

1.3 Basic functions used in biology

Reading: Otto and Day (2007), Primer 1.

Here we go through some of the single variable functions which regularly arise in all areas of biology.

1.3.1 Constants

$$y = f(x) = c$$

This function implies that x and y are independent.

1.3.2 Linear functions

$$y = f(x) = mx + c \tag{4}$$

There are a wide range of situations where we would like to describe how one variable increases as a function of another. For example, disease increases probability of dying, temperature increases number of offspring, smoking increases incidences of cancer etc. A positive slope $m > 0$ allows us to represent such relationships. A negative slope $m < 0$ describes a situation where one variable decreases as a function of another. For example, a treatment suppresses a disease, the expression of one gene can repress the expression of another etc.

Linear regression, one of the most commonly used operations in statistics, is based on the *assumption* that there are linear relationships between variables. As such, linear regression does not in itself test the hypothesis that a particular data set has a linear relationship.

We should be careful with linear relationships, because most biological variables can only take positive values. In which case, the function does not apply when we cross over the zero line.

1.3.3 Reciprocal functions

Another way of defining a negative relationship between two variables is using the reciprocal function:

$$y = f(x) = \frac{1}{ax + k} \quad (5)$$

This has the advantage over using a negative slope that as x tends to 0, y remains positive. Below we give different examples of equation 5.

A classic example in behavioural ecology of a reciprocal relationship is predator dilution. Your probability of being eaten by an attacking predator in a group of size n is equal to $1/n$.

1.3.4 Rational functions

Equations 4 and 5 can be combined to give

$$y = f(x) = \frac{mx + c}{ax + k} \quad (6)$$

To understand this function we should think about how it behaves for $x = 0$ and as $x \rightarrow \infty$. This allows us to plot equation 6.

This fractional function is well motivated on biological grounds. It can be set up to avoid strange negative values. This increased realism comes at a cost of adding complication. If m or respectively a are small then the reciprocal respectively linear function are more appropriate.

1.3.5 Quadratic function

Probably the simplest non-linear function is the quadratic function

$$y = f(x) = ax^2 + bx + c \tag{7}$$

These functions have (at most) a single maximum or minimum. By differentiating we see how this function behaves. The two examples for $a > 0$ and $a < 0$ are illustrated below.

A quadratic function behaves the "same" as $x \rightarrow \infty$ and $x \rightarrow -\infty$.

The points at which $x = 0$, known as the roots of $f(x)$, are given by the famous quadratic equation:

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (8)$$

The quadratic equation has a wide variety of important uses in biology. This is because it is the simplest model with a single maximum or minimum. In the form of the logistic growth model it describes the rate of growth of a population with limited resources (section 3.2). It is often used in problems of optimisation through natural selection ("climbing mount impossible"). For modelling gene frequencies in sexual populations it arises in describing allele frequencies.

1.3.6 Polynomial functions

The quadratic function can be generalised to higher order polynomials.

$$\begin{aligned}y = f(x) &= a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0 \\ &= \sum_{i=0}^n a_i x^i\end{aligned}$$

($x^0 = 1$). We can take the derivative of f by the usual rule of moving the power down.

1.3.7 Exponential functions

Lets define the following polynomial function with infinite powers

$$\begin{aligned}y = f(x) &= 1 + x + \frac{1}{2}x^2 + \frac{1}{2.3}x^3 + \frac{1}{2.3.4}x^4 + \dots + \frac{1}{(n-1)!}x^{n-1} + \frac{1}{n!}x^n + \dots(9) \\ &= \sum_{i=0}^{\infty} \frac{x^i}{i!}\end{aligned}\tag{10}$$

This function is quite interesting because it is its own derivative,

$$f'(x) = f(x)$$

The value of the derivative at any point is the function itself.

It turns out that this function, which at first sight is just a mathematical oddity, is probably the most important function in dynamic models of biology. It is known as the exponential function

$$y = f(x) = \exp(\alpha x) = e^{\alpha x} = \sum_{i=1}^n \frac{(\alpha x)^i}{i!} \quad (11)$$

Its most important application is in describing growth. It will be central in many of the differential equation models of growth presented in this course.

Below we show exponential growth (for $\alpha > 0$) and exponential decay (for $\alpha < 0$).

The inverse of the exponential is the natural logarithm.

$$x = \frac{\ln(y)}{\alpha}$$

Appendix A1.1 and A1.2 in the text book give basic rules for dealing with powers and logs. The basic idea is that in the logarithmic world multiplication becomes addition.

1.3.8 Power law functions

The function

$$y = f(x) = bx^a$$

describes a *scaling law* or *power law* between two variables. We can use logarithms to make this relationship clearer.

Power law scaling is observed in for example, metabolic rate which scales as approximately 3/4-power of mass of organism (Kleiber's law). Similarly, breathing and heart rates typically increase with mass raised to the power of 1/4

1.3.9 Sigmoidal functions

An important function in everything from biochemistry to ecology is the Hill function

$$y = f(x) = b \frac{x^n}{x^n + k^n} \quad (12)$$

where b is a saturation level or maximum possible value of y , k is the level of x at which y is half saturated and n describes the degree of 'co-operativity'. We plot this function below.

