INITIAL PARAMETER ESTIMATES FOR NONLINEAR REGRESSION MODELS

Michael P. Meredith

BU-887-M*

August, 1985

Abstract

Simple methods are presented for determining estimators to be used at the first stage of a nonlinear least squares or Maximum Likelihood estimation procedure. Focus is upon nonlinear models that are commonly used in biological studies. The exponential growth and decay, Logistic, Mitscherlich and Gompertz models are discussed. Curve "peeling" and the method of partial totals are discussed for models based upon sums of exponentials. Much of the development requires equally spaced time intervals, although this is not a great restriction in practice.

---

* In the Biometrics Unit, Cornell University, Ithaca, NY 14853.

Introduction

There has been considerable effort expended investigating the properties of nonlinear least squares (NLLS) estimators in the theoretical literature (Clarke, 1980; Hartley and Booker, 1965; Jennrich, 1969; Malinvaud, 1970). The actual task of computing parameter estimates for a nonlinear model typically employs an iterative scheme that may be computationally costly. Jennrich (1969) points out that NLLS converges with probability one to the "correct" estimate (i.e., the estimate which achieves a global minimum on the error sum of squares surface) only if the initial estimate is contained in a sufficiently small neighborhood of the correct estimate. He recommends a two-step procedure suggested by Hartley and Booker (1965) wherein consistent estimators of parameters are used as initial estimates to enter the iterative NLLS procedure.

In the following discussion it will be assumed that the y_i are observable random variables and define $Y_i = E[y_i] = f(x_i, \underset{\sim}{\theta})$, $i = 0,1,2,\cdots,n$. The x_is are fixed (such as time) and $\underset{\sim}{\theta}$ is a p-dimensional vector of unknown parameters. The error structure will be left unspecified and should be investigated upon fitting the model.

Computer software packages often suggest supplying the NLLS program with a p-dimensional grid of potential starting values. The element of the grid that results in the smallest residual sum of squares (RSS) is taken as the initial estimate for the parameter vector. Thus, if there are k_j values for each of the $j = 1,2,\cdots,p$ parameters, then a total of $\prod_{j=1}^{p} k_j$ residual sum of squares must be calculated before the iterative phase even begins. Clearly, supplying an extensive grid in the hope of obtaining good initial estimates is both naive and wasteful. Such an approach gives little credibility to the resultant estimates since they may represent local minima on the RSS surface. Indeed, the suggestion is often made that additional grids be tried to help establish the credibility of the estimates obtained from such a "hit or miss" procedure.

Discussion

Initial parameter estimates are determined for nonlinear growth models that are commonly used in biological research. The simplest initial estimates are found when the x_is are equally spaced. If the x_is are equally spaced then we may assume without loss of generality that $x_0=0$, $x_1=1,\cdots,x_n=n$. In general, denote $x_{i+1}-x_i = \Delta x_i$ and $y_{i+1}-y_i = \Delta y_i$.

Most of the models include a term which may be written as $\rho=\exp(-\alpha)$, where α represents a growth rate parameter from the defining differential equation. Since $\exp(-\alpha) = 1 - \alpha + \alpha^2/2 - \cdots$ then, for small α, we have $1-\rho \approx \alpha$. This relationship will prove useful in finding initial estimates that are derived from a first order approximation to the defining differential equation.

The methods for finding initial estimates may be placed into any of three categories: 1) Simple functions of ratios and differences of the observations; 2) linearization of the nonlinear model; or, 3) first order approximations to the defining differential equation.

These methods are simple and easily implemented by hand, hand calculator or a simple statistical software package such as Minitab. These initial estimates are then used to enter a sophisticated statistical software program such as BMDP3R, BMDPAR or PROC NLIN in SAS.

The requirement of equally spaced xs appears to be a severe restriction, although in practice this is not the case. Simply plot the observed response curve and use "eyeball interpolation" to insert enough data to allow for the calculation of initial estimates.

Exponential growth and decay

This model may be written as $Y_i = \beta e^{\alpha x_i}$. Equally spaced xs allow the model to be written as $Y_i = \beta\rho^i$.

(i) Note $Y_{i+1} = \rho Y_i$ so an obvious estimator of ρ is given by

$$\rho_0 = \sum_{i=0}^{n-1} y_i y_{i+1} \Big/ \sum_{i=1}^{n} y_i^2 \quad .$$

This will typically underestimate ρ.

(ii) Suppose $n = 2r-1$ and let $S_1 = \sum_{i=0}^{r-1} y_i$ and $S_2 = \sum_{i=r}^{2r-1} y_i$. Then, $E[S_1] = \beta\left(\frac{1-\rho^r}{1-\rho}\right)$ and $E[S_2] = \beta\rho^r\left(\frac{1-\rho^r}{1-\rho}\right)$ giving $S_2/S_1 \approx \rho^r$, or $\rho_0 = \left(\frac{S_2}{S_1}\right)^{1/r}$. This is a consistent estimator of ρ since it represents a function of independent consistent estimators. Consistency follows by an application of Slutzsky's Theorem (Cramér, 1946). This estimator is discussed by Cornell (1962).

(iii) Let $r_i = y_{i+1}/y_i$, $i=0,1,\cdots,n-1$, then choose $\rho_0 = \text{median}(r_i)$. Alternatively, consider $r_{[k]i} = \frac{y_{i+k+1}-y_{i+1}}{y_{i+k}-y_i}$ and proceed similarly. Clearly, other ratios of linear combinations of observations could be constructed. Choosing $\bar{r} = \frac{1}{n}\sum_{i=0}^{n-1} r_i$ results in an estimator sensitive to extreme rs. A trimmed mean would be a preferred alternative to the usual mean. Unfortunately these simplest of estimators are typically very unstable.

