmatrix}.$$

Of course this is equivalent to the vector of the expected values of the marginal PDFs. As an example, for the bivariate Gaussian PDF as given by (12.24) with W, Z replaced by X, Y, the marginals are $\mathcal{N}(\mu_X, \sigma_X^2)$ and $\mathcal{N}(\mu_Y, \sigma_Y^2)$ and hence the expected value or equivalently the mean of the random vector is

$$E_{X,Y}\left[\begin{bmatrix} X \\ Y \end{bmatrix}\right] = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$

as shown in Figure 12.17 for $\mu_X = \mu_Y = 1$.

We frequently require the expected value of a function of two jointly distributed random variables or of $Z = g(X, Y)$. By definition this is

$$E[Z] = \int_{-\infty}^{\infty} z p_Z(z) dz.$$

But as in the case for jointly distributed *discrete* random variables we can avoid the determination of $p_Z(z)$ by employing instead the formula

$$E[Z] = E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) p_{X,Y}(x, y) dx\, dy. \tag{12.27}$$

To remind us that the averaging PDF is $p_{X,Y}(x, y)$ we usually write this as

$$E_{X,Y}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) p_{X,Y}(x, y) dx\, dy. \tag{12.28}$$

If the function g depends on only one of the variables, say X, then we have

$$\begin{aligned} E_{X,Y}[g(X)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) p_{X,Y}(x, y) dx\, dy \\ &= \int_{-\infty}^{\infty} g(x) \underbrace{\int_{-\infty}^{\infty} p_{X,Y}(x, y) dy}_{p_X(x)}\, dx \\ &= E_X[g(X)]. \end{aligned}$$

As in the case of discrete random variables (see Section 7.7), the expectation has the following properties:

1. Linearity

$$E_{X,Y}[aX + bY] = aE_X[X] + bE_Y[Y]$$

and more generally

$$E_{X,Y}[ag(X, Y) + bh(X, Y)] = aE_{X,Y}[g(X, Y)] + bE_{X,Y}[h(X, Y)].$$

2. Factorization for independent random variables

If X and Y are independent random variables

$$E_{X,Y}[XY] = E_X[X]E_Y[Y] \tag{12.29}$$

and more generally

$$E_{X,Y}[g(X)h(Y)] = E_X[g(X)]E_Y[h(Y)]. \tag{12.30}$$

Also, in determining the variance for a sum of random variables we have

$$\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X,Y) \tag{12.31}$$

where $\operatorname{cov}(X,Y) = E_{X,Y}[(X - E_X[X])(Y - E_Y[Y])]$. If X and Y are independent, then by (12.30)

$$\begin{aligned}\operatorname{cov}(X,Y) &= E_{X,Y}[(X - E_X[X])(Y - E_Y[Y])] \\ &= E_X[(X - E_X[X])]E_Y[(Y - E_Y[Y])] \\ &= 0.\end{aligned}$$

The covariance can also be computed as

$$\operatorname{cov}(X,Y) = E_{X,Y}[XY] - E_X[X]E_Y[Y] \tag{12.32}$$

where

$$E_{X,Y}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy p_{X,Y}(x,y)dx\,dy. \tag{12.33}$$

An example follows.

Example 12.13 – Covariance for standard bivariate Gaussian PDF

For the standard bivariate Gaussian PDF of (12.3) we now determine $\operatorname{cov}(X,Y)$. We have already seen that the marginal PDFs are $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ so that $E_X[X] = E_Y[Y] = 0$. From (12.32) we have that $\operatorname{cov}(X,Y) = E_{X,Y}[XY]$ and using (12.33) and (12.3)

$$\operatorname{cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right]dx\,dy.$$

To evaluate this double integral we use iterated integrals and complete the square in the exponent of the exponential as was previously done in Example 12.2. This results in

$$Q = y^2 - 2\rho xy + x^2 = (y - \rho x)^2 + (1-\rho^2)x^2$$

and produces

$$\begin{aligned}&\operatorname{cov}(X,Y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}(y-\rho x)^2\right]\exp\left[-\frac{1}{2}x^2\right]dx\,dy \\ &= \int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}x^2\right]\int_{-\infty}^{\infty} y\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left[-\frac{1}{2(1-\rho^2)}(y-\rho x)^2\right]dy\,dx.\end{aligned}$$

The inner integral over y is just $E_Y[Y] = \int_{-\infty}^{\infty} y p_Y(y)dy$, where $Y \sim \mathcal{N}(\rho x, 1-\rho^2)$. Thus, $E_Y[Y] = \rho x$ so that

$$\begin{aligned}\operatorname{cov}(X,Y) &= \int_{-\infty}^{\infty} \rho x^2\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}x^2\right]dx \\ &= \rho E_X[X^2]\end{aligned}$$

where $X \sim \mathcal{N}(0,1)$. But $E_X[X^2] = \text{var}(X) + E_X^2[X] = 1 + 0^2 = 1$ and therefore we have finally that

$$\text{cov}(X,Y) = \rho.$$

$\diamondsuit$

With the result of the previous example we can now determine the correlation coefficient between X and Y for the standard bivariate Gaussian PDF. Since the marginal PDFs are $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$, the correlation coefficient between X and Y is

$$\begin{aligned}
\rho_{X,Y} &= \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} \\
&= \frac{\rho}{\sqrt{1 \cdot 1}} \\
&= \rho.
\end{aligned}$$

We have therefore established that *in the standard bivariate Gaussian PDF, the parameter ρ is the correlation coefficient.* This explains the orientation of the constant PDF contours shown in Figure 12.9. Also, we can now assert that if the correlation coefficient between X and Y is zero, i.e., $\rho = 0$, *and X and Y are jointly Gaussian distributed (i.e., a standard bivariate Gaussian PDF)*, then

$$\begin{aligned}
p_{X,Y}(x,y) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right] \\
&= \frac{1}{2\pi} \exp\left[-\frac{1}{2}(x^2 + y^2)\right] \\
&= \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right]}_{p_X(x)} \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}y^2\right]}_{p_Y(y)}
\end{aligned}$$

and *X and Y are independent.* This also holds for the general bivariate Gaussian PDF in which the marginal PDFs are $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. This result provides a *partial converse* to the theorem that if X and Y are independent, then the random variables are uncorrelated, *but only for this particular joint PDF.*

Finally, since $\rho = \rho_{X,Y}$ we have from (12.21) upon replacing W by X and Z by Y, that

$$\begin{aligned}
\mathbf{C} &= \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_Y\sigma_X & \sigma_Y^2 \end{bmatrix} \qquad (12.34) \\
&= \begin{bmatrix} \sigma_X^2 & \rho_{X,Y}\sigma_X\sigma_Y \\ \rho_{X,Y}\sigma_Y\sigma_X & \sigma_Y^2 \end{bmatrix} \\
&= \begin{bmatrix} \text{var}(X) & \text{cov}(X,Y) \\ \text{cov}(Y,X) & \text{var}(Y) \end{bmatrix}
\end{aligned}$$

is the covariance matrix. We have now established that $\mathbf{C}$ as given by (12.34) is actually the covariance matrix. Hence, the general bivariate Gaussian PDF is given in succinct form as (see (12.24))

$$p_{X,Y}(x,y) = \frac{1}{2\pi \det^{1/2}(\mathbf{C})} \exp\left(-\frac{1}{2}\begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}\right) \tag{12.35}$$

where $\mathbf{C}$ is given by (12.34) and is the covariance matrix (see Section also 9.5)

$$\mathbf{C} = \begin{bmatrix} \operatorname{var}(X) & \operatorname{cov}(X,Y) \\ \operatorname{cov}(Y,X) & \operatorname{var}(Y) \end{bmatrix}. \tag{12.36}$$

As previously mentioned, an extremely important property of the bivariate Gaussian PDF is that uncorrelated random variables implies independent random variables. Hence, if the covariance matrix in (12.36) is *diagonal*, then X and Y are independent. We have shown in Chapter 9 that it is always possible to diagonalize a covariance matrix by transforming the random vector using a linear transformation. Specifically, if the random vector $[X\,Y]^T$ is transformed to a new random vector $\mathbf{V}^T[X\,Y]^T$, where $\mathbf{V}$ is the modal matrix for the covariance matrix $\mathbf{C}$, then the transformed random vector will have a diagonal covariance matrix. Hence, the transformed random vector will have uncorrelated components. If furthermore, the transformed random vector also has a bivariate Gaussian PDF, then *its component random variables will be independent.* It is indeed fortunate that this is true—*a linearly transformed bivariate Gaussian random vector produces another bivariate Gaussian random vector*, as we now show. To do so it is more convenient to use a vector/matrix representation of the PDF. Let the linear transformation be