This function describes the rate at which one substance is taken up by another. In biochemistry x is a ligand and y is the rate at which it binds to a macromolecule. In ecology x can be a prey and y is the rate at which it is eaten by predators. In section 9.1 we will derive these responses and show how they are used in modelling.

Another sigmoidal function is the logistic function

$$y = f(x) = b \frac{e^{\alpha x}}{e^{\alpha x} + e^{\alpha k}} \quad (13)$$

As for the Hill function, b is a saturation level or maximum possible value of y and k is the level of x at which y is half saturated. The parameter α plays a similar role to n in the Hill function in determining the steepness of the sigmoidal shape.

When x is time, the logistic function describes the growth of a population. We will return to this in section 3.2.

The logistic function is commonly used in statistics in the form of 'logistic regression'. When $b = 1$, y is always between 0 and 1 and can therefore be interpreted as a probability. Re-arranging (inverting) equation 13 gives

$$\log\left(\frac{1-y}{y}\right) = \alpha k + \alpha x \quad (14)$$

Thus by transforming the observed probabilities it is possible to fit a straight line to them in terms of x . This function is widely used in medical studies, e.g. how the frequency of getting cancer y is related to the number of cigarettes smoked x . Like in linear regression, there is a strong assumption about how the probabilities are related to the factor said to cause them.

1.3.10 Bell-shaped functions

The most famous bell-shaped function is the Normal distribution

$$y = f(x) = c \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (15)$$

where μ is the mean of the distribution (i.e. the central point of the bell) and σ is the standard deviation. c determines the height of the bell. For the normal distribution

$$c = \frac{1}{\sqrt{2\pi\sigma^2}}$$

Roughly speaking, y can be interpreted as being proportional to the probability of seeing an observation of size x .

These last two examples should illustrate an important point: mathematics and statistics do not lie in two separate worlds. Statistics is the process of fitting to an (idealised) mathematical model.

1.4 A 'real' example

The extract from Harder et al. (2008) given on the next page gives an example of the problem of determining the degree of self-pollination needed to maximise fitness. We will leave the biological details to one side just now and look at the question the question of how the fitness function $W(p'_s)$ is maximised in terms of p'_s , the degree of self-pollination (Case I).

What you should recognise is that hidden in all the parameters is a rational function (equation 6). We will use this to obtain the same result as Harder without differentiating.

1.5 Exercises

Brush up on your algebraic manipulations. Try exercise A1.1.

ESS MATING SYSTEM WITH NO POLLEN DISCOUNTING

When self-pollination does not affect pollen export, individuals with either the resident or invading phenotypes realize equal out-cross siring success (i.e., the last term in eq. 3). Consequently, for a given fraction of exported pollen, p_x , the ESS maximizes maternal fitness, which equals the sum of the fitness gains from selfing and female outcrossing. For an invading phenotype $\mathbf{p}' = (p_s', p_x)$ in a resident population of phenotype $\mathbf{p} = (p_s, p_x)$, maternal fitness is

$$\hat{W}(\mathbf{p}', \mathbf{p}) = 2f_s(\mathbf{p}', \mathbf{p})g_s k(\mathbf{p}', \mathbf{p})d_s + f_{\zeta}(\mathbf{p}', \mathbf{p})g_x k(\mathbf{p}', \mathbf{p})d_x.$$

In the absence of pollen limitation of fertilization, an increase in self-fertilization can switch the constraint on seed production from resource to ovule limitation (Fig. 2A), so the three following cases must be considered to determine which selfing phenotype p_s' maximizes fitness.

Case 1: ($0 < g_s < g_x < m < 1$)

In this situation fewer embryos survive genetic death than can develop into seeds, so $k_s = k_x = 1$ regardless of the number of ovules fertilized. Based on this condition and the definitions of f_s and f_{ζ} given by equation 1a and b, respectively, total maternal fitness equals

$$\hat{W}(\mathbf{p}', \mathbf{p}) = \frac{2q_s p_s' g_s d_s + q_x p_x g_x d_x}{q_s p_s' + q_x p_x}.$$

The derivative of fitness with respect to p_s' ,

$$\frac{d\hat{W}}{dp_s'} = \frac{q_s q_x p_x (2g_s d_s - g_x d_x)}{(q_s p_s' + q_x p_x)^2},$$

indicates that total maternal fitness is a continuously increasing or decreasing function of p_s' , depending only on whether $2g_s d_s$ is greater or less than $g_x d_x$, regardless of the fixed value of p_x . If $2g_s d_s < g_x d_x$ a plant always benefits from decreased self-pollination. If $2g_s d_s > g_x d_x$ increased self-pollination is universally favored, within the constraint that plants allocate a fixed proportion of pollen to export. Because this latter result applies, regardless of p_x , selection should also favor a relaxation of the export constraint, leading to exclusive self-pollination. Consequently, if both p_s and p_x are free to evolve the optimal allocation of pollen to self-pollination and the resulting proportion of self-fertilized ovules are straightforward,

$$p_s^* = \begin{cases} 0, \\ 1, \end{cases} \quad \text{and} \quad f_s^* = \begin{cases} 0, & \text{if } 2g_s d_s < g_x d_x, \\ 1, & \text{if } 2g_s d_s \geq g_x d_x. \end{cases}$$