

Package ‘bootImpute’

June 1, 2023

Type Package

Title Bootstrap Inference for Multiple Imputation

Version 1.2.1

Author Jonathan Bartlett

Maintainer Jonathan Bartlett <jonathan.bartlett1@lshtm.ac.uk>

Description Bootstraps and imputes incomplete datasets. Then performs inference on estimates obtained from analysing the imputed datasets as proposed by von Hippel and Bartlett (2021) <[doi:10.1214/20-STS793](https://doi.org/10.1214/20-STS793)>.

Depends R (>= 2.10)

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Imports mice, smcfcs

Suggests testthat

NeedsCompilation no

Repository CRAN

Date/Publication 2023-06-01 10:30:02 UTC

R topics documented:

bootImpute	2
bootImputeAnalyse	3
bootMice	4
bootSmcfcs	5
ex_linquad	6
Index	7

 bootImpute

Bootstrap then impute an incomplete dataset

Description

Bootstraps an incomplete dataset and then imputes each bootstrap a number of times. The resulting list of bootstrapped then imputed datasets can be analysed with [bootImputeAnalyse](#).

Usage

```
bootImpute(
  obsdata,
  impfun,
  nBoot = 200,
  nImp = 2,
  nCores = 1,
  seed = NULL,
  ...
)
```

Arguments

obsdata	The data frame to be imputed.
impfun	A function which when passed an incomplete dataset will return a list of imputed data frames.
nBoot	The number of bootstrap samples to take. It is recommended that you use a minimum of 200. If you specify nCores>1, nBoot must be a multiple of the specified nCores value.
nImp	The number of times to impute each bootstrap sample. Two is recommended.
nCores	The number of CPU cores to use. If specified greater than one, bootImpute will impute using the number of cores specified.
seed	Random number seed.
...	Other parameters that are to be passed through to impfun, which will often include the argument that tells impfun to generate as many imputations as specified by the value passed to nImp.

Details

The impfun must be a function which when passed an incomplete datasets and possibly additional arguments, returns a list of (e.g. 2) imputed datasets. The number of imputed datasets that impfun returns should match the value you specify for the argument nImp. Depending on what your imputation function returns by default, you may need to write a small wrapper function that calls the imputation procedure and returns the list of nImp datasets. See the Example for an illustration with the mice package.

To improve computation times, bootImpute now supports multiple cores through the nCores argument which uses the parallel package.

Value

A list of imputed datasets.

Examples

```
#this example shows how you can use bootImpute to impute using the mice
#package. If you do want to impute using MICE you can instead use the
#bootMice function, which essentially contains the code below
library(mice)

#write a wrapper function to call mice generating M imputations
impM <- function(inputData,M) {
  miceImps <- mice::mice(inputData, m=M)
  imps <- vector("list", M)
  for (i in 1:M) {
    imps[[i]] <- mice::complete(miceImps,i)
  }
  imps
}

#bootstrap twice and impute each twice
#in practice you should bootstrap many more times, e.g. at least 200
#note you have to tell bootImpute how many imputations per bootstrap in
#nImp=2 and also pass through whatever your imp function argument is called
#for specifying number of imputations, which here is M=2.
imps <- bootImpute(ex_linquad, impM, nBoot=2, nImp=2, M=2, seed=564764)
```

bootImputeAnalyse

Analyse bootstrapped and imputed estimates

Description

Applies the user specified analysis function to each imputed dataset contained in `imps`, then calculates estimates, confidence intervals and p-values for each parameter, as proposed by von Hippel and Bartlett (2021).

Usage

```
bootImputeAnalyse(imps, analysisfun, nCores = 1, quiet = FALSE, ...)
```

Arguments

<code>imps</code>	The list of imputed datasets returned by <code>bootImpute</code>
<code>analysisfun</code>	A function which when applied to a single dataset returns the estimate of the parameter(s) of interest. The dataset to be analysed is passed to <code>analysisfun</code> as its first argument.
<code>nCores</code>	The number of CPU cores to use. If specified greater than one, <code>bootImputeAnalyse</code> will impute using the number of cores specified. The number of bootstrap samples in <code>imps</code> should be divisible by <code>nCores</code> .

quiet Specify whether to print a table of estimates, confidence intervals and p-values.
 ... Other parameters that are to be passed through to analysisfun.