(iv) An initial estimate for β may be written as $\beta_0 = \left(\frac{1-\rho_0}{1-\rho_0^{n+1}}\right)S \approx (1-\rho_0)S$, where $S = \sum_{i=0}^{n} y_i = S_1 + S_2$ and ρ_0 is determined from either i, ii or iii above.

(v) A less simple initial estimate of β is given by $\beta_0 = \frac{\sum \rho_0^i y_i}{\sum \rho_0^{2i}}$ derived by solving $y_i = \beta\rho_0^i$ for β via least squares.

(vi) A further easy estimate of β_0 is given by $\beta_0 = (1+\rho_0)\frac{\sum y_i^2}{\sum y_i}$. This follows since $\sum Y_i^2 = \beta^2\left(\frac{1-\rho^{2n}}{1-\rho^2}\right)$ and $\sum Y_i = \beta\left(\frac{1-\rho^n}{1-\rho}\right)$ gives $\frac{\sum Y_i^2}{\sum Y_i} = \beta\left(\frac{1-\rho^{2n}}{1-\rho^2}\right)\left(\frac{1-\rho}{1-\rho^n}\right) \approx \frac{\beta}{1+\rho}$. This estimator may be expected to have an upward bias that may be adjusted by dividing β_0 by $1+\rho_0^n$.

(vii) If the x_is are unequally spaced the usual approach is to fit a straight line to the logarithm of the responses. That is, fit $\log(y_i) = \log \beta + (\log \rho)x_i$ and choose $\beta_0 = \exp(\widehat{\log \beta})$ and $\rho_0 = \exp(\widehat{\log \rho})$, where $\widehat{\log \beta}$ and $\widehat{\log \rho}$ are the usual least squares estimators. This approach is reasonable if the errors are multiplicative on the original scale of measurement.

(viii) Another, perhaps simpler approach to unequally spaced x's is given by recognizing that $r_i = \frac{y_{i+1}}{y_i} \approx \rho^{\Delta x_i}$, or $\rho \approx (r_i)^{1/\Delta x_i}$ and choose $\rho_0 = \text{median}\left[(r_i)^{1/\Delta x_i}\right]$ and $\beta_0 = \frac{\sum \rho_0^{x_i} y_i}{\sum \rho_0^{2x_i}}$. As for equally spaced xs, this estimator is unstable.

(ix) In a manner similar to (viii) recognize that $\log(y_i) = \Delta x_i (\log \rho)$ giving $\log \rho \approx \frac{\Delta \log y_i}{\Delta x_i}$, and thus $\rho_0 = \text{median}\left[\exp\left(\frac{\Delta \log y_i}{\Delta x_i}\right)\right]$. Again, this estimator is unstable.

<u>Example</u>

Data (data set 1 in the appendix) have been generated from the model

$$y_i = 10(.8)^{i+\sigma z_i}, \qquad i=0,1,\cdots,19 \quad ,$$

where $z_i \sim$ iid $N(0,1)$ and $\sigma = 0.25$. Hence, $\beta=10$ and $\rho=0.8$. The results for methods (i) through (vii) follow.

(i) $\rho_0 = 0.781$.

(ii) $\rho_0 = 0.785$ using $r=10$, $S_1 = 45.441$ and $S_2 = 4.034$.

(iii) $\text{median}(r_i) = 0.874$, $\bar{r} = 0.857$, a 5% trimmed mean gives $\bar{r}_{.05} = 0.860$ and $\text{median}\left(r_{[2]i}\right) = 0.68$.

(iv) $\beta_0 = 10.84$ using ρ_0 from (i).
$\beta_0 = 10.63$ using ρ_0 from (ii).
$\beta_0 = 6.23$ using ρ_0 from (iii) as $\text{median}(r_i)$.

Similarly for (v) and (vi) we have:

(v)	$\beta_0 = 10.18$	(vi)	$\beta_0 = 10.22$
	$\beta_0 = 10.12$		$\beta_0 = 10.24$
	$\beta_0 = 7.84$		$\beta_0 = 10.75$

(vii) $\widehat{\log \beta} = 2.301$ giving $\beta_0 = 9.97$, and
$\widehat{\log \rho} = -0.2317$ giving $\rho_0 = 0.793$.

Mitscherlich model

This model may be written as $Y_i = \beta - \gamma e^{-\alpha x_i}$. Equal spacing of the x_is gives $Y_i = \beta - \gamma\rho^i$.

(i) Suppose $n = 3r-1$ and let $S_k = \sum_{i=(k-1)r}^{kr-1} y_i$, $k=1,2,3$. Then, it is easily shown that $\rho_0 = \left(\frac{S_3-S_2}{S_2-S_1}\right)^{1/r}$, $\beta_0 = \frac{S_2^2-S_1S_3}{r(S_3-2S_2+S_1)}$, and $\gamma_0 = \frac{1-\rho_0}{\left(1-\rho_0^r\right)^2}(S_2-S_1)$.

Notice that the numerators are not independent of denominators in the above expressions. The following estimator avoids this shortcoming to give a consistent estimator of ρ, but more data are required for any precision.

(ii) Suppose $n=4r-1$ and define $S_k = \sum_{i=(k-1)r}^{kr-1} y_i$, $k=1,2,3,4$. Then, $\rho_0 = \left(\frac{S_4-S_3}{S_2-S_1}\right)^{1/2r}$ provides a consistent estimator of ρ.

