Package ‘ZIPBayes’

November 25, 2021

Type Package

Title Bayesian Methods in the Analysis of Zero-Inflated Poisson Model

Version 1.0.2

Date 2021-09-27

Description Implementation of zero-inflated Poisson models under Bayesian framework using data augmentation as discussed in Chapter 5 of Zhang (2020) <https://hdl.handle.net/10012/16378>. This package is constructed in accommodating four different scenarios: the general scenario, the scenario with measurement error in responses, the external validation scenario, and the internal validation scenario.

License GPL (>= 2)

Imports Rcpp (>= 0.12.11), MASS

Depends R (>= 3.5.0)

LinkingTo Rcpp

NeedsCompilation yes

Author Qihuang Zhang [aut, cre, cph],
Grace Y. Yi [aut, ths]

Maintainer Qihuang Zhang <qihuang.zhang@uwaterloo.ca>

Repository CRAN

Date/Publication 2021-11-25 06:30:14 UTC

R topics documented:

ZIPBayes-package — *Bayesian Methods in Zero-inflated Poisson Model*

Description

Implementation of zero-inflated Poisson (ZIP) model in Bayesian methods with data augmentation strategy. The R package is general to different scenarios, including an ordinary scenario for zero-inflated Poisson Model, the scenario with measurement error in response, the scenario with internal or external validation data available.

Details

This package implemented the zero-inflated Poisson model based on a Bayesian framework. The method is implemented with a Monte Carlo Markov Chain (MCMC) approach with a data augmentation strategy. The package is integrated with C++ to improve the computing speed. It mainly contains four main functions. The function ZIPBayes corresponds to ordinary zero-inflated count data and no measurement error is considered. The function ZIPBayes_MErr considers the case where the response is subject to measurement error as a model by Qihuang Zhang (2020). The function ZIPBayes_Int and ZIPBayes_Ext are corresponding to the case where the internal or external validation data are available, respectively. Other helper functions are also contained in this packages, such as summarizing the trace from the MCMC algorithm, plotting the trace plot, etc.

Author(s)

Qihuang Zhang and Grace Y. Yi.

Maintainer: Qihuang Zhang <qihuang.zhang@uwaterloo.ca>

datasim — *Toy example data - main study only*

Description

This data set gives an example data for the illustration of usage of `ZIP` and `ZIPMErr` function. The dataset contains naivedata and the design matrices for the response model, measurement error model.

Usage

```
data(datasim)
```

Format

A data.frame of 6 columns. “Ystar” refers to the error-prone response. “Y” refers to the true count response. "X1" and "X2" are covariates in the response model. “Xplus” and “Xminus” are the covariates for the measurement error model.

`datasimExt` *Toy example data - main study and external validation study*

Description

This data set gives an example data for the illustration of usage of `ZIPExt` function. The dataset contains a list of main data and external validation data.

Usage

```
data(datasimExt)
```

Format

A list of two data.frames. The first data.frame, named "main", corresponds to the main data with 6 columns. Same as the `datasim`, "Ystar" refers to the error-prone response. "Y" refers to the true count response. "X1" and "X2" are covariates in the response model. "Zplus" and "Zplus" are the covariates for the measurement error model. The second data.frame corresponds to the validation data with 5 columns.

`datasimInt` *Toy example data - main study and internal validation study*

Description

This data set gives an example data for the illustration of usage of `ZIPInt` function. The dataset contains a list of main data and internal validation data.

Usage

```
data(datasimInt)
```

Format

A list of two data.frames. The first data.frame, named "main", corresponds to the main data with 6 columns. Same as the `datasim`, "Ystar" refers to the error-prone response. "Y" refers to the true count response. "X1" and "X2" are covariates in the response model. "Zplus" and "Zplus" are the covariates for the measurement error model. The second data.frame corresponds to the validation data with 7 columns.

