8.1 Why nonlinear models

1. (Comment) The world abounds in nonlinear models.

   (i) Mathematical model = differential equation

   (ii) diff. eqn. $\Rightarrow$ nonlinear model

2. (Growth eq.) $\frac{df}{dt} = \lambda f$

   Sol: $f(t) = \gamma e^{\lambda t} = \text{nonlinear model}$
3. (Unified Field Theory of Statistics)


4. (Non-standard applications)

Maximum likelihood estimation
Robust estimation
5. (Note) \( f = \gamma e^{\lambda t} \)
Is nonlinear because it is nonlinear in \( \lambda \) and **not** because it is nonlinear in \( t \).

6. (Least squares fitting) Minimize

\[
Q(\gamma, \lambda) = \sum_{i=1}^{n} (y_i - \gamma e^{\lambda t_i})^2
\]
8.2 An overview of PROC NLIN

Make a standard program immediately available.

Focus attention on aspects of the subject used in practice.
8.3 The nonlinear regression model

1. (Model) \[ y_i = f(x_i, \theta) + e_i \quad ; \quad i = 1, \ldots, n \]

\[ x_i = (x_{i1}, \ldots, x_{it}) \quad , \quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} \in \Theta \]

2. (Ex) \[ y_i = \gamma e^{\lambda t_i} + e_i \]

\[ x_i = t_i \quad , \quad \theta = \begin{pmatrix} \gamma \\ \lambda \end{pmatrix} \quad , \quad \Theta : \gamma \geq 0, \lambda \leq 0 \]

3. (Vector form) \[ y = f(\theta) + e \]

\[ f(\theta) = (f_i(\theta)) \quad , \quad f_i(\theta) = f(x_i, \theta) \]

4. (Linear case) \[ y = X\theta + e \quad , \quad f(\theta) = X\theta \]
8.4 Nonlinear least squares

1. (Def) $\hat{\theta}$ is a **least squares estimate** of $\theta$ iff $\theta = \hat{\theta}$ minimizes:

$$Q(\theta) = \sum_{i=1}^{n} (y_i - f(f(x_i, \theta)))^2 \quad (1)$$

$$= \|y - f(\theta)\|^2 , \quad \theta \in \Theta \quad (2)$$

2. (Observation space picture)
3.(Variable space picture)

4.(Normal equations) If $\hat{\theta}$ is an interior point of $\Theta$, then $\theta = \hat{\theta}$ satisfies

$$\sum_{i=1}^{n} \frac{\partial f_i}{\partial \theta_j}(y_i - f_i(\theta)) = 0, \ j = 1, \ldots, p$$

Pf: Differentiate $Q(\theta)$ w.r.t. $\theta_j$.

5.(Vector form) Let

$$\frac{df}{d\theta} = \left( \frac{\partial f_i}{\partial \theta_j} \right) = \text{Jacobian of } f(\theta)$$
The normal equations become

\[ \frac{df^T}{d\theta} (y - f(\theta)) = 0 \]

6. (Linear case) \( f(\theta) = X\theta \), \( \frac{df}{d\theta} = X \)
The normal equations are

\[ X^T (y - X\theta) = 0 \]
8.5 The Gauss-Newton algorithm

The algorithm

1. (Why)

(i) Works well (Bard, 1970)
(ii) Natural interface to linear theory
(iii) Basis for many other algorithms and most programs
(iv) Many non-standard applications
1. (Idea)

(i) Project $y - f(\theta)$ on tangent plane

(ii) Replace $\theta$ by $\theta + \Delta \theta$ and repeat

3. (Associated linear least squares problem) Minimize w.r.t. $\Delta \theta$

$$Q_\theta(\Delta \theta) = \| y - f(\theta) - \frac{df}{d\theta} \Delta \theta \|^2$$
4. (Gauss-Newton algorithm) The solution to 3 is

\[ \Delta \theta = \left( \frac{df^T}{d\theta} \frac{df}{d\theta} \right) \frac{df^T}{d\theta} (y - f(\theta)) \]
Standard modifications

Partial step modification:

1. (Partial step th) If \( \theta \) is an interior point of \( \Theta \) and \( \Delta \theta \) is defined and non-zero, then a small enough step in the direction of \( \Delta \theta \) will reduce \( Q(\theta) \).