(iii) Notice that $Y_i = \beta\rho^i(1-\rho)$ which suggests $r_i = \frac{\Delta y_i}{\Delta y_{i-1}} \approx \frac{\beta\rho^{i+1}(1-\rho)}{\beta\rho^i(1-\rho)} = \rho$, and choose $\rho_0 = \text{median}(r_i)$. Alternatively, let $d_{[k]i} = y_{i+k} - y_i \approx \beta\rho^i(1-\rho^k)$, and $r_{[k]i} = \frac{d_{[k]i+1}}{d_{[k]i}}$ and choose $\rho_0 = \text{median}(r_{[k]i})$. Once again, $\bar{r} = 1/(n-1)\ \Sigma\ r_i$ is particularly sensitive to outlying values. These estimators are found to be unstable in practice.

(iv) Notice that $Y_{i+1} = \beta(1-\rho) + \rho Y_i$ and proceed to regress y_{i+1} on y_i to give $\rho_0 = \hat{\rho}$ and $\beta_0 = \frac{\hat{\beta(1-\rho)}}{(1-\hat{\rho})}$, where the ^ indicates an ordinary least squares estimator. Properties for n=3 are discussed in Finney (1958).

(v) An estimation technique similar to the preceding was given by Hartley (1948), who suggested regressing $y_{i+1} - y_i$ on $y_{i+1} + y_i$, having noted that $Y_{i+1} - Y_i = 2\beta\left(\frac{1-\rho}{1+\rho}\right) + \left(\frac{1-\rho}{1+\rho}\right)(Y_{i+1} + Y_i)$. Identifying $z_i = y_{i+1} - y_i$, $w_i = y_{i+1} + y_i$, $b_0 = 2\beta\left(\frac{1-\rho}{1+\rho}\right)$ and $b_1 = -\left(\frac{1-\rho}{1+\rho}\right)$, proceed to fit $z_i = b_0 + b_1 w_i$ via ordinary least squares. Thus, determine $\rho_0 = \frac{1+\hat{b}_1}{1-\hat{b}_1}$ and $\beta_0 = \frac{-\hat{b}_0}{2\hat{b}_1}$. In practice, method (iv) is preferred as well as being more easily computed. Finney (1958) discusses properties of these estimators for small values of n.

(vi) For unequally spaced x_is the usual approach is to linearize the Mitscherlich equation by choosing $\beta_0 > \max_{0\le i\le n}(y_i)$ and writing $\log(\beta_0 - y_1) = \log\gamma - (\log\rho)x_i$. Then, γ_0 and ρ_0 are determined from the back transformed least squares estimates. In general, β_0 is too large which consequently affects the γ_0 and ρ_0 estimates as well.

<u>Example</u>

Data (data set 2 in the appendix) have been generated from the model $y_i = 10-10(.8)^i + \sigma z_i$, $i=0,1,\cdots,19$, $z_i \sim$ iid $N(0,1)$ and $\sigma = 0.5$. Hence, $\beta=10$, $\gamma=10$ and $\rho=0.8$. The results for methods (i) through (vi) follow.

(i) $\rho_0 = 0.840$, $\beta_0 = 11.00$ and $\gamma_0 = 10.38$ for r=6, $S_1 = 23.890$, $S_2 = 51.180$ and $S_3 = 60.780$. Note that the last two observations were discarded for these computations.

(ii) $\rho_0 = 0.830$ using $r=5$, $S_1 = 17.140$, $S_2 = 40.410$, $S_3 = 47.185$ and $S_4 = 50.785$.

(iii) median(r_i) = 0.50. The mean, $\bar{r}$, for these data was -3.9!

(iv) $\rho_0 = 0.797$ and $\beta_0 = 10.12$.

(v) $\rho_0 = 0.807$ and $\beta_0 = 10.25$.

(vi) Note that the maximum y_i was 10.55. Then, choosing $\beta_0 = 10.56$ gives $\gamma_0 = 11.01$ and $\rho_0 = 0.800$; choosing $\beta_0 = 10.65$ gives $\gamma_0 = 9.55$ and $\rho_0 = 0.829$; choosing $\beta_0 = 11.00$ gives $\gamma_0 = 8.71$ and $\rho_0 = 0.865$; and, choosing $\beta_0 = 11.55$ gives $\gamma_0 = 8.69$ and $\rho_0 = 0.892$.

Gompertz model

The model may be written as $Y_i = \gamma \exp(-\beta\rho^{x_i})$. On the logarithmic scale we have $\log(Y_i) = \log \gamma - \beta\rho^{x_i}$. In this form the log of the Gompertz model is recognized as the Mitscherlich model. Thus, the same methods developed for the Mitscherlich model may be used for determining initial parameter estimates for the Gompertz model.

Logistic model

The logistic model has received attention by many authors. Nair (1954) reviews methods proposed by Fisher, Yule, Hotelling and others for fitting the logistic curve $Y_i = \frac{\beta}{1+\gamma e^{-\alpha x_i}}$. Often, β is assumed to be known and occasionally γ as well. For equal spacing of the x_is the model may be written as $Y_i = \frac{\beta}{1+\gamma\rho^i}$.

(i) Note that we may write $\frac{1}{Y_{i+1}} = \left(\frac{1-\rho}{\beta}\right) + \rho\left(\frac{1}{Y_i}\right)$. Identify $\alpha = \frac{1-\rho}{\beta}$, then $\rho_0 = \hat{\rho}$ and $\beta_0 = \frac{1-\rho}{\alpha}$ from simple linear regression of $\frac{1}{y_{i+1}}$ on $\frac{1}{y_i}$. This method was suggested by Rhodes (1940).

