Likelihood inference for the exponential family

$\theta_i = \eta_i$ for canonical link

$$
\begin{tikzcd}
\theta_i \arrow[r, leftrightarrow, "E(Y_i) = b'(\theta_i)"] \arrow[rr, bend left, leftrightarrow] & \mu_i \arrow[r, leftrightarrow, "\substack{\eta_i = g(\mu_i) \\ \mu_i = h(\eta_i)}"] & \eta_i \arrow[r, leftrightarrow, "\eta_i = x_i^T\beta"] & \beta
\end{tikzcd}
$$

canonical parameter

mean of Y_i

TAVLA!

$$f(y_i) = \exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{\phi} w_i + c(y_i, \phi, w_i) \right\}$$

$$l_i(\beta) = \frac{y_i\theta_i - b(\theta_i)}{\phi} w_i + c(y_i, \phi, w_i)$$

not a function of θ_i

$$l(\beta) = \frac{1}{\phi} \sum_{i=1}^{n} (y_i\theta_i - b(\theta_i)) w_i + \sum_{i=1}^{n} c(y_i, \phi, w_i)$$

Canonical link: $\theta_i = x_i^T\beta$.

When ϕ and w_i are fixed constants, part of l involving both data and model parameters

$$\sum_i y_i x_i^T\beta = \sum y_i \sum \beta_j x_{ij} = \sum_j \beta_j \sum_i y_i x_{ij}$$

$$\Vert$$

$$
\begin{array}{c c}
y_1 & \\
\vdots & \\
y_n &
\end{array}
\begin{bmatrix}
x_{11} & \cdots & x_{1p} \\
\vdots & & \\
x_{n1} & & x_{np}
\end{bmatrix}
$$

Likelihood equations for the GLM

$$\frac{\partial l}{\partial \beta} = 0$$ ← set of p (nonlinear) equations to be solved

$\parallel$

$S(\beta)$.

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \underbrace{\frac{\partial l_i}{\partial \beta}}_{s_i(\beta)} = \sum_{i=1}^{n} s_i(\beta)$$

chain rule:

$$\frac{\partial l_i}{\partial \beta} = \frac{\partial l_i}{\partial \theta_i} \cdot \overbrace{\frac{\partial \theta_i}{\partial \mu_i}}^{(\mu_i = b'(\theta_i))} \cdot \overbrace{\frac{\partial \mu_i}{\partial \eta_i}}^{(\eta_i = g(\mu_i))} \cdot \frac{\partial \eta_i}{\partial \beta}$$

$$\frac{\partial x_i^T \beta}{\partial \beta} = x_i$$

$$l_i = \frac{y_i \theta_i - b(\theta_i)}{\phi} \cdot w_i + c(y_i, w_i, \phi)$$

$$(y_i - \overset{\mu_i}{\overset{\parallel}{b'(\theta_i)}}) \frac{w_i}{\phi}$$

depends on link function

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = Var(Y_i) \cdot \frac{w_i}{\phi} \quad - \quad \frac{\partial \theta_i}{\partial \mu_i} = \frac{\phi}{w_i} \cdot \frac{1}{Var(Y_i)}$$

$$b'(\theta_i) = \mu_i$$

$$\frac{\partial l_i}{\partial \beta} = \underbrace{(y_i - \mu_i) \cdot \frac{w_i}{\phi}}_{} \cdot \underbrace{\frac{\phi}{w_i} \, \frac{1}{Var(Y_i)}}_{} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_i$$

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \frac{(y_i - \mu_i) \cdot x_i}{Var(Y_i)} \cdot \frac{\partial \mu_i}{\partial \eta_i}$$

↖ depends on $g(\mu_i) = \eta_i$

$$S(\beta) = \sum_{i=1}^{n} \frac{(y_i - \mu_i) \cdot x_i \, h'(\eta_i)}{Var(Y_i)}$$

Agresti: $V = \text{diag}(\text{Var}(Y_i))$
$D = \text{diag}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)$

$$\left.\begin{matrix} V = \text{diag}(\text{Var}(Y_i)) \\ D = \text{diag}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\end{matrix}\right\} \quad \mathbb{X}^T D V^{-1}(y-\mu) = 0$$

where is β ?