Details

Multiple cores can be used by using the `nCores` argument, which may be useful for reducing computation times.

Value

A vector containing the point estimate(s), variance estimates, and degrees of freedom.

References

von Hippel PT, Bartlett JW. Maximum likelihood multiple imputation: faster, more efficient imputation without posterior draws. *Statistical Science*, 2021, 36(3):400-420. doi:10.1214/20STS793

Examples

```
library(mice)

set.seed(564764)

#bootstrap twice and impute each twice
#in practice you should bootstrap many more times, e.g. at least 200
imps <- bootMice(ex_linquad, nBoot=2, nImp=2)

#analyse estimates
#write a wrapper to analyse an imputed dataset
analyseImp <- function(inputData) {
  coef(lm(y~z+x+xsq,data=inputData))
}
ests <- bootImputeAnalyse(imps, analyseImp)
```

bootMice

Bootstrap then impute using mice

Description

Bootstraps an incomplete dataset and then imputes each bootstrap a number of times using the mice package. The resulting list of bootstrapped then imputed datasets can be analysed with [bootImputeAnalyse](#). To run this function requires the mice package to be installed.

Usage

```
bootMice(obsdata, nBoot = 200, nImp = 2, nCores = 1, seed = NULL, ...)
```

Arguments

obsdata	The data frame to be imputed.
nBoot	The number of bootstrap samples to take. It is recommended that you use a minimum of 200. If you specify nCores>1, nBoot must be a multiple of the specified nCores value.
nImp	The number of times to impute each bootstrap sample. Two is recommended.
nCores	The number of CPU cores to use. If specified greater than one, bootImpute will impute using the number of cores specified.
seed	Random number seed.
...	Other arguments that are to be passed to mice.

Value

A list of imputed datasets.

Examples

```
library(mice)

head(ex_linquad)

#bootstrap 10 times and impute each twice
imps <- bootMice(ex_linquad, nBoot=10, nImp=2, seed=564764)
```

 bootSmcfcs

Bootstrap then impute using smcfcs

Description

Bootstraps an incomplete dataset and then imputes each bootstrap a number of times using the smcfcs package. The resulting list of bootstrapped then imputed datasets can be analysed with [bootImputeAnalyse](#). To run this function requires the smcfcs package to be installed.

Usage

```
bootSmcfcs(obsdata, nBoot = 200, nImp = 2, nCores = 1, seed = NULL, ...)
```

Arguments

obsdata	The data frame to be imputed.
nBoot	The number of bootstrap samples to take. It is recommended that you use a minimum of 200. If you specify nCores>1, nBoot must be a multiple of the specified nCores value.
nImp	The number of times to impute each bootstrap sample. Two is recommended.

nCores	The number of CPU cores to use. If specified greater than one, bootImpute will impute using the number of cores specified.
seed	Random number seed.
...	Other arguments that are to be passed to smcfcs.

Value

A list of imputed datasets.

Examples

```
library(smcfcs)

head(ex_linquad)
#bootstrap twice and impute each twice
#in practice you should bootstrap many more times, e.g. at least 200
imps <- bootSmcfcs(ex_linquad, nBoot=2, nImp=2,
                  smtype="lm", smformula="y~z+x+xsq",
                  method=c("","","norm","x^2",""), seed=564764)
```

ex_linquad	<i>Simulated example data with continuous outcome and quadratic covariate effects</i>
------------	---

Description

A dataset containing simulated data where the outcome depends quadratically on a partially observed covariate.

Usage

```
ex_linquad
```

Format

A data frame with 1000 rows and 5 variables:

y Continuous outcome

z Fully observed covariate, with linear effect on outcome

x Partially observed normally distributed covariate, with quadratic effect on outcome

xsq The square of x, which thus has missing values also

v An auxiliary variable (i.e. not contained in the substantive model)

Index

* datasets

ex_linquad, 6

bootImpute, 2, 3

bootImputeAnalyse, 2, 3, 4, 5

bootMice, 4

bootSmcfcs, 5

ex_linquad, 6