(ii) Recall that the differential equation defining logistic growth may be written as

$$1/Y\ dY/dx = \alpha[1-(Y/\beta)] \quad .$$

Thus, an obvious approximation would be provided by the following difference equation

$$1/Y_i(\Delta Y_i) \approx \alpha[1-(Y_i/\beta)] \approx (1-\rho)[1-(Y_i/\beta)] \quad .$$

Let

$$z_i = (y_{i+1}-y_i)/y_i \quad ,$$

$$b_0 = 1-\rho \qquad \text{and} \qquad b_1 = (\rho-1)/\beta \quad .$$

Fit the equation $z_i = b_0 + b_1 y_i$ to determine the least squares estimates of b_0 and b_1. Then, find initial estimates for ρ and β as

$$\rho_0 = 1-\hat{b}_0 \qquad \text{and} \qquad \beta_0 = -\hat{b}_0/\hat{b}_1 \quad .$$

This method was used by Yule (1925).

(iii) Nair (1954) suggested regressing $\frac{1}{y_{i+1}} - \frac{1}{y_i}$ on $\frac{1}{y_{i+1}} + \frac{1}{y_i}$, since the following relationship holds:

$$\frac{1}{Y_{i+1}} - \frac{1}{Y_i} = \frac{\rho-1}{\rho+1}\left(\frac{2}{\beta}\right) + \frac{1-\rho}{1+\rho}\left(\frac{1}{Y_{i+1}} + \frac{1}{Y_i}\right) \quad .$$

Thus, initial estimates for ρ and β are given by

$$\rho_0 = \frac{1-\hat{b}_1}{1+\hat{b}_1} \qquad \text{and} \qquad \beta_0 = -2(\hat{b}_1/\hat{b}_0) \quad ,$$

where $\hat{b}_0$ and $\hat{b}_1$ are the least squares estimates of the intercept and slope, respectively, from the above equation.

(iv) Let $d_i = \frac{1}{y_i} - \frac{1}{y_{i+1}}$ and define $r_i = \frac{d_{i+1}}{d_i}$. Then, choose $\rho_0 = \text{median}(r_i)$. This estimator is found to be unstable in practice.

(v) Suppose $n=3r-1$ and define S_k as

$$S_k = \sum_{i=(k-1)r}^{kr-1} 1/y_i \qquad \text{for } k=1,2,3 \quad .$$

Then, $S_k \approx \frac{1}{\beta}\left[r + \gamma\rho^{(k-1)r}\left(\frac{1-\rho^r}{1-\rho}\right)\right]$ and hence $D_1 = S_1 - S_2 \approx \frac{\gamma}{\beta}\left(\frac{1-\rho^r}{1-\rho}\right)(1-\rho^r)$, and $D_2 = S_2 - S_3 \approx \frac{\gamma}{\beta}\left(\frac{1-\rho^r}{1-\rho}\right)(1-\rho^r)\rho^r$, giving $\rho^r \approx D_2/D_1$, or $\rho_0 = \left(\frac{D_2}{D_1}\right)^{1/r}$. Also, $\frac{r}{\beta_0} = S_1 - \frac{D_1^2}{D_1-D_2}$ which yields $\beta_0 = r\left[S_1 - \frac{D_1^2}{D_1-D_2}\right]^{-1}$. Finally, $\gamma_0 = \frac{\beta_0(1-\rho_0)}{1-\rho_0^r}\left[\frac{D_1^2}{D_1-D_2}\right]$.

(vi) If the xs are not equally spaced then choose $\beta_0 > \max_{0\le i\le n} y_i$ and let

$$z_i = \log\left(\frac{\beta_0}{y_i} - 1\right) = \log\gamma + (\log\rho)x_i \quad .$$

Find

$$\gamma_0 = \exp(\widehat{\log\gamma}) \qquad \text{and } \rho_0 = \exp(\widehat{\log\rho})$$

from least squares.

As in the Mitscherlich case, the possibility of overestimating β with β_0 is very likely.

<u>Example</u>

Data (data set 3 in the appendix) have been generated from the model

$$y_i = \frac{10}{1+5(.8)^i} + \sigma z_i, \qquad i=0,1,\cdots,19 \quad ,$$

$z_i \sim$ iid $N(0,1)$ and $\sigma=0.5$. Hence, $\beta=10$, $\gamma=5$ and $\rho=0.8$. The results for methods (i) through (vi) follow.

(i) $\rho_0 = 0.672$ and $\beta_0 = 7.73$.

(ii) $\rho_0 = 0.624$ and $\beta_0 = 9.79$ using $\Delta y_i/y_{i+1}$, whereas $\rho_0 = 0.762$ and $\beta_0 = 10.38$ when using $\Delta y_i/y_i$.

(iii) $\rho_0 = 0.690$ and $\beta_0 = 8.33$.

(iv) median$(r_i) = 0.12$ and $\bar{r} = 0.00$, not too useful.

(v) $\rho_0 = 0.748$, $\gamma_0 = 8.00$ and $\beta_0 = 9.16$ where $S_1 = 3.5160$, $S_2 = 1.1578$, $S_3 = 0.7432$, $D_1 = 2.3583$ and $D_2 = 0.4145$. The last two observations were discarded for these calculations.