`summary.ZIPBayes` *Summarizing the trace output from the MCMC algorithm*

Description

This function is a method for `ZIPBayes` object. It summarize the trace output from the main functions into interesting summary statistics, such as mean, median, confidence interval, and highest density region (HDR),

Usage

```
## S3 method for class 'ZIPBayes'

## S3 method for class 'ZIPBayes'
summary(object, burnin = 1, thinperiod = 1, confidence.level = 0.95, ...)
```

Arguments

`object`	the "ZIPBayes" object gotten from the main function.
`burnin`	the number of records to be discarded as the early period of MCMC algorithm. Default is 1, meaning the first data point will be discarded when calculating the summary statistics.
`thinperiod`	the number of period in the thining periord. The results will be picked every this number. Default is 1, meaning no thining will be done. See details.
`confidence.level`	the confidence level for the constructed confidence interval. Default is 0.95.
`...`	other arguments passed to the function.

Details

This function summarizes the tracing results produced by `ZIP`, `ZIPMErr`, `ZIPExt`, and `ZIPInt`. To diminish the influence of the starting values, we generally discard the first portion of each sequence and focus attention on the remaining. The argument `burnin` is set to control the number of steps to be discarded.

Another issue that sometimes arises, once approximate convergence has been reached, is whether to thin the sequences by keeping every k-th simulation draw from each sequence and discarding the rest. The argument `thinperiod` is used to set k here.

Value

`ZIPBayes`	a list of summary for each data set. "HDR_LB" and "HDR_UB" respectively respresents the lower and upper bound of the high density region.

Author(s)

Qihuang Zhang and Grace Y. Yi.

See Also

ZIP, ZIPMErr, ZIPExt, ZIPInt

Examples

```
## Please see the example in ZIP() function
```

ZIP *zero-inflated Poisson model*

Description

The function implements the MCMC algorithm with data augmentation to estimate the parameters in the zero-inflated Poisson model. The function returns the trace of the sampled parameters in each interaction. To obtain the summary estimation, use `summary()`.

Usage

```
ZIP(Y, Covarmainphi, Covarmainmu,
                          betaphi, betamu,
                          priorgamma,
                          propsigmaphi,  propsigmamu = propsigmaphi,
                          seed = 1, nmcmc = 500)
```

Arguments

Y	a count vector of length n specifying response in the zero-inflated Poisson model.
Covarmainphi	a $n \times p_1$ dimensional data.frame or matrix of data with respect to the probability component of the zero-inflated Poisson model.
Covarmainmu	a $n \times p_2$ dimensional data.frame or matrix of data with respect to the mean component of the zero-inflated Poisson model.
betaphi	a vector of length p_1 specifying the initial values of the parameters in the probability component of the zero-inflated Poisson model
betamu	a vector of length p_2 specifying the initial values of the parameters in the probability component of the zero-inflated Poisson model
priorgamma	a vector of length 2 specifying the two parameters of gamma prior
propsigmaphi	a vector of length p_1 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the probability component.
propsigmamu	a vector of length p_2 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the mean component.
seed	a numeric value specifying the seed for random generator
nmcmc	a integer specify the number of the generation of MCMC algorithm

Details

The zero-inflated Poisson model involves two components, the probability components and the mean compoenents (Zhang, 2020). Argument `Covarmainphi`, `betaphi`, `propsigmaphi` correspond to the probability compoenent; `Covarmainmu`, `betamu`, `propsigmamu` correspond to the mean compoenent.

Value

`BayesResults`	the list of trace of generated parameters for each component of the models. Data.frame "betaphi_trace" corresponds to the probability component of ZIP response model; "betamu_trace" refers to the mean component of the ZIP response model.

Author(s)

Qihuang Zhang and Grace Y. Yi

References

Zhang, Qihuang. "Inference Methods for Noisy Correlated Responses with Measurement Error." (2020).

