Package ‘zic’

January 2, 2012

Version 0.6.2

Date 2011-12-21

Title Bayesian Inference for Zero-Inflated Count Models

Author Markus Jochmann <markus.jochmann@ncl.ac.uk>

Maintainer Markus Jochmann <markus.jochmann@ncl.ac.uk>

Description This package provides MCMC algorithms for the analysis of zero-inflated count models. The case of stochastic search variable selection (SSVS) is also considered.
All MCMC samplers are coded in C++ for improved efficiency. A data set considering the demand for health care is also provided.

Repository CRAN

Date/Publication 2011-12-21 17:46:16

R topics documented:

`docvisits` *Demand for Health Care Data*

Description

This data set gives the number of doctor visits in the last three months for a sample of German male individuals in 1994. The data set is taken from Riphahn et al. (2003) and is a subsample of the German Socioeconomic Panel (SOEP). In contrast to Riphahn et al. (2003) only male individuals from the last wave are considered. See Jochmann (2009) for further details.

Usage

```
data(docvisits)
```

Format

This data frame contains 1812 observations on the following 22 variables:

docvisits number of doctor visits in last 3 months

age age

agesq age squared / 1000

health health satisfaction, 0 (low) - 10 (high)

handicap 1 if handicapped, 0 otherwise

hdegree degree of handicap in percentage points

married 1 if married, 0 otherwise

schooling years of schooling

hhincome household monthly net income, in German marks / 1000

children 1 if children under 16 in the household, 0 otherwise

self 1 if self employed, 0 otherwise

civil 1 if civil servant, 0 otherwise

bluec 1 if blue collar employee, 0 otherwise

employed 1 if employed, 0 otherwise

public 1 if public health insurance, 0 otherwise

addon 1 if add-on insurance, 0 otherwise

References

Jochmann, M. (2009). “What Belongs Where? Variable Selection for Zero-Inflated Count Models with an Application to the Demand for Health Care”, available at: `http://personal.strath.ac.uk/markus.jochmann`.

Riphahn et al. (2003), “Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation”, *Journal of Applied Econometrics*, 18, 387-405.

Wagner et al. (2007), “The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements”, *Schmollers Jahrbuch*, 127, 139-169.

zic *Bayesian Inference for Zero-Inflated Count Models*

Description

`zic` fits zero-inflated count models via Markov chain Monte Carlo methods.

Usage

```
zic(formula, data, bbar, dbar, ebar, fbar, n.burnin, n.mcmc, n.thin)
```

Arguments

`formula`	A symbolic description of the model to be fit specifying the response variable and covariates.
`data`	A data frame in which to interpret the variables in `formula`.
`bbar`	The diagonal elements of the prior variance matrix of β, a vector of length equal to the number of covariates.
`dbar`	The diagonal elements of the prior variance matrix of δ, a vector of length equal to the number of covariates.
`ebar`	The shape parameter for the inverse gamma prior on σ^2.
`fbar`	The inverse scale parameter the inverse gamma prior on σ^2.
`n.burnin`	Number of burn-in iterations of the sampler.
`n.mcmc`	Number of iterations of the sampler.
`n.thin`	Thinning interval.

Details

The considered zero-inflated count model is given by

$$y_i^* \sim \mathrm{Poisson}[\exp(\eta_i^*)],$$
$$\eta_i^* = x_i'\beta + \varepsilon_i,\ \varepsilon_i \sim \mathrm{N}(0, \sigma^2),$$
$$d_i^* = x_i'\delta + \nu_i,\ \nu_i \sim \mathrm{N}(0, 1),$$
$$y_i = 1(d_i^* > 0)y_i^*,$$

where y_i and x_i are observed. The assumed prior distributions are

$$\beta \sim \mathrm{N}(0, \bar{B}) \quad \text{with} \quad \bar{B} = \mathrm{diag}(\bar{b}_1, \ldots, \bar{b}_k),$$
$$\delta \sim \mathrm{N}(0, \bar{D}) \quad \text{with} \quad \bar{D} = \mathrm{diag}(\bar{d}_1, \ldots, \bar{d}_k),$$
$$\sigma^2 \sim \text{Inv-Gamma}\left(\bar{e}, \bar{f}\right).$$

The sampling algorithm developed by Jochmann (2009) is used.

Value

A list containing the following elements:

`beta`	The posterior draws for β.
`delta`	The posterior draws for δ.
`sigma2`	The posterior draws for σ^2.

References

Jochmann, M. (2009). "What Belongs Where? Variable Selection for Zero-Inflated Count Models with an Application to the Demand for Health Care". Available at: http://personal.strath.ac.uk/markus.jochmann.