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \mathbf{G}\begin{bmatrix} X \\ Y \end{bmatrix}$$

where $\mathbf{G}$ is an invertible 2×2 matrix. Assume for simplicity that $\mu_X = \mu_Y = 0$. Then, from (12.35)

$$p_{X,Y}(x,y) = \frac{1}{2\pi \det^{1/2}(\mathbf{C})} \exp\left(-\frac{1}{2}\begin{bmatrix} x \\ y \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} x \\ y \end{bmatrix}\right)$$

and using (12.18)

$$\begin{aligned} p_{W,Z}(w,z) &= p_{X,Y}\left(\mathbf{G}^{-1}\begin{bmatrix} w \\ z \end{bmatrix}\right)\left|\det(\mathbf{G}^{-1})\right| \\ &= \frac{1}{2\pi \det^{1/2}(\mathbf{C})} \exp\left(-\frac{1}{2}\begin{bmatrix} w \\ z \end{bmatrix}^T \mathbf{G}^{-1^T}\mathbf{C}^{-1}\mathbf{G}^{-1} \begin{bmatrix} w \\ z \end{bmatrix}\right)\left|\det(\mathbf{G}^{-1})\right|. \end{aligned}$$

But it can be shown that (see Section C.3 of Appendix C for matrix inverse and determinant formulas)

$$\mathbf{G}^{-1^T}\mathbf{C}^{-1}\mathbf{G}^{-1} = \mathbf{G}^{T^{-1}}\mathbf{C}^{-1}\mathbf{G}^{-1} = (\mathbf{G}\mathbf{C}\mathbf{G}^T)^{-1}$$

and

$$\begin{aligned} \left|\det(\mathbf{G}^{-1})\right| &= \frac{1}{|\det(\mathbf{G})|} \\ &= \frac{1}{(\det(\mathbf{G})\det(\mathbf{G}))^{1/2}} \\ &= \frac{1}{(\det(\mathbf{G})\det(\mathbf{G}^T))^{1/2}} \end{aligned}$$

so that

$$\begin{aligned} \frac{\left|\det(\mathbf{G}^{-1})\right|}{\det^{1/2}(\mathbf{C})} &= \frac{1}{\det^{1/2}(\mathbf{C})(\det(\mathbf{G})\det(\mathbf{G}^T))^{1/2}} \\ &= \frac{1}{(\det(\mathbf{C})\det(\mathbf{G})\det(\mathbf{G}^T))^{1/2}} \\ &= \frac{1}{(\det(\mathbf{G})\det(\mathbf{C})\det(\mathbf{G}^T))^{1/2}} \\ &= \frac{1}{\det^{1/2}(\mathbf{G}\mathbf{C}\mathbf{G}^T)}. \end{aligned}$$

Thus, we have finally that the PDF of the linearly transformed random vector is

$$p_{W,Z}(w,z) = \frac{1}{2\pi\det^{1/2}(\mathbf{G}\mathbf{C}\mathbf{G}^T)}\exp\left(-\frac{1}{2}\begin{bmatrix} w \\ z \end{bmatrix}^T (\mathbf{G}\mathbf{C}\mathbf{G}^T)^{-1}\begin{bmatrix} w \\ z \end{bmatrix}\right)$$

which is recognized as a bivariate Gaussian PDF with zero means and a covariance matrix $\mathbf{G}\mathbf{C}\mathbf{G}^T$. This also agrees with Property 9.4. We summarize our results in a theorem.

Theorem 12.7.1 (Linear transformation of Gaussian random variables) *If (X,Y) has the bivariate Gaussian PDF*

$$p_{X,Y}(x,y) = \frac{1}{2\pi\det^{1/2}(\mathbf{C})}\exp\left(-\frac{1}{2}\begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}^T \mathbf{C}^{-1}\begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}\right) \tag{12.37}$$

and the random vector is linearly transformed as

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \mathbf{G}\begin{bmatrix} X \\ Y \end{bmatrix}$$

where $\mathbf{G}$ *is invertible, then*

$$p_{W,Z}(w,z) = \frac{1}{2\pi \det^{1/2}(\mathbf{G}\mathbf{C}\mathbf{G}^T)} \exp\left(-\frac{1}{2}\begin{bmatrix} w-\mu_W \\ z-\mu_Z \end{bmatrix}^T (\mathbf{G}\mathbf{C}\mathbf{G}^T)^{-1} \begin{bmatrix} w-\mu_W \\ z-\mu_Z \end{bmatrix}\right)$$

where

$$\begin{bmatrix} \mu_W \\ \mu_Z \end{bmatrix} = \mathbf{G} \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$

is the transformed mean vector.

The bivariate Gaussian PDF with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$. Hence, the theorem may be paraphrased as follows—if $[X\,Y]^T \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$, then $\mathbf{G}[X\,Y]^T \sim \mathcal{N}(\mathbf{G}\boldsymbol{\mu}, \mathbf{G}\mathbf{C}\mathbf{G}^T)$. An example, which uses results from Example 9.4, is given next.

Example 12.14 – Transforming correlated Gaussian random variables to independent Gaussian random variables

Let $\mu_X = \mu_Y = 0$ and

$$\mathbf{C} = \begin{bmatrix} 26 & 6 \\ 6 & 26 \end{bmatrix}$$

in (12.37). The joint PDF and its constant PDF contours are shown in Figure 12.18. Now transform X and Y according to

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \mathbf{G} \begin{bmatrix} X \\ Y \end{bmatrix}$$

where $\mathbf{G}$ is the transpose of the modal matrix $\mathbf{V}$, which is given in Example 9.4. Therefore

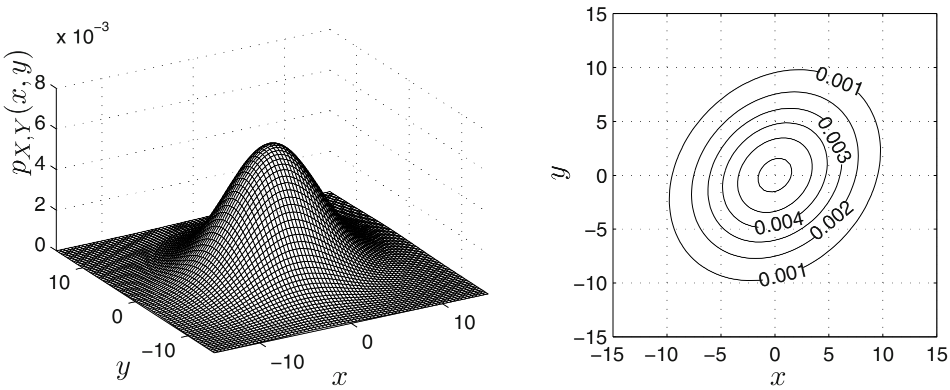
$$\mathbf{G} = \mathbf{V}^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

so that

$$\begin{aligned} W &= \frac{1}{\sqrt{2}}X - \frac{1}{\sqrt{2}}Y \\ Z &= \frac{1}{\sqrt{2}}X + \frac{1}{\sqrt{2}}Y. \end{aligned}$$

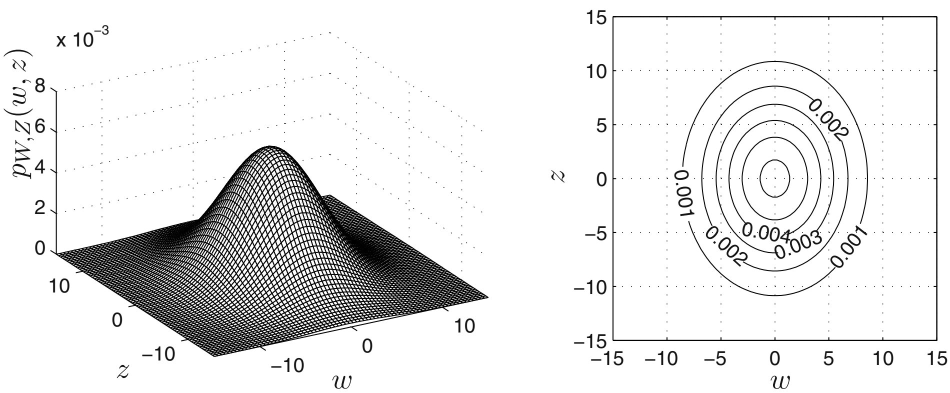
We have that

$$\mathbf{G}\mathbf{C}\mathbf{G}^T = \mathbf{V}^T\mathbf{C}\mathbf{V} = \begin{bmatrix} 20 & 0 \\ 0 & 32 \end{bmatrix}$$



(a) Joint PDF (b) Contours of constant PDF

Figure 12.18: Example of joint PDF for correlated Gaussian random variables.



