Package ‘bayesmcmc4m3’

January 6, 2022

Type Package

Title Reduction of bias caused by the misclassification of an exposure variable.

Version 1.0

Date 2022-01-05

Author Samiran Sinha

Maintainer Samiran Sinha <sinha@stat.tamu.edu>

Imports nnet

Description Bayesian MCMC technique has been used to estimate the model parameters of a logistic model when a set of instrumetal variables are observed along with the erroneous surrogate measurement for the exposure variable, and a set of error-free covariates are observed.

License GPL>=2

R topics documented:

`bayesmcmc4m3-package` *Reduction of bias caused by the misclassification of an exposure variable.*

Description

Bayesian MCMC technique has been used to estimate the model parameters of a logistic model when a set of instrumetal variables are observed along with the erroneous surrogate measurement for the exposure variable, and a set of error-free covariates are observed.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.
This package contains the function bmcmcm3 that carries out the Bayes analysis of the logistic model when a categorical exposure is misclassified. To reduce the bias caused by the misclassification, a set of instrumental variables must be available.

Author(s)

Samiran Sinha

Maintainer: Samiran Sinha <sinha@stat.tamu.edu>

References

Manuel, C., Sinha, S. and Wang, S. (2022). Reduction of bias due to misclassified exposures using instrumental variables. To appear in the Canadian Journal of Statistics.

See Also

<fmcmc>, <BayesianTools>, <MCMCpack>

Examples

```
# There is only one function in this package bmcmcm3. First, we simulate a dataset, and
# then apply the function to get the results.
nt=100
set.seed(nt)
n=500
# generation of error free covariates
z1=rbinom(n, 1, 0.23)
z2=rep(0, n)
z2=rbinom(n, 1, 0.31)
# Generation of instrumental variables. Here the instrumental variable is a nominal
# categorical variable with three levels.
x0star=t(rmultinom(n, 1, prob=c(0.74, 0.22, 0.04)))
x1star=ifelse(x0star==2, 1, 0)
x2star=ifelse(x0star==3, 1, 0)
# However, the instrumental variable could be numeric, and there could be more
# than one instrumental variable.
# Generation of X
## If needed change the coefficients of f(X|X^*, Z) of the following two lines
##
pr2= exp(0.25+2*x1star+1*x2star+0.3*z1-0.5*z2)
pr3= exp(0.5 + 2.5*x1star-1*x2star-0.3*z1+0.7*z2)
deno=1+pr2+pr3
pr1=1/deno
pr2=pr2/deno
pr3=pr3/deno
prob=cbind(pr1, pr2, pr3)
f1=function(i) sum(rmultinom(1, 1, prob[i, ])*(1:3))
myx=unlist(lapply(seq(1:n), f1))

# Generation of the response
## If needed change the coefficients of f(Y|X, Z) of the following line
eta=-2+ 1*as.numeric(myx==2) +0.7*as.numeric(myx==3) + 0.5*z1-0.25*z2
pry= 1/(1+exp(-eta))
y=rbinom(n, 1, pry)
### data where x is being recorded
originaldata=data.frame(myx, y, z1, z2)
head(originaldata)
### generation of the surrogate W
### For a different misclassification, change the entries in the next line
misclassmat=matrix(c(0.96, 0.025, 0.025, 0.1, 0.85, 0.05, 0.2, 0.1, 0.7),
 byrow=FALSE, ncol=3)
```

```
f2=function(i) sum(rmultinom(1, 1, misclassmat[,i])*(1:3))
myw=unlist(lapply(myx, f2))
# For checking you may print
table(myw, myx)
###
###
### data without x, rather we have w and instrumental variables
mydata=data.frame(myw, x1star, x2star, y, z1, z2)

response="y"
error.free.covariates=c("z1", "z2")
surrogate="myw"
IV=c("x1star", "x2star")
out1000=bmcmcm3(response, error.free.covariates, IV, surrogate, mydata,
nmcmc=40000, burnin=5000,  propscale=0.5)
print(out1000)
```

---

bmcmcm3 *This function returns the posterior mean, median, standard deviation and the 95 percent credible intervals of the parameters of the model.*

Description

This function returns the posterior mean, median, standard deviation and the 95 percent credible intervals of the parameters of the model. The MCMC samples are drawn through the random-walk Metropolis-Hastings algorithm.

Usage

```
bmcmcm3(response, error.free.covariates, IV, surrogate, mydata, nmcmc, burnin,
priorvarvec, propscale)
```

Arguments

`response`	It is the name of the binary response variable for which the logistic model is built. This variable should be one of the columns of the data frame, mydata.
`error.free.covariates`	It is the set of names of the error-free covariates that should be included as predictor variables in the logistic model. These variables should be identified as columns of the data frame, mydata.
`IV`	It is the set of names of the instrumental variables for the main exposure that is misclassified. These variables should be identified as columns of the data frame, mydata.
`surrogate`	The erroneous measurement of the underlying true exposure. This variable must be one of the columns of the data frame, mydata.
`mydata`	The data frame containing the above mentioned variables.
`nmcmc`	The number of MCMC iterations. The default value is set to 40,000.
`burnin`	The number of burn-in sample. The default value is min(5000, 0.25*nmcmc).

priorvarvec
: This a vector containing the variances of the prior distribution of the parameters. By default, each component of this vector is set to 2. Note that the length of this vector is the number of the parameters of the logistic regression plus the number of parameters of the multinomial model for the latent exposure X plus the parameters of the misclassification matrix. For instance, if the design matrix of the logistic regression has dimension n x 5 (including the intercept), the design matrix of the multinomial regression also has a dimension of n x 5, and the number of nominal categories of X is 3, and consequently the number of unknown parameters of the misclassification matrix is 3*(3-1)=6, and the length of *priorvarvec* is 5+ (3-1) *5 +6= 21. Here n denotes the sample size.

propscale
: This parameter is multipled with a term to calculate the variance of the random-walk proposal distribution. The default is set as one. Note that the proposal variance increases with this parameter.

Value

M2
: Estimates and standard error (SE) of the logistic regression parameters under the naive method. This is the frequentist result.

M3
: Posterior summaries of the logistic regression parameters under the proposed method. The posterior summary includes the posterior mean, median, standard error and the 95

Manuel, C., Sinha, S. and Wang, S. (2022). Reduction of bias due to misclassified exposures using instrumental variables. To appear in the Canadian Journal of Statistics.

Samiran Sinha

See the previous example.

Index