Ch:8 NONLINEAR REGRESSION

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: $f(t) = \gamma e^{\lambda t}$ = nonlinear model

3.(Unified Field Theory of Statistics)

Stat = Linear reg. + Nonlinear reg.

4.(Non-standard applications)

Maximum likelihood estimation

Robust estimation

5.(Note) $f = \gamma e^{\lambda t}$

Is nonlinear because it is nonlinear in λ 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) $y_i = f(x_i, \theta) + e_i \quad ; \quad i = 1, \cdots, n$

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

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) $\widehat{\theta}$ is a **least squares estimate** of θ iff $\theta = \widehat{\theta}$ minimizes:

$$\begin{aligned} Q(\theta) &= \sum_{i=1}^{n} (y_i - f(f(x_i, \theta))^2 && (1) \\ &= \|y - f(\theta)\|^2 \quad , \quad \theta \in \Theta && (2) \end{aligned}$$

2.(Observation space picture)

3.(Variable space picture)

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

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

Pf: Differentiate $Q(\theta)$ w.r.t. θ_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 θ 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 θ is an interior point of Θ 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) $\Delta\theta$, $\Delta\theta/2$, $\Delta\theta/4$, $\cdots$

Other modifications:

1.(Constraints)

$$a_i \le \theta_i \le b_i \quad , \quad i = 1, \cdots, p$$

2.(Weights)

$$\begin{aligned} Q_w(\theta) &= \sum w_i (y_i - f_i(\theta))^2 \\ &= \sum (\sqrt{w_i} y_i - \sqrt{w_i} f_i(\theta))^2 \end{aligned}$$

The Gauss-Newton algorithm becomes

$$\Delta\theta = \left(\frac{df^T}{d\theta} W \frac{df}{d\theta} \right)^{-1} \frac{df^T}{d\theta} W (y - f(\theta))$$

$$W = \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{pmatrix}$$

Using the computer

1.(Logistic growth model)

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

2.(Ex 8.1) U.S. population

$$\frac{\partial f}{\partial M} = \frac{e^{\alpha+\beta t}}{1+e^{\alpha+\beta t}} \stackrel{df}{=} p$$

$$\frac{\partial f}{\partial \alpha} = f(1-p)$$

$$\frac{\partial f}{\partial \beta} = f(1-p)t$$

Initial values:

$M = 300 =$ 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)

$y_1, \cdots, y_n$ independent

$E\, y_i = f_i(\theta)$ unbiased

$var(\sqrt{w_i} y_i) = \sigma^2$

There is no formal advantage to the assumption of normality.

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

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

3.(Distribution)

$$\widehat{\theta} \overset{o}{\sim} N\left(\theta, \sigma^2 \left(\frac{df^T}{d\theta} W \frac{df}{d\theta}\right)^{-1}\right)$$

Conditions p. 269

4.(Residual mean square) Let

$$\widehat{\sigma}^2 = Q_w(\widehat{\theta})/(n-p) \approx \sigma^2$$

5.(Standard error estimates)

$$\widehat{\mathrm{cov}}\,\widehat{\theta} = \widehat{\sigma}^2 \left(\frac{\widehat{df}^T}{d\theta} W \frac{\widehat{df}}{d\theta} \right)^{-1}$$

6.(Functions of parameters)

$$\widehat{\mathrm{var}}\, g(\widehat{\theta}) = \frac{\widehat{dg}}{d\theta} (\widehat{\mathrm{cov}}\,\widehat{\theta}) \frac{\widehat{dg}^T}{d\theta}$$

Motivation:

$$g(\widehat{\theta}) \approx g(\theta) + \tfrac{dg}{d\theta}(\widehat{\theta} - \theta)$$

$$\mathrm{var}\, g(\widehat{\theta}) \approx \mathrm{var}\left(\tfrac{dg}{d\theta} \widehat{\theta} \right) = \tfrac{dg}{d\theta} (\mathrm{cov}\,\widehat{\theta}) \tfrac{dg^T}{d\theta}$$

7.(Intervals and tests)

$$\frac{g(\widehat{\theta}) - g(\theta)}{\widehat{\mathsf{std}}\, g(\widehat{\theta})} \overset{o}{\sim} t(n-p)$$

8.(Corr)

$$\frac{\widehat{\theta}_i - \theta_i}{\widehat{\mathsf{std}}\, \widehat{\theta}_i} \overset{o}{\sim} t(n-p)$$

$$\widehat{\mathsf{var}}\, \widehat{\theta}_i = \widehat{\sigma}^2 \left(\frac{\widehat{df}^T}{d\theta} W \frac{\widehat{df}}{d\theta} \right)^{-1}_{ii}$$

9.(Ex 8.2) Maize , Table 8.2

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

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

The problem specification is as in Fig 8.17

9.(Ex 8.2 Cont) Note that

$$\beta = \widehat{\beta} \pm t_0 \widehat{\mathsf{std}}\, \widehat{\beta} = 0.113 \pm .010 = \text{output}$$

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

$$\log f = \widehat{\log} f \pm t_0 \widehat{\mathsf{std}}(\widehat{\log} f) = 2.906 \pm .1943$$

$$15.1 < f < 22.2$$

For: $g(\theta) = M\beta/4 =$ max growth rate

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

$$\widehat{\text{var}}\, \widehat{g} = \widehat{\frac{dg}{d\theta}}(\widehat{\text{cov}}\, \widehat{\theta})\widehat{\frac{dg}{d\theta}^T}$$

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

$$= 0.1035$$

Trick:

Add case: 1 .

Add code: Fig 8.24

9.(Ex 8.2 cont)

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

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

Note: $(\widehat{\text{std}}\, \widehat{g})^2 = .1036 =$ value above

Goodness of fit and other hypotheses

1.(Restricted model)

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

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

2.(Approximate F)

$$\frac{n-p}{p-q} \cdot \frac{Q_w(\hat{\hat{\theta}}) - Q_w(\hat{\theta})}{Q_w(\hat{\theta})} \overset{\circ}{\sim} F(p-q, n-p)$$

Conditions: Text, p276

3.(Other notation)

$$\frac{n-p}{p-q} \cdot \frac{\text{RSS}_0 - \text{RSS}}{\text{RSS}} \overset{\circ}{\sim} F(p-q, n-p)$$

4.(Determining dimension)

$$f = \alpha_1 e^{\lambda_1 t} + \alpha_2 e^{\lambda_2 t} \quad , \quad p = 4$$

The hypotheses $\lambda_2 = 0$ gives

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

The hypothesis $\alpha_2 = 0$ gives

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

Note also

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

has the restricted form

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

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

$$y \;=\; \alpha_1 e^{\lambda_1 t} + \alpha_2 e^{\lambda_2 t} + e \quad , \quad \text{RSS} = 0.815$$

$$y = \alpha e^{\lambda t} + e \quad , \quad \text{RSS}_0 = 47.69$$

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

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).