24. Bivariate functions and their extrema: a graphical approach

So far we have met three kinds of functions, namely, ordinary functions, index functions and sequences. In all three cases, function means a rule for unambiguously assigning LABELS (in a range) to THINGS (in a domain). For ordinary functions, both THINGS and LABELS are real numbers. For sequences, the THINGS are integers and the LABELS are real numbers. For index functions, the THINGS are function-subdomain pairs and the LABELS are again real numbers. Even if the LABELS are real numbers, however, not every function belongs to one of the above three categories. In this lecture we meet a fourth kind of function, for which THINGS are number pairs. We will call such functions **bivariate**.

A bivariate function, like an ordinary function, is most readily defined in terms of its graph, which is still a plot of all possible (THING, LABEL) pairs. But because each THING is now a pair of numbers, each (THING, LABEL) pair is now a triad of numbers. Each of these numbers is called a **coordinate**. The first coordinate is measured along a horizontal axis, the second coordinate along a horizontal axis perpendicular to the first axis, and the third coordinate along a vertical axis. The first two coordinates represent THING, the third coordinate represents LABEL. Thus THING corresponds to a point in a horizontal plane, and (THING, LABEL) to a point distance LABEL vertically above it, with due regard for sign (meaning vertical distance |LABEL| below the plane if LABEL < 0). In other words, the graph of a bivariate function is a two-dimensional surface in three-dimensional space (whereas the graph of an ordinary function is a one-dimensional curve in two-dimensional space). If x, y, z are the three coordinates and P denotes the function, we write $z = P(x, y)$ for the graph of P. Its domain is most often a rectangle, of the form $\{(x, y) \mid a \le x \le b, c \le y \le d\}$ — for which the standard shorthand, to be used henceforth, is $[a, b] \times [c, d]$ — but it could also be any other region in the horizontal plane (e.g., a circular disk).

Consider, for example, the energy consumption rate, or power, required to sustain a bird in steady flight at the speed that minimizes fuel consumption over a given distance. The heavier the bird, the greater the power; the wider the wings, the lesser the power. Thus, all other things being equal, required power depends on mass and wing span. It is therefore represented by a bivariate function.

BIRD		MASS, x (kg)	SPAN, y (m)	POWER, z (kW)
Sterna hirundo	Common tern	0.118	0.83	0.00135
Crex crex	Corncrake	0.155	0.48	0.00484
Garrulus glandarius	Jay	0.160	0.54	0.00428
Columba palumbus	Woodpigeon	0.495	0.75	0.01717

Table 24.1 Some representative avian masses and wing spans. Source: Rayner (1977, p. 442)

Let P be this function, and let the power required to sustain an x-kilogram bird with a y-meter wing span be z kilowatts. Typical avian masses and wing spans appear in Table 3, alongside the corresponding required powers from a model devised by Rayner (1977). Thus, for example, a corncrake can sustain flight by consuming energy at less than 5 watts, but a woodpigeon requires at least 17. Because every mass or wingspan in Table 1 lies between 0.1 and 1, let us restrict P's domain to the rectangle $[0.1, 1] \times [0.1, 1]$. Then Figure 1 shows $z = P(x, y)$ according to Rayner's model. The line of sight is toward the point with coordinates (0, 0, 0), or **origin** of coordinates (in such a way that if the origin were one corner of a cube, then the viewpoint would be the opposite one). Note that the surface is highest where $(x, y) = (1, 0.1)$ and lowest where $(x, y) = (0.1, 1)$. So, because the definitions of maximum and minimum are exactly the same for bivariate as for ordinary functions, the global maximum and minimum of P are $P(1, 0.1) = 1.138$ kW and $P(0.1, 1) = 0.7756 \times 10^{-3}$ kW (= 0.7756 watts), respectively.