See Also

glm

Examples

```
data(datasim)
set.seed(0)
example_ZIP <- ZIP( Y = datasim$Ystar,
                         Covarmainphi = datasim[,c("intercept","X1")],
                         Covarmainmu = datasim[,c("intercept","X2")],
                         betaphi = c(-0.7,0.7), betamu = c(1,-0.5),
                         priorgamma = rep(1,1), propsigmaphi = c(0.05,0.05),
                         nmcmc = 100)

summary(example_ZIP)
```

`ZIPExt` *Zero-inflated Poisson model under measurement error and external validation data are available*

Description

The function implements the MCMC algorithm with data augmentation to estimate the parameters in the zero-inflated Poisson model while correcting for the measurement error arising from the responses. The function returns the trace of the sampled parameters in each interaction. To obtain the summary estimation, use `summary()`.

Usage

```
ZIPExt (Ystar, Covarmainphi, Covarmainmu, Covarplus, Covarminus,
           Ystarval, Yval, Covarvalplus, Covarvalminus,
           betaphi, betamu, alphaplus, alphaminus,
           Uibound = c(7,11),
           priorgamma, priormu, priorSigma,
           propsigmaphi,  propsigmamu = propsigmaphi,
           propsigmaplus = propsigmaphi, propsigmaminus = propsigmaphi,
           seed = 1, nmcmc = 500)
```

Arguments

`Ystar`	a count vector of length n specifying the error-prone response in the zero-inflated Poisson model.
`Covarmainphi`	a $n \times p_1$ dimensional data.frame or matrix of the covariate data with respect to the probability component of the zero-inflated Poisson model
`Covarmainmu`	a $n \times p_2$ dimensional data.frame or matrix of the covariate data with respect to the mean component of the zero-inflated Poisson model
`Covarplus`	a $n \times q_1$ dimensional data.frame or matrix of the covariate data for the measurement error model of the add-in error process
`Covarminus`	a $n \times q_2$ dimensional data.frame or matrix of the covariate data for the measurement error model of the leave-out error process
`Ystarval`	a count vector of length m specifying the error-prone response in the validation data.
`Yval`	a count vector of length m specifying the precisely measured response in the validation data.
`Covarvalplus`	a $m \times q_1$ dimensional data.frame or matrix of the covariate for validation data of the add-in error process
`Covarvalminus`	a $m \times q_2$ dimensional data.frame or matrix of the covariate for validation data of the leave-out error process
`betaphi`	a vector of length p_1 specifying the initial values of the parameters in the probability component of the zero-inflated Poisson model
`betamu`	a vector of length p_2 specifying the initial values of the parameters in the probability component of the zero-inflated Poisson model
`alphaplus`	a vector of length q_1 specifying the initial values of the parameters for the measurement error model of the add-in error process
`alphaminus`	a vector of length q_2 specifying the initial values of the parameters in the probability component of the leave-out error process
`Uibound`	a vector of length 2 specifying the maximum number of the count in the inverse sampling method
`priorgamma`	a vector of length 2 specifying the two parameters of gamma prior
`priormu`	a vector of length q_2 specifying the mean vector of the normal prior for the measurement error model of the leave-out error process

priorSigma	a vector of length q_2 specifying the standard erros of the normal prior for the measurement error model of the leave-out error process
propsigmaphi	a vector of length p_1 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the probability component.
propsigmamu	a vector of length p_2 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the mean component.
propsigmaplus	a vector of length q_1 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the add-in error process.
propsigmaminus	a vector of length q_1 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the leave-out error process.
seed	a numeric value specifying the seed for random generator
nmcmc	a integer specify the number of the generation of MCMC algorithm

Details

Comparing to the ZIPMErr function, this function has an addition component – validation data. Here, the argument "Ystarval", "Yval", "Covarvalplus", "Covarvalminus", are new for the sceanrio with external validation.

