

Description

A zero-truncated negative binomial model count data.

Implementation

The file `zt.nb.sim.R` simulates data according to the model statement presented below, and `zt.nb.mcmc.R` contains the MCMC algorithm for model fitting.

Derivation of zero-truncated negative binomial distribution

The probability mass function of the (non-truncated) negative binomial distribution is:

$$[z] = \left(\frac{\Gamma(z + \alpha)}{\Gamma(\alpha) \Gamma(z + 1)} \right) \left(\frac{\alpha}{\alpha + \lambda} \right)^\alpha \left(1 - \frac{\alpha}{\lambda + \alpha} \right)^z \quad (1)$$

It follows that the probability that $z = 0$ is

$$[z | z = 0] = \left(\frac{\Gamma(\alpha)}{\Gamma(\alpha) \Gamma(1)} \right) \left(\frac{\alpha}{\alpha + \lambda} \right)^\alpha \left(1 - \frac{\alpha}{\lambda + \alpha} \right)^0 \quad (2)$$

$$= \left(\frac{\alpha}{\alpha + \lambda} \right)^\alpha, \quad (3)$$

and thus the probability that $z > 0$ is $1 - [z | z = 0] = 1 - \left(\frac{\alpha}{\alpha + \lambda} \right)^\alpha$. We arrive at the density function of the zero-truncated negative binomial distribution by excluding the probability that $z = 0$ from the standard negative binomial distribution (Eq. 1). This is accomplished by dividing Eq. 1 by $[z | z = 0]$:

$$[z | z > 0] = \left(\frac{\Gamma(z + \alpha)}{\Gamma(\alpha) \Gamma(z + 1)} \right) \left(\frac{\alpha}{\alpha + \lambda} \right)^\alpha \left(1 - \frac{\alpha}{\lambda + \alpha} \right)^z \left(1 - \left(\frac{\alpha}{\alpha + \lambda} \right)^\alpha \right)^{-1} \quad (4)$$

$$= \text{NB}(z | \lambda, \alpha) \left(1 - \left(\frac{\alpha}{\alpha + \lambda} \right)^\alpha \right)^{-1}. \quad (5)$$

We abbreviate the density function for the zero-truncated negative binomial distribution as $\text{ZTNB}(\lambda, \alpha)$.

Model statement

Let z_i , for $i = 1, \dots, n$, be observed non-zero count data (i.e., z_i are integers greater than 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.

$$\begin{aligned} z_i &\sim \text{ZTNB}(\lambda_i, \alpha) \\ \log(\lambda_i) &= \mathbf{x}_i' \boldsymbol{\beta} \\ \boldsymbol{\beta} &\sim \mathcal{N}(\mathbf{0}, \sigma_\beta^2 \mathbf{I}) \\ \alpha &\sim \text{Gamma}(a, b), \end{aligned}$$

where $E[z_i] = \lambda_i$ and $\text{Var}[z_i] = \lambda_i + \frac{\lambda_i^2}{\alpha}$.

Full conditional distributions

Regression coefficients (β):

$$\begin{aligned} [\beta \mid \cdot] &\propto \prod_{i=1}^n [z_i \mid \beta, \alpha] [\beta] \\ &\propto \prod_{i=1}^n \text{ZTNB}(z_i \mid \mathbf{x}'_i \beta, \alpha) \mathcal{N}(\beta \mid \mathbf{0}, \sigma_\beta^2 \mathbf{I}) . \end{aligned}$$

The update for β proceeds using Metropolis-Hastings.

Dispersion (i.e., size) parameter (α):

$$\begin{aligned} [\alpha \mid \cdot] &\propto \prod_{i=1}^n [z_i \mid \beta, \alpha] [\alpha] \\ &\propto \prod_{i=1}^n \text{ZTNB}(z_i \mid \mathbf{x}'_i \beta, \alpha) \text{Gamma}(\alpha \mid a, b) . \end{aligned}$$

The update for α proceeds using Metropolis-Hastings.