$\mu_i = g^{-1}(\eta_i)$

$x_i^T\beta$

Fahrmeir et al:

$\Sigma = \text{diag}(\text{Var}(Y_i))$

$D = \text{diag}(h'(\eta_i))$

$$S(\beta) = \mathbb{X}^T D \Sigma^{-1}(y-\mu) \qquad \text{[Box 5.6 p 305]}$$

Observe: $s(\beta) = 0$ depends on the distribution of Y_i only through μ_i and $\text{Var}(Y_i)$

And - often? - $\text{Var}(Y_i)$ depends on μ_i

$$\begin{aligned} \text{Var}(Y_i) &= \mu_i \quad \text{for Poisson} \\ &= \mu_i(1-\mu_i)\cdot\tfrac{1}{n_i} \quad \text{for } \underline{\text{bin}} \text{ proportion} \\ &= \sigma^2 \quad \text{(constant) for Normal} \\ &= \frac{\mu_i^2}{\nu} \quad \text{for Gamma} \end{aligned}$$

For exp. fam the relationship between the mean and variance characterize the distribution.
(Agresti s 125)

If : $\text{Var}(Y_i) = \mu_i \Rightarrow$ must be Poisson.

The expected Fisher information matrix

Remember $\hat{\beta}_{ML} \approx N(\beta, F^{-1}(\beta))$

also $\approx N(\beta, F^{-1}(\hat{\beta}))$

so, we "need a formula" for $F(\beta)$ for the exp. fam GLM

For this the "prefered" solution is to look at

$$F(\beta) = E\left(-\frac{\partial^2 l}{\partial\beta\partial\beta^T}\right) = \sum_{i=1}^{n} E\left(-\frac{\partial^2 l_i}{\partial\beta\partial\beta^T}\right) \quad \left[\begin{array}{l} E(\text{sum}) \\ = \text{sum}(E) \\ \text{always.} \end{array}\right]$$

and for the exp. fam. the following result holds

$$E\left(-\frac{\partial^2 l_i}{\partial\beta_h\partial\beta_l}\right) = E\left(\frac{\partial l_i}{\partial\beta_h}\cdot\frac{\partial l_i}{\partial\beta_l}\right)$$

because $= Cov(s(\beta)) = E(s(\beta)\, s(\beta)^T)$

We have already

calculated $\displaystyle \frac{\partial l_i}{\partial\beta_h} = \frac{(y_i-\mu_i)\cdot x_{ih}}{Var(Y_i)} \cdot \underbrace{\frac{\partial\mu_i}{\partial\eta_i}}_{h'(\eta_i)}$

so element (h,l) of $F_i(\beta)$ $\qquad F(\beta) = \sum_{i=1}^{n} F_i(\beta)$

$$E\left(\frac{\partial l_i}{\partial\beta_h}\cdot\frac{\partial l_i}{\partial\beta_l}\right) = E\left[\frac{(Y_i-\mu_i)\,x_{ih}}{Var(Y_i)}\cdot h'(\eta_i)\,\frac{(Y_i-\mu_i)\,x_{il}}{Var(Y_i)}\,h'(\eta_i)\right]$$

$$= E\left[(Y_i-\mu_i)^2\, x_{ih}\, x_{il}\, [h'(\eta_i)]^2 \cdot \frac{1}{Var(Y_i)^2}\right]$$

$$= x_{ih} \cdot x_{il} \, [h'(\eta_i)]^2 \cdot \frac{\overbrace{E\left((Y_i-\mu_i)^2\right)}^{Var(Y_i)}}{Var(Y_i)^2}$$

$$= \frac{x_{ih} \cdot x_{il} \, [h'(\eta_i)]^2}{Var(Y_i)}$$

$$\underset{[h,l]}{F(\beta)} = \sum_{i=1}^{n} \underset{[h,l]}{F_i(\beta)} = \sum_{i=1}^{n} \frac{x_{ih} \, x_{il} \, [h'(\eta_i)]^2}{Var(Y_i)}$$

Let $W = \text{diag}\left(\frac{h'(\eta_i)^2}{Var(Y_i)} \right)$

then $F(\beta) = \mathbb{X}^T W \mathbb{X}$

and when $\hat{\beta} \to \hat{\mu} \to \dots$ then $\underline{\hat{W}}$ evaluated at $\hat{\beta}$

$$F(\hat{\beta}) = \mathbb{X}^T \hat{W} \mathbb{X}$$

and $\widehat{Cov}(\hat{\beta}) = (\mathbb{X}^T \hat{W} \mathbb{X})^{-1}$

When ϕ needs to be estimated $\rightarrow$ how about $Cov(\hat{\beta})$ then?

