

$$\hat{f}(x_0) = \sum_{i=1}^N l_i(x_0) Y_i, \quad Y_i = f(x_i) + \varepsilon_i, \quad E(\varepsilon_i) = 0, \quad \text{Var}(\varepsilon_i) = \sigma^2$$

$$\rightarrow E(Y_i) = f(x_i), \quad \text{Var}(Y_i) = \text{Var}(\varepsilon_i) = \sigma^2$$

We have:

$$E(\hat{f}(x_0)) = E\left(\sum_{i=1}^N l_i(x_0) Y_i\right) = \sum_{i=1}^N l_i(x_0) \overbrace{E(Y_i)}^{=f(x_i)}$$

$$= \sum_{i=1}^N l_i(x_0) f(x_i)$$

The Taylor expansion of  $f(x_i)$  around  $x_0$  gives:

$$f(x_i) \approx f(x_0) + f'(x_0)(x_i - x_0) + \frac{1}{2} f''(x_0)(x_i - x_0)^2$$

We get

$$E(\hat{f}(x_0)) \approx \sum_{i=1}^N l_i(x_0) \left( f(x_0) + f'(x_0)(x_i - x_0) + \frac{1}{2} f''(x_0)(x_i - x_0)^2 \right)$$

$$= f(x_0) \sum_{i=1}^N l_i(x_0) + f'(x_0) \sum_{i=1}^N l_i(x_0)(x_i - x_0) + \frac{1}{2} f''(x_0) \sum_{i=1}^N l_i(x_0)(x_i - x_0)^2$$

It may be shown that

$$\sum_{i=1}^N l_i(x_0) = 1 \quad \text{and} \quad \sum_{i=1}^N l_i(x_0)(x_i - x_0) = 0$$

So that

$$E(\hat{f}(x_0)) \approx f(x_0) + \frac{1}{2} f''(x_0) \sum_{i=1}^N l_i(x_0)(x_i - x_0)^2$$

Since  $l_i(x_0)$  tends to be large when  $x_i$  is close to  $x_0$ , the bias of  $\hat{f}(x_0)$  is typically small.

$$\hat{\mu} = \bar{y} = \frac{e^{\hat{\alpha}}}{1 + e^{\hat{\alpha}}} \longrightarrow \hat{\alpha} = \log \frac{\hat{\mu}}{1 - \hat{\mu}} = \log \frac{\bar{y}}{1 - \bar{y}}$$

Iterated least squares for ordinary logistic regression:

$$\beta^{(t+1)} = (X^T W^{(t)} X)^{-1} X^T W^{(t)} z^{(t)},$$

with the adjusted response

$$z^{(t)} = X \beta^{(t)} + (W^{(t)})^{-1} (y - p^{(t)})$$

and

$W^{(t)}$  is the diagonal matrix with diagonal entries  $p^{(t)}(x_i)(1 - p^{(t)}(x_i))$

$$z_i^{(t)} = \underbrace{x_i^T \beta}_{= \eta_i^{(t)}} + \frac{y_i - p^{(t)}(x_i)}{p^{(t)}(x_i)(1 - p^{(t)}(x_i))} = \eta_i^{(t)} + \frac{y_i - p^{(t)}(x_i)}{p^{(t)}(x_i)(1 - p^{(t)}(x_i))}$$