Pf: Recall that if \( A \) is positive definite and \( x \neq 0 \), then \( x^T A x > 0 \).

\[
\frac{dQ}{d\theta} = -2(y - f)^T \frac{df}{d\theta}
\]

thus

\[
\lim_{\alpha \to 0} \frac{Q(\theta + \alpha \Delta \theta) - Q(\theta)}{\alpha} = \frac{dQ}{d\theta} \Delta \theta
\]

\[
= -2(y - f(\theta))^T \frac{df}{d\theta} \left( \frac{df^T}{d\theta} \frac{df}{d\theta} \right)^{-1} \frac{df^T}{d\theta} (y - f(\theta)) > 0
\]

2. (Step halving) \( \Delta \theta, \Delta \theta/2, \Delta \theta/4, \ldots \)
Other modifications:

1. (Constraints)
   \[ a_i \leq \theta_i \leq b_i \quad , \quad i = 1, \ldots, p \]

2. (Weights)
   \[ Q_w(\theta) = \sum w_i (y_i - f_i(\theta))^2 = \sum (\sqrt{w_i}y_i - \sqrt{w_i}f_i(\theta))^2 \]

The Gauss-Newton algorithm becomes

\[ \Delta \theta = \left( \frac{d f^T}{d \theta} W \frac{d f}{d \theta} \right)^{-1} \frac{d f^T}{d \theta} W (y - f(\theta)) \]
\[ W = \begin{pmatrix} w_1 & \cdots & 0 \\ 0 & \cdots & w_n \end{pmatrix} \]
Using the computer

1. (Logistic growth model)

\[ f(t, \theta) = M \frac{e^{\alpha + \beta t}}{1 + e^{\alpha + \beta t}}, \quad \theta = (\alpha, \beta, M)' \]
2. (Ex 8.1) U.S. population

\[
\frac{\partial f}{\partial M} = \frac{e^{\alpha + \beta t}}{1 + e^{\alpha + \beta t}} \frac{df}{dt} = \rho
\]
\[
\frac{\partial f}{\partial \alpha} = f(1 - \rho)
\]
\[
\frac{\partial f}{\partial \beta} = f(1 - \rho)t
\]

Initial values:

\[M = 300 = \text{asymptote guess}\]

Fitting at \(t = 1, 20\) (Text)

\[\alpha = -4.6, \; \beta = 0.29\]
Log model:

\[ \log y = \log f + e \]

This makes exponential growth plot linearly on \( t \). The derivatives are easy.

\[ \frac{\partial}{\partial \theta} \log f = \frac{\partial f}{\partial \theta} / f \]
Statistical properties

1. (Assumptions)

\[ y_1, \cdots, y_n \text{ independent} \]
\[ E \ y_i = f_i(\theta) \text{ unbiased} \]
\[ \text{var}(\sqrt{w_i}y_i) = \sigma^2 \]

There is no formal advantage to the assumption of normality.

2. (Weighted least squares) \( \hat{\theta} \) minimizes

\[ Q_w(\theta) = \sum w_i(y_i - f_i(\theta))^2 \]

3. (Distribution)

\[ \hat{\theta} \sim N \left( \theta, \sigma^2 \left( \frac{df^T}{d\theta} W \frac{df}{d\theta} \right)^{-1} \right) \]

Conditions p. 269
4. (Residual mean square) Let
\[ \hat{\sigma}^2 = Q_w(\hat{\theta})/(n - p) \approx \sigma^2 \]

5. (Standard error estimates)
\[ \text{cov} \hat{\theta} = \hat{\sigma}^2 \left( \frac{\hat{f}'^T}{d\theta} W \frac{\hat{f}}{d\theta} \right)^{-1} \]

6. (Functions of parameters)
\[ \text{var } g(\hat{\theta}) = \frac{\hat{g}}{d\theta} \left( \text{cov} \hat{\theta} \right) \frac{\hat{g}'^T}{d\theta} \]

Motivation:
\[ g(\hat{\theta}) \approx g(\theta) + \frac{dg}{d\theta}(\hat{\theta} - \theta) \]
\[ \text{var } g(\hat{\theta}) \approx \text{var } \left( \frac{dg}{d\theta}(\hat{\theta} - \theta) \right) = \frac{dg}{d\theta} \left( \text{cov} \hat{\theta} \right) \frac{g'^T}{d\theta} \]
7. (Intervals and tests)
\[
\frac{g(\hat{\theta}) - g(\theta)}{\text{std } g(\hat{\theta})} \sim t(n - p)
\]

8. (Corr)
\[
\frac{\hat{\theta}_i - \theta_i}{\text{std } \hat{\theta}_i} \sim t(n - p)
\]
\[
\text{var } \hat{\theta}_i = \hat{\sigma}^2 \left( \frac{d\hat{f}^T}{d\theta} W \frac{df}{d\theta} \right)_{ii}^{-1}
\]
9. (Ex 8.2) Maize, Table 8.2

\[ \log y = \log f(t, \theta) + e \]

\[ f(t, \theta) = M \frac{e^{\alpha + \beta t}}{1 + e^{\alpha + \beta t}}, \quad \theta = (\alpha, \beta, M)' \]

