Working with Data

MATH 469, Texas A&M University

Spring 2020

Data for a Hare-Lynx Interaction

Let's start with a set of hare and lynx populations collected by the Hudson Bay Company for the years 1900-1920.

Year	1900	1901	1902	1903	1904	1905	1906
# Hare	30.0	47.2	70.2	77.4	36.3	20.6	18.1
# Lynx	4.0	6.1	9.8	35.2	59.4	41.7	19.0
Year	1907	1908	1909	1910	1911	1912	1913
# Hare	21.4	22.0	25.4	27.1	40.3	57.0	76.6
# Lynx	13.0	8.3	9.1	7.4	8.0	12.3	19.5
Year	1914	1915	1916	1917	1918	1919	1920
# Hare	52.3	19.5	11.2	7.6	14.6	16.2	24.7
# Lynx	45.7	51.1	29.7	15.8	9.7	10.1	8.6

Table: Hare and Lynx Populations in thousands, 1900-1920.

The Hare-Lynx Interaction

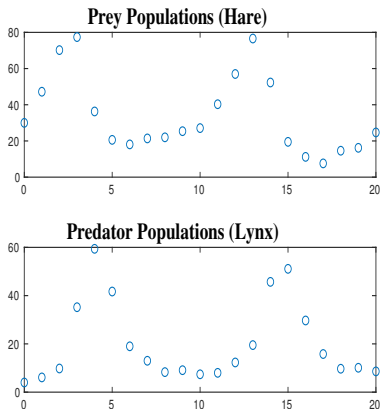As a start, let's plot the two populations together.

Figure: Hare and lynx populatons plotted together.

Fitting a Model to Data

Let's model the hare-lynx interaction with the predator-prey model

$$y_{1_{t+1}} - y_{1_t} = ay_{1_t}(1 - \frac{y_{1_t}}{K}) - by_{1_t}y_{2_t}$$

$$y_{2_{t+1}} - y_{2_t} = -ry_{2_t} + cy_{1_t}y_{2_t}.$$

Recall that this is a two-step process. First, we need to find some way of writing the system so that we can use linear regression to obtain initial values for the parameters, and second, we need to use these values to initialize a nonlinear regression calculation.

For the first, let's start with the predator equation, which we can express as

$$y_{2_{t+1}} - y_{2_t} = y_{2_t}(-r + cy_{1_t}) \implies \frac{y_{2_{t+1}} - y_{2_t}}{y_{2_t}} = -r + cy_{1_t},$$

for $t = 0, 1, \ldots 19$. (E.g. $t = 0$ corresponds with 1900.)

Fitting a Model to Data

We can plot the values $\frac{y_{2_{t+1}} - y_{2_t}}{y_{2_t}}$ vs y_{1_t}, and we obtain c as the slope of the fit and $-r$ as the y-intercept.
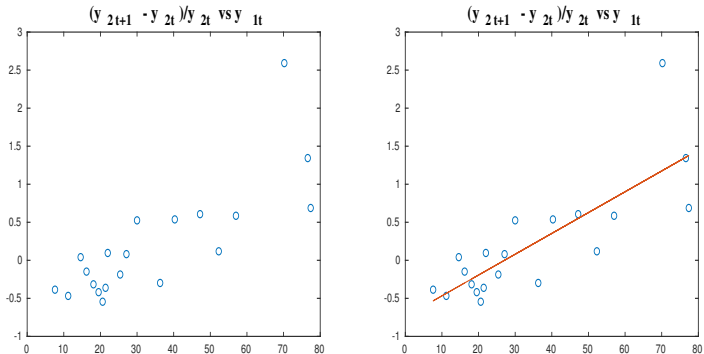

Figure: Linear fit for the lynx equation.

Fitting a Model to Data

We find $c = .0274$ and $r = .7411$.

Similarly, for the hare equation

$$y_{1_{t+1}} - y_{1_t} = ay_{1_t}(1 - \frac{y_{1_t}}{K}) - by_{1_t}y_{2_t},$$

we can write

$$\frac{y_{1_{t+1}} - y_{1_t}}{y_{1_t}} = a - \frac{a}{K}y_{1_t} - by_{2_t}.$$

In this case, we don't have a line, but we still have linear (multivariate) regression. I.e., we can fit $\frac{y_{1_{t+1}} - y_{1_t}}{y_{1_t}}$ as a function of y_{1_t} and y_{2_t}. We find $a = .6794$, $K = 114.9335$, and $b = .0190$.

Fitting a Model to Data

Now, we're ready to carry out the nonlinear regression. Since MATLAB will do all the work, let's be clear about what it will be doing. Let

$$\vec{y}_t(a, b, c, K, r) = \begin{pmatrix} y_{1_t}(a, b, c, K, r) \\ y_{2_t}(a, b, c, K, r) \end{pmatrix},$$

denote the vector solution at time t, where we're emphasizing the dependence of this solution on the parameter values. Let's also introduce vector notation for our data: $\vec{p}_t$. E.g., $\vec{p}_0 = \begin{pmatrix} 30.0 \\ 4.0 \end{pmatrix}$, $\vec{p}_1 = \begin{pmatrix} 47.2 \\ 6.1 \end{pmatrix}$, etc. For the nonlinear regression, we need to minimize the sum of squared residuals (SSR)

$$E(a, b, c, K, r) = \sum_{t=0}^{20} |\vec{y}_t(a, b, c, K, r) - \vec{p}_t|^2.$$

Fitting a Model to Data

In order to compute E, we take as input values for a, b, c, K, r and we (numerically) solve the system

$$y_{1_{t+1}} - y_{1_t} = ay_{1_t}(1 - \frac{y_{1_t}}{K}) - by_{1_t}y_{2_t} \quad y_{1_0} = 30$$

$$y_{2_{t+1}} - y_{2_t} = -ry_{2_t} + cy_{1_t}y_{2_t}, \quad y_{2_0} = 4.$$

The solution is a collection of 21 numerical (vector) values

$$\{\vec{y}_t(a, b, c, K, r)\}_{t=0}^{20}.$$

For a particular data vector $\vec{p}_t$,

$$|\vec{y}_t(a, b, c, K, r) - \vec{p}_t|^2$$
$$= (y_{1_t}(a, b, c, K, r) - p_{1_t})^2 + (y_{2_t}(a, b, c, K, r) - p_{2_t})^2.$$

We compute $E(a, b, c, K, r)$ by summing these values up for $t = 0, 1, \ldots, 20$.

Fitting a Model to Data

This gives us a value for

$$E(a, b, c, K, r) = \sum_{t=0}^{20} |\vec{y_t}(a, b, c, K, r) - \vec{p_t}|^2.$$

Using MATLAB to minimize E, we find $a = 1.4974$, $K = 82.3206$, $b = .0425$, $r = .5820$, and $c = .0239$. Using these values, we plot our model along with data on the next slide.
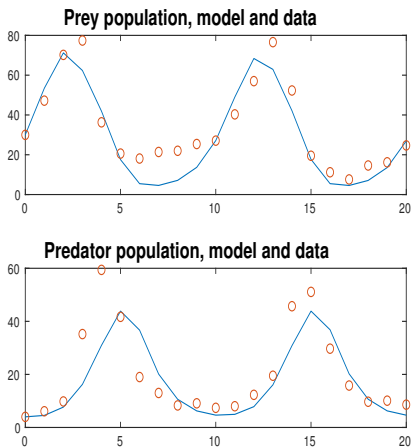
Fitting a Model to Data

Figure: Hare-lynx model along with data.