Non linear regression

Before $y = X\beta + \varepsilon \quad , \varepsilon \sim N(0, \sigma^2)$

Now $y = \underbrace{f(X;\beta)}_{\text{maybe nonlinear in } \beta} + \varepsilon \, , \quad \varepsilon \sim N(0, \sigma^2)$

Still need Basic Assumptions to hold if using Least Squares to fit.

NLS (nonlinear least squares)

$$\min_{\beta} \sum (y_i - f(x_i;\beta))^2 = \|y - f(X;\beta)\|^2$$

$$\Rightarrow \frac{\partial}{\partial \beta_j} \| \quad \|^2 = -2 \sum (y_i - f(x_i;\beta)) \underbrace{\frac{\partial f(x_i;\beta)}{\partial \beta_j}}_{\text{gradient } (\nabla f)_j} = 0 \quad \forall j$$

$$\Rightarrow \sum_i y_i \frac{\partial f(x_i;\beta)}{\partial \beta_j} = \sum_i f(x_i;\beta) \frac{\partial f(x_i;\beta)}{\partial \beta_j} \quad \forall j$$

$\Rightarrow$ On matrix form $\boxed{(\nabla f)' y = (\nabla f)' f}$

where both f and ∇f may depend on β in a complex fashion

$(\nabla f)' y = (\nabla f)' f$ may not be easy to solve for β, at least no to obtain a closed-form solution.

$\Longrightarrow$ Solution: turn the nonlinear problem into a series of linear ones.

$$y = \underbrace{f(X;\beta)}_{\substack{\text{look at } f \text{ near} \\ \text{a point } \beta^0}} + \varepsilon$$

$\Rightarrow$ 1st order approximation

linear term

$$f(X;\beta) \approx f(X;\beta^0) + \sum_j (\beta_j - \beta_j^0) \left.\frac{\partial f}{\partial \beta_j}\right|_{\beta^0}$$

f evaluated @ β^0

gradient evaluated @ β^0

$$= W^0 + Z^0 \beta$$

$$W^0 = \underbrace{f(X;\beta^0) - \sum_j \beta_j^0 Z_j^0}_{\substack{\text{depends only on} \\ \text{fixed value } \beta^0}}$$

$$\underbrace{Z_j^0 = \left.\frac{\partial f}{\partial \beta_j}\right|_{\beta^0}}_{\substack{\text{new "X" variables =} \\ \text{gradients evaluated @ } \beta^0}}$$

So, near β^0 we can write

$$y \simeq W^0 + Z^0\beta + \varepsilon$$

fixed intercept

$$\Rightarrow \quad (y - W^0) \simeq Z^0\beta + \varepsilon$$ This is a linear model

$\Rightarrow$ Least squares solution

$$\boxed{\hat{\beta} = (Z^{0\prime} Z^0)^{-1} Z^{0\prime}(y - W^0)}$$

Now we have a new point $\hat{\beta}$ to linearize around.
Set $\beta^0 = \hat{\beta}$ and redo the approximation & linear fit above.

Repeat until convergence

Note ① We are searching locally (near a β^0)
so we may not converge to the true optimum $\Rightarrow$ try many different starting values.

② Or use a stochastic search to avoid converging to a local solution

Inference

Some things are just like for the linear case

→ Since $\varepsilon \sim N(0, \sigma^2)$ is the error assumption

→ $RSS \sim \chi^2$ ⇒ F-tests can be used for testing & model selection

⇒ Can also use AIC & BIC since they are based on $\varepsilon \sim N(0, \sigma^2)$ assumption too.

BUT Testing & CI for parameters β_j not so easy.

$\hat{\beta}_j$ is no longer linear in y, so no longer $\sim N$

⇒ $\dfrac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)}$ no longer t-distributed.

+ What is $SE(\hat{\beta}_j)$?? In linear case $SE(\hat{\beta}_j) = \hat{\sigma}\sqrt{(X'X)^{-1}_{jj}}$ but here ?

