Package 'robustsae'

December 6, 2016

Type Package

Title Robust Bayesian Small Area Estimation

Version 0.1.0

Date 2016-12-05

Author Malay Ghosh, Jiyoun Myung, Fernando Moura

Maintainer Jiyoun Myung <jiyoun@ufl.edu>

Description Functions for Robust Bayesian Small Area Estimation.

License GPL-3

LazyData true

Depends R(>= 3.1.0), MCMCpack, coda, lattice, mvtnorm, pscl

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-12-06 07:52:43

R topics documented:

---

robustsae-package   *Robust Bayesian Small Area Estimation*

1

Description

The package provides a function `robustsae` for full non-subjective Bayesian analysis for the general area level small area models. This considers small area modeling of both the population means and the population variances. This is possible due to the availability of additional data purported to estimate the error variances. Also, in order to induce some robustness of the procedure, t-prior for the random effects is used. When the data set includes true values for interest parameter, this function returns the comparison criteria.

Details

Package:	robustsae
Type:	Package
Version:	1.0
Date:	2016-12-05
License:	GPL-3

This package provides function for full Bayesian analysis of small area models.

Author(s)

Malay Ghosh, Jiyoun Myung, Fernando Moura

Maintainer: Jiyoun Myung <jiyoun@ufl.edu>

References

Chip, S., and Green berg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, **49**, 327-335.

Rao, J. N. K. (2003) *Small Area Estimation*. John Wiley and Sons.

You, Y. and Chapman, B. (2006) Small Area Estimation Using Area Level Models and Estimated Sampling Variances. *Survey Methodology,* **32:** 97-103.

Malay Ghosh, Jiyoun Myung, and Fernando Moura. (submitted) Robust Bayesian Small Area Estimation.

BZdata   *Brazilian data*

Description

The data set is selected by a 10% random sampling of households in each area from a test demographic census completed in one municipality in Brazil consisting of 140 enumeration districts. This data set includes two centered auxiliary covariates, sampling means, sampling variances and true means for all areas. The contained information is available only at the area level.

Usage

```
data("BZdata")
```

Format

A data frame with 140 observations on the following 6 variables.

ni: sample size for each district.

X1: respective small area population means of the educational attainment of the head of household, centered auxiliary covariate.

X2: respective average number of rooms in households, centered auxiliary covariate

S2: respective sampling variances income of head of the household.

y: respective average mean income of head of the household.

truemean: respective true mean income of head of the household.

---

corndata   *Corn data in 8 counties in Iowa.*

Description

Survey and satellite data for corn and soy beans in 12 Iowa counties, obtained from the 1978 June Enumerative Survey of the U.S. Department of Agriculture and from land observatory satellites (LANDSAT) during the 1978 growing season.

Usage

```
data("corndata")
```

Format

A data frame with 8 observations on the following 6 variables.

County: county names.

ni: sample size for each county.

Xi: mean of reported hectares of corn from the survey, direct survey estimate.

Z1i: mean of pixels of corn for each, from satellite data.

Z2i: mean of pixels of soy bean for each county, from satellite data.

Si: square root of sample variance of reported hectares of corn from the survey.

Details

While the original dataset includes survey and satellite data for corn in 12 Iowa counties, this dataset contains only 8 counties' information where sample sizes are greater than 1.

Source

- Battesse, G.E., Harter, R.M. and Fuller, W.A. (1988). An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, **83**, 28-36.

- You, Y. and Chapman, B. (2006) Small Area Estimation Using Area Level Models and Estimated Sampling Variances. *Survey Methodology*, **32**, 97-103.

---

robustsae	*Robust Small Area Estimation Modeling Both Means and Variances*

Description

This function provides full Bayesian Analysis for specific area-level small area models when data are provided for modeling both the mean and the variance.

Usage

```
robustsae(formula, S2, ni, nsim = 1000, burnin = 500, data, truemean)
```

Arguments

formula	a symbolic description of the model to be fitted. The details of model specification are given under Details.
S2	a vector contain the sampling variances which are given for estimating the true variances.
ni	a vector containing the sample sizes for each area.
nsim	user-specified number of MCMC draws. See German (2006).
burnin	the number of burning iterations for the sampler. See German (2006).
data	an optional data frame containing the variables named in formula, S2 and ni.
truemean	true mean values for each area.

Details

Let θ_i denotes interest parameter for each area i, x_i the available area-specific auxiliary data, β the regression coefficients and m the number of small areas. A typical area level model is given by

$$y_i = x_i\beta + u_i + e_i, (i = 1, \ldots, m).$$

Assume that the random effects u_i and the sampling errors e_i are to be independently distributed with the $u_i \, N(0, \sigma^2)$ and the $e_i \, N(0, v_i)$. To foster robustness in small area estimation procedures, student t distribution is used for the random effects. Also, due to the availability of additional data purported to estimate the error variances, this considers modeling of both the means and the variances.

The robust Bayesian small area estimation model is

$$y_i | \theta_i \, N(\theta_i, v_i)$$

$$S_i^2 | v_i \ Gamma((n_i - 1)/2, 1/(2v_i))$$

$$\theta_i | \beta, \sigma^2, df \ t(x_i\beta, \sigma^2, df)$$

, where df is degrees of freedom parameter. For a full Bayesian analysis, this function uses the modified Jeffrey' prior which is the product of the general Jeffrey' prior and $e^{(-a/(2*\sigma^2))}$ where a is chosen as 1:

$$\pi(\beta) \ 1$$

$$\pi(v_i) \ 1/v_i$$

$$\sigma^2 \ Inv - Gamma(p/2, a/2), for a > 0$$

$$\pi(df) \ df^{-1/2}(df + 1)^{p/2-1}(df + 3)^{-p/2-1/2}$$

The estimates of interest parameters are obtained by Rao-Balackwellization with Gibbs sampling with Metropolis-Hastings algorithm.

Value

The function returns a object of class "robustsae" containing the following components:

mean   Rao-Balackwellization estimates of theta's

variance   Rao-Balackwellization estimates of v's

Criteria   a list containing the following comparison criteria : Returns NA if truemean is not provided.

- ASD: average squared deviation, defined as $1/m \sum_{i=1}^{m} (\hat{\theta}_i - \theta_i)^2$
- AAB: average absolute bias, defined as $1/m \sum_{i=1}^{m} |\hat{\theta}_i - \theta_i|$
- ASRB: average squared relative bias, defined as $1/m \sum_{i=1}^{m} ((\hat{\theta}_i - \theta_i)/\theta_i)^2$
- ARB: average relative bias, defined as $1/m \sum_{i=1}^{m} |(\hat{\theta}_i - \theta_i)/\theta_i|$

Author(s)

Malay Ghosh, Jiyoun Myung, Fernando Moura

References

Rao, J. N. K. (2003) *Small Area Estimation*. John Wiley and Sons.

Chip, S., and Green berg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, **49**, 327-335.

Examples

```
# If there is truemean data,
# load data set
data(BZdata)
attach(BZdata)

result <- robustsae(y ~ X1 + X2, S2, ni = BZdata$ni, nsim = 1000, burnin = 500,
                        data = BZdata, truemean = truemean)
result

detach(BZdata)

# If there is no truemean data,
#load data set
data(corndata)
attach(corndata)

result2 <- robustsae(Xi ~ Z1i, Si^2, ni=corndata$ni, data = corndata) # no truemean
result2$mean
result2$variance

detach(corndata)
```

Index