

We are often interested in finding the most probable value for some quantity. In the case of probability distributions over discrete variables this poses little problem. However, for continuous variables there is a subtlety arising from the nature of probability densities and the way they transform under non-linear changes of variable.

Consider first the way a function $f(x)$ behaves when we change to a new variable y where the two variables are related by $x = g(y)$. This defines a new function of y given by

$$\tilde{f}(y) = f(g(y)). \quad (2)$$

Suppose $f(x)$ has a mode (i.e. a maximum) at \hat{x} so that $f'(\hat{x}) = 0$. The corresponding mode of $\tilde{f}(y)$ will occur for a value \hat{y} obtained by differentiating both sides of (2) with respect to y

$$\tilde{f}'(\hat{y}) = f'(g(\hat{y}))g'(\hat{y}) = 0. \quad (3)$$

Assuming $g'(\hat{y}) \neq 0$ at the mode, then $f'(g(\hat{y})) = 0$. However, we know that $f'(\hat{x}) = 0$, and so we see that the locations of the mode expressed in terms of each of the variables x and y are related by $\hat{x} = g(\hat{y})$, as one would expect. Thus, finding a mode with respect to the variable x is completely 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 ((1.27)). Let us write $g'(y) = s|g'(y)|$ where $s \in \{-1, +1\}$. Then ((1.27)) can be written

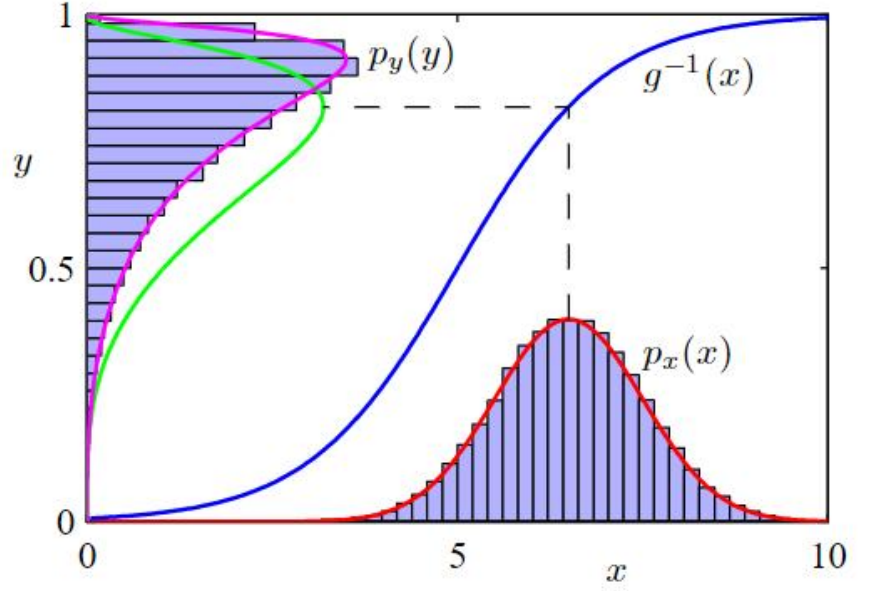
$$p_y(y) = p_x(g(y))sg'(y).$$

Differentiating both sides with respect to y then gives

$$p'_y(y) = sp'_x(g(y))\{g'(y)\}^2 + sp_x(g(y))g''(y). \quad (4)$$

Due to the presence of the second term on the right hand side of (4) the relationship $\hat{x} = g(\hat{y})$ 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. In the case of linear transformation, the second term on

Example of the transformation of the mode of a density under a non-linear change of variables, illustrating the different behaviour compared to a simple function. See the text for details.



the right hand side of (4) vanishes, and so the location of the maximum transforms according to $\hat{x} = g(\hat{y})$.

This effect can be illustrated with a simple example, as shown in Figure 1. We begin by considering a Gaussian distribution $p_x(x)$ over x with mean $\mu = 6$ and standard deviation $\sigma = 1$, shown by the red curve in Figure 1. 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 non-linear change of variables from x to y given by

$$x = g(y) = \ln(y) - \ln(1 - y) + 5. \quad (5)$$

The inverse of this function is given by

$$y = g^{-1}(x) = \frac{1}{1 + \exp(-x + 5)} \quad (6)$$

which is a *logistic sigmoid* function, and is shown in Figure 1 by the blue curve.

If we simply transform $p_x(x)$ as a function of x we obtain the green curve $p_x(g(y))$ shown in Figure 1, and we see that the mode of the density $p_x(x)$ is transformed via the sigmoid function to the mode of this curve. However, the density over y transforms instead according to (1.27) and is shown by the magenta curve on the left side of the diagram. Note that this has its mode shifted relative to the mode of the green curve.

To confirm this result we take our sample of 50,000 values of x , evaluate the corresponding values of y using (6), and then plot a histogram of their values. We see that this histogram matches the magenta curve in Figure 1 and not the green curve!