Because a vertical plane intersects a surface in a two-dimensional curve, a bivariate function reduces to an ordinary function when we fix the first or second coordinate. In particular, if we fix mass then power becomes an ordinary function of wing span, whereas if we fix wing span then power becomes an ordinary function of mass. To illustrate, in Figure 2 we show the ordinary-function graphs that correspond to the edges of P's graph, namely, $z = P(x, 0.1)$, $z = P(0.1, y)$, $z = P(x, 1)$ and $z = P(1, y)$.

Bivariate functions often arise in the context of fitting cumulative distribution functions to data. For example, in Exercise 19.14 we introduced a distribution whose c.d.f. on $[0, \infty)$ is defined by

$$F(t) = \begin{cases} At(c-t) + \dfrac{\theta + Ac^2}{\theta+1}\left(\dfrac{t}{c}\right) & \text{if } 0 \le t \le c \\ 1 - \dfrac{1-Ac^2}{\theta+1}\left(\dfrac{c}{t}\right)^{\theta} & \text{if } c \le t < \infty \end{cases} \tag{24.1}$$

where θ is a positive integer and $Ac^2 \le 1$ (otherwise, $F(t)$ would exceed 1). We found that the corresponding probability density function, defined on $[0, \infty)$ by

$$f(t) = \begin{cases} 2A(c-t) + \dfrac{\theta(1-Ac^2)}{(\theta+1)c} & \text{if } 0 \le t \le c \\ \dfrac{\theta(1-Ac^2)}{(\theta+1)c}\left(\dfrac{c}{t}\right)^{\theta+1} & \text{if } c \le t < \infty \end{cases} \tag{24.2}$$

has a corner at $t = c$ unless

$$\theta = \frac{2Ac^2}{1-Ac^2}. \tag{24.3}$$

In these circumstances, it is unlikely that f is smooth, because the right-hand side of (3) is unlikely to be an integer. Although θ had to be an integer in Lecture 19, because otherwise c^{θ} and $t^{-\theta}$ in (1) would have had no general meaning, circumstances have since changed: Lecture 22 permits an exponent to be any real number. So, if we believe that f should be smooth, then we can arrange it simply by insisting on (3). It is then most convenient to eliminate θ from (1) and write

$$F(t) = \begin{cases} \dfrac{4Act}{1+Ac^2} - At^2 & \text{if } 0 \le t \le c \\ 1 - \dfrac{\{1-Ac^2\}^2}{1+Ac^2}\left(\dfrac{c}{t}\right)^{\frac{2Ac^2}{1-Ac^2}} & \text{if } c \le t < \infty \end{cases} \tag{24.4}$$

YEAR OF DEATH	FREQUENCY	YEAR OF DEATH	FREQUENCY
1	312	5	12
2	101	6	8
3	74	≥ 7	0
4	35		

Table 24.2 Prairie dog lifespans

To illustrate how bivariate functions arise in the context of data fitting, we will fit the c.d.f. above to Lecture 19's data on prairie-dog lifespans, repeated in Table 2 for ease of reference. As in Lecture 19, if

$$X = \text{AGE AT DEATH OF PRAIRIE DOG} \tag{24.5}$$

chosen randomly from the sample and

$$P_n = \mathrm{Prob}(X \le n), \tag{24.6}$$

then our measure of the misfit between F and the data is the sum of squared errors

$$\Delta = \sum_{n=1}^{6} \{F(n) - P_n\}^2. \tag{24.7}$$

The smaller the value of Δ, the better the fit. So we choose A and c to minimize Δ.