(vi) Note that the maximum y_i was 9.57. Then, choosing $\beta_0 = 9.6$ gives $\gamma_0 = 9.52$ and $\rho_0 = 0.732$; choosing $\beta_0 = 10$ gives $\gamma_0 = 7.55$ and $\rho_0 = 0.777$; and choosing $\beta_0 = 11$ gives $\gamma_0 = 7.13$ and $\rho_0 = 0.811$.

Linear combinations of exponentials

Linear combinations of exponentials arise naturally in the study of decay and tracer problems. Pharmacokinetic and compartmental models are common examples. When determining parameter estimates for these models it is critical to have good initial estimates upon entering an iterative nonlinear optimization routine. Two methods will be discussed here.

The first method is known as "curve peeling" (Foss, 1969). This is an intuitive procedure that requires practice and subjectivity. A justification for the peeling technique will be outlined. The second method is called the "method of partial totals" (Cornell, 1962). The method of partial totals is straightforward, produces consistent estimators, but requires equally spaced xs.

The model may be written as:

$$Y_i = \sum_{k=1}^{p} \beta_k e^{-\alpha_k x_i} ,$$

or, in the case of equally spaced xs:

$$Y_i = \sum_{k=1}^{p} \beta_k \rho_k^i \quad .$$

Curve peeling

Curve peeling is a process whereby estimates of β_k and α_k are successively determined. If the αs are ordered such that $\alpha_1 > \alpha_2 > \cdots > \alpha_p$, then it is intuitively clear that for large x all but the exponential term in α_p have decayed to zero. Thus, when plotting $\log(y_i)$ vs. x the responses corresponding to the largest xs will tend to fall on a straight line. It is from this part of the curve that the estimates of β_p and α_p are "peeled" off.

The procedure may be described in the following steps.

Step 1: Plot $\log(y_i)$ vs. x_i and determine the linear segment associated with the m_p largest xs (preferably m_p should be greater than two). Regress $\log(y_i)$ on x_i for the m_p points to determine the least squares estimates $\hat{\alpha}_p$ and $\log(\hat{\beta}_p)$. Thus, determine $\hat{\beta}_p = \exp[\log(\hat{\beta}_p)]$. Now, calculate $y_{i,p} = y_i - \hat{\beta}_p \exp(-\hat{\alpha}_p x_i)$ for $i=0,1,2,\cdots,n$. Discard at least m_p $y_{i,p}$s associated with the m_p observations used to calculate $\hat{\alpha}_p$ and $\hat{\beta}_p$.

Step 2: Plot $\log(y_{i,p})$ vs. x_i, $i=0,1,2,\cdots,n-m_p$. Regress $\log(y_{i,p})$ on the m_{p-1} largest xs to obtain $\hat{\alpha}_{p-1}$ and $\hat{\beta}_{p-1}$ as in Step 1. Calculate $y_{i,p-1} = y_{i,p} - \hat{\beta}_{p-1} \exp(-\hat{\alpha}_{p-1} x_i)$, $i=0,1,2,\cdots,n$. Again, discard the $y_{i,p-1}$s associated with the largest $m_p + m_{p-1}$ xs.

Step 3: Continue to "peel off" estimates until $\hat{\beta}_1$ and $\hat{\alpha}_1$ are calculated from $\log(y_{i,2})$ vs. x_i. The $y_{i,1}$s will be the residuals.

Note: Clearly, we must have a minimum of $n=2p$ observations to obtain estimates of all βs and αs. In practice many more observations are desired.

Justification of the "peeling" procedure

A justification for the method is examined for p=2. The results easily extend for any p.

Suppose, without loss of generality, that $\alpha_1 > \alpha_2 > 0$. Write $Y_i = \beta_1 e^{-\alpha_1 x_i} + \beta_2 e^{-\alpha_2 x_i}$, then

$$\log(Y_i) = \log\left\{\beta_2 e^{-\alpha_2 x_i}\left[1 + \frac{\beta_1}{\beta_2} e^{-(\alpha_1-\alpha_2)x_i}\right]\right\}$$

$$= \log \beta_2 - \alpha_2 x_i + \log\left[1 + \frac{\beta_1}{\beta_2} e^{-(\alpha_1-\alpha_2)x_i}\right] .$$

Recall that $\log(1+z) = z - \frac{1}{2} z^2 + \frac{1}{3} z^3 - \cdots$ for $|z| < 1$. Hence, for curve peeling to be successful observations are required at large enough levels of x so that $\frac{\beta_1}{\beta_2} e^{-(\alpha_1-\alpha_2)x_i} \ll 1$. If this is so then we may write $\log(Y_i) = \log \beta_2 - \alpha_2 x_i + \left\{\frac{\beta_1}{\beta_2} e^{-(\alpha_1-\alpha_2)x_i} - \frac{\beta_1^2}{2\beta_2^2} e^{-2(\alpha_1-\alpha_2)x_i} + \cdots\right\}$. For sufficiently large x the bracketed series is negligible. Thus, regressing $\log(Y_i)$ on x will yield viable estimates of α_2 and β_2.

How does the bias effect the estimates of α_2 and β_2 if observations are not taken at sufficiently large values of x? Clearly the bias is positive and depends upon x. In fact, a linear approximation to the bias may be given by $k_1 - k_2 x_i$ where $k_1, k_2 > 0$. Thus, the equation for $\log(Y_i)$ including the linear approximation for bias becomes:

$$\log(Y_i) \approx (\log \beta_2 + k_1) - (\alpha_2 + k_2) x_i \quad .$$

Consequently, one would expect $\hat{\alpha}_2$ and $\hat{\beta}_2$ to be biased upwards. This is indeed the case in the following example.