Examples

```
# library( zic )
# data( docvisits )

# prior parameters and formula
# bbar <- rep( 10.0, 16 )
# dbar <- rep( 10.0, 16 )
# ebar <- 3.0
# fbar <- 2.0
# formula <- docvisits ~ age + agesq + health + handicap + hdegree +
#            married + schooling + hhincome + children + self +
#            civil + bluec + employed + public + addon

# set seed and run MCMC sampler
# set.seed(1)
# results <- zic( formula, docvisits, bbar, dbar, ebar, fbar, 10000, 100000, 10 )

# print posterior means for beta
# apply( results$beta, 2, mean )
```

---

zic.ssvs *SSVS for Zero-Inflated Count Models*

Description

`zic.ssvs` applies SSVS to zero-inflated count models

Usage

```
zic.ssvs(formula, data, tausq0bar, tausq1bar, omegasq0bar, omegasq1bar,
         ebar, fbar, pbar, qbar, n.burnin, n.mcmc, n.thin)
```

Arguments

`formula`	A symbolic description of the model to be fit specifying the response variable and covariates.
`data`	A data frame in which to interpret the variables in `formula`.
`tausq0bar`	The "small" prior variances of β, a vector of length equal to the number of covariates.
`tausq1bar`	The "large" prior variances of β, a vector of length equal to the number of covariates.
`omegasq0bar`	The "small" prior variances of δ, a vector of length equal to the number of covariates.
`omegasq1bar`	The "large" prior variances of δ, a vector of length equal to the number of covariates.
`ebar`	The shape parameter for the inverse gamma prior on σ^2.
`fbar`	The inverse scale parameter the inverse gamma prior on σ^2.
`pbar`	The prior inclusion probabilties of β, a vector of length equal to the number of covariates.
`qbar`	The prior inclusion probabilties of δ, a vector of length equal to the number of covariates.
`n.burnin`	Number of burn-in iterations of the sampler.
`n.mcmc`	Number of iterations of the sampler.
`n.thin`	Thinning interval.

Details

The considered zero-inflated count model is given by

$$y_i^* \sim \text{Poisson}[\exp(\eta_i^*)],$$

$$\eta_i^* = x_i'\beta + \varepsilon_i,\ \varepsilon_i \sim \text{N}(0, \sigma^2),$$

$$d_i^* = x_i'\delta + \nu_i,\ \nu_i \sim \text{N}(0, 1),$$

$$y_i = 1(d_i^* > 0)y_i^*,$$

where y_i and x_i are observed. The SSVS prior is given by

$$\beta_j \sim (1 - \gamma_j)\text{N}\left(0, \overline{\tau}_{0j}^2\right) + \gamma_j \text{N}\left(0, \overline{\tau}_{1j}^2\right), \quad j = 1, \ldots, k,$$

$$\delta_j \sim (1 - \kappa_j)\text{N}\left(0, \overline{\omega}_{0j}^2\right) + \kappa_j \text{N}\left(0, \overline{\omega}_{1j}^2\right), \quad j = 1, \ldots, k,$$

$$\text{P}(\gamma_j = 1) = 1 - \text{P}(\gamma_j = 0) = \overline{p}_j, \quad j = 1, \ldots, k,$$

$$\text{P}(\kappa_j = 1) = 1 - \text{P}(\kappa_j = 0) = \overline{q}_j, \quad j = 1, \ldots, k,$$

$$\sigma^2 \sim \text{Inv-Gamma}\left(\bar{e}, \bar{f}\right).$$

The sampling algorithm described in Jochmann (2009) is used.

Value

A list containing the following elements:

beta	The posterior draws for β.
delta	The posterior draws for δ.
sigma2	The posterior draws for σ^2.
gamma	The posterior draws for γ.
kappa	The posterior draws for κ.

References

Jochmann, M. (2009). "What Belongs Where? Variable Selection for Zero-Inflated Count Models with an Application to the Demand for Health Care". Available at: http://personal.strath.ac.uk/markus.jochmann.

Examples

```
# library( zic )
# data( docvisits )
# formula <- docvisits ~ age + agesq + health + handicap + hdegree +
#            married + schooling + hhincome + children + self +
#            civil + bluec + employed + public + addon

# prior parameters for preliminary run
# bbar <- rep( 10.0, 16 )
# dbar <- rep( 10.0, 16 )
# ebar <- 3.0
# fbar <- 2.0

# set seed and run preliminary MCMC sampler
# set.seed(1)
# prelim <- zic( formula, docvisits, bbar, dbar, ebar, fbar, 10000, 100000, 10 )

# prior parameters SSVS run
# varbeta <- apply( prelim$beta, 2, var )
# vardelta <- apply( prelim$delta, 2, var )
# tausq0bar <- 0.01 * varbeta
# tausq1bar <- 100.0 * varbeta
# omegasq0bar <- 0.01 * vardelta
# omegasq1bar <- 100.0 * vardelta
# ebar <- 3.0
# fbar <- 2.0
# pbar <- rep( 0.5, 16 )
# qbar <- rep( 0.5, 16 )

# set seed and run preliminary MCMC sampler
# set.seed(2)
# results <- zic.ssvs( formula, docvisits, tausq0bar, tausq1bar,
#                      omegasq0bar, omegasq1bar, ebar, fbar,
#                      pbar, qbar, 10000, 100000, 10 )
```

```
# print posterior inclusion probabilities means for delta
# apply( results$kappa, 2, mean )
```

Index