Package 'credsubs'

June 15, 2020

Title Credible Subsets

Version 1.1.0

Date 2020-06-15

Author Patrick Schnell,
Brad Carlin

Maintainer Patrick Schnell <schnell.31@osu.edu>

Description Functions for constructing simultaneous credible bands and identifying subsets via the ``credible subsets'' (also called ``credible subgroups'') method.

Suggests ff, R.rsp, shiny

License GPL-3

RoxygenNote 7.1.0

VignetteBuilder R.rsp

LazyData true

NeedsCompilation no

Repository CRAN

Date/Publication 2020-06-15 21:00:02 UTC

R topics documented:

| alzheimers | *Clinical trial data from 369 Alzheimer's disease patients.* |

Description

A dataset containing the treatment assignment, baseline measurements, and outcomes of 369 patients enrolled in four clinical trials for treatments of Alzheimer's disease. Patients assigned to the test treatments are excluded, leaving patients assigned to either a placebo or the palliative standard of care.

Usage

```
alzheimers
```

Format

A data frame with 369 rows and 6 variables:

Treatment treatment assignment

Severity baseline severity (ADAS-Cog 11)

Decline decline in score on Mini Mental State Exam (MMSE) divided by number of years since onset of symptoms

Sex male or female

Carrier carrier status of ApoE4 allele

Improvement baseline severity minus severity at 12 weeks (positive is good)

Source

AbbVie, Inc.

| build.shiny.calc | *Build a credible subset calculator* |

Description

This function builds a `shiny` application in the specified directory that gives the maximum credible level at an entered covariate point.

Usage

```
build.shiny.calc(
  credsubs.level,
  cov.space,
  name = "calc",
  dir = ".",
  title = "Credible Subsets Calculator",
  instructions = "Select a covariate point."
)
```

Arguments

`credsubs.level`	An object of class `credsubs.level`.
`cov.space`	A data frame whose rows are human-readable covariate points corresponding to the entries of `credsubs.level$level`.
`name`	A character string indicating the name of the application.
`dir`	The directory in which to place the application.
`title`	A character string to be displayed as the application title.
`instructions`	A character string to be displayed as instructions. HTML allowed.

Details

The calculator creates a subdirectory according to `name` in the directory specified by `dir`, and places in it files `server.R`, `ui.R`, and `config.RData`. This application requires the `shiny` package to run, and can be executed by passing the directory path to `run.shiny.calc()`. The produced application directory may be moved from its original location.

---

credsubs	*Constructs a credible subset pair*

Description

`credsubs` returns a credible subset pair over a finite set of covariate points given either a sample from the posterior of the regression surface or a function `FUN(x,params)` and a sample from the posterior of the parameters.

Usage

```
credsubs(
  params,
  design = NULL,
  FUN = function(x, params) {      params %*% t(x) },
  cred.level = 0.95,
  threshold = 0,
  method = c("asymptotic", "quantile"),
```

```
    step.down = TRUE,
    sides = c("both", "exclusive", "inclusive"),
    est.FUN = mean,
    var.FUN = sd,
    point.estimate = NULL,
    track = numeric(0),
    verbose = FALSE
)
```

Arguments

params	A numeric matrix whose rows are draws from the posterior distribution of either the regression surface or the parameter vector.
design	(Optional) A numeric matrix whose rows are covariate points over which the band is to be constructed.
FUN	(Optional) a function of the form function(x,params) that takes a row of design and the entire params matrix and returns a vector of the same length of x representing the regression surface.
cred.level	Numeric; the credible level.
threshold	Numeric; the value of FUN above which a covariate is included in the target subset.
method	Either "asymptotic" (default) or "quantile"; see details.
step.down	Logical (default TRUE); should the step-down procedure be used?
sides	One of "both" (default), "exclusive", or "inclusive". Which bounds should be constructed?
est.FUN	The function used to produce estimates of the regression surface. Default is mean.
var.FUN	The function used to quantify the variability of the regression surface posterior. Default is sd.
point.estimate	If not null, replaces the mean and sets the reference around which the standard error is computed. Useful for bootstrapping methods. Treated as a row of the params matrix.
track	A numeric vector of indices indicating which rows (default none) of the design matrix should have the sample of the corresponding FUN(x,params) returned.
verbose	Logical (default FALSE); print progress?

Details

If design is NULL (default), it is taken to be the identity matrix of dimension ncol(params), so that the rows of params are treated as draws from the posterior FUN(x, params).

