This procedure for transforming densities can be very powerful. Any density $p(y)$ can be obtained from a fixed density $q(x)$ that is everywhere non-zero by making a nonlinear change of variable $y = f(x)$ in which $f(x)$ is a monotonic function so that $0 \leqslant f'(x) < \infty$.

Exercise 2.19

One consequence of the transformation property (2.71) is that the concept of the maximum of a probability density is dependent on the choice of variable. Suppose $f(x)$ has a mode (i.e., a maximum) at $\widehat{x}$ so that $f'(\widehat{x}) = 0$. The corresponding mode of $\widetilde{f}(y)$ will occur for a value $\widehat{y}$ obtained by differentiating both sides of (2.70) with respect to y:

$$\widetilde{f}'(\widehat{y}) = f'(g(\widehat{y}))g'(\widehat{y}) = 0. \tag{2.72}$$

Assuming $g'(\widehat{y}) \neq 0$ at the mode, then $f'(g(\widehat{y})) = 0$. However, we know that $f'(\widehat{x}) = 0$, and so we see that the locations of the mode expressed in terms of each of the variables x and y are related by $\widehat{x} = g(\widehat{y})$, as one would expect. Thus, finding a mode with respect to the variable x is equivalent to first transforming to the variable y, then finding a mode with respect to y, and then transforming back to x.

Now consider the behaviour of a probability density $p_x(x)$ under the change of variables $x = g(y)$, where the density with respect to the new variable is $p_y(y)$ and is given by (2.71). To deal with the modulus in (2.71) we can write $g'(y) = s|g'(y)|$ where $s \in \{-1, +1\}$. Then (2.71) can be written as

$$p_y(y) = p_x(g(y))sg'(y)$$

where we have used $1/s = s$. Differentiating both sides with respect to y then gives

$$p'_y(y) = sp'_x(g(y))\{g'(y)\}^2 + sp_x(g(y))g''(y). \tag{2.73}$$

Due to the presence of the second term on the right-hand side of (2.73), the relationship $\widehat{x} = g(\widehat{y})$ no longer holds. Thus, the value of x obtained by maximizing $p_x(x)$ will not be the value obtained by transforming to $p_y(y)$ then maximizing with respect to y and then transforming back to x. This causes modes of densities to be dependent on the choice of variables. However, for a linear transformation, the second term on the right-hand side of (2.73) vanishes, and so in this case the location of the maximum transforms according to $\widehat{x} = g(\widehat{y})$.

This effect can be illustrated with a simple example, as shown in Figure 2.12. We begin by considering a Gaussian distribution $p_x(x)$ over x shown by the red curve in Figure 2.12. Next we draw a sample of $N = 50{,}000$ points from this distribution and plot a histogram of their values, which as expected agrees with the distribution $p_x(x)$. Now consider a nonlinear change of variables from x to y given by

$$x = g(y) = \ln(y) - \ln(1 - y) + 5. \tag{2.74}$$

The inverse of this function is given by

$$y = g^{-1}(x) = \frac{1}{1 + \exp(-x + 5)}, \tag{2.75}$$

which is a *logistic sigmoid* function and is shown in Figure 2.12 by the blue curve.