Deterministic functions for Modelling Biological Data

Michael Noonan

Biol 520C: Statistical modelling for biological data

Table of contents

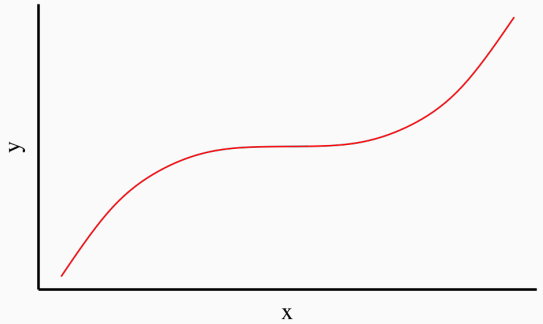
The ‘Linear’ in Linear Regression

Linear Regression

All of the models we've been working with so far are categorised as 'linear' deterministic functions. We say linear not because of the shape of the relationship, but because our regression parameters β_n are linear combinations of one another.

$$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3$$

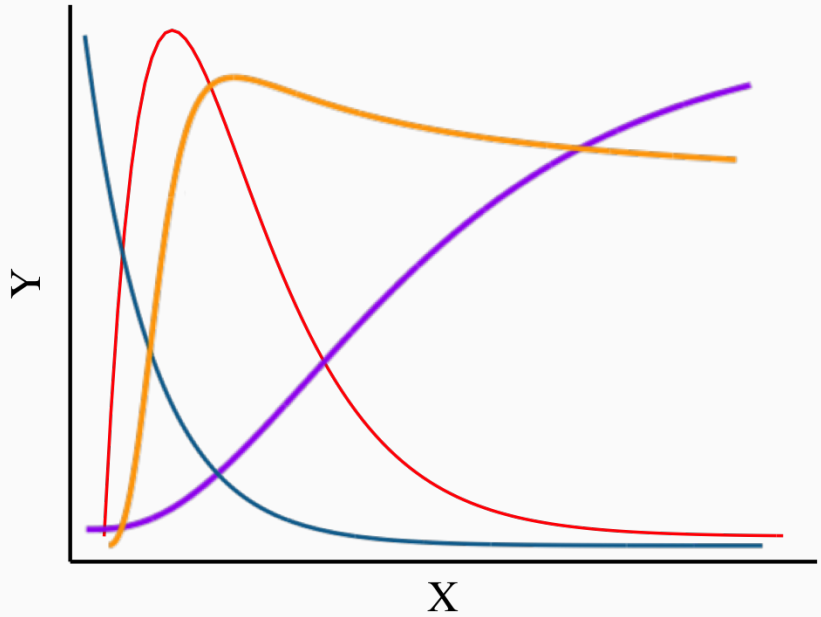


$$\mu = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Even if we add polynomial terms, or fit GLMs with link functions the βs still combine linearly.

Biology is not always linear

Biological systems are not always linear, and you will need to become familiar with a wide range of deterministic functions.



Today we will learn about tools that can help you understand new functions, and explore a range of different functions that get used routinely.

Finding Out About Functions Numerically and Analytically

Finding out about functions

When you encounter a new function you need to be able to look at it's form and get a feeling for its behaviour. There are two main ways you can do this:

1. Numerically (plugging in numbers and observing changes)

2. Analytically (evaluate it mathematically using e.g., limits, and derivatives)

Finding out about functions numerically

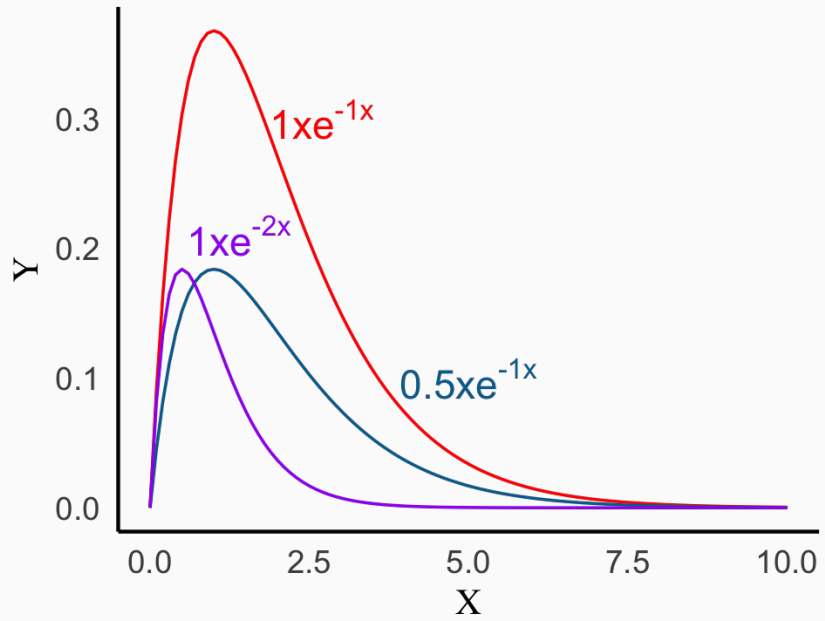
R can be a useful too for evaluating functions numerically.

Start by building a `function()` to describe the model, next define parameter values, run your function on a range of x values, plot the results.

You would then repeat this process for multiple parameter values.

For the Ricker function ($y = axe^{-bx}$), the process would look like this:

```
ricker <- function(x, a = 1, b = 1) {
  a * x * exp(-b * x)
  }

x <- seq(0,10, 0.1)

y <- ricker(x)
y2 <- ricker(x, a = 0.5)
y3 <- ricker(x, b = 2)
```



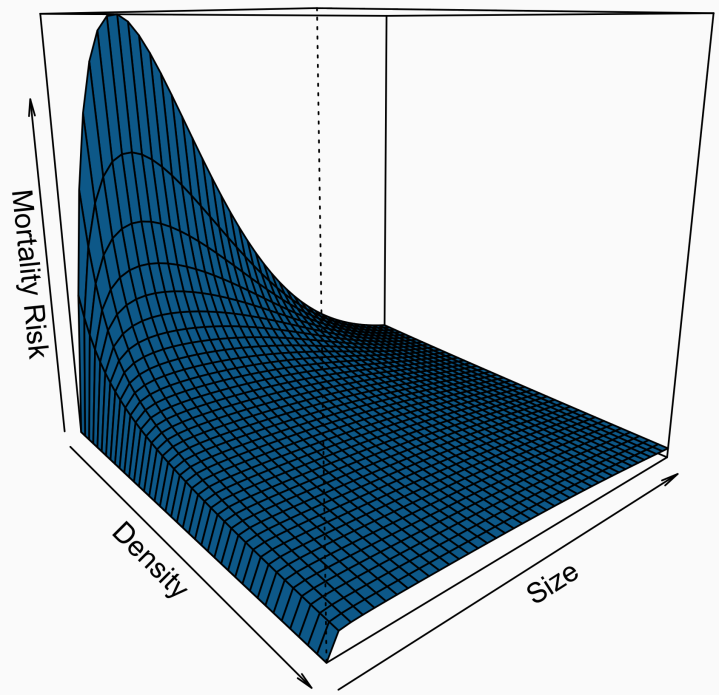
Numerical evaluations cont.

If you have multiple parameters you can use surface plots to explore how different inputs change outcomes.

... but 3D plotting in `R` is limited, and >2 params becomes unwieldy. The `curve3d` function from the `emdbook` package (for Bolker's book) allows you to relatively easily create surface plots.