The 'asymptotic' method assumes that the marginal posteriors of the FUN(x, params) are asymptotically normal and is usually significantly faster and less memory-intensive than the 'quantile' method, which makes no such assumption.

Value

An object of class credsubs, which contains:

exclusive A logical vector indicating membership in the exclusive credible subset.

inclusive A logical vector indicating membership in the inclusive credible subset.

cred.level As provided.

threshold As provided.

method As provided.

step.down As provided.

sides As provided.

est Posterior estimate of the regression surface.

est.FUN As provided.

var Summary of posterior variability of the regression surface.

var.FUN As provided.

W An estimate of the extremal errors.

W.crit The critical quantile of W.

trace The posterior samples of the regression surface indicated by the track argument.

Examples

```
### Sample from regression surface posterior
reg.surf.sample <- matrix(rnorm(1000, mean=1:10), ncol=2, byrow=TRUE)
credsubs(reg.surf.sample, cred.level=0.80)

### Parametric case
design <- cbind(1, 1:10)
params <- matrix(rnorm(200, mean=1:2), ncol=2, byrow=TRUE)
credsubs(params, design)

### With custom function
params.sd <- cbind(1 / rgamma(100, 1), params)
FUN.sd <- function(x, params) { params[, -1] %*% t(x) / params[, 1] }
credsubs(params.sd, design, FUN.sd, threshold=1)
```

---

credsubs.level	*Compute the maximum credible levels at which conclusions may be drawn*

Description

For each covariate point, credsubs.level computes the maximum credible level at which a conclusion may be drawn at each point, and what that conclusion is.

Usage

```
credsubs.level(
  params,
  design = NULL,
  FUN = function(x, params) {      params %*% t(x) },
  threshold = 0,
  method = c("asymptotic", "quantile"),
  step.down = TRUE,
  sides = c("both", "exclusive", "inclusive"),
  est.FUN = mean,
  var.FUN = sd,
  point.estimate = NULL,
  track = numeric(0),
  verbose = FALSE,
  z.store = c("ram", "recompute", "disk")
)
```

Arguments

params	A numeric matrix whose rows are draws from the posterior distribution of either the regression surface or the parameter vector.
design	(Optional) A numeric matrix whose rows are covariate points over which the band is to be constructed.
FUN	(Optional) a function of the form function(x,params) that takes a row of design and the entire params matrix and returns a vector of the same length of x representing the regression surface.
threshold	Numeric; the value of FUN above which a covariate is included in the target subset.
method	Either "asymptotic" (default) or "quantile"; see details.
step.down	Logical (default TRUE); should the step-down procedure be used?
sides	One of "both" (default), "exclusive", or "inclusive". Which bounds should be constructed?
est.FUN	The function used to produce estimates of the regression surface. Default is mean.
var.FUN	The function used to quantify the variability of the regression surface posterior. Default is sd.
point.estimate	If not null, replaces the mean and sets the reference around which the standard error is computed. Useful for bootstrapping methods. Treated as a row of the params matrix.
track	A numeric vector of indices indicating which rows (default none) of the design matrix should have the sample of the corresponding FUN(x,params) returned.
verbose	Logical (default FALSE); print progress?
z.store	How should certain intermediate computations be handled? See details.

Details

If design is NULL (default), it is taken to be the identity matrix of dimension ncol(params), so that the rows of params are treated as draws from the posterior FUN(x, params).

The 'asymptotic' method assumes that the marginal posteriors of the FUN(x, params) are asymptotically normal and is usually significantly faster and less memory-intensive than the 'quantile' method, which makes no such assumption.

By default (z.store = "ram"), the maximum credible level computation stores a potentially very large amount of intermediate computation results in memory. If not enough memory is available, z.store = "disk" uses the ff package to store the intermediate results on disk, which can still be fairly quick if the storage is fast (e.g. a local SSD). Alternatively, z.store = "recompute" discards the intermediate results and recomputes whenever needed. This uses minimal memory, but is usually the slowest option.

Value

An object of class credsubs.level, which contains:

level A numeric vector indicating the maximum credible level at which a conclusion may be drawn at each covariate point.

sign A numeric vector indicating the which credible subsets of which each covariate point is a member at the credible level indicated by level. Exclusive and inclusive: 1, inclusive only: 0, neither: -1.

threshold As provided.

method As provided.

step.down As provided.

sides As provided.

est Posterior estimate of the regression surface.

est.FUN As provided.

var Summary of posterior variability of the regression surface.

var.FUN As provided.

trace The posterior samples of the regression surface indicated by the track argument.

