Nonlinear regression models

Linear statistical models.

Recall that in regression we started with a normal model:

\[ Y \sim \text{normal}(\mu, \sigma^2) \]

We allowed the mean parameter in the model to depend on one or more predictor variables (covariates):

\[ \mu = E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k. \]

Such a model is not necessarily a linear function of the predictor variables, for example:

\[ \mu = E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2. \]

These models are called linear statistical models, in that the parameters \( \beta_0, \beta_1, \beta_2 \ldots \) in the mean enter the mean linearly.

For such models, the maximum likelihood estimates of parameters in the mean are the least squares estimates given by the solution to a system of linear equations:

\[ \hat{\beta} = (X'X)^{-1}X'y. \]
Nonlinear statistical models.

Frequently models are formulated in scientific applications that are nonlinear functions of parameters. When data are available for estimation of the parameters (fitting the model to data), the statistical problem of nonlinear regression results.

Examples of nonlinear statistical models:

(biochemistry) Michaelis-Menten model of the rate of an enzyme-catalyzed reaction.

\[ v = \frac{v_{\text{max}}s}{k + s}, \]

where \( v \) is rate of product formation, \( s \) is substrate concentration, \( v_{\text{max}}, k \) are parameters.

(population ecology) Logistic model of population growth.

\[ n_t = \frac{k}{1 + \left(\frac{k-n_0}{n_0}\right)e^{-rt}}, \]

where \( n_t \) is population size at time \( t \), and \( r, k \) (and possibly \( n_0 \)) are parameters.

Statistical model for nonlinear regression:

\[ Y_i \sim \text{normal}(\mu_i(\theta), \sigma^2), \]

where \( \theta \) is a vector of parameters.
ex. Michaelis-Menten.

data: \((s_1, v_1), (s_2, v_2), \ldots, (s_n, v_n)\)

\[
\mu_i(k, v_{\text{max}}) = \frac{v_{\text{max}} s_i}{k + s_i}.
\]

ex. Logistic.

data: \((n_0, 0), (n_1, t_1), (n_2, t_2), \ldots, (n_q, t_q)\)

\[
\mu_i(k) = \frac{k}{1 + \left(\frac{k-n_0}{n_0}\right)e^{-rt_i}}.
\]

ML estimate of \(\theta\) minimizes

\[
SS(\theta) = \sum_{i=1}^{n} [y_i - \mu_i(\theta)]^2.
\]

Generally there are no closed formulas for the ML/LS parameter estimates. Instead, the estimates must be computed with numerical minimization of \(SS(\theta)\). PROC NLIN in SAS is a nonlinear regression procedure that will perform the iterative calculations.

Standard errors, confidence intervals, correlations of parameter estimates, and hypothesis tests are based on large-sample theory of ML estimates and likelihood ratio testing.
ex. predation rate model.

Data from Messier (1994 *Ecology*): $x$ is moose density (number of moose per 1000 square kilometers), $y$ is average feeding rate of a wolf (average number of moose killed per wolf per 100 days).

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.17</td>
<td>0.37</td>
</tr>
<tr>
<td>0.23</td>
<td>0.47</td>
</tr>
<tr>
<td>0.23</td>
<td>1.90</td>
</tr>
<tr>
<td>0.26</td>
<td>2.04</td>
</tr>
<tr>
<td>0.37</td>
<td>1.12</td>
</tr>
<tr>
<td>0.42</td>
<td>1.74</td>
</tr>
<tr>
<td>0.66</td>
<td>2.78</td>
</tr>
<tr>
<td>0.80</td>
<td>1.85</td>
</tr>
<tr>
<td>1.11</td>
<td>1.88</td>
</tr>
<tr>
<td>1.30</td>
<td>1.96</td>
</tr>
<tr>
<td>1.37</td>
<td>1.80</td>
</tr>
<tr>
<td>1.41</td>
<td>2.44</td>
</tr>
<tr>
<td>1.73</td>
<td>2.81</td>
</tr>
<tr>
<td>2.49</td>
<td>3.75</td>
</tr>
</tbody>
</table>

model is “Type II functional response”:

$$y = \frac{ax}{b + x} \quad \text{(Michaelis-Menten!)}$$

Statistical model uses mean

$$\mu_i(a, b) = \frac{ax_i}{b + x_i}.$$
ML/LS estimates: $\hat{a} = 3.37$  $\hat{b} = 0.47$