Probit Regression using the EM Algorithm

STA 250 Fall 2013, Prof. Baines (11/10/13)

We want to fit the following model:

$$Y_i|\beta \sim \text{Bin}\left(1, g^{-1}(x_i^T\beta)\right), \qquad i = 1, \ldots, n, \tag{1}$$

where $g^{-1}(u) = \Phi(u)$. Note that $g^{-1}(u) = \exp(u)/(1+\exp(u))$ corresponds to logistic regression. The more general setting with Binomial counts (rather than Bernoulli) can be derived similarly. We first note that a complete data model can be formed as:

$$Y_i|Z_i, \beta, \sim \mathcal{I}_{\{Z_i>0\}} \qquad Z_i|\beta \sim N(x_i^T\beta, 1), \tag{2}$$

where $\mathcal{I}_A$ is the indicator function for the event A i.e., $\mathcal{I}_A = 1$ if the event A occurs, $= 0$ o/w. Note that this model preserves the observed data likelihood i.e.,

$$\mathbb{P}\left(Y_i = 1|\beta\right) = \int \mathbb{P}(Y_i = 1, Z_i|\beta)dZ_i = \mathbb{P}(Z_i > 0|\beta) = \mathbb{P}(Z_i - x_i^T\beta > -x_i^T\beta) = 1 - \Phi(-x_i^T\beta) = \Phi(x_i^T\beta).$$

Probit EM

Note that here we have the following correspondence:

- $Y_{obs}: \left\{y = (y_1, \ldots, y_n)^T\right\}$
- $Y_{mis}: \left\{Z = (Z_1, \ldots, Z_n)^T\right\}$
- θ: β

The Q-function is therefore:

$$Q(\beta|\beta^{(t)}) = \mathbb{E}\left[-\frac{1}{2}\sum_{i=1}^{n}\left(z_i - x_i^T\beta\right)^2 |y, \beta^{(t)}\right] + \text{ const }, \tag{3}$$

$$= -\frac{1}{2}\mathbb{E}\left[\left(Z - X\beta\right)^T\left(Z - X\beta\right)|y, \beta^{(t)}\right] + \text{ const }, \tag{4}$$

where we drop all terms not involving θ (since they do not impact the maximization). To compute this, we need to first derive the conditional distribution required for the expectation: $Z_i|y_i, \beta^{(t)}$. Here it is striaghtforward to see that:

$$Z_i|y_i = 0, \beta^{(t)} \sim TN\left(x_i^T\beta^{(t)}, 1; (\infty, 0]\right) \tag{5}$$

$$Z_i|y_i = 1, \beta^{(t)} \sim TN\left(x_i^T\beta^{(t)}, 1; [0, \infty)\right). \tag{6}$$

To maximize Q in (3) we note that:

$$\frac{dQ}{d\beta} = \mathbb{E}\left[Z|y, \beta^{(t)}\right]^T X - X^T X\beta.$$

Denoting:

$$Z^{(t+1)} = \mathbb{E}\left[Z|y, \beta^{(t)}\right],$$

we see that:

$$\beta^{(t+1)} = (X^T X)^{-1} X^T Z^{(t+1)}, \tag{7}$$

i.e., the least squares estimate when regression $Z^{(t+1)}$ on X. All that remains is to actually compute $\mathbb{E}\left[Z|y, \beta^{(t)}\right]$ where $Z_i|y, \beta^{(t)}$ has the distribution given in (5) and (6). For a truncated normal distribution we can derive the expected value using simple properties of the moment generating function (left as an exercise). Let $U \sim TN(\mu, \sigma^2; (-\infty, b))$, $V \sim TN(\mu, \sigma^2; (a, \infty))$ and $W \sim TN(\mu, \sigma^2; (a, b))$ then:

$$\begin{aligned}
\mathbb{E}\left[U\right] &= \mu - \sigma \frac{\phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma})}, \\
\mathbb{E}\left[V\right] &= \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{a-\mu}{\sigma})}, \\
\mathbb{E}\left[W\right] &= \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}.
\end{aligned}$$

For the probit regression setting we then obtain that:

$$Z_i^{(t+1)} = \begin{cases} x_i^T \beta^{(t)} - \frac{\phi(x_i^T \beta^{(t)})}{\Phi(-x_i^T \beta^{(t)})} & \text{if } y_i = 0 \\ x_i^T \beta^{(t)} + \frac{\phi(x_i^T \beta^{(t)})}{1-\Phi\left(-x_i^T \beta^{(t)}\right)} & \text{if } y_i = 1 \end{cases}. \tag{8}$$

Therefore, using a relative error stopping rule with tolerance $\epsilon > 0$, the EM algorithm can be summarized as follows:

1. Select starting value $\beta(0)$ and set $t = 0$.

2. **E-Step:** Compute $Z^{(t+1)}$ using (8).

3. **M-Step:** Compute $\beta^{(t+1)}$ using (7).

4. If $\frac{\|\beta^{(t+1)} - \beta^{(t)}\|}{\|\beta^{(t)}\|} < \epsilon$ then declare $\hat{\beta} = \beta^{(t+1)}$, else increment $t \mapsto t + 1$ and return to step 2.

Obviously other stopping rules can be used in place of the relative error criteria shown above. The EM algorithm above is implemented in the course GitHub repo.