OPTIMISATION OF BIVARIATE FUNCTIONS

Unconstrained Optimisation of a Bivariate Function

In the case of a continuous function $f(x)$ of a single variable x, a maximum or a minimum point is a stationary point on the graph of the function where the first derivative of the function is zero-valued. The function $f(x)$ defines a one-dimensional entity which is plotted on a two-dimensional graph.

A continuous bivariate function $z = f(x, y)$ can be envisaged as a surface—which is a two-dimensional entity—residing in a three-dimensional space. A maximum or a minimum point of the function—which is an extreme value in other words—corresponds to a peak or a trough in the surface.

These situations are easily envisaged since they can be represented in terms of the topology of our everyday experience. Thus, the task of maximising a continuous bivariate function of an unknown mathematical form can be compared to that of climbing a hill in a dense fog without the benefit of a map in the faith that the ground is smooth and that there are no pitfalls or cliffs to endanger the unwary climber. However, as befits a scientific expedition, we may assume that a compass is provided and that the directions taken and the distances travelled are accurately recorded.

The intuition which can be derived by studying the case of a bivariate function can be employed in cases of multivariate functions; and, therefore, it is appropriate to consider the bivariate case in detail.

Much of the analysis of multivariate functions is conducted by treating a succession of their univariate components. Thus, in the case of the bivariate function, we may begin the analysis by taking two cross sections of the function which are at right angles. These give rise to a pair of ordinary two-dimensional graphs.

It simplifies matters if the sections are taken in directions parallel to the zx and zy planes. Then the two cross-sectional functions may be denoted by $z = f(x, b)$ and $z = f(b, y)$, where a and b are fixed values of x and y respectively.

The derivatives of these functions are called partial derivatives. The partial derivative of the function $f(x, y)$ in respect of x is denoted by $f_x(x, y)$ or by $\partial f/\partial x$. The value of $f_x(x, y)$ when $(x, y) = (a, b)$ is denoted by $f_x(a, b)$; and it is defined formally by

$$f_x(a, b) = \lim_{h\to 0} \frac{f(a + h, b) - f(a, b)}{h}. \tag{1}$$

The partial derivative in respect of y, which pertains to the other cross section, is defined similarly.

Successive partial derivatives may be defined as in the case of the derivatives of a function of a single variable. However, since $\partial z/\partial x$ is a function of two variables, there will be two second-order derivatives. These are

$$(2) \qquad f_{xx}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) \quad \text{and} \quad f_{yx}(x,y) = \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right).$$

They may be abbreviated as $\partial^2 f/\partial x^2$ and $\partial^2 f/\partial x \partial x$ respectively.

We might expect a bivariate function to have four distinct partial derivatives denoted by f_{xx}, f_{yx}, f_{xy} and f_{yy}. However, it is a fundamental theorem of multivariate analysis that

(3) If the partial derivatives $f_{xy}(x,y)$ and $f_{yx}(x,y)$ of the bivariate function $f(x,y)$ exist and are continuous everywhere in an open rectangle containing the point (a,b), then their values are equal at (a,b).

A proof of this may be found in most books of mathematical analysis of at intermediate or advanced level.

Conditions for an Optimum

A sure way of climbing a hill in the fog is to proceed in a fixed direction until one ceases to ascend. Then, if one has not yet reached the top, one may continue the ascent in a direction at right angles to the previous direction. One can imagine that, if the hill were shaped more or less as a inverted pudding basin, then, regardless of the original direction, we should need to turn through a right angle only once before reaching the summit. Moreover, with a hill of this shape, we would know that we had reached the summit simply by noting that the ground falls away in both directions along the *N–S* and the *E–W* axes. On a hill of a more complicated shape, in would not be sufficient to check these two axes alone—one would have to be assured that the ground falls away in every direction.

The conditions for having reached the summit can be expressed easily in mathematical form. Let the hill be described by the function $z = f(x,y)$ and let the *N–S* and the *E–W* directions correspond to the directions of the x and the y axes. Let the point at which we are standing be denoted by (a,b). Then

(4) For the point (a,b) to correspond to a maximum of the function $f(x,y)$, it is necessary, (but not sufficient) that (i) $f_x(a,b) = 0$, $f_{xx}(a,b) < 0$ and that (ii) $f_y(a,b) = 0$, $f_{yy}(a,b) < 0$.

The conditions under (i) and (ii) are simply the conditions for the maximisation of the univariate functions $z = f(x,b)$ and $z = f(a,y)$ respectively.