(a) Joint PDF (b) Contours of constant PDF

Figure 12.19: Example of joint PDF for transformed correlated Gaussian random variables. The random variables are now uncorrelated and hence independent.

$$
\begin{aligned}
p_{W,Z}(w,z) &= \frac{1}{2\pi\sqrt{20\cdot 32}} \exp\left(-\frac{1}{2}\begin{bmatrix} w \\ z \end{bmatrix}^T \begin{bmatrix} 1/20 & 0 \\ 0 & 1/32 \end{bmatrix} \begin{bmatrix} w \\ z \end{bmatrix}\right) \\
&= \frac{1}{\sqrt{2\pi\cdot 20}} \exp\left[-\frac{1}{2}\frac{w^2}{20}\right] \cdot \frac{1}{\sqrt{2\pi\cdot 32}} \exp\left[-\frac{1}{2}\frac{z^2}{32}\right]
\end{aligned}
$$

which is the factorization of the joint PDF of W and Z into the marginal PDFs $W \sim \mathcal{N}(0,20)$ and $Z \sim \mathcal{N}(0,32)$. Hence, W and Z are now *independent* random variables, each with a marginal Gaussian PDF. The joint PDF $p_{W,Z}$ is shown in Figure 12.19. Note the rotation of the contour plots in Figures 12.18b and 12.19b. This rotation was asserted in Example 9.4 (see also Problem 12.48).

◇

12.8 Joint Moments

For jointly distributed continuous random variables the k-lth joint moments are defined as $E_{X,Y}[X^k Y^l]$. They are evaluated as

$$E_{X,Y}[X^k Y^l] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k y^l p_{X,Y}(x,y)dx\,dy. \tag{12.38}$$

An example for $k = l = 1$ and for a standard bivariate Gaussian PDF of $E_{X,Y}[XY]$ was given Example 12.13. The k-lth joint *central* moments are defined as $E_{X,Y}[(X - E_X[X])^k (Y - E_Y[Y])^l]$ and are evaluated as

$$E_{X,Y}[(X-E_X[X])^k(Y-E_Y[Y])^l] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-E_X[X])^k(y-E_Y[Y])^l p_{X,Y}(x,y)dx\,dy. \tag{12.39}$$

Of course, the most important case is for $k = l = 1$ for which we have the $\text{cov}(X,Y)$. For independent random variables the joint moments factor as

$$E_{X,Y}[X^k Y^l] = E_X[X^k]E_Y[Y^l]$$

and similarly for the joint central moments.

12.9 Prediction of Random Variable Outcome

In Section 7.9 we described the prediction of the outcome of a discrete random variable based on the observed outcome of another discrete random variable. We now examine the prediction problem for jointly distributed continuous random variables, and in particular, for the case of a bivariate Gaussian PDF. First we plot a scatter diagram of the outcomes of the random vector $[X\,Y]^T$ in the x-y plane. Shown in Figure 12.20 is the result for a random vector with a zero mean and a covariance matrix

$$\mathbf{C} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}. \tag{12.40}$$

Note that the correlation coefficient is given by

$$\begin{aligned} \rho_{X,Y} &= \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} \\ &= \frac{0.9}{\sqrt{1 \cdot 1}} = 0.9. \end{aligned}$$

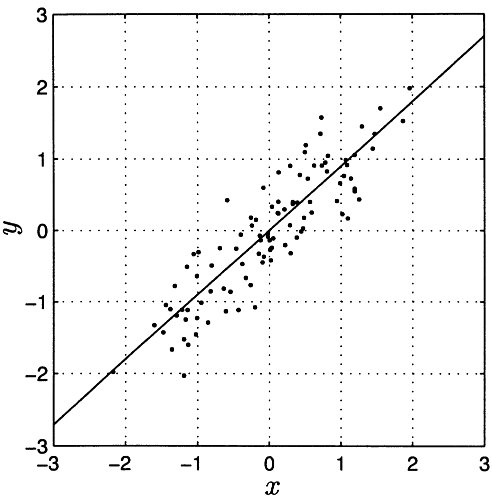


Figure 12.20: 100 outcomes of bivariate Gaussian random vector with zero means and covariance matrix given by (12.40). The best prediction of Y when $X = x$ is observed is given by the line.

It is seen from Figure 12.20 that knowledge of X should allow us to predict the outcome of Y with some accuracy. To do so we adopt as our error criterion the minimum mean square error (MSE) and use a linear predictor or $\hat{Y} = aX + b$. From Section 7.9 the best linear prediction when $X = x$ is observed is

$$\hat{Y} = E_Y[Y] + \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}(x - E_X[X]). \tag{12.41}$$

For this example the best linear prediction is

$$\hat{Y} = 0 + \frac{0.9}{1}(x - 0) = 0.9x \tag{12.42}$$

and is shown as the line in Figure 12.20. Note that the error $\epsilon = Y - 0.9X$ is also a random variable and can be shown to have the PDF $\epsilon \sim \mathcal{N}(0, 0.19)$ (see Problem 12.49). Finally, note that the predictor, which was constrained to be linear (actually affine but the use of the term linear is commonplace), cannot be improved upon by resorting to a *nonlinear* predictor. This is because it can be shown that *the optimal predictor, among all predictors, is linear if* (X, Y) *has a bivariate Gaussian PDF* (see Section 13.6). Hence, in this case the prediction of (12.42) is optimal among all linear *and* nonlinear predictors.

12.10 Joint Characteristic Functions

The joint characteristic function for two jointly continuous random variables X and Y is defined as

$$\phi_{X,Y}(\omega_X, \omega_Y) = E_{X,Y}[\exp[j(\omega_X X + \omega_Y Y)]]. \tag{12.43}$$

It is evaluated as

$$\phi_{X,Y}(\omega_X, \omega_Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{X,Y}(x,y)\exp[j(\omega_X x + \omega_Y y)]\,dx\,dy \tag{12.44}$$

and is seen to be the two-dimensional Fourier transform of the PDF (with a $+j$ instead of the more common $-j$ in the exponential). As in the case of discrete random variables, the joint moments can be found from the characteristic function using the formula

$$E_{X,Y}[X^k Y^l] = \frac{1}{j^{k+l}} \left.\frac{\partial^{k+l}\phi_{X,Y}(\omega_X,\omega_Y)}{\partial\omega_X^k \partial\omega_Y^l}\right|_{\omega_X=\omega_Y=0}. \tag{12.45}$$

Another important application is in determining the PDF for the sum of two independent continuous random variables. As shown in Section 7.10 for discrete random variables and also true for jointly continuous random variables, if X and Y are independent, then the characteristic function of the sum $Z = X + Y$ is

$$\phi_Z(\omega) = \phi_X(\omega)\phi_Y(\omega). \tag{12.46}$$

If we were to take the inverse Fourier transform of both sides of (12.46), then the PDF of $X+Y$ would result. Hence, the procedure to determine the PDF of $X+Y$, where X and Y are independent random variables, is

1. Find the characteristic function $\phi_X(\omega)$ by evaluating the Fourier transform $\int_{-\infty}^{\infty} p_X(x)\exp(j\omega x)dx$ and similarly for $\phi_Y(\omega)$.

2. Multiply $\phi_X(\omega)$ and $\phi_Y(\omega)$ together to form $\phi_X(\omega)\phi_Y(\omega)$.

3. Finally, find the inverse Fourier transform to yield the PDF for the sum $Z = X + Y$ as

$$p_Z(z) = \int_{-\infty}^{\infty} \phi_X(\omega)\phi_Y(\omega)\exp(-j\omega z)\frac{d\omega}{2\pi}. \tag{12.47}$$

Alternatively, one could convolve the PDFs of X and Y using the convolution *integral* of (12.14) to yield the PDF of Z. However, the convolution approach is seldom easier. An example follows.