Example of curve peeling for p=2

Data (data set 4 in the appendix) are generated from the equation

$$Y_i = 10 \exp(-1.0\ x_i) + 2 \exp(-0.2\ x_i), \qquad i=0,1,2,\cdots,24 \quad .$$

The xs lie between 0 and 6 in increments of 0.25. No random errors have been added to the data.

The following table illustrates the behavior of the estimates due to regressing $\log(Y_i)$ on the last m xs. The estimates of α_1 and β_1 were calculated using $Y_{i,2} = Y_i - \hat{\beta}_2 \exp(-\hat{\alpha}_2 x_i)$, $i=0,1,\cdots,9$.

m	$\hat{\beta}_2$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\alpha}_2$
2	2.57	.235	9.61	1.086
3	2.62	.238	9.57	1.094
4	2.69	.243	9.52	1.103
5	2.76	.247	9.47	1.114
6	2.84	.252	9.42	1.127
7	2.93	.258	9.35	1.141
8	2.98	.264	9.29	1.143
9	3.15	.271	9.21	1.177
10	3.28	.279	9.12	1.200
11	3.43	.287	9.03	1.227
12	3.60	.297	8.92	1.260
25	8.25	.479	–	–
exact	2	0.2	10	1.0

As indicated in the previous discussion the estimates of α_2 and β_2 have a positive bias. In fact, we know that

$$\text{bias} = \log\left[1 + \frac{\beta_1}{\beta_2} e^{-(\alpha_1-\alpha_2)x_i}\right] = \log\left(1 + 5e^{-0.8x_i}\right) \quad .$$

For example, the following data were used to estimate k_1 and k_2 in order to adjust the estimates for m=5.

x_i	Bias		
5.00	0.087624	$k_1 = 0.322$	$(R^2 = 0.995)$
5.25	0.074980	$k_2 = 0.047$	
5.50	0.061387		
5.75	0.050259		
6.00	0.041149		

Thus, the revised estimates are given by

$$\hat{\alpha}_2 = 0.247 - 0.047 = 0.200, \quad \text{and}$$

$$\hat{\beta}_2 = \exp(0.322)2.76 = 2.00 \quad .$$

Unfortunately we do not know the true bias in practice. However, it is prudent to apply the peeling procedure several times choosing different m_ks each time. The estimates yielding the smallest residual sum of squares should be used as the initial parameter estimates. It may also be wise to deflate the estimates of α_p and β_p before continuing the peeling process since we know that they are likely biased upwards.

Curve peeling to determine p

Occasionally the observations must be used to determine p, the number of exponentials in the sum. Curve peeling provides an easy way to make this determination at each step of the peeling process. Recall that at each stage the residuals $y_{i,k}$ are computed. Thus, simply examine the scatter plot of these residuals versus x_i at each stage of peeling. When the residual plot reflects only random variability then choose p=k. There may be an obvious trend in the residuals reflecting variance heterogeneity, but there should be no evidence of a linear or curvilinear trend remaining that could be explained by x. Additionally, if the mean of the random residuals is nonzero (positive) then this provides evidence that the process under study is decaying to some constant background other than zero. This case is discussed shortly.

Method of partial totals

The method of partial totals has been discussed at length by Cornell (1962). The method requires equally spaced xs and 2pm observations to estimate 2p parameters. See Cornell (1962) for derivations and proof of consistency of the method. Related papers include Agha (1971), Della Corte, Buricchi and Romano (1974).

The model will be written as

$$Y_i = \sum_{k=1}^{p} \beta_k \rho_k^i \qquad \text{for } i=0,2,\cdots,2pm-1 \quad .$$

It is assumed that $\beta_k \neq 0$ for all k and $\rho_k \neq \rho_{k'}$ for all $k \neq k'$. Remember that $\rho_k = \exp(-\alpha_k)$, or equivalently that $\alpha_k = -\log \rho_k$.

Define 2p independent partial totals (sums), each of which is the sum of m observations as:

$$S_q = \sum_{i=(q-1)m}^{qm-1} y_i, \qquad q=1,2,\cdots,2p \quad .$$

Then, the estimators of the ρ_k and β_k are found in the following two steps.

Step 1: Let R and R_{p-r+1} be p×p matrices whose u,vth elements are defined by:

$$(R)_{uv} = S_{u+v-1} \quad ,$$

$$(R_{p-r+1})_{uv} = \begin{cases} S_{u+v-1} & \text{if} \quad u \leq p-r \\ S_{u+v} & \text{if} \quad u > p-r \end{cases} ,$$

for $r=1,2,\cdots,p$. Define $L_r = \det(R_{p-r+1})/\det(R)$, for $r=1,2,\cdots,p$.

The estimators of ρ_k^m are found to be the roots of the pth degree polynomial given by:

$$z^p - L_1 z^{p-1} + L_2 z^{p-2} - \cdots + (-1)^p L_p = 0 \quad .$$

Calculate $\hat{\rho}_k = \left[\hat{\rho}_k^m\right]^{1/k}$, $k=1,2,\cdots,p$.

