Lecture 23

Recap: fitting ZIP models

Y_i count variable w/ too many zeros

latent variable (unobserved) Z_i

$$Z_i \sim \text{Bernoulli}(p_i)$$

$$Y_i \mid (Z_i = 1) \equiv 0 \quad \text{(point mass at 0)}$$

$$Y_i \mid (Z_i = 0) \sim \text{Poisson}(\lambda_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \gamma^T X_i$$

$$\log(\lambda_i) = \beta^T X_i$$

want to estimate γ and β (but we don't get Z_is)

If we observed Z_i, then the complete-data log likelihood is

$$\ell(\gamma, \beta) = \underbrace{\sum_{i=1}^{n} \left(Z_i \log(p_i) + (1-Z_i) \log(1-p_i) \right)}_{\ell(\gamma)}$$

$$+ \underbrace{\sum_{i=1}^{n} (1-Z_i)\left[-\lambda_i + Y_i \log \lambda_i\right] - \sum_{i=1}^{n} (1-Z_i) \log(Y_i!)}_{\ell(\beta)}$$

If we know Z_i, we can separately maximize $\ell(\gamma)$ and $\ell(\beta)$

(But, we don't know Z_i)

Suppose we know γ and β. We want a guess for what Z_i would be.

We will calculate

$$P(Z_i=1 \mid Y_i, \gamma, \beta)$$

If $Y_i > 0$: $\quad P(Z_i=1 \mid Y_i, \gamma, \beta) = 0$

If $Y_i = 0$: $\quad P(Z_i=1 \mid Y_i=0, \gamma, \beta) = \dfrac{P(Y_i=0 \mid Z_i=1, \gamma, \beta)\, P(Z_i=1 \mid \gamma, \beta)}{\left(P(Y_i=0 \mid Z_i=1, \gamma, \beta)\, P(Z_i=1 \mid \gamma, \beta) + P(Y_i=0 \mid Z_i=0, \gamma, \beta)\, P(Z_i=0 \mid \gamma, \beta)\right)}$

$$P(Y_i=0 \mid Z_i=1, \gamma, \beta) = 1$$

$$P(Y_i=0 \mid Z_i=0, \gamma, \beta) = e^{-\lambda_i}$$

$$P(Z_i=1 \mid \gamma, \beta) = p_i \qquad \Rightarrow P(Z_i=0 \mid \gamma, \beta) = 1-p_i$$

$$\Rightarrow P(Z_i=1 \mid Y_i, \gamma, \beta) = \begin{cases} 0 & Y_i > 0 \\ \dfrac{p_i}{p_i + e^{-\lambda_i}(1-p_i)} & Y_i = 0 \end{cases}$$

So: ① If we know Y_i, γ, β, then we can calculate

$$P(Z_i = 1 \mid Y_i, \gamma, \beta) = \mathbb{E}[Z_i \mid Y_i, \gamma, \beta]$$

② If we know Z_i, Y_i we can calculate $\hat{\gamma}, \hat{\beta}$ (maximizing $\ell(\gamma)$ and $\ell(\beta)$)

E-M algorithm:

(expectation) E-step: Given current estimate $\gamma^{(k)}, \beta^{(k)}$

$$Z_i^{(k)} = \mathbb{E}[Z_i \mid Y_i, \gamma^{(k)}, \beta^{(k)}]$$

(maximization) M-step: Given current $Z_i^{(k)}$,

$$\gamma^{(k+1)} = \underset{\gamma}{\arg\max}\ \ell(\gamma; Y, Z^{(k)})$$

$$\beta^{(k+1)} = \underset{\beta}{\arg\max}\ \ell(\beta; Y, Z^{(k)})$$

iterate E & M steps until convergence

EM algorithm

E step

$$Z_i^{(k)} = \mathbb{E}\left[Z_i \mid Y_i, \gamma^{(k)}, \beta^{(k)}\right]$$

$$\hat{p}_i = \frac{\exp\{\gamma^{(k)T} x_i\}}{1 + \exp\{\gamma^{(k)T} x_i\}} \qquad \hat{\lambda}_i = \exp\{\beta^{(k)T} x_i\}$$

$$\Rightarrow \quad Z_i^{(k)} = \begin{cases} 0 & Y_i > 0 \\ \dfrac{\hat{p}_i}{\hat{p}_i + e^{-\hat{\lambda}_i}(1-\hat{p}_i)} & Y_i = 0 \end{cases}$$

M step

M-Step for β :

$$\beta^{(k+1)} = \underset{\beta}{\operatorname{argmax}} \; \ell(\beta; Y, Z^{(k)})$$

$$= \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{n} (1 - Z_i^{(k)}) \left[-\lambda_i + Y_i \log \lambda_i \right]$$

weighted Poisson: maximize $\sum_{i=1}^{n} w_i \left[Y_i \log \lambda_i - \lambda_i \right]$

