Package ‘Kernelheaping’

April 27, 2015

Type Package
Title Kernel Density Estimation for Heaped Data
Version 0.5
Date 2015-04-24
Depends R (>= 2.15.0), plyr, evmix, MASS, corpcor
Author Marcus Gross
Maintainer Marcus Gross <marcus.gross@fu-berlin.de>
Description In self-reported data the user often encounters heaped data, i.e. data which are rounded to a different degree of coarseness. While this is mostly a minor problem in parametric density estimation the bias can be very large for non-parametric methods such as kernel density estimation. This package implements a partly Bayesian algorithm treating the true unknown values as additional parameters and estimates the rounding parameters to give a corrected kernel density estimate. It supports various standard bandwidth selection methods. Additionally varying rounding probabilities (with the true value) and asymmetric rounding is estimable as well.
License GPL-2 | GPL-3
NeedsCompilation no
Repository CRAN
Date/Publication 2015-04-27 13:42:48

R topics documented:

createSim.Kernelheaping ................................................. 2
dheaping ............................................................... 2
Kernelheaping ........................................................... 4
plot.Kernelheaping ..................................................... 4
Rprx ................................................................. 5
sim.Kernelheaping .................................................... 6
simSummary.Kernelheaping ........................................... 7
students ............................................................... 7
createSim.Kernelheaping

Create heaped data for Simulation

Description

Create heaped data for Simulation

Usage

createSim.Kernelheaping(n, distribution, rounds, thresholds, offset = 0,
downbias = 0.5, Beta = 0, ...)

Arguments

- **n**: sample size
- **distribution**: name of the distribution where random sampling is available, e.g. "norm"
- **rounds**: rounding values
- **thresholds**: rounding thresholds (for Beta=0)
- **offset**: certain value added to all observed random samples
- **downbias**: bias parameter
- **Beta**: acceleration parameter
- **...**: additional attributes handed over to "rdistribution" (i.e. rnorm, rgamma,...)

Value

List of heaped values, true values and input parameters

dheaping

Kernel density estimation for heaped data

Description

Kernel density estimation for heaped data

Usage

dheaping(xheaped, rounds, burnin = 5, samples = 10, setBias = FALSE,
bw = "nrd0", boundary = FALSE, unequal = FALSE, adjust = 1)
**Arguments**

- `xheaped`: heaped values from which to estimate density of x
- `rounds`: rounding values
- `burnin`: burn-in sample size
- `samples`: sampling iteration size
- `setBias`: if TRUE a rounding Bias parameter is estimated. For values above 0.5, the respondents are more prone to round down, while for values < 0.5 they are more likely to round up
- `bw`: bandwidth selector method, defaults to "nrd0" see `density` for more options
- `boundary`: TRUE for positive only data (no positive density for negative values)
- `unequal`: if TRUE a probit model is fitted for the rounding probabilities with log(true value) as regressor
- `adjust`: as in `density`, the user can multiply the bandwidth by a certain factor such that $bw = \text{adjust} \times bw$

**Value**

The function returns a list object with the following objects (besides all input objects):

- `meanPostDensity`: Vector of Mean Posterior Density
- `gridx`: Vector Grid on which density is evaluated
- `resultDensity`: Matrix with Estimated Density for each iteration
- `resultRR`: Matrix with rounding probability threshold values for each iteration (on probit scale)
- `resultBias`: Vector with estimated Bias parameter for each iteration
- `resultBeta`: Vector with estimated Beta parameter for each iteration
- `resultX`: Matrix of true latent values X estimates

**Examples**

```
# Real Data Example
####
# Student learning hours per week
data(students)
xheaped <- as.numeric(na.omit(students$StudyHrs))
## Not run: est <- dheaping(xheaped, rounds=c(1,2,5,10), boundary=TRUE, unequal=TRUE, burnin=20, samples=50)
plot(est)
summary(est)
## End (Not run)
```

```
# Simulate Data
####
Sim1 <- createSim.Kernelheaping(n=500, distribution="norm", rounds=c(1,10,100),
```
In self-reported data the user often encounters heaped data, i.e. data which are rounded to a different degree of coarseness. While this is mostly a minor problem in parametric density estimation the bias can be very large for non-parametric methods such as kernel density estimation. This package implements a partly Bayesian algorithm treating the true, unknown values as additional parameters and estimates the rounding parameters to give a corrected kernel density estimate. It supports various standard bandwidth selection methods. Additionally varying rounding probabilities (with the true value) and asymmetric rounding is estimable as well.

The most important function is `dheaping`. See the help and the attached examples on how to use the package.