Example 12.15 – PDF for sum of independent Gaussian random variables

If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ and X and Y are independent, we wish to determine the PDF of $Z = X + Y$. A convolution approach is explored in

Problem 12.51. Here we use (12.47) to accomplish the same task. First we need the characteristic function of a Gaussian PDF. From Table 11.1 if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\phi_X(\omega) = \exp\left(j\omega\mu - \frac{1}{2}\sigma^2\omega^2\right).$$

Thus, the characteristic function for $X + Y$ is

$$\begin{aligned}\phi_{X+Y}(\omega) &= \exp\left(j\omega\mu_X - \frac{1}{2}\sigma_X^2\omega^2\right)\exp\left(j\omega\mu_Y - \frac{1}{2}\sigma_Y^2\omega^2\right) \\ &= \exp\left(j\omega(\mu_X + \mu_Y) - \frac{1}{2}(\sigma_X^2 + \sigma_Y^2)\omega^2\right).\end{aligned}$$

Since this is again the characteristic function of a Gaussian random variable, albeit with different parameters, we have that $X+Y \sim \mathcal{N}(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$. (Recognizing that the characteristic function is that of a known PDF allows us to avoid inverting the characteristic function according to (12.47).) Hence, the PDF of the sum of independent Gaussian random variables is again a Gaussian random variable whose mean is $\mu = \mu_X + \mu_Y$ and whose variance is $\sigma^2 = \sigma_X^2 + \sigma_Y^2$. The Gaussian PDF is therefore called a *reproducing PDF*. By the same argument it follows that *the sum of any number of independent Gaussian random variables is again a Gaussian random variable with mean equal to the sum of the means and variance equal to the sum of the variances.* In Problem 12.53 it is shown that the Gamma PDF is also a reproducing PDF.

◇

The result of the previous example could also be obtained by appealing to Theorem 12.7.1. If we let

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \end{bmatrix}$$

then by Theorem 12.7.1, W and $Z = X+Y$ are bivariate Gaussian distributed. Also, we know that the marginals of a bivariate Gaussian PDF are Gaussian PDFs and therefore the PDF of $Z = X + Y$ is Gaussian. Its mean is $\mu_X + \mu_Y$ and its variance is $\sigma_X^2 + \sigma_Y^2$, the latter because X and Y are independent and hence uncorrelated.

12.11 Computer Simulation of Jointly Continuous Random Variables

For an arbitrary joint PDF the generation of continuous random variables is most easily done using ideas from conditional PDF theory. In Chapter 13 we will see how this is done. Here we will consider only the generation of a bivariate Gaussian random vector. The approach is based on the following properties:

1. Any affine transformation of two jointly Gaussian random variables results in two new jointly Gaussian random variables. A special case, the linear transformation, was proven in Section 12.7 and the general result summarized in Theorem 12.7.1. We will now consider the affine transformation

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \mathbf{G} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} a \\ b \end{bmatrix}. \tag{12.48}$$

2. The mean vector and covariance matrix of $[W\,Z]^T$ transform according to

$$\begin{aligned} E\left[\begin{bmatrix} W \\ Z \end{bmatrix}\right] &= \mathbf{G}E\left[\begin{bmatrix} X \\ Y \end{bmatrix}\right] + \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{(see Problem 9.22)} && (12.49) \\ \mathbf{C}_{W,Z} &= \mathbf{G}\mathbf{C}_{X,Y}\mathbf{G}^T \quad \text{(see (Theorem 12.7.1))} && (12.50) \end{aligned}$$

where we now use subscripts on the covariance matrices to indicate the random variables.

The approach to be described next assumes that X and Y are standard Gaussian and independent random variables whose realizations are easily generated. In MATLAB the command `randn(1,1)` can be used. Otherwise, if only $\mathcal{U}(0,1)$ random variables are available, one can use the Box-Mueller transform to obtain X and Y (see Problem 12.54). Then, to obtain any bivariate Gaussian random variables (W, Z) with a given mean $[\mu_W\,\mu_Z]^T$ and covariance matrix $\mathbf{C}_{W,Z}$, we use (12.48) with a suitable $\mathbf{G}$ and $[a\,b]^T$ so that

$$\begin{aligned} E\left[\begin{bmatrix} W \\ Z \end{bmatrix}\right] &= \begin{bmatrix} \mu_W \\ \mu_Z \end{bmatrix} \\ \mathbf{C}_{W,Z} &= \begin{bmatrix} \sigma_W^2 & \rho\sigma_W\sigma_Z \\ \rho\sigma_Z\sigma_W & \sigma_Z^2 \end{bmatrix}. \end{aligned} \tag{12.51}$$

Since it is assumed that X and Y are zero mean, from (12.49) we choose $a = \mu_W$ and $b = \mu_Z$. Also, since X and Y are assumed independent, hence uncorrelated, and with unit variances, we have

$$\mathbf{C}_{X,Y} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}.$$

It follows from (12.50) that $\mathbf{C}_{W,Z} = \mathbf{G}\mathbf{G}^T$. To find $\mathbf{G}$ if we are given $\mathbf{C}_{W,Z}$, we could use an eigendecomposition approach based on the relationship $\mathbf{V}^T\mathbf{C}_{W,Z}\mathbf{V} = \mathbf{\Lambda}$ (see Problem 12.55). Instead, we next explore an alternative approach which is somewhat easier to implement in practice. Let $\mathbf{G}$ be a *lower triangular* matrix

$$\mathbf{G} = \begin{bmatrix} a & 0 \\ b & c \end{bmatrix}.$$

Then, we have that

$$\mathbf{G}\mathbf{G}^T = \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} = \begin{bmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{bmatrix}. \tag{12.52}$$

The numerical procedure of decomposing a covariance matrix into a product such as $\mathbf{G}\mathbf{G}^T$, where $\mathbf{G}$ is lower triangular, is called the *Cholesky decomposition* [Golub and Van Loan 1983]. Here we can do so almost by inspection. We need only equate the elements of $\mathbf{C}_{W,Z}$ in (12.51) to those of $\mathbf{G}\mathbf{G}^T$ as given in (12.52). Doing so produces the result

$$\begin{aligned} a &= \sigma_W \\ b &= \rho\sigma_Z \\ c &= \sigma_Z\sqrt{1-\rho^2}. \end{aligned}$$

Hence, we have that

$$\mathbf{G} = \begin{bmatrix} \sigma_W & 0 \\ \rho\sigma_Z & \sigma_Z\sqrt{1-\rho^2} \end{bmatrix}.$$

In summary, to generate a realization of a bivariate Gaussian random vector we first generate two independent standard Gaussian random variables X and Y and then transform according to

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \begin{bmatrix} \sigma_W & 0 \\ \rho\sigma_Z & \sigma_Z\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} \mu_W \\ \mu_Z \end{bmatrix}. \tag{12.53}$$

As an example, we let $\mu_W = \mu_Z = 1$, $\sigma_W = \sigma_Z = 1$, and $\rho = 0.9$. The constant PDF contours as well as 500 realizations of $[W\, Z]^T$ are shown in Figure 12.21. To verify that the mean vector and covariance matrix are correct, we can estimate these quantities using (9.44) and (9.46) which are

$$\widehat{E_{W,Z}}\left[\begin{bmatrix} W \\ Z \end{bmatrix}\right] = \frac{1}{M}\sum_{m=1}^{M} \begin{bmatrix} w_m \\ z_m \end{bmatrix}$$

$$\widehat{\mathbf{C}_{W,Z}} = \frac{1}{M}\sum_{m=1}^{M} \left(\begin{bmatrix} w_m \\ z_m \end{bmatrix} - \widehat{E_{W,Z}}\left[\begin{bmatrix} W \\ Z \end{bmatrix}\right]\right)\left(\begin{bmatrix} w_m \\ z_m \end{bmatrix} - \widehat{E_{W,Z}}\left[\begin{bmatrix} W \\ Z \end{bmatrix}\right]\right)^T$$

where $[w_m\, z_m]^T$ is the mth realization of $[W\, Z]^T$. The results and the true values for $M = 2000$ are

$$\widehat{E_{W,Z}}\left[\begin{bmatrix} W \\ Z \end{bmatrix}\right] = \begin{bmatrix} 1.0326 \\ 1.0252 \end{bmatrix} \quad E_{W,Z}\left[\begin{bmatrix} W \\ Z \end{bmatrix}\right] = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\widehat{\mathbf{C}_{W,Z}} = \begin{bmatrix} 0.9958 & 0.9077 \\ 0.9077 & 1.0166 \end{bmatrix} \quad \mathbf{C}_{W,Z} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}.$$

The MATLAB code used to generate the realizations and to estimate the mean vector and covariance matrix is given next.

