Package 'HDCurves'

August 18, 2022

Type Package

Title Hierarchical Derivative Curve Estimation

Version 0.1.1

Description

A procedure that fits derivative curves based on a sequence of quotient differences. In a hierarchical setting the package produces estimates of subject-specific and group-specific derivative curves. In a non-hierarchical setting the package produces a single derivative curve.

License GPL

Depends R (>= 3.5.0)

NeedsCompilation yes

Author Garritt L. Page [aut, cre],
Maria Xose Rodriguez-Alvarez [aut],
Dae-Jin Lee [ctb],
S. McKay Curtis [ctb],
Radford M. Neal Neal [ctb]

Maintainer Garritt L. Page <page@stat.byu.edu>

Repository CRAN

Date/Publication 2022-08-18 15:00:02 UTC

R topics documented:

---

emp_deriv	*Estimate of empirical derivative based on sequence of quotient differences.*

Description

`emp_deriv` Provides an estimate of an empirical derivative curve using one of the two sequence of quotient differences detailed in Page, Rodriguez-Alvarez, Lee (2020).

Usage

```
emp_deriv(y, t, D, u, ztype = 0)
```

Arguments

y	numeric vector (response variable).
t	numeric vector (time variable).
D	integer indicating the bin width when estimating empirical derivatives
u	real indicating the bandwidth with regards to smoothing when estimating empirical derivatives.
ztype	integer indicating which sequence of empirical derivatives will be employed. (0 indicates that which is parameterized by D and u while 1 indicates that which is parametrized by u)

Value

The vector which as the same length as the input variable "y" that contains empirical derivative estimates for each entry of "y".

References

Page, G. L.; Rodriguez-Alvarez, M. X.; Lee, DJ; (2020) "Bayesian Hierarchical Modeling of Growth Curve Derivatives via Sequences of Quotient Differences" to appear

Examples

```
# Our R-package
library(HDCurves)

data(growth)

y <- growth$hgtm[,1]
t <- growth$age

fprime <- emp_deriv(y=y, t=t, D=10, u=1, ztype=0)
```

```
plot(t,y, type="b")
plot(t, fprime, type="b")
```

---

growth *Berkeley Growth Study data.*

Description

A list containing the heights of 39 boys and 54 girls from age 1 to 18 and the ages at which they were collected.

Format

This list contains the following components:

hgtm a 31 by 39 numeric matrix giving the heights in centimeters of 39 boys at 31 ages.

hgtf a 31 by 54 numeric matrix giving the heights in centimeters of 54 girls at 31 ages.

age a numeric vector of length 31 giving the ages at which the heights were measured.

Details

These data are included here for convenience as they are also available in the "fda" R-package.

Source

Ramsay, James O., and Silverman, Bernard W. (2006,ISBN 978-0-387-22751-1), *Functional Data Analysis, 2nd ed.*, Springer, New York.

Ramsay, James O., and Silverman, Bernard W. (2002, ISBN 978-0-387-22465-7), *Applied Functional Data Analysis*, Springer, New York, ch. 6.

Tuddenham, R. D., and Snyder, M. M. (1954) "Physical growth of California boys and girls from birth to age 18", *University of California Publications in Child Development*, 1, 183-364.

Ramsay, James O., Wickham, Hadley, Graves, Spencer, and Hooker, Giles (2018)<https://CRAN.R-project.org/package=fda>, *fda: Functional Data Analysis*,R package version 2.4.8,

---

| HDCurves | *R wrapper that accesses C code to fit hierarchical model for derivative curves based on sequence of quotient differences* |

Description

HDCurves is the main function used to fit the the hierarchical model for derivative curves. This function also fits single derivative curves. Derivative curves are estimated based on a sequence of empirical derivatives.

Usage

```
HDCurves(y, t, trt, ids, balanced = TRUE, tpred = NULL, ztype = 0,
  p = 0.5, au = 1, bu = 1, A = 2, as = 1, bs = 1, atau = 1,
  btau = 1/0.005, ndx = 10, q = 3, verbose = FALSE, draws = 1100,
  burn = 100, thin = 1)
```

Arguments

y	numeric vector (response variable). Must be in long format.
t	numeric vector (time variable). Must be in long format.
trt	numeric vector indicating to which treatment each subject belongs. The length of this vector is the same as "y" and "t" and must begin with 1. The remaining treatment labels must form a contiguous sequence of integers. See example below.
ids	numeric vector containing subject id number. The length of this vector is the same as "y" and "t". Subject id numbers must begin with 1 and subsequent subject id numbers need to remain a contiguous sequence of integers. See example below.
balanced	logical with TRUE corresponding to a balanced case, and FALSE to an unbalanced case
tpred	numeric (time points for prediction). If not supplied, the predictions at each observed time point are returned for subject-specific derivative curves. For treatment-specific derivative curves a sequence of time points is produced based on the mininum and maximum observed time points.
ztype	integer indicating which sequence of empirical derivatives will be employed. (0 indicates that which is parameterized by D and u while 1 indicates that which is parametrized by u)
p	prior parameter for D, default is 0.5. This value is not used if ztype=1.
au	shape prior parameter for u, default is 1
bu	scale prior parameter for u, default is 1
A	parameter that defines upper bound for lambda. Default is 2. This parameter is not used in the case that a single derivative curve is being estimated.