⇒ Base inference on final linear approximation.

Let $F_{ij} = \dfrac{\partial f(x_i ; \hat{\beta}^{final})}{\partial \beta_j}$, $F = \{F_{ij}\}$ $n \times p$ matrix

The gradients are the x-variables in the last linear approximation — so F is the the design matrix X in a linear model

$\Rightarrow V(\hat{\beta}^{final}) \approx \hat{\sigma}^2 (F'F)^{-1}$

$\Rightarrow CI(\beta_j) = [\hat{\beta}_j \pm t_{n-p}(1-\alpha/2) \hat{\sigma} \sqrt{(F'F)^{-1}_{jj}}\,]$ (*)

Problems? ◦ Well, linear approx @ $\hat{\beta}^{final}$ may be a poor approximation in which case the CI above can be misleading

If the f is very nonlinear @ $\hat{\beta}^{final}$ perhaps the CI should be non-symmetric etc.

◦ Also $(F'F)$ can have large off-diagonal components if the gradients wrt β_j and $\beta_{j'}$ are correlated $\Rightarrow$ instability & correlated $\hat{\beta}$ s.

What to do?

(1) Check if linear approximation is adequate.

Profile function

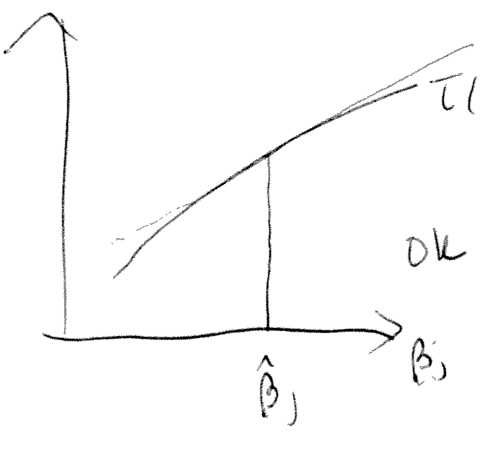
$$\tau(\beta_j) = \operatorname{sign}(\beta_j - \hat{\beta}_j)\sqrt{F(\beta_j)}$$

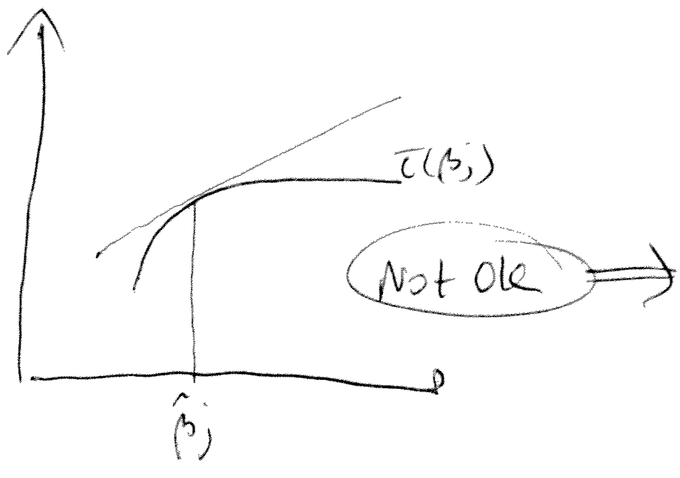
where $F(\beta_j) = \dfrac{RSS(\beta_j) - RSS(\hat{\beta}_j)}{\hat{\sigma}^2}$

That is, compute the RSS near the final solution $\hat{\beta}_j$ and see how it deviates.

* You vary just one component β_j, but hold all other β's fixed at the final solution

If $\sqrt{F(\beta_j)}$ is ~linear in β_j, then $RSS(\beta_j)$ is near quadratic in β_j and the linear approximation is quite good. Then t-based CI as in (*) on page 5 is OK.





If $\tau(\beta_j)$ not linear near $\hat{\beta}_j$ $\Rightarrow$ cannot trust the linear approximation based CIs

$\Longrightarrow$ Bootstrap!