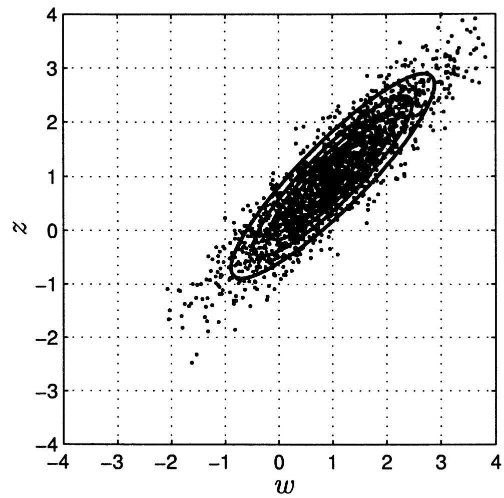


Figure 12.21: 500 outcomes of bivariate Gaussian random vector with mean $[1\,1]^T$ and covariance matrix given by (12.40).

```
randn('state',0) % set random number generator to initial value
G=[1 0;0.9 sqrt(1-0.9^2)]; % define G matrix
M=2000; % set number of realizations
for m=1:M
  x=randn(1,1);y=randn(1,1); % generate realizations of two
                             % independent N(0,1) random variables
  wz=G*[x y]'+[1 1]'; % transform to desired mean and covariance
  WZ(:,m)=wz; % save realizations in 2 x M array
end
Wmeanest=mean(WZ(1,:)); % estimate mean of W
Zmeanest=mean(WZ(2,:)); % estimate mean of Z
WZbar(1,:)=WZ(1,:)-Wmeanest; % subtract out mean of W
WZbar(2,:)=WZ(2,:)-Zmeanest; % subtract out mean of Z
Cest=[0 0;0 0];
for m=1:M
  Cest=Cest+(WZbar(:,m)*WZbar(:,m)')/M; % compute estimate of
                                        % covariance matrix
end
Wmeanest % write out estimate of mean of W
Zmeanest % write out estimate of mean of Z
Cest % write out estimate of covariance matrix
```

12.12 Real-World Example – Optical Character Recognition

An important use of computers is to be able to scan a document and automatically read the characters. For example, bank checks are routinely scanned to ascertain the account numbers, which are usually printed on the bottom. Also, scanners are used to take a page of alphabetic characters and convert the text to a computer file that can later be edited in a computer. In this section we briefly describe how this might be done. A more comprehensive description can be found in [Trier, Jain, and Taxt 1996]. To simplify the discussion we consider recognition of the digits $0, 1, 2, \ldots, 9$ that have been generated by a printer (as opposed to handwritten, the recognition of which is much more complex due to the potential variations of the characters). An example of these characters is shown in Figure 12.22. They were obtained by printing the characters from a computer to a laser printer and then scanning them back into a computer. Note that each digit consists of an 80×80

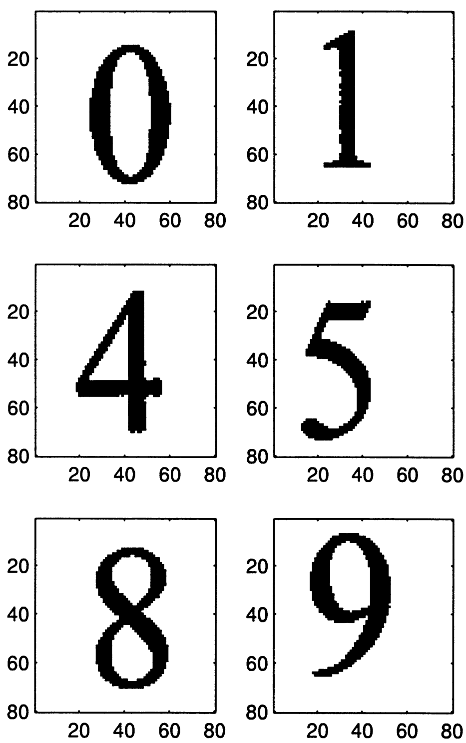


Figure 12.22: Scanned digits for optical character recognition.

array of pixels and each pixel is either black or white. This is termed a *binary* image. A magnified version of the digit "1" is shown in Figure 12.23, where the "pixelation" is clearly evident. Also, some of the black pixels have been omitted due to errors in the scanning process. In order for a computer to be able to recognize and decode

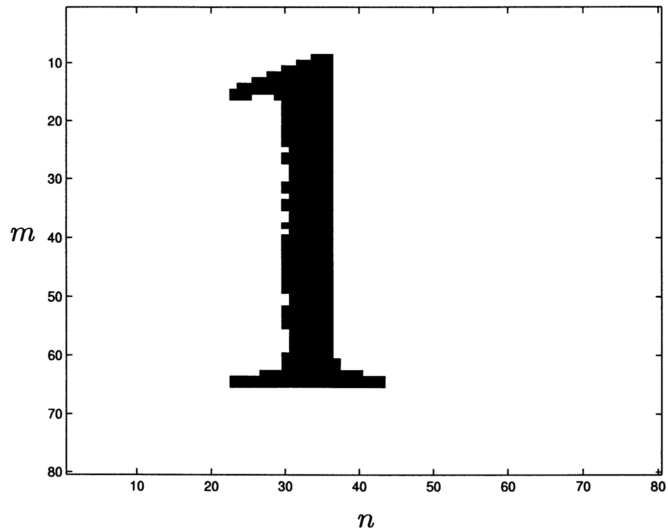


Figure 12.23: Magnified version of the digit "1".

the digits it is necessary to reduce each 80×80 image to a number or set of numbers that characterize the digit. These numbers are called the *features* and they compose a *feature vector* that must be different for each digit. This will allow a computer to distinguish between the digits and be less susceptible to noise effects as is evident in the "1" image. For our example, we will choose only two features, although in practice many more are used. A typical feature based on the geometric character of the digit images is the *geometric moments.* It attempts to measure the distribution of the black pixels and is completely analogous to our usual joint moments. (Recall our motivation of the expected value using the idea of the center of mass of an object in Section 11.3.) Let $g[m,n]$ denote the pixel value at location $[m,n]$ in the image, where $m = 1,2,\ldots,80$, $n = 1,2,\ldots,80$ and either $g[m,n] = 1$ for a black pixel or $g[m,n] = 0$ for a white pixel. Note from Figure 12.23 that the indices for the $[m,n]$ pixel are specified in matrix format, where m indicates the row and n indicates the column. The geometric moments are defined as

$$\mu'[k,l] = \frac{\sum_{m=1}^{80}\sum_{n=1}^{80} m^k n^l g[m,n]}{\sum_{m=1}^{80}\sum_{n=1}^{80} g[m,n]}. \tag{12.54}$$

If we were to define

$$p[m,n] = \frac{g[m,n]}{\sum_{m=1}^{80}\sum_{n=1}^{80} g[m,n]} \qquad m = 1,2,\ldots,80; n = 1,2,\ldots,80$$

then $p[m,n]$ would have the properties of a joint PMF, in that it is nonnegative and sums to one. A somewhat better feature is obtained by using the *central* geometric moments which will yield the same number even as the digit is translated in the horizontal and vertical directions. This may be seen to be of value by referring to Figure 12.22, in which the center of the digits do not all lie at the same location. Using central geometric moments alleviates having to center each digit. The definition is

$$\mu[k,l] = \frac{\sum_{m=1}^{80}\sum_{n=1}^{80}(m-\bar{m})^k(n-\bar{n})^l g[m,n]}{\sum_{m=1}^{80}\sum_{n=1}^{80} g[m,n]} \tag{12.55}$$

where

$$\begin{aligned}
\bar{m} &= \mu'[1,0] = \frac{\sum_{m=1}^{80}\sum_{n=1}^{80} mg[m,n]}{\sum_{m=1}^{80}\sum_{n=1}^{80} g[m,n]} \\
\bar{n} &= \mu'[0,1] = \frac{\sum_{m=1}^{80}\sum_{n=1}^{80} ng[m,n]}{\sum_{m=1}^{80}\sum_{n=1}^{80} g[m,n]}
\end{aligned}$$

The coordinate pair $(\bar{m},\bar{n})$ is the center of mass of the character and is completely analogous to the mean of the "joint PDF" $p[m,n]$.