Value

BayesResults	the list of trace of generated parameters for each component of the models. Data frame "betaphi_trace" corresponds to the probability component of ZIP response model; "betamu_trace" refers to the mean component of the ZIP response model. Data frames "alphaplus_trace" and "alphaminus_trace", respectively, correspond to the add-in error and leave-out error process in the measruement error model.

Author(s)

Qihuang Zhang and Grace Y. Yi

References

Zhang, Qihuang. "Inference Methods for Noisy Correlated Responses with Measurement Error." (2020).

See Also

glm

Examples

```
## load data
data(datasimExt)
set.seed(0)
example_ZIP_Ext <-  ZIPExt (Ystar = datasimExt$main$Ystar,
                            Covarmainphi = datasimExt$main[,c("intercept","X1")],
```

```
                                  Covarmainmu = datasimExt$main[,c("intercept","X2")],
                                  Covarplus = datasimExt$main[,c("intercept","Zplus")],
                                  Covarminus = datasimExt$main[,c("intercept","Zminus")],
                                  Ystarval = datasimExt$validation$Ystar,
                                  Yval = datasimExt$validation$Y,
                                  Covarvalplus = datasimExt$validation[,3:4],
                                  Covarvalminus = datasimExt$validation[,3:4],
                                  betaphi = c(0.7,-0.7), betamu = c(1,-1.5),
                                  alphaplus = c(0,0), alphaminus=c(0,0),
                                  priorgamma = c(0.001,0.001), priormu = c(0,0),
                                  priorSigma = c(1,1), propsigmaphi = c(0.05,0.05),
                                  nmcmc = 10)

summary(example_ZIP_Ext)
```

`ZIPInt` *Zero-inflated Poisson model under measurement error and internal validation data are available*

Description

The function implements the MCMC algorithm with data augmentation to estimate the parameters in the zero-inflated Poisson model while correcting for the measurement error arising from the responses. The function returns the trace of the sampled parameters in each interaction. To obtain the summary estimation, use `summary()`.

Usage

```
ZIPInt (Ystar, Covarmainphi, Covarmainmu, Covarplus, Covarminus,
     Ystarval, Yval, Covarvalmainphi, Covarvalmainmu, Covarvalplus, Covarvalminus,
       betaphi, betamu, alphaplus, alphaminus,
       Uibound = c(7,11),
       priorgamma, priormu, priorSigma,
       propsigmaphi,  propsigmamu = propsigmaphi,
            propsigmaplus = propsigmaphi,  propsigmaminus = propsigmaphi,
       seed = 1, nmcmc = 500)
```

Arguments

`Ystar`	a count vector of length n specifying the error-prone response in the zero-inflated Poisson model.
`Covarmainphi`	a $n \times p_1$ dimensional data.frame or matrix of the covariate data with respect to the probability component of the zero-inflated Poisson model
`Covarmainmu`	a $n \times p_2$ dimensional data.frame or matrix of the covariate data with respect to the mean component of the zero-inflated Poisson model
`Covarplus`	a $n \times q_1$ dimensional data.frame or matrix of the covariate data for the measurement error model of the add-in error process

Covarminus
: a $n \times q_2$ dimensional data.frame or matrix of the covariate data for the measurement error model of the leave-out error process

Ystarval
: a count vector of length m specifying the error-prone response in the validation data.

Yval
: a count vector of length m specifying the precisely measured response in the validation data.

Covarvalmainphi
: a $m \times p_1$ dimensional data.frame or matrix of the covariate for validation data with respect to the probability component of ZIP model

Covarvalmainmu
: a $m \times p_2$ dimensional data.frame or matrix of the covariate for validation data with respect to the mean component of ZIP model

Covarvalplus
: a $m \times q_1$ dimensional data.frame or matrix of the covariate for validation data of the add-in error process