Step 2: Once the estimates of the ρ_k have been determined, then the $\hat{\beta}_k$s may be found as the unique solution to the system of equations: $\underset{\sim}{S} = W\underset{\sim}{\hat{\beta}}$, where

$$\underset{\sim}{S}' = (S_1, S_2, \cdots, S_p) \quad ,$$

$$\underset{\sim}{\hat{\beta}}' = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p) \quad ,$$

and W is a p×p matrix with u,vth element defined by

$$(W)_{uv} = \frac{1}{1-\hat{\rho}_v}\left[\hat{\rho}_v^{(u-1)m}(1-\hat{\rho}_v^m)\right] \quad .$$

Alternative Step 2: An obvious alternative procedure to determine estimates of the β_k is to define the n×p matrix X as:

$$(X)_{uv} = \exp(-\hat{\alpha}_v x_u) \qquad \begin{array}{l} u=0,1,\cdots,n \\ v=1,2,\cdots,p \end{array} \quad .$$

Proceed to calculate

$$\underset{\sim}{\hat{\beta}} = (X'X)^{-1}X'\underset{\sim}{y} \quad ,$$

where $\underset{\sim}{y}$ is the n×1 vector of observed responses. Any multiple linear regression program may be used to find $\underset{\sim}{\hat{\beta}}$. Remember to specify the no-intercept option of the computing package.

Example of the method of partial totals for p=2

The same model used for the curve peeling example will be used for this example. Use x_is for i=0,1,···,23, giving 2p=4 partial totals with m=6 observations per group. Thus, calculate

$$S_1 = 45.7494 \qquad S_3 = 7.58167$$

$$S_2 = 15.7104 \qquad S_4 = 4.71143 \quad .$$

Step 1

$$\det(R) = \det\begin{pmatrix} S_1 & S_2 \\ S_2 & S_3 \end{pmatrix} = S_1S_3 - S_2^2 = 100.0405$$

$$\det(R_1) = \det\begin{pmatrix} S_2 & S_3 \\ S_3 & S_4 \end{pmatrix} = S_2S_4 - S_3^2 = 16.53668$$

$$\det(R_2) = \det\begin{pmatrix} S_1 & S_2 \\ S_3 & S_4 \end{pmatrix} = S_1S_4 - S_2S_3 = 96.4341 \quad .$$

Hence,

$$L_1 = \det(R_1)/\det(R) = 0.96395 \quad ,$$

and

$$L_2 = \det(R_1)/\det(R) = 0.16530 \quad .$$

The p=2 degree polynomial is given by $z^2 - 0.96395z + 0.16530$. The roots are 0.22313 and 0.74082. Consequently,

$$\hat{\rho}_1^6 = 0.22313 \qquad \text{giving} \qquad \hat{\rho}_1 = 0.77880$$

$$\hat{\rho}_2^6 = 0.74082 \qquad \text{giving} \qquad \hat{\rho}_2 = 0.95123 \quad .$$

In order to calculate $\hat{\alpha}_1$ and $\hat{\alpha}_2$, note that $x_i = 0.25i$, $i=0,1,\cdots,23$. Thus,

$$\hat{\rho}_k^i = e^{-\hat{\alpha}_k(0.25)i} \quad ,$$

or

$$-4 \log(\hat{\rho}_k) = \hat{\alpha}_k \quad .$$

Calculate

$$\hat{\alpha}_1 = -4 \log(0.77880) = 1.0000 \quad ,$$

and

$$\hat{\alpha}_2 = -4 \log(0.95123) = 0.2000 \quad .$$

<u>Step 2</u>: Note that $\underset{\sim}{S} = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} = \begin{pmatrix} 45.7494 \\ 15.7104 \end{pmatrix}$ and $\hat{\underset{\sim}{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$. Calculate

$$(W)_{11} = \frac{1-\hat{\rho}_1^6}{1-\hat{\rho}_1} = 3.5121 \quad ,$$

$$(W)_{21} = \frac{\hat{\rho}_1^6(1-\hat{\rho}_1^6)}{1-\hat{\rho}_1} = 0.78365 \quad ,$$

$$(W)_{12} = \frac{1-\hat{\rho}_2^6}{1-\hat{\rho}_2} = 5.3143 \quad ,$$

and

$$(W)_{22} = \frac{\hat{\rho}_2^6(1-\hat{\rho}_2^6)}{1-\hat{\rho}_2} = 3.9369 \quad .$$

Thus,

$$W^{-1} = \begin{bmatrix} 0.40745 & -0.55001 \\ -0.08110 & 0.36349 \end{bmatrix} \quad ,$$

and

$$\underset{\sim}{\beta} = W^{-1}\underset{\sim}{S} = \begin{pmatrix} 10.000 \\ 2.000 \end{pmatrix} \quad .$$

So, the estimates found from steps 1 and 2 are:

$$\hat{\alpha}_1 = 1.0000, \qquad \hat{\beta}_1 = 10.000$$

and

$$\hat{\alpha}_2 = 0.20000, \qquad \hat{\beta}_2 = 2.000 \quad .$$

<u>Alternative Step 2</u>: Calculate $X'X = \begin{pmatrix} 2.54148 & 3.85542 \\ 3.85542 & 9.55504 \end{pmatrix}$, $X'\underset{\sim}{y} = \begin{pmatrix} 33.1256 \\ 57.6642 \end{pmatrix}$, $(X'X)^{-1} = \begin{pmatrix} 1.01437 & -0.40929 \\ -0.40929 & 0.26980 \end{pmatrix}$ and $\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y} = \begin{pmatrix} 10.000 \\ 2.000 \end{pmatrix}$. Again, $\hat{\beta}_1 = 10$ and $\hat{\beta}_2 = 2$, as expected.

The estimates are exact to within rounding error because no random error was added to the model.

In practice both curve peeling and the method of partial totals will provide initial estimates of the parameters in the model. The method of partial totals is straightforward, whereas the curve peeling technique requires some artistic subjectivity on the part of the investigator.