To demonstrate the procedure by which optical character recognition is accomplished we will add noise to the characters. To simulate a "dropout", in which a black pixel becomes a white one (see Figure 12.23 for an example), we change each black pixel to a white one with a probability of 0.4, and make no change with probability of 0.6. To simulate spurious scanning marks we change each white pixel to a black one with probability of 0.1, and make no change with probability of 0.9. An example of the corrupted digits is shown in Figure 12.24. As a feature vector we will use the pair $(\mu[1,1],\mu[2,2])$. For the digits "1" and "8", 50 realizations of the feature vector are shown in Figure 12.25a. The black square indicates the center of mass $(\bar{m},\bar{n})$ for each digit's feature vector. Note that we could distinguish between the two characters without error if we recognize an outcome as belonging to a "1" if we are below the line boundary shown and as a "8" otherwise. However, for the digits "1" and "3" there is an overlap region where the outcomes could belong to either character as seen in Figure 12.25b. For these digits we could not separate the digits without a large error. The latter is more typically the case and can only be resolved by using a larger dimension feature vector. The interested reader should consult [Duda, Hart, and Stork 2001] for a further discussion of pattern recognition (also called *pattern classification*). Also, note that the digits "3" and "8" would produce outcomes that would overlap greatly. Can you explain why? You might consider some typical scanned digits as shown in Figure 12.26 that have been designed to make recognition easier!

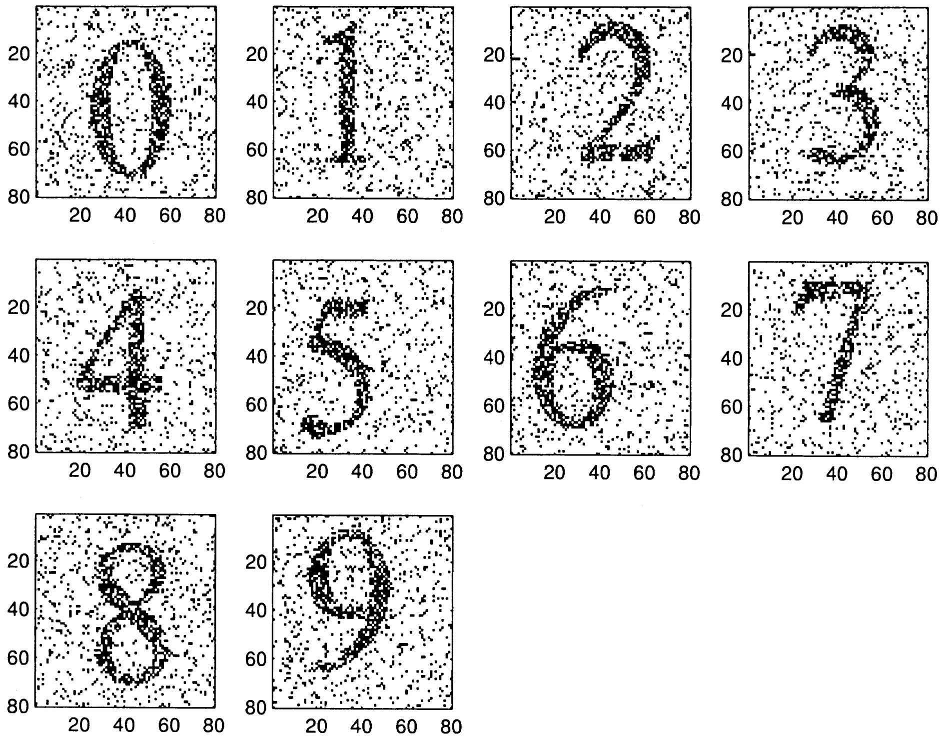


Figure 12.24: Realization of corrupted digits.

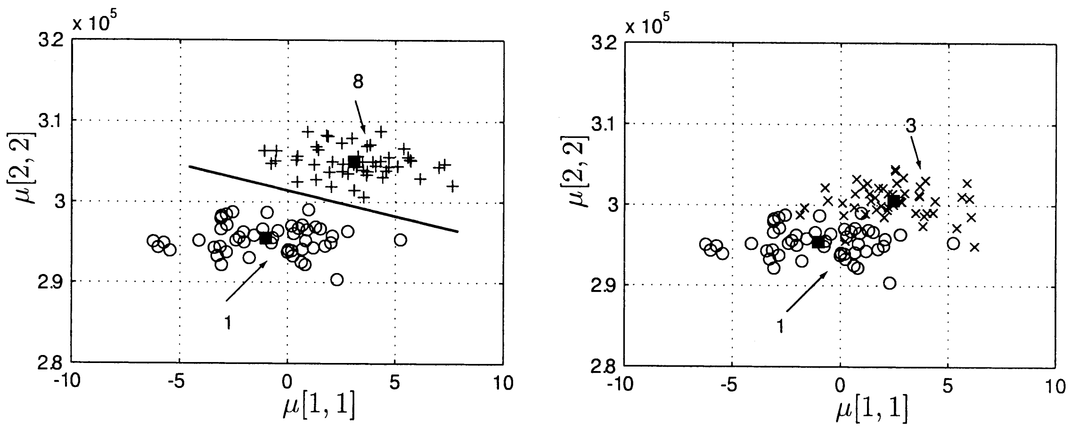


Figure 12.25: 50 realizations of feature vector for two competing digits.

Figure 12.26: Some scanned digits typically used in optical character recognition. They were scanned into a computer, which accounts for the obvious errors.

References

Duda, R.O., P.E. Hart, D.G. Stork, *Pattern Classification, 2nd Ed.*, John Wiley & Sons, New York, 2001.

Golub, G.H., C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1983.

Trier, O.D., A.K. Jain, T. Taxt, "Feature Extraction Methods for Character Recognition – A Survey," *Pattern Recognition*, Vol. 29, No. 4, pp. 641–662, 1996, Pergamon, Elsevier Science.

Problems

12.1 ($\ddot\smile$) **(w)** For the dartboard shown in Figure 12.1 determine the probability that the novice dart player will land his dart in the outermost ring, which has radii $3/4 \leq r \leq 1$. Do this by using geometrical arguments and also using double integrals. Hint: For the latter approach convert to polar coordinates (r, θ) and remember to use $dx\, dy = r dr\, d\theta$.

12.2 (c) Reproduce Figure 12.2a by letting $X \sim \mathcal{U}(-1,1)$ and $Y \sim \mathcal{U}(-1,1)$, where X and Y are independent. Omit any realizations of (X, Y) for which $\sqrt{X^2+Y^2} > 1$. Explain why this produces a uniform distribution of points in the unit circle. See also Problem 13.23 for a more formal justification of this procedure.

12.3 ($\ddot\smile$) **(w)** For the novice dart player is $P[0 \leq R \leq 0.5] = 0.5$ (R is the distance from the center of the dartboard)? Explain your results.

12.4 (w) Find the volume of a cylinder of height h and whose base has radius r by using a double integral evaluation.

12.5 ($\ddot\smile$) **(c)** In this problem we estimate π using probability arguments. Let $X \sim \mathcal{U}(-1,1)$ and $Y \sim \mathcal{U}(-1,1)$ for X and Y independent. First relate $P[X^2+Y^2 \leq 1]$ to the value of π. Then generate realizations of X and Y and use them to estimate π.

12.6 (f) For the joint PDF

$$p_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} & x^2+y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

find $P[|X| \leq 1/2]$. Hint: You will need

$$\int \sqrt{1-x^2}\,dx = \frac{1}{2}x\sqrt{1-x^2} + \frac{1}{2}\arcsin(x).$$

12.7 (⌣̈) (f) If a joint PDF is given by

$$p_{X,Y}(x,y) = \begin{cases} \frac{c}{\sqrt{xy}} & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

find c.

12.8 (w) A point is chosen at random from the sample space $\mathcal{S} = \{(x,y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Find $P[Y \leq X]$.

12.9 (f) For the joint PDF $p_{X,Y}(x,y) = \exp[-(x+y)]u(x)u(y)$, find $P[Y \leq X]$.

12.10 (⌣̈) (w,c) Two persons play a game in which the first person thinks of a number from 0 to 1, while the second person tries to guess player one's number. The second player claims that he is telepathic and knows what number the first player has chosen. In reality the second player just chooses a number at random. If player one also thinks of a number at random, what is the probability that player two will choose a number whose difference from player one's number is less than 0.1? Add credibility to your solution by simulating the game and estimating the desired probability.

12.11 (⌣̈) (f) If (X,Y) has a standard bivariate Gaussian PDF, find $P[X^2+Y^2 = 10]$.