If mortality risk can be modelled as a function of density (N) and size dependent attack rates ($\alpha(s)$) using a Holling Type II functional response ($y = \alpha(s)/(1 + \alpha(s)HN)$) (Vonesh & Bolker, 2005), then:

```
library(emdbook)

mortrisk <- function(N, size, H = .84) {
  a <- ricker(size)
  a/(1 + a * N * H)
}

curve3d(mortrisk(N = x, size = y),
        to = c(60,6),theta = 50,
        xlab = "Density",
        ylab = "Size",
        zlab = "Mortality Risk")
```



Finding out about functions analytically

Exploring functions numerically can be quick and easy, and doesn't require much math, but it can be limited/involve a lot of guessing.

Fully understanding how functions work requires evaluating them analytically.

Today we'll cover two ways for doing this:

1. Taking limits

2. Using derivatives

Limiting behaviour

It's often useful to know how certain functions are expected to behave at either end. This is known as their 'limiting behaviour'.

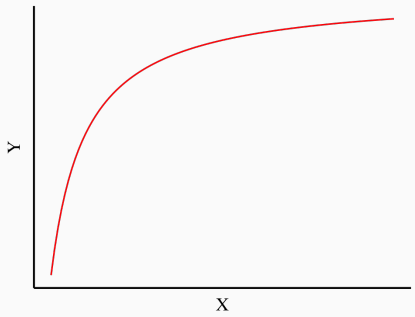
To understand the limiting behaviour of a function we look at what happens when x gets large ($x \to \infty$), or when x gets small ($x \to 0$ or $x \to -\infty$)

As x tends towards these limits terms get dropped and we get a feel for what the tail ends of the function would look like.

Michaelis-Menten limits

The Michaelis-Menten function is a well known function the limits to an asymptote and is given by:

$$y = \frac{ax}{b+x}$$



What happens in the limit where $x \to \infty$?

$$\frac{ax}{b+x} \qquad x \gg b, \text{ so } b + x \approx x \qquad \frac{ax}{x}, \frac{a\cancel{x}}{\cancel{x}} = a$$

What happens in the limit where $x \to 0$?

$$\frac{ax}{b+x} \qquad \frac{a0}{b+0} \qquad \frac{0}{b} = 0$$

So we now know the limits of this function are 0 and a.

Ricker limits

The Ricker function is a common function for modelling density dependence and is given by:

$$y = axe^{-bx}$$

What happens in the limit where $x \to 0$?

$$y = axe^{-bx} \qquad ax \to 0, \text{ and } -bx \to 0, \qquad \text{so we get } 0e^0 = 0$$

What happens in the limit where $x \to \infty$?

$$y = axe^{-bx} \qquad ax \to \infty, \text{ and } -bx \to -\infty, \qquad \text{so we get } \infty e^{-\infty}$$

Exponents are stronger than powers, and powers are stronger than linear terms, so $e^{-\infty}$ outweighs ∞ and $axe^{-bx} \to 0$ as $x \to \infty$

So we now know the limits of this function are 0 and 0.

Derivatives

Knowing the limiting behaviour of functions is useful, but it's also good to know how the functions increase/decrease towards them.

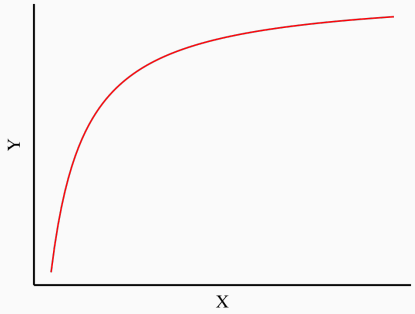
To understand this we need to take derivatives of our functions with respect to x.

Usefully, we can work out simple derivatives in `R` with the `D()` function.