Sum of exponentials plus a constant

Suppose the model is written as

$$Y_i = \gamma + \sum_{k=1}^{p} \beta_k e^{-\alpha_k x_i} .$$

If k=1 the model is simply

$$Y_i = \gamma + \beta e^{-\alpha x_i}$$

and methods similar to those for the Mitscherlich model may be used. Recognize that γ is an asymptote to which the response now decays rather than grows as in the Mitscherlich model.

If $k \geq 2$ then one of several methods may be employed. An extension of the method of partial totals is discussed by Cornell (1962). A simpler approach is to take $\gamma_0 = \min_{0 \leq i \leq n} y_i$ and proceed to work with the new data as $y_i - \gamma_0$ via the techniques developed for sums of exponentials. It may be wise to choose $\gamma_0 = f \min_{0 \leq i \leq n} y_i$ where $0 < f \leq 1$ since bias due to the exponential terms may be non-negligible.

Summary

All of the nonlinear models considered in this report have at least one thing in common. They are all solutions to first order differential equations (or a system of first order DEs in the case of the sum of exponentials). It is clear that initial parameter estimates are easily found using linear least squares once an estimate of the parameter ρ (or the ρ_ks) has been determined. Similar approaches may be applied to nonlinear models not discussed here.

Probably the single most important key to successful initial estimates is common sense. It is common sense to crudely interpolate the observed responses so that simple (often consistent) methods developed for equally spaced xs may be used. It is also common sense that directs one to use estimators based upon a suitably linearized version of the model as well as past experience. These will always prove superior to "hit or miss" (sometimes aptly termed "shotgun") procedures.

References

Agha, M. (1971). A direct method for fitting linear combinations of exponentials. *Biometrics* **27**, 399-413.

Clarke, G.P.Y. (1980). Moments of the least squares estimators in a nonlinear regression model. *J.R.S.S. (B)* **42**, 227-237.

Cornell, R.G. (1962). A method for fitting linear combinations of exponentials. *Biometrics* **18**, 104-113.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press, Princeton, NJ.

Della Corte, M., Buricchi, L. and Romano, S. (1974). On the fitting of linear combinations of exponentials. *Biometrics* **30**, 367-369.

Finney, D.J. (1958). The efficiencies of alternative estimators for an asymptotic regression equation. *Biometrika* **45**, 370-388.

Foss, S.D. (1969). A method for obtaining initial estimates of parameters in exponential curve fitting. *Biometrics* **25**, 580-584.

Hartley, H.O. (1948). The estimation of nonlinear parameters by 'internal least squares.' *Biometrika* **35**, 32-45.

Hartley, H.O. and Booker, A. (1965). Nonlinear least squares estimation. *Ann. Math. Statist.* **36**, 638-650.

Jennrich, R.I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40**, 633-643.

Malinvaud, E. (1970). The consistency of nonlinear regressions. *Ann. Math. Statist.* **41**, 956-969.

Nair, K.R. (1954). The fitting of growth curves. In *Statistics and Mathematics in Biology.* O. Kempthorne, T.A. Bancroft, J.W. Gowen and J.L. Lush, editors. Hafner, New York.

Rhodes, E.C. (1940). Population mathematics — III. *J.R.S.S.* **103**, 362-387.

Yule, G.U. (1925). The growth of population and the factors which control it. *J.R.S.S.* **88**, 1-58.

Appendix: Data sets used in the examples

Set 1

i	x_i	y_i
1	0	9.24197
2	1	5.97771
3	2	7.30348
4	3	7.30334
5	4	3.96406
6	5	5.23950
7	6	1.79433
8	7	2.04656
9	8	1.12700
10	9	1.44355
11	10	0.64757
12	11	0.72615
13	12	0.52754
14	13	0.49025
15	14	0.63366
16	15	0.26805
17	16	0.24347
18	17	0.19073
19	18	0.21324
20	19	0.10275

Set 2

i	x_i	y_i
1	0	0.2347
2	1	1.8788
3	2	3.8143
4	3	5.1841
5	4	6.0284
6	5	6.7501
7	6	7.4583
8	7	9.0566
9	8	8.6888
10	9	8.4538
11	10	8.8172
12	11	8.7051
13	12	9.4145
14	13	10.0972
15	14	10.1524
16	15	10.1484
17	16	10.4178
18	17	10.5485
19	18	9.9981
20	19	9.6715

Set 3

i	x_i	y_i
1	0	0.89988
2	1	1.14275
3	2	1.97427
4	3	2.71334
5	4	3.21267
6	5	2.91246
7	6	4.56462
8	7	4.61599
9	8	4.15900
10	9	5.59675
11	10	6.48571
12	11	6.72315
13	12	7.59674
14	13	7.47629
15	14	7.52931
16	15	7.93789
17	16	9.57398
18	17	8.72673
19	18	8.99235
20	19	9.48426

Set 4

i	x_i	y_i
1	0.00	12.0000
2	0.25	9.6905
3	0.50	7.8750
4	0.75	6.4451
5	1.00	5.3163
6	1.25	4.4226
7	1.50	3.7129
8	1.75	3.1471
9	2.00	2.6940
10	2.25	2.3292
11	2.50	2.0339
12	2.75	1.7932
13	3.00	1.5955
14	3.25	1.4318
15	3.50	1.2951
16	3.75	1.1799
17	4.00	1.0818
18	4.25	0.9975
19	4.50	0.9242
20	4.75	0.8600
21	5.00	0.8031
22	5.25	0.7524
23	5.50	0.7066
24	5.75	0.6651