12.12 (f,c) Plot the values of (x,y) for which $x^2 - 2\rho xy + y^2 = 1$ for $\rho = -0.9$, $\rho = 0$, and $\rho = 0.9$. Hint: Solve for y in terms of x.

12.13 (w,c) Plot the standard bivariate PDF in three dimensions for $\rho = 0.9$. Next examine your plot if $\rho \to 1$ and determine what happens. As $\rho \to 1$, can you predict Y based on $X = x$?

12.14 (f) If $p_{X,Y}(x,y) = \exp[-(x+y)]u(x)u(y)$, determine the marginal PDFs.

12.15 (⌣̈) (f) If

$$p_{X,Y}(x,y) = \begin{cases} 2 & 0 < x < 1, 0 < y < x \\ 0 & \text{otherwise} \end{cases}$$

find the marginal PDFs.

12.16 (t) Assuming that $(x, y) \neq (0,0)$, prove that $x^2 - 2\rho xy + y^2 > 0$ for $-1 < \rho < 1$.

12.17 (f) If $p_X(x) = (1/2)\exp[-(1/2)x]u(x)$ and $p_Y(y) = (1/4)\exp[-(1/4)y]u(y)$, find the joint PDF of X and Y.

12.18 (⌣̇̇) (f) Determine the joint CDF if X and Y are independent with

$$p_X(x) = \begin{cases} \frac{1}{2} & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

$$p_Y(y) = \begin{cases} \frac{1}{4} & 0 < y < 4 \\ 0 & \text{otherwise.} \end{cases}$$

12.19 (f) Determine the joint CDF corresponding to the joint PDF

$$p_{X,Y}(x,y) = \begin{cases} xy\exp\left[-\frac{1}{2}(x^2+y^2)\right] & x \geq 0, y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Next verify Properties 12.1–12.6 for the CDF.

12.20 (t) Prove that (12.10) is true if (12.11) is true and vice versa. Hint: Let $A = \{a \leq x \leq b\}$ and $B = \{y : c \leq y \leq d\}$ for the first part and let $A = \{x : x_0 - \Delta x/2 \leq x \leq x_0 + \Delta x/2\}$ and $B = \{y : y_0 - \Delta y/2 \leq y \leq y_0 + \Delta y/2\}$ with x_0 and y_0 arbitrary for the second part.

12.21 (t) Prove that (12.11) and (12.12) are equivalent.

12.22 (w) Two independent speech signals are added together. If each one has a Laplacian PDF with parameter σ^2, what is the power of the resultant signal?

12.23 (⌣̇̇) (w) Lightbulbs fail with a time to failure modeled as an exponential random variable with a mean time to failure of 1000 hours. If two lightbulbs are used to illuminate a room, what is the probability that both bulbs will fail before 2000 hours? Assume that the failure time of one bulb does not affect the failure time of the other bulb.

12.24 (f) If a joint PDF is given as $p_{X,Y}(x,y) = 6\exp[-(2x+3y)]u(x)u(y)$, what is the probability of $A = \{(x,y) : 0 < x < 2, 0 < y < 1\}$? Are the two random variables independent?

12.25 (⌣̇̇) (w) A joint PDF is uniform over the region $\{(x,y) : 0 \leq y < x, 0 \leq x < 1\}$ and zero elsewhere. Are X and Y independent?

12.26 (⌣̇̇) (w) The temperature in Antarctica is modeled as a random variable $X \sim \mathcal{N}(20, 1500)$ degrees Fahrenheit, while that in Ecuador is modeled also as a random variable with $Y \sim \mathcal{N}(100, 100)$ degrees Fahrenheit. What is the probability that it will be hotter in Antarctica than in Ecuador? Assume the random variables are independent.

12.27 (w,c) In Section 2.3 we discussed the outcomes resulting from adding together two random variables uniform on $(0,1)$. We claimed that the probability of 500 outcomes in the interval $[0,0.5]$ and 500 outcomes in the interval $[1.5,2]$ resulting from a total of 1000 outcomes is

$$\binom{1000}{500}\left(\frac{1}{8}\right)^{1000} \approx 2.2 \times 10^{-604}.$$

Can you now justify this result? What assumptions are implicit in its calculation? Hint: For each trial consider the 3 possible outcomes $(0,0.5)$, $[0.5,1.5)$, and $[1.5,2)$. Also, see Problem 3.48 on how to evaluate expressions with large factorials.

12.28 (f) Find the PDF of $X = U_1 + U_2$, where $U_1 \sim \mathcal{U}(0,1)$, $U_2 \sim \mathcal{U}(0,1)$, and U_1, U_2 are independent. Use a convolution integral to do this.

12.29 (w) In this problem we show that the ratio of areas for the linear transformation

$$\begin{bmatrix} w \\ z \end{bmatrix} = \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\boldsymbol{\xi}}$$

is $|\det(\mathbf{G})|$. To do so let $\boldsymbol{\xi} = [x\,y]^T$ take on values in the region $\{(x,y) : 0 \le x \le 1, 0 \le y \le 1\}$ as shown by the shaded area in Figure 12.27. Then, consider a point in the unit square to be represented as $\boldsymbol{\xi} = \alpha \mathbf{e}_1 + \beta \mathbf{e}_2$, where $0 \le \alpha \le 1$, $0 \le \beta \le 1$, $\mathbf{e}_1 = [1\,0]^T$, and $\mathbf{e}_2 = [0\,1]^T$. The transformed vector is

$$\begin{aligned}
\mathbf{G}\boldsymbol{\xi} &= \mathbf{G}(\alpha \mathbf{e}_1 + \beta \mathbf{e}_2) \\
&= \alpha \mathbf{G}\mathbf{e}_1 + \beta \mathbf{G}\mathbf{e}_2 \\
&= \alpha \begin{bmatrix} a \\ c \end{bmatrix} + \beta \begin{bmatrix} b \\ d \end{bmatrix}.
\end{aligned}$$

It is seen that the natural basis vectors $\mathbf{e}_1, \mathbf{e}_2$ map into the vectors $[a\,c]^T$, $[b\,d]^T$, which appear as shown in Figure 12.27. The region in the w-z plane that results from mapping the unit square is shown as shaded. The area of the parallelogram can be found from Figure 12.28 as BH. Determine the ratio of areas to show that

$$\frac{\text{Area in } w\text{-}z \text{ plane}}{\text{Area in } x\text{-}y \text{ plane}} = ad - bc = \det(\mathbf{G}).$$

The absolute value is needed since if for example $a < 0, b < 0$, the parallelogram will be in the second quadrant and its determinant will be negative. The absolute value sign takes care of all the possible cases.

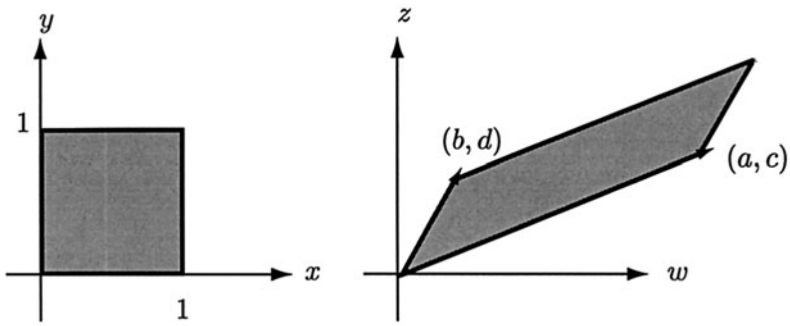


Figure 12.27: Mapping of areas for linear transformation.

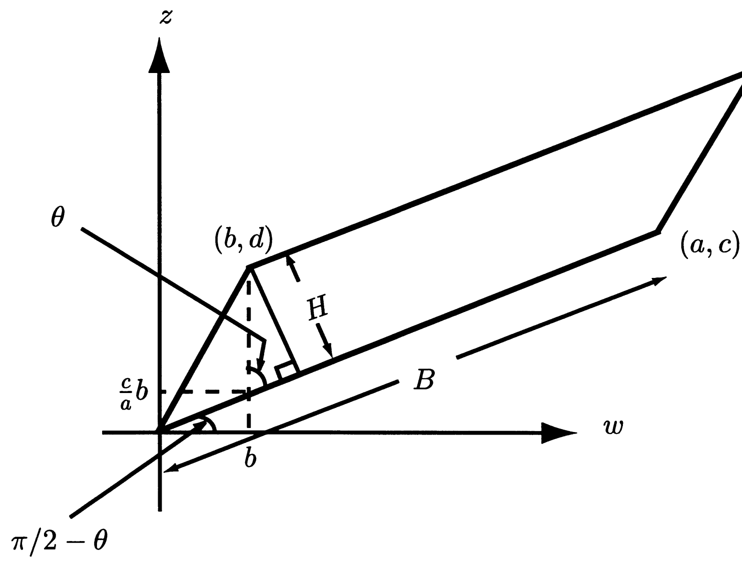


Figure 12.28: Geometry to determine area of parallelogram.

