We introduce the models for zero inflated count data. Define the ZINB model for the response variable y assuming only non-negative values as follows:

$$f_{ZINB}\left(y\right) = \phi I\left(y=0\right) + \left(1-\phi\right) f_D\left(y\right),$$

where ϕ denotes the mixture proportion and $I\left(.\right)$ is the indicator variable defined as

$$I\left(y\right) = \begin{cases} 1, & \text{if } y=0 \\ 0, & \text{if } y \sim f_D(.) \end{cases},$$

and $f_D\left(\cdot\right)$ is the density for Negative Binomial/Poisson distribution given by

$$\begin{aligned} f_{NB}\left(y\right) &= \frac{\Gamma\left(y+1/\alpha\right)}{y!\Gamma\left(1/\alpha\right)} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\lambda}\right)^{\alpha^{-1}} \left(\frac{\lambda}{\alpha^{-1}+\lambda}\right)^{y} \\ f_{Pois}\left(y\right) &= \frac{e^{-\lambda}\lambda^y}{y!} \end{aligned}$$

where $\lambda\left(>0\right)$ is the mean of y and $Var\left(y\right) = \lambda$ for Poisson and $\lambda + a\lambda^2$ for negative Binomial, $a \geq 0$ being the dispersion parameter for negative Binomial distribution. It is easy to see that the mean and variance of y under the ZINB model is given by

$$E\left(y\right) = \left(1-\phi\right)\lambda,\ Var\left(y\right) = \left[\left(1-\phi\right)\lambda\right]\left[1+\lambda\left(\phi+\alpha\right)\right]$$

while those for the ZIP model is

$$E\left(y\right) = \left(1-\phi\right)\lambda,\ Var\left(y\right) = \left[\left(1-\phi\right)\lambda\right]\left[1+\lambda\phi\right]$$

The mean parameter λ is related to the covariates $\boldsymbol{x}_i = \left(x_0,\ x_1, \ldots,\ x_p\right)_i$ by the log link function given by

$$\log\left(\lambda_i\right) = \sum_{j=0}^{p} x_j\beta_j = \boldsymbol{x}_i'\boldsymbol{\beta}$$

where $\boldsymbol{\beta} = \left(\beta_0,\ \beta_1, \ldots,\ \beta_p\right)$ is the $p+1$ dimensional vector of unknown regression coefficients. Finally the mixture proportion ϕ is linked to the set of zero inflated covariates $\boldsymbol{z} = \left(z_0,\ z_1, \ldots,\ z_q\right)$ and the $q+1$ vector of regression coefficients $\boldsymbol{\gamma} = \left(\gamma_0,\ \gamma_1, \ldots,\ \gamma_q\right)$ by the logit link function as in the following

$$logit\left(\phi\right) = \sum_{k=0}^{q} z_k\gamma_k = \boldsymbol{z}'\boldsymbol{\gamma}.$$

The z_0 is assumed to be 1 corresponding to the intercept γ_0.