Plot Kernel density estimate of heaped data naively and corrected by partly bayesian model.
Usage

```r
## S3 method for class 'Kernelheaping'
plot(x, trueX = NULL, ...)
```

Arguments

- `x`: Kernelheaping object produced by `dheaping` function
- `trueX`: optional, if true values X are known (in simulations, for example) the 'Oracle' density estimate is added as well
- `...`: additional arguments given to standard plot function

Value

plot with Kernel density estimates (Naive, Corrected and True (if provided))

---

Rprx

*Conditional Posterior for \( R_i \) given X, beta, a, tau*

Description

Conditional Posterior for \( R_i \) given X, beta, a, tau

Usage

```r
Rprx(rounds, Bias, RR, beta, gridx, postMatrix)
```

Arguments

- `rounds`: rounding values
- `Bias`: Bias parameter (on probit scale)
- `RR`: Threshold values for rounding parameters
- `beta`: acceleration parameter
- `gridx`: grid on which density is evaluated
- `postMatrix`: empty matrix with nrow=2*rounds and ncol=length(gridx) and posterior probabilities to fill

Value

List with Probabilities
**Simulation of heaping correction method**

**Description**

Simulation of heaping correction method

**Usage**

```r
sim.Kernelheaping(simRuns, n, distribution, rounds, thresholds,
  downbias = 0.5, setBias = FALSE, Beta = 0, unequal = FALSE,
  burnin = 5, samples = 10, bw = "nrd0", offset = 0, boundary = FALSE,
  adjust = 1, ...)```

**Arguments**

- `simRuns`: number of simulations runs
- `n`: sample size
- `distribution`: name of the distribution where random sampling is available, e.g. "norm"
- `rounds`: rounding values
- `thresholds`: rounding thresholds (for Beta=0)
- `downbias`: Bias parameter used in the simulation
- `setBias`: if TRUE a rounding Bias parameter is estimated. For values above 0.5, the respondents are more prone to round down, while for values < 0.5 they are more likely to round up
- `Beta`: Parameter of the probit model for rounding probabilities used in simulation
- `unequal`: if TRUE a probit model is fitted for the rounding probabilities with log(true value) as regressor
- `burnin`: burn-in sample size
- `samples`: sampling iteration size
- `bw`: bandwidth selector method, defaults to "nrd0" see density for more options
- `offset`: location shift parameter used simulation in simulation
- `boundary`: TRUE for positive only data (no positive density for negative values)
- `adjust`: as in density, the user can multiply the bandwidth by a certain factor such that bw=adjust*bw
- `...`: additional attributes handed over to createSim.Kernelheaping

**Value**

List of estimation results
Examples

```r
## not run: S1 <- sim.Kernelheaping(simRuns=2, n=500, distribution="norm",
rounds=c(1,10,100), thresholds=c(0.3,0.4,0.3), sd=100)
## End(Not run)
```

Simulation Summary

**Description**

Simulation Summary

**Usage**

```r
simSummary.Kernelheaping(sim, coverage = 0.9)
```

**Arguments**

- `sim` Simulation object returned from `sim.Kernelheaping`
- `coverage` probability for computing coverage intervals

**Value**

list with summary statistics

---

**students**

*Student0405*

**Description**

Data collected during 2004 and 2005 from students in statistics classes at a large state university in the northeastern United States.

**Author(s)**

Jessica M. Utts, Robert F. Heckard

**Source**

http://mathfaculty.fullerton.edu/mori/Math120/Data/readme
**summary.Kernelheaping**  
*Prints some descriptive statistics (means and quantiles) for the estimated rounding, bias and acceleration (beta) parameters*

**Description**

Prints some descriptive statistics (means and quantiles) for the estimated rounding, bias and acceleration (beta) parameters

**Usage**

```r
## S3 method for class 'Kernelheaping'
summary(object, ...)
```

**Arguments**

- `object`: Kernelheaping object produced by `dheaping` function
- `...`: unused

**Value**

Prints summary statistics

---

**tracePlots**  
*Plots some trace plots for the rounding, bias and acceleration (beta) parameters*

**Description**

Plots some trace plots for the rounding, bias and acceleration (beta) parameters

**Usage**

```r
tracePlots(x, ...)
```

**Arguments**

- `x`: Kernelheaping object produced by `dheaping` function
- `...`: additional arguments given to standard plot function

**Value**

Prints summary statistics
Index

createSim.Kernelheaping, 2

dheaping, 2, 4

Kernelheaping, 4
Kernelheaping-package (Kernelheaping), 4

plot.Kernelheaping, 4

Rprx, 5

sim.Kernelheaping, 6
simSummary.Kernelheaping, 7
students, 7
summary.Kernelheaping, 8

tracePlots, 8