Zero-inflated Poisson Generalized Linear Model for Count Data

Brian M. Brost

06 April 2016

Description

A zero-inflated generarlized linear model for count data.

Implementation

The file zi.poisson.glm.sim.R simulates data according to the model statement presented below, and zi.poisson.glm.mcmc.R contains the MCMC algorithm for model fitting.

Model statement

Let z_i, for $i = 1, \ldots, n$, be observed count data (i.e., z_i are integers greater than or equal to 0). Also let $\mathbf{x}_i$ be a vector of covariates associated with z_i for which inference is desired, and the vector $\boldsymbol{\beta}$ be the corresponding coefficients.

$$\begin{aligned}
y_i &\sim \begin{cases} \text{Pois}\left(\lambda_i\right), & z_i = 1 \\ 0, & z_i = 0 \end{cases} \\
z_i &\sim \text{Bern}\left(p\right) \\
\log\left(\lambda_i\right) &= \mathbf{x}_i'\boldsymbol{\beta} \\
\boldsymbol{\beta} &\sim \mathcal{N}(\mathbf{0}, \sigma_\beta^2\mathbf{I}) \\
p &\sim \text{Beta}\left(\alpha_1, \alpha_2\right)
\end{aligned}$$

Full conditional distributions

Latent mixture component indicator variable (z_i):

$$\begin{aligned}
[z_i \mid \cdot] &\propto [y_i \mid \lambda_i, z_i]\,[z_i \mid p] \\
&\propto \text{Pois}(y_i \mid \lambda_i)^{z_i} 1_{\{y_i=0\}}^{1-z_i} p^{z_i}\left(1-p\right)^{z_i-1} \\
&\propto \left(\frac{\lambda_i^{y_i}\exp\left(-\lambda_i\right)}{y_i!}\right)^{z_i} p^{z_i}\left(1-p\right)^{z_i-1} \\
&\propto \left(\exp\left(-\lambda_i\right)\right)^{z_i} p^{z_i}\left(1-p\right)^{z_i-1} \\
&\propto \left(p \times \exp\left(-\lambda_i\right)\right)^{z_i}\left(1-p\right)^{z_i-1} \\
&= \text{Bern}\left(\tilde{p}\right),
\end{aligned}$$

where

$$\tilde{p} = \frac{p \times \exp\left(-\lambda_i\right)}{p \times \exp\left(-\lambda_i\right) + 1 - p}.$$

Note that z_i is only updated for instances where $y_i = 0$ ($z_i = 1$ when $y_i > 0$).

Probability associated with the mixture component indicator variables (p):

$$\begin{aligned}
[p \mid \cdot] &\propto \prod_{i=1}^{n} [z_i \mid p] [p] \\
&\propto \prod_{i=1}^{n} p^{z_i} (1-p)^{1-z_i} \, p^{\alpha_1 - 1} (1-p)^{\alpha_2 - 1} \\
&\propto p^{\sum_{i=1}^{n} z_i} (1-p)^{n - \sum_{i=1}^{n} z_i} \, p^{\alpha_1 - 1} (1-p)^{\alpha_2 - 1} \\
&= \text{Beta}\left(\sum_{i=1}^{n} z_i + \alpha_1, n - \sum_{i=1}^{n} z_i + \alpha_2 \right)
\end{aligned}$$

Regression coefficients ($\boldsymbol{\beta}$):

$$\begin{aligned}
[\boldsymbol{\beta} \mid \cdot] &\propto \prod_{i=1}^{n} [y_i \mid \lambda_i, z_i] [\boldsymbol{\beta}] \\
&\propto \prod_{i=1}^{n} \text{Pois}\,(y_i \mid \lambda_i)^{z_i} \, 1_{\{y_i = 0\}}^{1-z_i} \mathcal{N}\left(\boldsymbol{\beta} \mid \mathbf{0}, \sigma_{\beta}^2 \mathbf{I}\right) \\
&\propto \prod_{i=1}^{n} \text{Pois}\,(y_i \mid \lambda_i)^{z_i} \, \mathcal{N}\left(\boldsymbol{\beta} \mid \mathbf{0}, \sigma_{\beta}^2 \mathbf{I}\right)
\end{aligned}$$

The update for $\boldsymbol{\beta}$ proceeds using Metropolis-Hastings.