Examples

```
### Sample from regression surface posterior
reg.surf.sample <- matrix(rnorm(1000, mean=1:10), ncol=2, byrow=TRUE)
credsubs.level(reg.surf.sample)

### Parametric case
design <- cbind(1, 1:10)
params <- matrix(rnorm(200, mean=1:2), ncol=2, byrow=TRUE)
credsubs(params, design)

### With custom function
params.sd <- cbind(1 / rgamma(100, 1), params)
FUN.sd <- function(x, params) { params[, -1] %*% t(x) / params[, 1] }
```

```
credsubs(params.sd, design, FUN.sd, threshold=1)
```

---

run.shiny.calc   *Run a calculator*

Description

Runs the specified calculator using the shiny package. Calculators must be built using build.shiny.calc.

Usage

```
run.shiny.calc(
  app.dir = system.file("shiny", "alzheimers", package = "credsubs")
)
```

Arguments

app.dir   A character string pointing to the application directory.

Details

The app.dir argument need not exactly match the value recommended by build.shiny.calc(), as long as it points to the correct directory. For example, "./calc/", "calc/", and "calc" are all equivalent. If no value is supplied, an example is run.

---

sim.cred.band   *Constructs a simultaneous credible band*

Description

sim.cred.band returns a simultaneous band over a finite set of covariate points given either a sample from the posterior of the regression surface or a function FUN(x,params) and a sample from the posterior of the parameters.

Usage

```
sim.cred.band(
  params,
  design = NULL,
  FUN = function(x, params) {      params %*% t(x) },
  cred.level = 0.95,
  method = c("asymptotic", "quantile"),
  sides = c("both", "upper", "lower"),
  est.FUN = mean,
  var.FUN = sd,
```

```
    point.estimate = NULL,
    track = numeric(0),
    verbose = FALSE
)
```

Arguments

`params`	A numeric matrix whose rows are draws from the posterior distribution of either the regression surface or the parameter vector.
`design`	(Optional) A numeric matrix whose rows are covariate points over which the band is to be constructed.
`FUN`	(Optional) a function of the form `function(x,params)` that takes a row of `design` and the entire `params` matrix and returns a vector of the same length of x representing the regression surface.
`cred.level`	Numeric; the credible level.
`method`	Either "asymptotic" (default) or "quantile"; see details.
`sides`	One of "both" (default), "upper", or "lower". Which bounds should be constructed?
`est.FUN`	The function used to produce estimates of the regression surface. Default is `mean`.
`var.FUN`	The function used to quantify the variability of the regression surface posterior. Default is `sd`.
`point.estimate`	If not null, replaces the mean and sets the reference around which the standard error is computed. Useful for bootstrapping methods. Treated as a row of the `params` matrix.
`track`	A numeric vector of indices indicating which rows (default none) of the design matrix should have the sample of the corresponding FUN(x,params) returned.
`verbose`	Logical (default `FALSE`); print progress?

Details

If `design` is `NULL` (default), it is taken to be the identity matrix of dimension `ncol(params)`, so that the rows of params are treated as draws from the posterior FUN(x,params).

The 'asymptotic' method assumes that the marginal posteriors of the FUN(x,params) are asymptotically normal and is usually significantly faster and less memory-intensive than the 'quantile' method, which makes no such assumption.

Value

An object of class `sim.cred.band`, which contains:

upper A numeric vector of upper bounds.

lower A numeric vector of lower bounds.

cred.level As provided.

method As provided.

sides As provided.

est Posterior estimate of the regression surface.

est.FUN As provided.

var Summary of posterior variability of the regression surface.

var.FUN As provided.

W An estimate of the extremal errors.

W.crit The critical quantile of W.

trace The posterior samples of the regression surface indicated by the track argument.

Examples

```
### Sample from regression surface posterior
reg.surf.sample <- matrix(rnorm(1000, mean=1:10), ncol=2, byrow=TRUE)
sim.cred.band(reg.surf.sample, cred.level=0.80)

### Parametric case
design <- cbind(1, 1:10)
params <- matrix(rnorm(200, mean=1:2), ncol=2, byrow=TRUE)
sim.cred.band(params, design)

### With custom function
params.sd <- cbind(1 / rgamma(100, 1), params)
FUN.sd <- function(x, params) { params[, -1] %*% t(x) / params[, 1] }
sim.cred.band(params.sd, design, FUN.sd)
```

Index