12.30 (⌣̈) (w,c) The champion dart player described in Section 12.3 is able to land his dart at a point (x, y) according to the joint PDF

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/64 & 0 \\ 0 & 1/64 \end{bmatrix}\right)$$

with some outcomes shown in Figure 12.2b. Determine the probability of a bullseye. Next simulate the game and plot the outcomes. Finally estimate the probability of a bullseye using the results of your computer simulation.

12.31 (t) Show that (12.19) can be written as (12.20).

12.32 (t) Consider the nonlinear transformation $w = g(x, y), z = h(x, y)$. Use a tangent approximation to both functions about the point (x_0, y_0) to express $[w\, z]^T$ as an approximate affine function of $[x\, y]^T$, and use matrix/vector notation. For example,

$$w = g(x,y) \approx g(x_0, y_0) + \left.\frac{\partial g}{\partial x}\right|_{\substack{x=x_0\\y=y_0}} (x - x_0) + \left.\frac{\partial g}{\partial y}\right|_{\substack{x=x_0\\y=y_0}} (y - y_0)$$

and similarly for $z = h(x, y)$. Compare the matrix to the Jacobian matrix of (12.23).

12.33 (f) If a joint PDF is given as $p_{X,Y}(x,y) = (1/4)^2 \exp[-\frac{1}{2}(|x| + |y|)]$ for $-\infty < x < \infty, -\infty < y < \infty$, find the joint PDF of

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

12.34 (f) If a joint PDF is given as $p_{X,Y}(x,y) = \exp[-(x+y)]u(x)u(y)$, find the joint PDF of $W = XY, Z = Y/X$.

12.35 (w,c) Consider the nonlinear transformation

$$\begin{aligned} W &= X^2 + 5Y^2 \\ Z &= -5X^2 + Y^2. \end{aligned}$$

Write a computer program to plot in the x-y plane the points (x_i, y_j) for $x_i = 0.95 + (i-1)/100$ for $i = 1, 2, \ldots, 11$ and $y_j = 1.95 + (j-1)/100$ for $j = 1, 2, \ldots, 11$. Next transform all these points into the w-z plane using the given nonlinear transformation. What kind of figure do you see? Next calculate the area of the figure (you can use a rough approximation based on the computer generated figure output) and finally take the ratio of the areas of the figures in the two planes. Does this ratio agree with the Jacobian factor

$$\left|\det\left(\frac{\partial(w,z)}{\partial(x,y)}\right)\right|$$

when evaluated at $x = 1$, $y = 2$?

12.36 (⌣̈) **(f)** Find the marginal PDFs of the joint PDF given in (12.25).

12.37 (f) Determine the marginal PDFs for the joint PDF given by

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 & 4 \\ 4 & 6 \end{bmatrix}\right).$$

12.38 (⌣) **(f)** If X and Y have the joint PDF

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\right)$$

find the joint PDF of the transformed random vector

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

12.39 (t) Prove that the PDF of $Z = Y/X$, where X and Y are independent, is given by

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(x) p_Y(xz)|x|dx.$$

12.40 (t) Prove that the PDF of $Z = XY$, where X and Y are independent is given by

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(x) p_Y(z/x) \frac{1}{|x|} dx.$$

12.41 (c) Generate outcomes of a Cauchy random variable using Y/X, where $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$ and X and Y are independent. Can you explain what happens when the Cauchy outcome becomes very large in magnitude?

12.42 (t) Prove that $s(t) = A\cos(2\pi F_0 t) + B\sin(2\pi F_0 t)$ can be written as $s(t) = \sqrt{A^2 + B^2}\cos(2\pi F_0 t - \arctan(B/A))$. Hint: Convert (A, B) to polar coordinates.

12.43 (⌣) **(w)** A particle is subject to a force in a random force field. If the velocity of the particle is modeled in the x and y directions as $V_x \sim \mathcal{N}(0, 10)$ and $V_y \sim \mathcal{N}(0, 10)$ meters/sec, and V_x and V_y are assumed to be independent, how far will the particle move on the average in 1 second?

12.44 (f) Prove that if X and Y are independent standard Gaussian random variables, then $X^2 + Y^2$ will have a χ^2_2 PDF.

12.45 (⌣) **(w,f)** Two independent random variables X and Y have zero means and variances of 1. If they are linearly transformed as $W = X + Y, Z = X - Y$, find the covariance between the transformed random variables. Are W and Z uncorrelated? Are W and Z independent?

12.46 (f) If

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)$$

determine the mean of $X + Y$ and the variance of $X + Y$.

12.47 (⌣) **(w)** The random vector $[X\,Y]^T$ has a covariance matrix

$$\mathbf{C} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Find a 2×2 matrix $\mathbf{G}$ so that $\mathbf{G}[X\,Y]^T$ is a random vector with uncorrelated components.

12.48 (t) Prove that if a random vector has a covariance matrix

$$\mathbf{C} = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$$

then the matrix

$$\mathbf{G} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

can always be used to diagonalize it. Show that the effect of this matrix transformation is to rotate the point (x, y) by 45° and relate this back to the contours of a standard bivariate Gaussian PDF.

12.49 (f) Find the MMSE estimator of Y based on observing $X = x$ if (X, Y) has the joint PDF

$$p_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{0.19}} \exp\left[-\frac{1}{2(0.19)}(x^2 - 1.8xy + y^2)\right].$$

Also, find the PDF of the error $Y - \hat{Y} = Y - (aX + b)$, where a, b are the optimal values. Hint: See Theorem 12.7.1.

12.50 (w,c) A random signal voltage $V \sim \mathcal{N}(1, 1)$ is corrupted by an independent noise sample N, where $N \sim \mathcal{N}(0, 2)$, so that $V + N$ is observed. It is desired to estimate the signal voltage as accurately as possible using a linear MMSE estimator. Assuming that V and N are independent, find this estimator. Then plot the constant PDF contours for the random vector $(V + N, V)$ and indicate the estimated values on the plot.

12.51 (f) Using a convolution integral prove that if X and Y are independent standard Gaussian random variables, then $X + Y \sim \mathcal{N}(0, 2)$.

12.52 (⌣) **(f)** If

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

find $P[X + Y > 2]$.

12.53 (t) Prove that if $X \sim \Gamma(\alpha_X, \lambda)$ and $Y \sim \Gamma(\alpha_Y, \lambda)$ and X and Y are independent, then $X + Y \sim \Gamma(\alpha_X + \alpha_Y, \lambda)$.

12.54 (f) To generate two independent standard Gaussian random variables on a computer one can use the *Box-Mueller transform*

$$\begin{aligned} X &= \sqrt{-2\ln U_1}\cos(2\pi U_2) \\ Y &= \sqrt{-2\ln U_1}\sin(2\pi U_2) \end{aligned}$$

where U_1, U_2 are both uniform on $(0,1)$ and independent of each other. Prove that this result is true. Hint: To find the inverse transformation use a polar coordinate transformation.

12.55 (t) Prove that by using the eigendecomposition of a covariance matrix or $\mathbf{V}^T\mathbf{C}\mathbf{V} = \mathbf{\Lambda}$ that one can factor $\mathbf{C}$ as $\mathbf{C} = \mathbf{G}\mathbf{G}^T$, where $\mathbf{G} = \mathbf{V}\sqrt{\mathbf{\Lambda}}$, and $\sqrt{\mathbf{\Lambda}}$ is defined as the matrix obtained by taking the positive square roots of all the elements. Recall that $\mathbf{\Lambda}$ is a diagonal matrix with positive elements on the main diagonal. Next find $\mathbf{G}$ for the covariance matrix

$$\mathbf{C} = \begin{bmatrix} 26 & 6 \\ 6 & 26 \end{bmatrix}$$

and verify that $\mathbf{G}\mathbf{G}^T$ does indeed produce $\mathbf{C}$.

12.56 (c) Simulate on the computer realizations of the random vector

$$\begin{bmatrix} W \\ Z \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}\right).$$

Plot these realizations as well as the contours of constant PDF on the same graph.