

measure zero. Because the exceptions occupy a negligible amount of space, they can be safely ignored for many applications. Some important results in probability theory hold for all discrete values but only hold “almost everywhere” for continuous values.

Another technical detail of continuous variables relates to handling continuous random variables that are deterministic functions of one another. Suppose we have two random variables,  $\mathbf{x}$  and  $\mathbf{y}$ , such that  $\mathbf{y} = g(\mathbf{x})$ , where  $g$  is an invertible, continuous, differentiable transformation. One might expect that  $p_y(\mathbf{y}) = p_x(g^{-1}(\mathbf{y}))$ . This is actually not the case.

As a simple example, suppose we have scalar random variables  $x$  and  $y$ . Suppose  $y = \frac{x}{2}$  and  $x \sim U(0, 1)$ . If we use the rule  $p_y(y) = p_x(2y)$  then  $p_y$  will be 0 everywhere except the interval  $[0, \frac{1}{2}]$ , and it will be 1 on this interval. This means

$$\int p_y(y)dy = \frac{1}{2}, \quad (3.43)$$

which violates the definition of a probability distribution. This is a common mistake. The problem with this approach is that it fails to account for the distortion of space introduced by the function  $g$ . Recall that the probability of  $\mathbf{x}$  lying in an infinitesimally small region with volume  $\delta\mathbf{x}$  is given by  $p(\mathbf{x})\delta\mathbf{x}$ . Since  $g$  can expand or contract space, the infinitesimal volume surrounding  $\mathbf{x}$  in  $\mathbf{x}$  space may have different volume in  $\mathbf{y}$  space.

To see how to correct the problem, we return to the scalar case. We need to preserve the property

$$|p_y(g(x))dy| = |p_x(x)dx|. \quad (3.44)$$

Solving from this, we obtain

$$p_y(y) = p_x(g^{-1}(y)) \left| \frac{\partial x}{\partial y} \right| \quad (3.45)$$

or equivalently

$$p_x(x) = p_y(g(x)) \left| \frac{\partial g(x)}{\partial x} \right|. \quad (3.46)$$

In higher dimensions, the derivative generalizes to the determinant of the **Jacobian matrix**—the matrix with  $J_{i,j} = \frac{\partial x_i}{\partial y_j}$ . Thus, for real-valued vectors  $\mathbf{x}$  and  $\mathbf{y}$ ,

$$p_x(\mathbf{x}) = p_y(g(\mathbf{x})) \left| \det \left( \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right) \right|. \quad (3.47)$$