Michaelis-Menten derivative

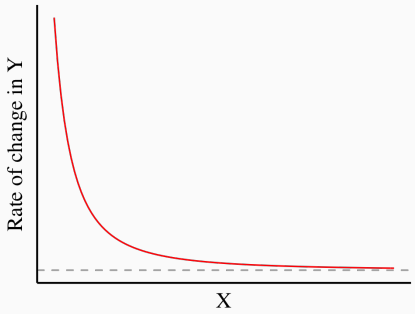
We know the Michaelis-Menten function limits to 0 and a:

$$y = \frac{ax}{b+x}$$



But how does it approach its limits?

```
michmen = expression(a * x / (b + x))
D(michmen, "x")

a/(b + x) - a * x/(b + x)^2
```



The rate of change in Y is greatest as $X \to 0$, and limits to 0 as $X \to \infty$ (called a saturating function)

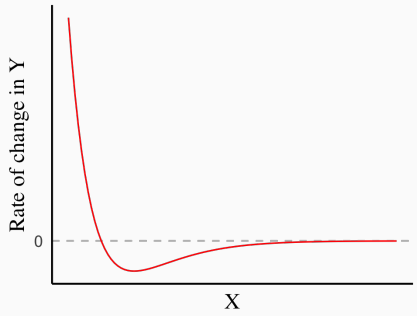
Ricker derivative

We know the limits of the Ricker function are 0 and 0.

$$y = axe^{-bx}$$

But how does it approach its limits?

```
Ricker = expression(a * x * exp(-b * x))
Ricker_deriv <- D(f,"x")

a * exp(-b * x) - a * x * (exp(-b * x) * b)
```



The rate of change in Y is greatest as $X \to 0$, it starts off growing but hits an inflection point, it then decays and the rate of change limits to 0 as $X \to \infty$ (saturating)

Take homes

If you know how to find out about a function's behaviour (numerically or analytically), you can get a better feeling of what each of the parameters are doing.

You can then change their values, shift or scale the functions, match them to biological parameters (remember they will all have units), etc.

All of this helps you get a better feel for how you can build functions to match the system you're modelling.

Deterministic Functions for Modelling Biological Data

Polynomial functions

Polynomial functions have the general form $y = \sum_{i=1}^{n} \beta_i x^i$

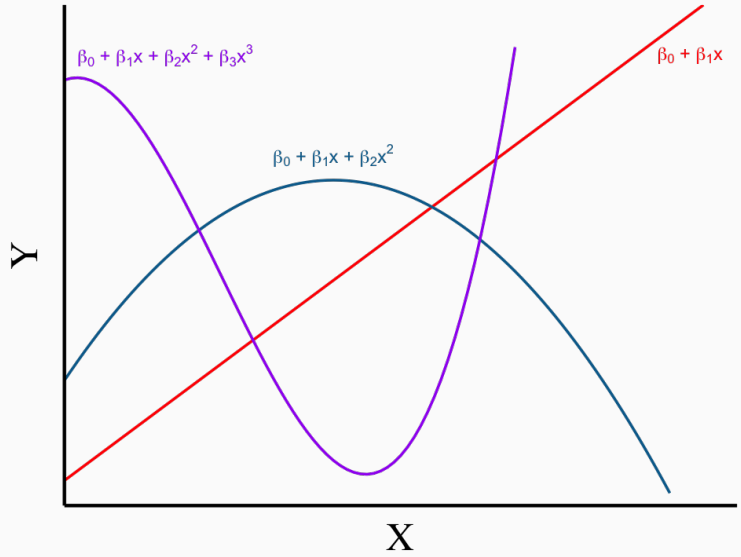
Examples:

Linear: $f(x) = \beta_0 + \beta_1 x$

Quadratic: $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$

Cubic: $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

Range: $-\infty, \infty$



Advantages: Easy to understand; easy to reduce; easy to extend to higher orders; can fit arbitrarily complex data.

Disadvantage: Anything beyond 2nd order polynomials are hard to justify mechanistically; they don't level off as $X \to \infty$ or $-\infty$ (extrapolations are often unrealistic); higher order polynomials can be unstable.

Piecewise polynomial functions

Polynomial functions can be made more flexible by using them as *piecewise* functions where different functions apply over different ranges of your predictor (x).