The problem specification is as in Fig 8.17
9. (Ex 8.2 Cont) Note that

$$\beta = \hat{\beta} \pm t_0 \text{std} \hat{\beta} = 0.113 \pm 0.010 = \text{output}$$

To estimate \( f(t, \theta) \) at \( t = 204 \):

$$\log f = \log f \pm t_0 \text{std}(\log f) = 2.906 \pm 0.1943$$

$$15.1 < f < 22.2$$

For: \( g(\theta) = M\beta/4 = \text{max growth rate} \)

$$\frac{dg}{d\theta} = (0, M/4, \beta/4)$$

$$\text{var} \: \tilde{g} = \frac{dg}{d\theta} (\text{cov} \: \hat{\theta}) \frac{dg^T}{d\theta}$$

$$= \frac{1}{16} (\hat{M}, \hat{\beta}) \begin{pmatrix} \text{var}(\hat{\beta}) & \text{cov}(\hat{\beta}, \hat{M}) \\ \text{cov}(\hat{M}, \hat{\beta}) & \text{var}(\hat{M}) \end{pmatrix} \begin{pmatrix} \hat{M} \\ \hat{\beta} \end{pmatrix}$$

$$= 0.1035$$
Trick:

Add case: 1.

Add code: Fig 8.24
9. (Ex 8.2 cont)

$$f_{n+1} = M \beta / 4 = g$$

$$\hat{g} = \hat{y}_{n+1} = 3.389, \; \text{std} \; \hat{g} = .3219$$

Note: $(\text{std} \; \hat{g})^2 = .1036 = \text{value above}$
Goodness of fit and other hypotheses

1. (Restricted model)

\[ y = f(\theta) + e, \quad \theta \in \Theta_0 \subset \Theta \]

\[ \hat{\theta} = \text{restricted LS est}, \quad q = \text{dim } \Theta_0 \]

2. (Approximate F)

\[ \frac{n-p}{p-q} \cdot \frac{Q_w(\hat{\theta}) - Q_w(\hat{\theta})}{Q_w(\hat{\theta})} \sim F(p - q, n - p) \]

Conditions: Text, p276

3. (Other notation)

\[ \frac{n-p}{p-q} \cdot \frac{RSS_0 - RSS}{RSS} \sim F(p - q, n - p) \]

4. (Determining dimension)

\[ f = \alpha_1 e^{\lambda_1 t} + \alpha_2 e^{\lambda_2 t}, \quad p = 4 \]
The hypotheses $\lambda_2 = 0$ gives

$$f_0 = \alpha_1 e^{\lambda_1 t} + \alpha_2, \quad q = 3$$

The hypothesis $\alpha_2 = 0$ gives

$$f_0 = \alpha_1 e^{\lambda_1 t}, \quad q = 2$$

Note also

$$f = M \frac{e^{\alpha + \beta t}}{1 + e^{\alpha + \beta t}}, \quad p = 3$$

has the restricted form

$$f_0 = \gamma e^{\beta t}, \quad q = 2$$

**Pf:** Let $M = \gamma e^{-\alpha}$. As $\alpha \to -\infty$

$$f = \frac{\gamma e^{\beta t}}{1 + e^{\alpha + \beta t}} \to \gamma e^{\beta t}$$

5. (Ex 8.3, p277) Sulfate data, $n=21$

$$y = \alpha_1 e^{\lambda_1 t} + \alpha_2 e^{\lambda_2 t} + e, \quad RSS = 0.815$$
\[ y = \alpha e^{\lambda t} + e \quad , \quad \text{RSS}_0 = 47.69 \]

\[ F = \frac{21 - 4}{4 - 2} \cdot \frac{47.69 - 0.815}{0.815} = 489^{***} \]

6. (Ex 8.4, p278) Maize data, \( n = 14 \). Let

\[ f = M \frac{e^{\alpha + \beta t}}{1 + e^{\alpha + \beta t}} \quad , \quad p = 3 \]

Test

\[ f_0 = \gamma e^{\lambda t} \quad , \quad q = 2 \]

Using a log-transform

\[ \text{RSS} = .3860 \quad , \quad \text{Fig 8.3} \]

\[ \text{RSS}_0 = 5.1016 \quad , \quad \text{Simple linear reg} \]

\[ F = \frac{14 - 3}{3 - 2} \cdot \frac{5.1016 - .3860}{.3860} = 134.4^{***} \]

as suggested by the plot in Fig 8.22 (p273).