as	shape prior parameter associated with subject-specific variance (sigma2), default is 1
bs	rate prior parameter associated with subject-specific variance (sigma2), default is 1
atau	shape prior parameter associated with penalty parameter (tau2), default is 1
btau	rate prior parameter associated with penalty parameter (tau2), default is 1/0.005
ndx	number of segments (default is 10)
q	degree of the B-spline (default is 3)
verbose	If TRUE, then details associated with data being fit are printed to screen along with MCMC iterate counter
draws	number of MCMC iterates to be collected. default is 1100
burn	number of MCMC iterates discared as burn-in. default is 100
thin	number by which the MCMC chain is thinne. default is 1

Value

A list containing arrays filled with MCMC iterates corresponding to model parameters, sequence of empirical derivatives, and derivative curves. In order to provide more detail, in what follows let

"T" - be the number of MCMC iterates collected,

"M" - be the number of subjects,

"N" - be the total number of measured responses,

"K" - be the maximum number of observations measured across the M subjects,

"V" - the number of prediction points in the user supplied vector tpred,

"P" - the number of P-spline basis coefficients, and

"G" - the number of treatment groups.

The output list contains the following

z	an array of dimension (K, M, T) containing MCMC iterates associated with sequence of empirical derivatives for each subject. In the case of an unbalanced design, for subjects with less than K observations the values in the array beyond the number of measured responses for the subject are of no use.
beta	an array of dimension (M, P, T) containing MCMC iterates assocated with subject-specific B-spline basis coefficients.
tpred	vector of size V containing the prediction points supplied in the "tpred" argument of the function.
fprime	an array of dimension (V, M, T) if "tpred" is supplied or (K, M, T) if not that contains MCMC iterates associated with derivative curves for each subject. In the case of an unbalanced design and "tpred" not being supplied, for subjects with less than K observations the values in the array beyond the number of measured responses for the subject are of no use.
u	a matrix of dimension (T, M) containing MCMC interates associated with "u".

d	a matrix of dimension (T, M) containing MCMC interates associated with "d". If ztype=1, then this array is of no use.
sig2	a matrix of dimension (T, M) containing MCMC interates associated with "sigma2".
theta	an array of dimension (G, P, T) containing MCMC iterates associated with group-specific B-spline basis coefficients ("theta"). In the case of a single subject, this array is of no use.
fgprime	an array of dimension (V, M, T) if "tpred" is supplied or (K, M, T) if not that contains MCMC iterates associated with group-specific derivative curves. In the case of a single subject, this array is of no use.
lam	a matrix of dimension (T, G) containing MCMC iterates associated with "lambda" the upper bound of "sigma2". In the case of a single subject, this array is of no use.
tau2	a matrix of dimesino (T, G) containing MCMC iterates associated with "tau2" the penalty parameters the the P-spline specification.
HmatBySubject	an array that contains the basis design matrix for each subject. If the study is balanced, then this is a single matrix of dimensino (M, P). If the design is not balanced, this is a one-dimensional array of size P x (the total number of observations).
Htheta	a matrix of dimension (M, P) that contains the basis design matrix used for each treatment group.

References

Page, G. L.; Rodriguez-Alvarez, M. X.; Lee, DJ; (2020) "Bayesian Hierarchical Modeling of Growth Curve Derivatives via Sequences of Quotient Differences" to appear

Examples

```
# Our R-package
library(HDCurves)

data(growth)

nmale <- ncol(growth$hgtm)
nfemale <- ncol(growth$hgtf)
ntime <- nrow(growth$hgtf)

dat <- data.frame(age=rep(growth$age, nmale + nfemale),
      height = c(growth$hgtm, growth$hgtf),
      tmt = c(rep(1, nmale*ntime),
               rep(2, nfemale*ntime)),
      case = rep(1:(nmale+nfemale), each=ntime))


############
# MCMC Specs
###########
ni <- 20000
```

```
nb <- 10000
nt <- 10
nout <- (ni - nb)/nt

# B-spline details
ndx <- 40   # number of segments (inner knots)
bdeg <- 3  # Degree of the B-spline
tt <- sort(unique(dat$age))


# This takes about 5 minutes to run
date();
fit <- HDCurves(y = dat$height, t = dat$age, tpred = tt, trt = dat$tmt,
                ids = dat$case, ztype = 0,  p = 0.9, A = 1,  au = 1, bu = 1,
                as = 1, bs = 1, atau = 1, btau = 1/0.5, ndx = ndx, q = bdeg,
                balanced = TRUE,draws=ni, burn=nb, thin=nt)
date();

# subject-specific derivative curve means
fpmn <- apply(fit$fprime,c(1,2),mean)


# group-specific derivative curve means
gfpmn <- apply(fit$fgprime,c(1,2),mean)
gfpci <- apply(fit$fgprime,c(1,2),
      function(x) quantile(x, c(0.0275, 0.975)))

plot(tt,gfpmn[,1], type="n", ylim=range(c(gfpmn)),
    ylab="Height  Rate of Change", xlab="age",
    xaxp=c(2,18,n=8),cex.axis=1.45, cex.lab=1.45)
lines(tt, gfpmn[,1], type='o', col='blue', pch=20)
lines(tt, gfpci[1,,1],  col='blue', pch=20,lty=2)
lines(tt, gfpci[2,,1],  col='blue', pch=20,lty=2)

lines(tt, gfpmn[,2], type='o', col='red', pch=20)
lines(tt, gfpci[1,,2],  col='red', pch=20,lty=2)
lines(tt, gfpci[2,,2],  col='red', pch=20,lty=2)
legend(x="topright", lty=1, pch=20, cex=1.45,
      legend=c("Girls","Boys"), col=c("red","blue"))
```

Index