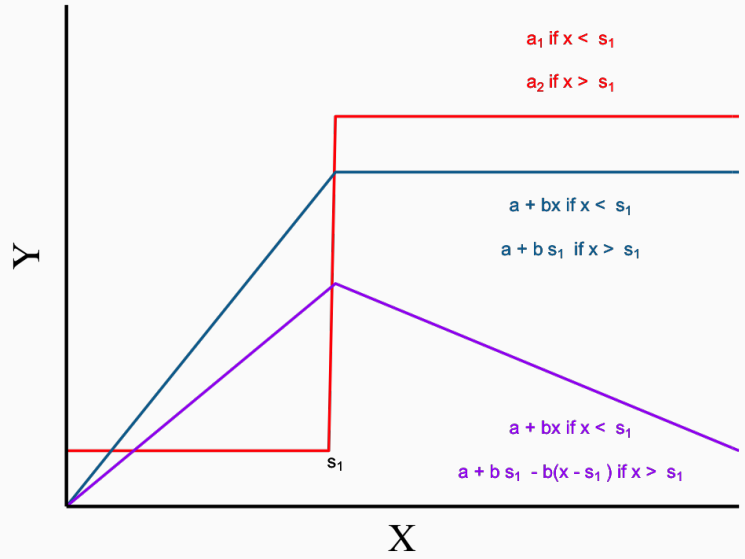
Examples:

Threshold:
$f(x) = a_1 \text{ if } x < s_1, \ a_2 \text{ if } x > s_1$

Hockey stick:
$f(x) = a + bx \text{ if } x < s_1, \ a + bs_1 \text{ if } x > s_1$

Piecewise linear:
$f(x) = a + bx \text{ if } x < s_1, \ a + b_1 s_1 + b_2(x - s_1) \text{ if } x > s_1$

Cubic splines:
$f(x)$ is complicated.



Range: $-\infty, \infty$

Advantages: Make sense and give added flexibility if there is a biological switching point.

Disadvantage: Hard to fit; discontinuous derivatives that may not make biological sense.

Rational functions

Rational functions are ratios of polynomials with the general form: $\frac{\sum a_i x^i}{\sum b_j x^j}$

Examples:

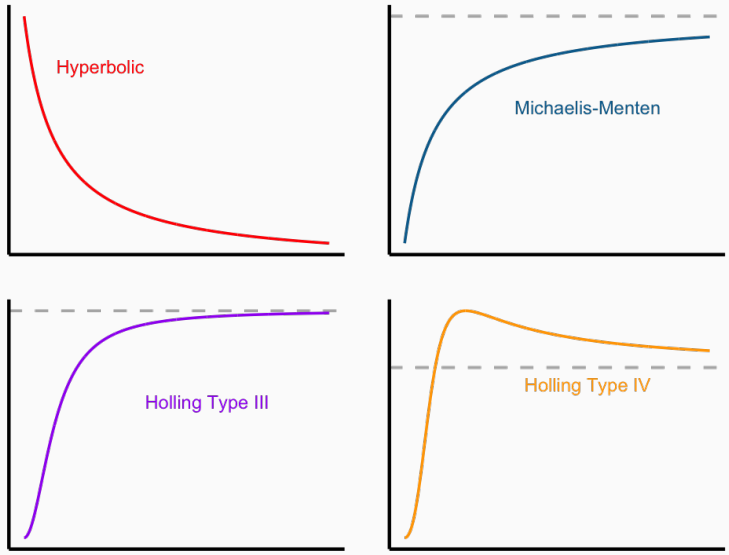
Hyperbolic: $f(x) = \frac{a}{b+x}$

Michaelis-Menten: $f(x) = \frac{ax}{b+x}$

Holling type III: $f(x) = \frac{ax^2}{b^2+x^2}$

Holling type IV (c<0): $f(x) = \frac{ax^2}{b+cx+x^2}$

Range: $0, \infty$

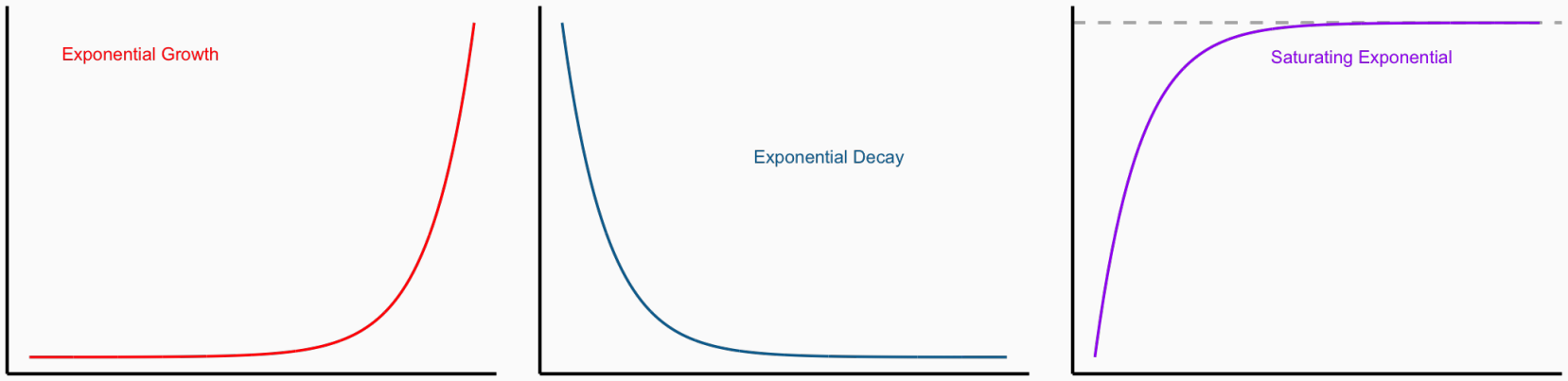


Advantages: Very flexible; often make mechanistic sense; simple to estimate parameters; easy to reduce down; reach finite asymptotes so are more stable than polynomials.

Disadvantage: Complicated derivatives; Approach their asymptotes very slowly, which can make these hard to estimate.

Exponential functions

Exponential functions are based on exponential growth (ae^{bx}), exponential decay (ae^{-bx}), or saturating exponential functions ($a(1 - e^{-bx})$). Their range is $0, \infty$.



Advantages: Exponential growth/decay occur commonly in nature and these function make mechanistic sense; can be used to calculate doubling times or half-lives.

Disadvantage: Similar behaviour to rational functions which makes it hard to distinguish between them without a lot of data.

Comb. of exponentials with other fncs.

Exponential functions can be combined with other functions to increase their flexibility. Their ranges are $0, \infty$.

Examples:

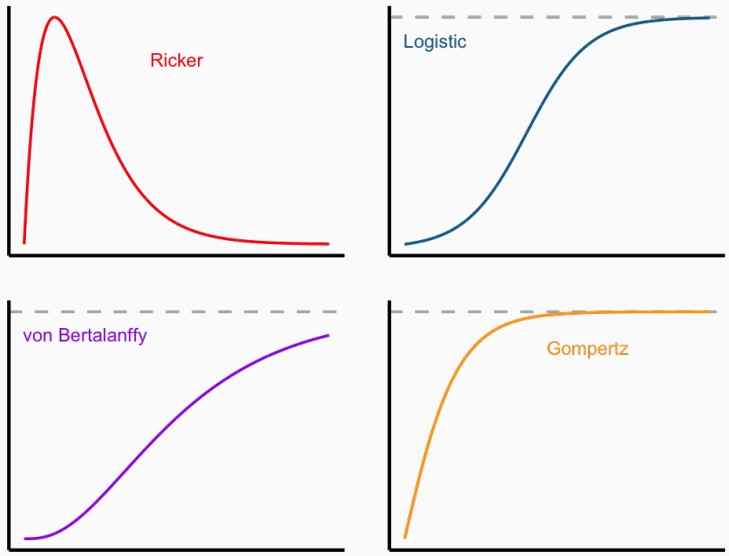
Ricker: $f(x) = axe^{-bx}$

Logistic: $f(x) = \frac{e^{a+bx}}{1+e^{a+bx}}$

von Bertalanffy:
$f(x) = a(1 - e^{-k(1-d)(x-x_0)})^{(\frac{1}{1-d})}$

Gompertz: $f(x) = e^{-ae^{-bx}}$

Range: $0, \infty$



Advantages: Very flexible; often make mechanistic sense; simple to estimate parameters; easy to reduce down; reach finite asymptotes so are more stable than polynomials.

Disadvantage: Similar behaviour to other functions which makes it hard to distinguish between them without a lot of data.

Fitting Non-Linear Models in R

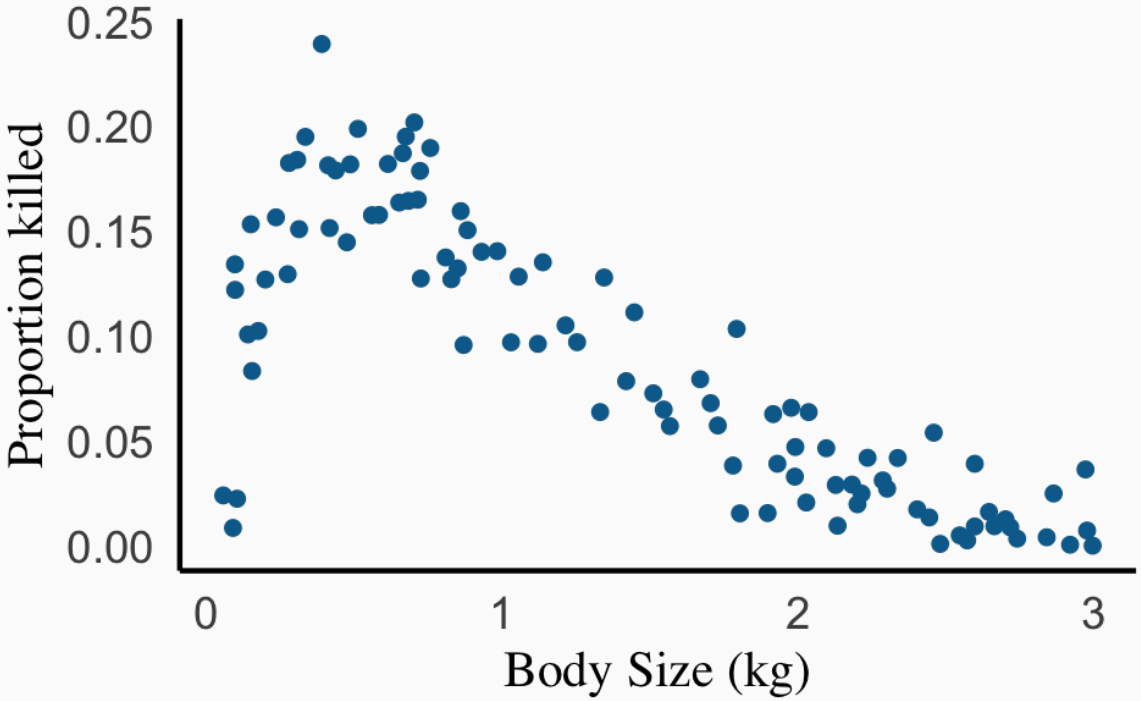
Fitting Non-Linear Models in R

Fitting non-linear models in R is only slightly more challenging than fitting linear regressions.

The `nls()` function allows fitting of non-linear relationships between a response variable and one or more explanatory variables using non-linear least squares.

Kill rate data

We'll work with a simulated dataset describing size-dependent predation rate.