For some GLMs (normal, gamma) our parameter vector is partitioned into $(\beta, \phi) = \beta^*$

We should then solve $\frac{\partial l}{\partial \beta^*} = 0$ to find $(\hat{\beta}, \hat{\phi})$

and then
$$\begin{bmatrix} \hat{\beta} \\ \hat{\phi} \end{bmatrix} \approx N\left(\begin{bmatrix} \beta \\ \phi \end{bmatrix}, F^{-1}(\hat{\beta}^*) \right)$$

where here
$$F(\beta^*) = \begin{bmatrix} F_{\beta\beta} & F_{\beta\phi} \\ F_{\phi\beta} & F_{\phi\phi} \end{bmatrix}$$

$F_{\beta\beta}$ has elements from $E\left(\frac{\partial l}{\partial \beta_h} \frac{\partial l}{\partial \beta_r}\right)$ ← we have so far only looked at this

$F_{\beta\phi}$ $\quad E\left(\frac{\partial l}{\partial \beta_h} \cdot \frac{\partial l}{\partial \phi}\right)$

$F_{\phi\beta}$ $\quad E\left(\frac{\partial l}{\partial \phi} \frac{\partial l}{\partial \beta_h}\right)$

$F_{\phi\phi}$ $\quad E\left(\frac{\partial l}{\partial \phi} \frac{\partial l}{\partial \phi}\right)$

What is $F_{\beta\phi}$ $\qquad (F_{\beta\phi} = F_{\phi\beta}^T)$

$$\frac{\partial l}{\partial \beta_h} = \sum_{i=1}^{n} \frac{(y_i-\mu_i)\, x_{ih}}{\underbrace{Var(Y_i)}_{b''(\theta_i)\cdot\frac{\phi}{w_i}}} \frac{\partial \mu_i}{\partial \eta_i} = \sum_{i=1}^{n} \frac{(y_i-\mu_i) x_{ih}}{b''(\theta_i)\cdot \frac{\phi}{w_i}} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)$$

$[f_{\beta\phi}]_h$

$$\frac{\partial^2 l}{\partial\phi\,\partial\beta_h} = \sum_{i=1}^{n} w_i \cdot \frac{(y_i-\mu_i)\, x_{ih}}{b'(\theta_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) \cdot \underbrace{\frac{\partial}{\partial \phi}\left(\phi^{-1}\right)}_{-\frac{1}{\phi^2}}$$

$$E(\;\cdot\;) = \sum_{i=1}^{n} w_i \frac{x_{ih}}{b'(\theta_i)} \frac{\partial \mu_i}{\partial \eta_i}\left(-\frac{1}{\phi^2}\right) \cdot \underbrace{E(Y_i - \mu_i)}_{0} = 0$$

Therefore

$$f = \begin{bmatrix} f_{\beta\beta} & 0 \\ 0 & f_{\phi\phi} \end{bmatrix} \text{ and}$$

$$f^{-1} = \begin{bmatrix} f_{\beta\beta}^{-1} & 0 \\ 0 & f_{\phi\phi}^{-1} \end{bmatrix}$$

and

$$\begin{bmatrix} \hat\beta \\ \hat\phi \end{bmatrix} \sim N\left[\begin{pmatrix} \beta \\ \phi \end{pmatrix}, \begin{bmatrix} \overset{f(\beta)^{-1}}{\overset{\parallel}{f_{\beta\beta}^{-1}}} & 0 \\ 0 & f_{\phi\phi}^{-1} \end{bmatrix}\right]$$

so $\hat\beta \approx N(\beta, f(\beta)^{-1})$ is the same whether ϕ is fixed or estimated

We say that the parameters β and ϕ are orthogonal and for our exp. fam GLM this is always the case.

Remark : observed fisher info not so?

Estimation of ϕ

$$Var(Y_i) = \phi \frac{b''(\theta_i)}{w_i} \quad \text{and} \quad V(\mu_i) = b''(\theta_i)$$

variance function

since $\mu_i = b'(\theta_i)$ then $b''(\theta_i)$ depends on μ_i

$$\hat{\phi} = \frac{1}{G-p} \sum_{i=1}^{G} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)/n_i}$$

with data grouped "as much as possible" $\rightarrow G, n_i$