n	P_n	$F(n)$	$F(n) - P_n$
0	0	0	
1	$\frac{312}{545}$	$\frac{A(4c-1-Ac^2)}{1+Ac^2}$	$\frac{A(4c-1-Ac^2)}{1+Ac^2} - \frac{312}{545}$
2	$\frac{413}{545}$	$1 - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{2}\right)^{\frac{2Ac^2}{1-Ac^2}}$	$\frac{132}{545} - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{2}\right)^{\frac{2Ac^2}{1-Ac^2}}$
3	$\frac{487}{545}$	$1 - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{3}\right)^{\frac{2Ac^2}{1-Ac^2}}$	$\frac{58}{545} - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{3}\right)^{\frac{2Ac^2}{1-Ac^2}}$
4	$\frac{522}{545}$	$1 - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{4}\right)^{\frac{2Ac^2}{1-Ac^2}}$	$\frac{23}{545} - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{4}\right)^{\frac{2Ac^2}{1-Ac^2}}$
5	$\frac{537}{545}$	$1 - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{5}\right)^{\frac{2Ac^2}{1-Ac^2}}$	$\frac{8}{545} - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{5}\right)^{\frac{2Ac^2}{1-Ac^2}}$
6	1	$1 - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{6}\right)^{\frac{2Ac^2}{1-Ac^2}}$	$-\frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{6}\right)^{\frac{2Ac^2}{1-Ac^2}}$

Table 24.3 Calculation of Δ for prairie-dog lifespans

To calculate Δ from (7), we need F(n) for n = 1, ..., 6. So, from (4), we must know whether $n \in [0, c]$ or $n \in [c, \infty)$ for n = 1, ..., 6. For the sake of definiteness, we assume

$$1 \le c \le 2. \tag{24.8}$$

Then $1 \in [0, c]$, but $n \in [c, \infty)$ if $1 \le n \le 6$. From (7) and Table 3, we obtain

$$\Delta = S(A,c), \tag{24.9}$$

where the bivariate function S is defined by

$$S(A,c) = \left(\frac{A(4c-1-Ac^2)}{1+Ac^2} - \frac{312}{545}\right)^2 + \left(\frac{132}{545} - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{2}\right)^{\frac{2Ac^2}{1-Ac^2}}\right)^2$$
$$+ \left(\frac{58}{545} - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{3}\right)^{\frac{2Ac^2}{1-Ac^2}}\right)^2 + \left(\frac{23}{545} - \frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{4}\right)^{\frac{2Ac^2}{1-Ac^2}}\right)^2$$

$$+\left(\frac{8}{545}-\frac{\{1-Ac^2\}^2}{1+Ac^2}\left(\frac{c}{5}\right)^{\frac{2Ac^2}{1-Ac^2}}\right)^2+\frac{\{1-Ac^2\}^4}{\{1+Ac^2\}^2}\left(\frac{c}{6}\right)^{\frac{4Ac^2}{1-Ac^2}}. \qquad (24.10)$$

Why do we assume $1 \le c \le 2$, as opposed to, say, $2 \le c \le 3$? The answer is that (8) is the best choice, in the sense that it will yield the lowest minimum value for Δ. You can verify this assertion by a method similar to that which led to Tables 19.3 and 19.5 (but doing so now would needlessly distract us).

In Figure 3, the graph of S is plotted on subdomain $[0.12, 0.15] \times [1.7, 2]$, where it has a unique global minimum; but it is difficult to see where this minimum lies, because the surface is relatively flat in its vicinity. So in Figure 4 we have drawn a contour map. In this diagram, horizontal coordinates of points at selected altitudes are joined by continuous curves. From innermost to outermost, the four closed contours are at altitudes S = 0.006575, S = 0.0066, S = 0.00665 and S = 0.0067; the two almost closed contours are at S = 0.00675 and S = 0.0068; and thereafter S increases by 0.005 on each successive contour as one moves away either to the northwest or to the southeast, so that the contours in the top-right and bottom-left corners are both at S = 0.0268. For a fixed increment, the closer together the contours, the steeper the slope of the surface; thus the surface rises more steeply toward the southwest than toward the northeast (as Figure 3 confirms). Moreover, the smaller the length of a closed contour, the nearer it lies to the minimum. The shortest contour of all – not marked – is a degenerate contour of length zero, a point at the center of the innermost contour. If we denote its coordinates by $(\hat{A},\hat{c})$, then $(\hat{A},\hat{c})$ is the global minimizer of S, and the corresponding altitude $S(\hat{A},\hat{c})$ is the least sum of squared errors. In this case, $(\hat{A},\hat{c}) = (0.13146, 1.8457)$ and $S(\hat{A},\hat{c}) = 0.0065526$ (to 5 s.f.).