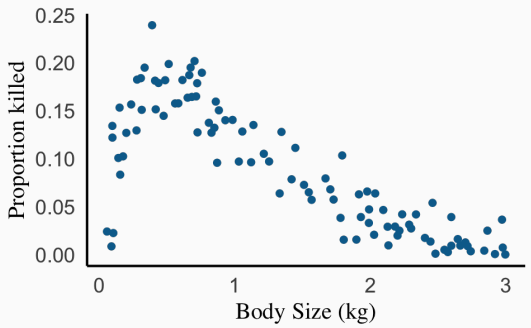


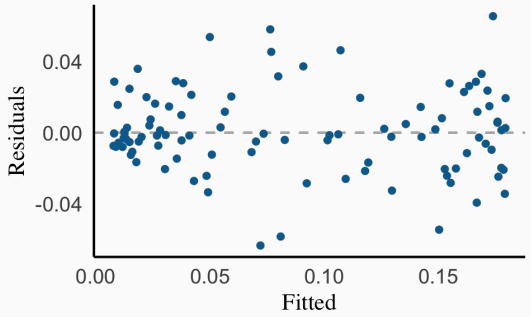
Does a linear model look like a good option here? What deterministic function do think we should try?

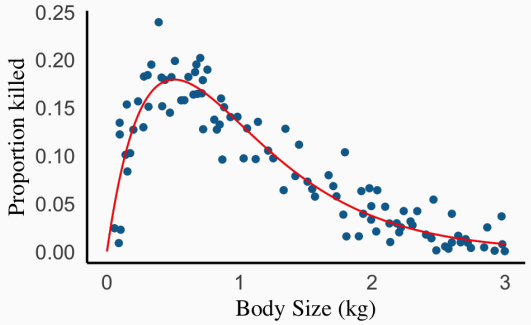
Kill rate data cont.

We can fit a Ricker function to the data using the `nls()` function.

```
#Need to provide starting values for the parameters
FIT <- nls(prop ~ a * size * exp(-b * size),
           start = list(a = 1,
                        b = 2),
           data = DATA)

summary(FIT)

Formula: prop ~ a * size * exp(-b * size)

Parameters:
  Estimate Std. Error t value Pr(>|t|)
a  0.96502    0.04175   23.11   <2e-16 ***
b  1.97821    0.05528   35.79   <2e-16 ***
---
Residual standard error: 0.02352 on 98 degrees of freedom


#To plot fitted model need to code up the Ricker function
ricker <- function(x) {
  coef(FIT)[1] * x * exp(-coef(FIT)[2] * x)
}
x <- seq(0,3, 0.01)
y <- ricker(x)
```







Technical considerations

The parameters of non-linear models can be very tricky to estimate.

Without good starting values the models can fail to fit properly (this will happen more than you like in practice). The better you are at understanding how the parameters of a particular function work, the easier it will be to eyeball reasonable starting values.

Unlike `lm()`, `nls()` requires that the formula includes all of the parameters you want to fit, including an intercept if you want one fitted. E.g., in `lm()` you would write the formula for linear regression as:

```
y ~ x
```

but to fit the same model in `nls()` you would have to write this as:

```
y ~ a + b*x
```

Take homes

We covered some of the most common functions, but the full list of possibilities is infinite.

The better you get building a working knowledge of deterministic functions, the better you will get at building models to fit and make theoretical predictions (very useful knowledge to have in your tool-belt).

If you're interested in learning more about the range of functions commonly used in biological modelling, absolutely do read Chapter 3 of Bolker's book (p. 88 is especially useful).

If you combine these functions with a stochastic model and maximum likelihood estimation you can fit any model you can write down to data (Chapter 8 of Bolker's book).

References

Vonesh, J.R. & Bolker, B.M. (2005). Compensatory larval responses shift trade-offs associated with predator-induced hatching plasticity. *Ecology*, 86, 1580–1591.

Bolker, B. M. (2008). Ecological models and data in R. Princeton University Press.