$\Rightarrow$ M-Step for β : weighted Poisson regression where

$$w_i = 1 - Z_i^{(k)}$$

(we can use glm function in R w/ weights to estimate β)

$U(\beta) = \sum_i w_i (Y_i - \lambda_i) X_i \qquad \mathcal{I}(\beta) = \sum_i w_i \lambda_i X_i X_i^T$

M-Step for γ :

$$\gamma^{(k+1)} = \underset{\gamma}{\operatorname{argmax}}\ \ell(\gamma; Y, Z^{(k)})$$

$$= \underset{\gamma}{\operatorname{argmax}} \sum_{i=1}^{n} \left(Z_i^{(k)} \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i)\right)$$

Logistic regression: $\underset{\gamma}{\operatorname{argmax}} \sum_{i=1}^{n} \left(Y_i^* \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i)\right)$

where $Y_i^* \in \{0,1\}$

but $Z_i^{(k)} \in [0,1]$
(not necessarily binary)

weighted logistic: $\underset{\gamma}{\operatorname{argmax}} \sum_{i=1}^{n} \left(w_i Y_i^* \log\left(\frac{p_i}{1-p_i}\right) + w_i \log(1-p_i)\right)$

Goal: make $\gamma^{(k+1)}$ update look like weighted logistic regression

$$\underbrace{\sum_{i=1}^{n} Z_i^{(k)} \log\left(\frac{p_i}{1-p_i}\right)} + \underbrace{\sum_{i=1}^{n} \log(1-p_i)}$$

$$= \sum_{i=1}^{n} Z_i^{(k)} \log\left(\frac{p_i}{1-p_i}\right) \qquad = \sum_{i=1}^{n} Z_i^{(k)} \log(1-p_i) + \sum_{i=1}^{n} (1-Z_i^{(k)}) \log(1-p_i)$$

$$+ \sum_{i=1}^{n} 0 \cdot (1-Z_i^{(k)}) \log\left(\frac{p_i}{1-p_i}\right)$$

$$\boxed{\begin{aligned} &\text{Also:} \quad Z_i^{(k)} = 0 \quad \text{if} \quad Y_i > 0 \\ &\Rightarrow \quad Z_i^{(k)} = \mathbb{1}\{Y_i = 0\} Z_i^{(k)} \end{aligned}}$$

$$\Rightarrow \sum_{i=1}^{n} \mathbb{1}\{Y_i = 0\} Z_i^{(k)} \log\left(\frac{p_i}{1-p_i}\right) + \sum_{i=1}^{n} 0 \cdot (1-Z_i^{(k)}) \log\left(\frac{p_i}{1-p_i}\right)$$

$$+ \sum_{i=1}^{n} Z_i^{(k)} \log(1-p_i) + \sum_{i=1}^{n} (1-Z_i^{(k)}) \log(1-p_i)$$

$$\sum_{i=1}^{n} \mathbb{1}\{Y_i=0\} Z_i^{(k)} \log\left(\frac{p_i}{1-p_i}\right) + \sum_{i=1}^{n} 0 \cdot (1-Z_i^{(k)}) \log\left(\frac{p_i}{1-p_i}\right)$$

$$+ \sum_{i=1}^{n} Z_i^{(k)} \log(1-p_i) + \sum_{i=1}^{n} (1-Z_i^{(k)}) \log(1-p_i)$$

Let $Y^* = \begin{bmatrix} \mathbb{1}\{Y_1=0\} \\ \mathbb{1}\{Y_2=0\} \\ \vdots \\ \mathbb{1}\{Y_n=0\} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} n \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n \end{matrix}$ $\in \mathbb{R}^{2n}$

$X^* = \begin{bmatrix} X_1^T \\ \vdots \\ X_n^T \\ X_1^T \\ \vdots \\ X_n^T \end{bmatrix} \in \mathbb{R}^{2n \times p}$

weights: $w = \begin{bmatrix} Z_1^{(k)} \\ \vdots \\ Z_n^{(k)} \\ 1-Z_1^{(k)} \\ \vdots \\ 1-Z_n^{(k)} \end{bmatrix} \in \mathbb{R}^{2n}$

$\Rightarrow \gamma^{(k+1)} = \underset{\gamma}{\operatorname{argmax}}$ of

$$\sum_{i=1}^{2n} w_i Y_i^* \log\left(\frac{p_i}{1-p_i}\right) + \sum_{i=1}^{n} w_i \log(1-p_i)$$

Summarize:

E-step: plug in $\gamma^{(k)}, \beta^{(k)}$ to calculate $P(Z_i = 1 \mid Y_i, \gamma^{(k)}, \beta^{(k)})$

M-step:

For β: fit weighted Poisson regression

For γ: fit weighted logistic regression