Another way to find the global minimum is to take vertical sections through the graph of S, parallel to a coordinate axis, and to locate the minimum of S along each section. Then the least of all section minima is the global minimum. This method is illustrated by Figure 5, which shows sections parallel to the A-axis at c = 1.7, c = 1.8, c = 1.9 and c = 2. Each such curve is the graph of an ordinary function, say f, defined by

$$f(A) \;=\; S(A,c). \qquad (24.11)$$

Here, the right-hand side is identical to (10) except that c is a fixed parameter, and so A is the only independent variable.

Let A^* be the minimizer of f on any given section. Then A^* depends on c, as illustrated by the dots in Figure 5. Furthermore, the dependence is unambiguous, i.e., it defines a function, and so we write $A^* = A^*(c)$. For example, from Figure 5, $A^*(1.7) = 0.147$, $A^*(1.8) = 0.136$, $A^*(1.9) = 0.1265$ and $A^*(2) = 0.12$. Because the minimizer depends unambiguously on c, the section minimum – S^*, say – also depends unambiguously on c, and so it defines another function. In fact, from (11) and the definition of A^*,

$$S^*(c) \;=\; f(A^*(c)) \;=\; S(A^*(c),c). \qquad (24.12)$$

For example, from Figure 5, $S^*(1.7) = 0.00681$, $S^*(1.8) = 0.00658$, $S^*(1.9) = 0.00658$ and $S^*(2) = 0.00687$. These data suggest that S^* has a minimum between c = 1.8 and c = 1.9, at $c = \hat{c}$, say. The suggestion is confirmed by Figure 6, where the graph of S^* is plotted on [1.7, 2]: $\hat{c} = 1.8457$ and $S^*(\hat{c}) = 0.0065526$, in agreement with Figure 4.

There is no particular reason to take sections parallel to the A-axis, however; parallel to the c-axis works just as well. The least of all section minima is still the global minimum of S. This alternative is illustrated by Figure 7, which shows vertical sections parallel to the c-axis at A = 0.12, A = 0.13, A = 0.14 and A = 0.15. As before, each such curve is the graph of an ordinary function, say g, defined by

$$g(c) \;=\; S(A,c). \qquad (24.13)$$

Again, the right-hand side is identical to (10), but now A is the fixed parameter, and c is the independent variable. Let c^* be the minimizer of g on any given section. Then c^* depends on A, as illustrated by Figure 7, and this dependence defines a function, or $c^* = c^*(A)$. For example, $c^*(0.12) = 1.973$, $c^*(0.13) = 1.861$, $c^*(0.14) = 1.765$ and $c^*(0.15) = 1.7$. Now, in place of (11), we obtain

$$S^*(A) \;=\; g(c^*(A)) \;=\; S(A,c^*(A)) \qquad (24.14)$$

and, from Figure 8, where S^* is plotted against A, the least section minimum is at $A = \hat{A} = \hat{A} = 0.13146$ with $S^*(\hat{A}) = 0.0066526$, agreeing again with Figure 4.

