Statistical Theory for Generalized Linear Models

The generalized linear model extends beyond the normal distribution to the exponential family of distributions,

$$p(y;\theta,\phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right],$$

where $p(y;\theta,\phi)$ is the pmf or pdf of Y, $a(.), b(.)$, and $c(.)$ are known functions, $\theta = g_c(\mu)$ is the canonical parameter, and ϕ is the dispersion parameter. Table 15.9 in the text lists the values of $a(.), b(.)$, and $c(.)$ for several well-known distributions. The log-likelihood function for a sample is then

$$\log L(\boldsymbol{\theta},\phi;\mathbf{y}) = \sum_{i=1}^{n}\left[\frac{Y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(Y_i,\phi)\right].$$

Using the results that

$$E\left(\frac{\partial \log L(\theta,\phi;y)}{\partial\theta}\right) = 0 \text{ and}$$
$$E\left(\frac{\partial^2 \log L(\theta,\phi;y)}{\partial\theta^2}\right) + E\left(\frac{\partial \log L(\theta,\phi;y)}{\partial\theta}\right)^2 = 0$$

we can show that

$$E(Y) = \mu = b'(\theta) \text{ and}$$
$$V(Y) = a(\phi)b''(\theta) = a(\phi)v(\mu).$$

Maximum Likelihood Estimation

For a model with

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik},$$

we find the MLE's by taking the derivative of each term in the sum for $l = \log L(\boldsymbol{\theta},\phi;\mathbf{y})$ with respect to a parameter β_j using the chain rule,

$$\frac{\partial l_i}{\partial\beta_j} = \frac{\partial l_i}{\partial\theta_i} \times \frac{d\theta_i}{d\mu_i} \times \frac{d\mu_i}{d\eta_i} \times \frac{\partial\eta_i}{\partial\beta_j} \text{ for each } j.$$

We have

$$\begin{aligned}
\frac{\partial l_i}{\partial \theta_i} &= \frac{y_i - b'(\theta_i)}{a_i(\phi)} = \frac{y_i - \mu_i}{a_i(\phi)}, \\
\frac{d\theta_i}{d\mu_i} &= \frac{1}{\partial \mu_i / \partial \theta_i} = \frac{1}{\partial b'(\theta_i)/\partial \theta_i} = \frac{1}{b''(\theta_i)} = \frac{1}{v(\mu_i)}, \\
\frac{\partial \eta_i}{\partial \beta_j} &= x_{ij},
\end{aligned}$$

leading to the estimating equations

$$\sum_{i=1}^{n} \frac{Y_i - \mu_i}{a_i v(\mu_i)} \times \frac{d\mu_i}{d\eta_i} \times x_{ij} = 0.$$

These equations can be simplified if $g(.)$ is the canonical link, but in most circumstances iterative methods are required to solve for the MLE's.

IWLS for solving for the MLE's

Let

$$\begin{aligned}
Z_i &\equiv \eta_i + (Y_i - \mu_i)\frac{d\eta_i}{d\mu_i} \\
&= \eta_i + (Y_i - \mu_i) g'(\mu_i).
\end{aligned}$$

Then

$$\begin{aligned}
&E(Z_i) = \eta_i = \beta_0 + \beta_1 X_{i1} \ + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}, \text{ and} \\
&V(Z_i) = \left[g'(\mu_i)\right]^2 a_i v(\mu_i).
\end{aligned}$$

These equations are the motivation for the iteratively (re)weighted least squares procedure for computing the maximum likelihood estimates:

Step 1: Start with initial estimates of $\widehat{\mu_i}$ and $\widehat{\eta_i}$, call them $\widehat{\mu_i}^{(0)}$ and $\widehat{\eta_i}^{(0)}$.

Step 2: For iteration l, compute the working response variable $Z_i^{(l-1)}$ as follows,

$$Z_i^{(l-1)} = \eta_i^{(l-1)} + (Y_i - \mu_i^{(l-1)}) g'(\mu_i^{(l-1)})$$

with weights

$$W_i^{(l-1)} = \frac{1}{\left[g'(\mu_i^{(l-1)})\right]^2 a_i v(\mu_i^{(l-1)})}.$$

Step 3: Calculate the weighted least squares estimate of $\mathbf{b}^{(l)}$,

$$\mathbf{b}^{(l)} = \left(\mathbf{X}'\mathbf{W}^{(l-1)}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{W}^{(l-1)}\mathbf{z}^{(l-1)}.$$

Step 4: Repeat steps 2 and 3 until convergence.

For canonical links, IWLS implements a Newton-Raphson algorithm, but more generally is equivalent to Fisher's method of scoring.

The dispersion parameter ϕ is typically estimated via

$$\widetilde{\phi} = \frac{1}{n-k-1} \sum_{i=1}^{n} \frac{(Y_i - \widehat{\mu}_i)^2}{a_i v(\widehat{\mu}_i)},$$

and the estimated asymptotic covariance matrix of $\mathbf{b}$ is,

$$\widehat{V}(\mathbf{b}) = \widetilde{\phi} \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}.$$

We can use the estimation approach developed above by just specifying first and second moment relationships for Y without specifying a conditional distribution. Viewed this way, the estimates are called quasi-likelihood estimates.