Covarvalminus
: a $m \times q_2$ dimensional data.frame or matrix of the covariate for validation data of the leave-out error process

betaphi
: a vector of length p_1 specifying the initial values of the parameters in the probability component of the zero-inflated Poisson model

betamu
: a vector of length p_2 specifying the initial values of the parameters in the probability component of the zero-inflated Poisson model

alphaplus
: a vector of length q_1 specifying the initial values of the parameters for the measurement error model of the add-in error process

alphaminus
: a vector of length q_2 specifying the initial values of the parameters in the probability component of the leave-out error process

Uibound
: a vector of length 2 specifying the maximum number of the count in the inverse sampling method

priorgamma
: a vector of length 2 specifying the two parameters of gamma prior

priormu
: a vector of length q_2 specifying the mean vector of the normal prior for the measurement error model of the leave-out error process

priorSigma
: a vector of length q_2 specifying the standard erros of the normal prior for the measurement error model of the leave-out error process

propsigmaphi
: a vector of length p_1 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the probability component.

propsigmamu
: a vector of length p_2 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the mean component.

propsigmaplus
: a vector of length q_1 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the add-in error process.

propsigmaminus
: a vector of length q_1 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the leave-out error process.

seed
: a numeric value specifying the seed for random generator

nmcmc
: a integer specify the number of the generation of MCMC algorithm

Details

Comparing to the ZIPExt function for the external validation study, this function has an addition component – covariates of response model in the validation data. Here, the argument "Covarvalmainphi" and "Covarvalmainmu" are new for the sceanrio with external validation.

Value

`BayesResults` the list of trace of generated parameters for each component of the models. Data frame "betaphi_trace" corresponds to the probability component of ZIP response model; "betamu_trace" refers to the mean component of the ZIP response model. Data frames "alphaplus_trace" and "alphaminus_trace", respectively, correspond to the add-in error and leave-out error process in the measruement error model.

Author(s)

Qihuang Zhang and Grace Y. Yi

References

Zhang, Qihuang. "Inference Methods for Noisy Correlated Responses with Measurement Error." (2020).

See Also

`glm`

Examples

```
data(datasimInt)
set.seed(0)
result <-  ZIPInt(Ystar = datasimInt$main$Ystar,
                  Covarmainphi = datasimInt$main[,c("intercept","X1")],
                  Covarmainmu = datasimInt$main[,c("intercept","X2")],
                  Covarplus = datasimInt$main[,c("intercept", "Zplus")],
                  Covarminus = datasimInt$main[,c("intercept", "Zminus")],
                  Ystarval = datasimInt$val$Ystar,
                  Yval = datasimInt$val$Y,
                  Covarvalmainphi = datasimInt$val[,c("intercept","X1")],
                  Covarvalmainmu = datasimInt$val[,c("intercept","X2")],
                  Covarvalplus = datasimInt$val[,c("intercept", "Zplus")],
                  Covarvalminus = datasimInt$val[,c("intercept", "Zminus")],
                  betamu =  c(-0.5,0.5), betaphi = c(0.5,0),
                  alphaplus = c(0,0), alphaminus= c(0,0),
                  priorgamma = c(1,1), priormu = c(0,0), priorSigma = c(1,1),
                     propsigmaphi = c(0.05,0.05), nmcmc = 10)

summary(result)
```

ZIPMErr *Zero-inflated Poisson model with measurement error*

Description

The function implements the MCMC algorithm with data augmentation to estimate the parameters in the zero-inflated Poisson model while correcting for the measurement error arising from the responses. The function returns the trace of the sampled parameters in each iteraction. To obtain the summary estimation, use `summary()`.