Either way, for the resultant cumulative distribution function, (4) implies

$$F(t) \;=\; \begin{cases} \dfrac{4\hat{A}\hat{c}t}{1+\hat{A}\hat{c}^2} - \hat{A}t^2 & \text{if } 0 \le t \le \hat{c} \\ 1 - \dfrac{\{1-\hat{A}\hat{c}^2\}^2}{1+\hat{A}\hat{c}^2}\left(\dfrac{\hat{c}}{t}\right)^{\frac{2\hat{A}\hat{c}^2}{1-\hat{A}\hat{c}^2}} & \text{if } \hat{c} \le t < \infty \end{cases} \qquad (24.15)$$

$$= \begin{cases} 0.6703t - 0.1315t^2 & \text{if } 0 \le t \le 1.846 \\ 1 - 0.5691t^{-1.622} & \text{if } 1.846 \le t < \infty \end{cases} \qquad (24.16)$$

and

$$f(t) \;=\; F'(t) \;=\; \begin{cases} 0.6703 - 0.2629t & \text{if } 0 \le t \le 1.846 \\ 0.923t^{-2.622} & \text{if } 1.846 \le t < \infty \end{cases} \qquad (24.17)$$

Both F and f are plotted in Figure 9. The fit is nowhere near as good as for the Weibull model of Lecture 19, i.e., for

$$F(t) \;=\; 1 - e^{-(t/s)^c} \qquad (24.18)$$

$$f(t) \;=\; F'(t) \;=\; \frac{c}{s}(t/s)^{c-1}e^{-(t/s)^c} \qquad (24.19)$$

with c = 1 and s = 1.286. In practice, therefore, we prefer the Weibull.

The point of introducing (4), however, was not to obtain a better fit; rather, it was to demonstrate how a better fit can be obtained. In particular, we can use our method to obtain an even better Weibull model: now that exponents are allowed to be arbitrary real numbers, there is no particular reason why c in (18)-(19) should be an integer. In fact, choosing

$$(s,c) \;=\; (\hat{s},\hat{c}) \;=\; (1.247, 0.9173) \qquad (24.20)$$

in (18) reduces total error for the Weibull model from 0.220×10^{-2} in Figure 19.1 to 0.155×10^{-2} in Figure 10, where F and f defined by

$$F(t) \;=\; 1 - e^{-(t/\hat{s})^{\hat{c}}} \;=\; 1 - e^{-0.8168t^{0.9173}} \qquad (24.21)$$

$$f(t) \;=\; \frac{\hat{c}}{\hat{s}}(t/\hat{s})^{\hat{c}-1}e^{-(t/\hat{s})^{\hat{c}}} \;=\; 0.7493t^{-0.08269}e^{-0.8168t^{0.9173}} \qquad (24.22)$$

are both plotted. See Exercise 2.

Reference

Rayner, Jeremy (1977) The Intermittent Flight of Birds. In: Pedley, T.J. (ed), Scale Effects in Animal Locomotion, pp. 437-443. Academic Press, London

Exercises 24

24.1* The bivariate function S defined in Appendix 2A by

$$S(\alpha,\beta) = 3433.26 - 15599\alpha + 18006.5\alpha^2 - 275.6\beta + 667\alpha\beta + 7\beta^2$$

has a global minimum at $(\hat{\alpha}, \hat{\beta})$. Estimate $(\hat{\alpha}, \hat{\beta})$ and, hence, $S(\hat{\alpha}, \hat{\beta})$ by drawing a contour map of S on subdomain $[0.55, 0.65] \times [-11, -6]$.

24.2 What bivariate function must be minized to fit (18) to the data in Table 2? Use the methods of this lecture to establish that (20) is its global minimizer, thus verifying that (22) is the best-fit Weibull p.d.f.

Solutions to Selected Exercises

24.2 By analogy with Table 3, the bivariate function to be minimized is ε defined by

$$\varepsilon(s,c) = \left(\frac{233}{545} - e^{-(1/s)^c}\right)^2 + \left(\frac{132}{545} - e^{-(2/s)^c}\right)^2 + \left(\frac{58}{545} - e^{-(3/s)^c}\right)^2 + \left(\frac{23}{545} - e^{-(4/s)^c}\right)^2 + \left(\frac{8}{545} - e^{-(5/s)^c}\right)^2 + e^{-2(6/s)^c}$$