The first-order condition of (i) assures us that, for small departures from (a, b) in either direction along the x axis, we should find that $dz = f_x(x, b)dx = 0$. Likewise the first-order condition of (ii) insures that, along of the y axis, we should find that $dz = f_y(a, y)dy = 0$. However, for (a, b) to be a stationary point, it is required that $dz = 0$ for small departures in every direction. This is also assured, since it follows from the existing first-order conditions that

$$dz = \frac{\partial z(a,b)}{\partial x}dx + \frac{\partial z(a,b)}{\partial y}dy = 0 \tag{5}$$

for all dx, dy in the neighbourhood of zero.

The second-order conditions given under (4) are clearly insufficient to insure that (a, b) is a maximising point. Thus, when we think of the problem of reaching the top of a hill, we recognise that, although the ground may fall away along both the N–S and the E–W axes, it may yet be rising in the direction of one or other of intermediate points of the compass. In the corresponding mathematical framework, what is required is that the second derivative of a univariate cross-sectional function should be negative regardless of the direction in which the section is taken. In terms of the differential dz, the condition is that $d(dz) < 0$ for all values of dx and dy. The second-order differential is given by

$$\begin{aligned} d(dz) &= \frac{\partial(dz)}{\partial x}dx + \frac{\partial(dz)}{\partial y}dy \\ &= \frac{\partial}{\partial x}\big(f_x dx + f_y dy\big)dx + \frac{\partial}{\partial y}\big(f_x dx + f_y dy\big)dy \\ &= f_{xx}d^2x + 2f_{xy}dxdy + f_{yy}d^2y. \end{aligned} \tag{6}$$

This is a quadratic function in the variables dx and dy. The condition that the point (a, b) corresponds to a maximum is the condition that, when the various second-order derivatives are fixed at the values attained at (a, b), the quadratic is negative-valued for all values of dx and dy.

To make this condition more tractable, we must must reform the final expression of (6) by completing the square. It can be see that

$$\begin{aligned} &f_{xx}d^2x + 2f_{xy}dxdy + f_{yy}d^2y \\ &\quad= f_{xx}\Big(d^2x + 2\frac{f_{xy}}{f_{xx}}dxdy\Big) + f_{yy}d^2y \\ &\quad= f_{xx}\Big(d_{xx} + \frac{f_{xy}}{f_{xx}}dy\Big)^2 + \frac{1}{f_{xx}}\Big(f_{xx}f_{yy} - f_{xy}^2\Big)d^2y. \end{aligned} \tag{7}$$

This is a sum comprising two squared terms.

In order for the expression to have a negative value for all values of dx and dy, it is necessary and sufficient that the coefficients associated with the squares should both be negative. Therefore, the first requirement is that $f_{xx} < 0$. If this condition is met, then, also, $1/f_{xx} < 0$; and so the second requirement is that $f_{xx}f_{yy} - f_{xy}^2 > 0$. Observe, however, that f_{xx} and $-f_{xy}^2$ are both negative; and so, if the requirement it to have any chance of being met, it is necessary that $f_{yy} < 0$. This reasoning leads us to the following conclusion:

(8) For the point (a, b) to correspond to a *maximum* of the function $f(x, y)$, it is necessary and sufficient that, when they are evaluated at this point, the derivatives should fulfil the following conditions:

$$\begin{aligned}&\text{(i)} \quad f_x = f_y = 0 \quad \text{and}\\&\text{(ii)} \quad f_{xx},\ f_{yy} < 0 \quad \text{and} \quad f_{xx}f_{yy} - f_{xy}^2 > 0.\end{aligned}$$

It is notable that conditions which are both necessary and sufficient are obtained by supplementing the necessary conditions of (4) by the single requirement that $f_{xx}f_{yy} - f_{xy}^2 > 0$. At a later stage, we shall be able to express this requirement as the condition that a certain two-by-two matrix of partial derivatives fufills a condition of positive definiteness.

So far, we have dealt exclusively with the problem of *maximising* a function $f(x, y)$. It is as important to derive the conditions for *minimising* a bivariate function. This can be achieved by making a single modification to the arguments above. Now, in place of the condition that $d(dz) < 0$, we require, for a minimum, that it should the case that $d(dz) > 0$ for all dx and dz in a neighbourhood of zero. It follows immediately from the final expression of equation (7) that

(9) For the point (a, b) to correspond to a *minimum* of the function $f(x, y)$, it is necessary and sufficient that, when they are evaluated at this point, the derivatives should fulfil the following conditions:

$$\begin{aligned}&\text{(i)} \quad f_x = f_y = 0 \quad \text{and}\\&\text{(ii)} \quad f_{xx},\ f_{yy} > 0 \quad \text{and} \quad f_{xx}f_{yy} - f_{xy}^2 > 0.\end{aligned}$$

Notice that, in comparison with (8), it is only the conditions affecting f_{xx} and f_{yy} on their own which have changed.