Usage

```
ZIPMErr (Ystar, Covarmainphi, Covarmainmu, Covarplus, Covarminus,
        betaphi, betamu, alphaplus, alphaminus,
        Uibound = c(7,11),
      priorgamma, priormu, priorSigma, propsigmaphi,  propsigmamu = propsigmaphi,
        propsigmaplus = propsigmaphi,  propsigmaminus = propsigmaphi,
        seed = 1, nmcmc = 500)
```

Arguments

`Ystar`	a count vector of length n specifying the error-prone response in the zero-inflated Poisson model.
`Covarmainphi`	a $n \times p_1$ dimensional data.frame or matrix of the covariate data with respect to the probability component of the zero-inflated Poisson model
`Covarmainmu`	a $n \times p_2$ dimensional data.frame or matrix of the covariate data with respect to the mean component of the zero-inflated Poisson model
`Covarplus`	a $n \times q_1$ dimensional data.frame or matrix of the covariate data for the measurement error model of the add-in error process
`Covarminus`	a $n \times q_1$ dimensional data.frame or matrix of the covariate data for the measurement error model of the leave-out error process
`betaphi`	a vector of length p_1 specifying the initial values of the parameters in the probability component of the zero-inflated Poisson model
`betamu`	a vector of length p_2 specifying the initial values of the parameters in the probability component of the zero-inflated Poisson model
`alphaplus`	a vector of length q_1 specifying the initial values of the parameters for the measurement error model of the add-in error process
`alphaminus`	a vector of length q_2 specifying the initial values of the parameters in the probability component of the leave-out error process
`Uibound`	a vector of length 2 specifying the maximum number of the count in the inverse sampling method
`priorgamma`	a vector of length 2 specifying the two parameters of gamma prior
`priormu`	a vector of length q_2 specifying the mean vector of the normal prior for the measurement error model of the leave-out error process

priorSigma	a vector of length q_2 specifying the standard erros of the normal prior for the measurement error model of the leave-out error process
propsigmaphi	a vector of length p_1 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the probability component.
propsigmamu	a vector of length p_2 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the mean component.
propsigmaplus	a vector of length q_1 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the add-in error process.
propsigmaminus	a vector of length q_1 specifying the standard error of the Gaussian proposal distribution for the parameters corresponds to the leave-out error process.
seed	a numeric value specifying the seed for random generator
nmcmc	a integer specify the number of the generation of MCMC algorithm

Details

The zero-inflated Poisson model involves two components, the probability components and the mean compoenents (Zhang, 2020). Although there are might arugments involved in the functions, they can be summarized to four sources in the model. The response model (zero-inflated Poisson model) involves two components: the probability component and the mean count component. The measurement error models contains two process: the add-in process and leave-out process. The arguements end with "-phi" corresponds to the probability component of the response model. The arguements end with "-mu" corresponds to the mean component of the response model. The arguements end with "-plus" corresponds to the add-in error process in the measurment error model. The arguements end with "-minus" corresponds to the leave-out process of the measurement error model.

Value

BayesResults	the list of trace of generated parameters for each component of the models. Data frame "betaphi_trace" corresponds to the probability component of ZIP response model; "betamu_trace" refers to the mean component of the ZIP response model. Data frames "alphaplus_trace" and "alphaminus_trace", respectively, correspond to the add-in error and leave-out error process in the measruement error model.

Author(s)

Qihuang Zhang and Grace Y. Yi

References

Zhang, Qihuang. "Inference Methods for Noisy Correlated Responses with Measurement Error." (2020).

See Also

glm

Examples

```
## load data
data(datasim)
set.seed(0)
example_ZIP_MErr <-  ZIPMErr (Ystar = datasim$Ystar,
                          Covarmainphi = datasim[,c("intercept","X1")],
                          Covarmainmu = datasim[,c("intercept","X2")],
                          Covarplus = datasim[,c("intercept","Xplus")],
                          Covarminus = datasim[,c("intercept","Xminus")],
                          betaphi = c(0.7,-0.7), betamu = c(1,-1.5),
                          alphaplus = c(0,0), alphaminus=c(0,0),
                          priorgamma = c(0.001,0.001), priormu = c(0,0),
                          priorSigma = c(1,1), propsigmaphi = c(0.05,0.05),
                          nmcmc = 10)

summary(example_ZIP_MErr)
```

Index