Recall: Nonlinear (Regression) Models

Example: \[ B(t, \theta_1, \theta_2) = \frac{\theta_1}{\theta_1 - \theta_2} \left[ \exp(\theta_1 t) - \exp(\theta_2) \right] \]

We might observe \( y_i = B(t_i, \theta_1, \theta_2) + \varepsilon_i \) at various \( t_i \) and want to estimate \( \theta_1, \theta_2 \).

We'll consider the model:

\[ y_i = \underbrace{f(x_i, \beta)}_{\text{dimension is constant but irrelevant}} + \varepsilon_i \]

Least squares (maximum likelihood when the \( \varepsilon_i \) are constant variance independent normal variables) is minimization of:

\[ g(\beta) = \sum_{i=1}^{n} (y_i - f(x_i, \beta))^2 \]

Here in the general nonlinear model there is no closed-form solution for \( \beta_{OLS} \), a parameter vector minimizing \( g(\beta) \), and there is no exact distribution theory for \( \beta_{OLS} \). This supports inference. So we're driven to numerical analysis to compute \( \beta_{OLS} \) and large sample theory to give approximate inference from it.
A few details about finding \( \mathbf{b} \) s: The typical method is to set all partials of \( g(\mathbf{b}) \) equal to 0 and try to solve (iteratively) — i.e. try to solve

\[
\frac{\partial g}{\partial \beta_1} \bigg|_{\beta = \mathbf{b}_{0 \times s}} = 0 \\
\vdots \\
\frac{\partial g}{\partial \beta_s} \bigg|_{\beta = \mathbf{b}_{0 \times s}} = 0
\]

simultaneously.

Some calculus and algebra can be used to show that these can be written

\[
\mathbf{0} = \mathbf{D}^\top \left( \mathbf{Y} - \mathbf{f}(\mathbf{X}, \mathbf{b}) \right)
\]

where

\[
\mathbf{D} = \left( \frac{\partial f(x_i, \beta)}{\partial \beta_j} \right)_{i=1}^{n \times 1} \\
\mathbf{Y} = \left( y_1 \right)_{1 \times 1} \\
\mathbf{f} = \left( f(x_1, \beta) \right)_{1 \times 1}
\]

What is this when \( f(x_i, \beta) = \mathbf{x}_i^\top \beta \)?
Under this set of circumstances
\[ D = \left( \begin{array}{ccc}
\frac{\partial \beta_1}{\partial \beta_1} & \cdots & \frac{\partial \beta_1}{\partial \beta_k} \\
\vdots & \ddots & \vdots \\
\frac{\partial \beta_k}{\partial \beta_1} & \cdots & \frac{\partial \beta_k}{\partial \beta_k}
\end{array} \right) = (\alpha_{ij}) = X
\]

So the set of estimating equations is
\[ \hat{\theta} = X' (Y - X \hat{\theta}) 
\]
i.e.
\[ X'Y = X'X \hat{\theta} \]

The meta-qirlly of exactly what iterative method I use to solve the estimating equations is of no interest to me—there are many possibilities—see Kocher —Gauss-Newton Method —Fisher Scoring Method—

The general idea is to iterate until values converge—or the error sum of squares (deviance) converges—\( \text{R \ int(\ )} \) will do this for us

How do I do inference? Two versions of using large sample theory:
1) use of approximate theory for ML
2) use theory for the shape of the likelihood surface
(related to LR testing)

Methods Based on Large n Approximate Dns for MLE's

For large n standard theory (Stat 543 stuff) suggests

1) \[ b \overset{\text{ML}}{\sim} \text{MVN}_k \left( \beta, \sigma^2 (D' D)^{-1} \right) \]

\[ \text{when } D = \left( \frac{\partial^2 (y; b)}{\partial b} \bigg|_{b = \beta} \right) \]

\[ \text{typically gets small as } n \text{ gets big} \]

2) \[ \text{MSE} = \frac{\text{SSE}}{n - k} \approx \sigma^2 \]

or \( n ? \)

3) \[ (D' D)^{-1} \approx (\hat{D}' \hat{D})^{-1} \]

\[ \hat{D} = \left( \frac{\partial^2 (y; b)}{\partial b} \right) \]

\[ \text{for } \hat{D} \text{ (smooth/differentiable) function} \]

\[ \hat{h} \left( \mathbb{R}^k \rightarrow \mathbb{R}^l \right) \]
\[ \hat{h}(b_{\text{est}}) \sim \text{MVN}_{q}(\hat{h}(\beta), \sigma^2 G (D' D)^{-1} G') \]

where

\[ G = \left( \begin{array}{c} \frac{\partial h_i(b)}{\partial b_j} \\ \vdots \\ \frac{\partial h_i(b)}{\partial b_j} \end{array} \right)_{j \times q} \]

This is based on 1) and the "delta method" method.

Taylor Theorem (see Appendix 7.2 of class outline)