We introduce the models for zero inflated count data following Zeng (2013). Define the ZIP model for the response variable y assuming only non-negative values as follows:

$$f_{ZIP}(y) = \phi I(y=0) + (1-\phi) f_{pois}(y),$$

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

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

where $f_{pois}(\cdot)$ is the density of Poisson variable given as

$$f_{pois}(y) = \frac{e^{-\lambda}\lambda^y}{y!},$$

$\lambda\,(>0)$ being the mean and variance of y, i.e. $E(y) = Var(y) = \lambda$ which is related to the covariates or the explanatory variables $\boldsymbol{x} = (x_0,\ x_1, \ldots,\ x_p)$ by the log link function given by

$$log(\lambda) = \sum_{j=0}^{p} x_j \beta_j = \boldsymbol{x}'\boldsymbol{\beta}$$

where $\boldsymbol{\beta} = (\beta_0,\ \beta_1, \ldots,\ \beta_p)$ is the $p+1$ dimensional vector of unknown regression coefficients. The x_0 is assumed to be 1 corresponding to the intercept term β_0. It can be easily verified that the mean and variance of the ZIP model is

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

In the similar pattern the ZINB model is written as

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

where the Negative Binomial density $f_{NB}(\cdot)$ is given by

$$f_{NB}(y) = \frac{\Gamma(y + 1/\alpha)}{y!\Gamma(1/\alpha)} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\lambda}\right)^{\alpha^{-1}} \left(\frac{\lambda}{\alpha^{-1}+\lambda}\right)^{y}$$

where $\lambda\,(>0)$ is the mean of y and $Var(y) = \lambda + a\lambda^2$, $a \geq 0$ being the dispersion parameter. The λ is related to the covariates and the regression coefficients by the log link function as given in the Poisson setting. Again it is easy to see that the mean and variance of y under the ZINB model is given by

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

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

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

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