Package ‘gss’

May 4, 2018

Version 2.1-9
Date 2018-04-22
Title General Smoothing Splines
Author Chong Gu <chong@purdue.edu>
Maintainer Chong Gu <chong@purdue.edu>
Depends R (>= 2.14.0), stats
Description A comprehensive package for structural multivariate function estimation using smoothing splines.
License GPL (>= 2)
NeedsCompilation yes
Repository CRAN
Date/Publication 2018-05-03 23:07:51 UTC

R topics documented:

aids ................................................................. 2
bacteriuria ......................................................... 3
buffalo .............................................................. 3
cdsscden ........................................................... 4
cdssden ............................................................. 5
clim ................................................................. 6
ColoCan ............................................................. 6
dsscaden ............................................................ 8
dssden .............................................................. 9
esc ................................................................. 10
eyetrack ............................................................ 10
fitted.ssanova .................................................... 11
gastric ............................................................. 12
gauss.quad ........................................................ 12
gssanova .......................................................... 13
gssanova0 ......................................................... 16
hzdrate.sshzd ....................................................... 19
AIDS Incubation

A data set collected by Centers for Disease Control and Prevention concerning AIDS patients who were infected with the HIV virus through blood transfusion.

Usage

data(aids)

Format

A data frame containing 295 observations on the following variables.

- `incu` Time from HIV infection to AIDS diagnosis.
- `infe` Time from HIV infection to end of data collection (July 1986).
- `age` Age at time of blood transfusion.
**Source**


---

**bacteriuria**

*Treatment of Bacteriuria*

---

**Description**

Bacteriuria patients were randomly assigned to two treatment groups. Weekly binary indicator of bacteriuria was recorded for every patient over 4 to 16 weeks. A total of 72 patients were represented in the data, with 36 each in the two treatment groups.

**Usage**

data(bacteriuria)

**Format**

A data frame containing 820 observations on the following variables.

- **id**: Identification of patients, a factor.
- **trt**: Treatments 1 or 2, a factor.
- **time**: Weeks after randomization.
- **infect**: Binary indicator of bacteriuria (bacteria in urine).

**Source**


**References**


---

**buffalo**

*Buffalo Annual Snowfall*

---

**Description**


**Usage**

data(buffalo)
cdsscden

Format

A vector of 63 numerical values.

Source


cdsscden

Evaluating Conditional PDF, CDF, and Quantiles of Smoothing Spline Conditional Density Estimates

Description

Evaluate conditional pdf, cdf, and quantiles of \( f(y_1|x,y_2) \) for smoothing spline conditional density estimates \( f(y|x) \).

Usage

\[
\text{cdsscden}(\text{object, } y, x, \text{cond, int=NULL)} \\
\text{cpsscden}(\text{object, } q, x, \text{cond)} \\
\text{cqsscden}(\text{object, } p, x, \text{cond)}
\]

Arguments

- **object**: Object of class "sccd1" or "sccd11".
- **x**: Data frame of x values on which conditional density \( f(y_1|x,y_2) \) is to be evaluated.
- **y**: Data frame or vector of y1 points on which conditional density \( f(y_1|x,y_2) \) is to be evaluated.
- **cond**: One row data frame of conditioning variables y2.
- **q**: Vector of points on which cdf is to be evaluated.
- **p**: Vector of probabilities for which quantiles are to be calculated.
- **int**: Vector of normalizing constants.

Details

The arguments x and y are of the same form as the argument newdata in \texttt{predict.lm}, but y in \texttt{cdsscden} can take a vector for 1-D y1.

\texttt{cpsscden} and \texttt{cqsscden} naturally only work for 1-D y1.
Value

cdssden returns a list object with the following components.

pdf Matrix or vector of conditional pdf \( f(y_1|x,y_2) \), with each column corresponding to a distinct x value.

int Vector of normalizing constants.

cpssden and cqssden return a matrix or vector of conditional cdf or quantiles of \( f(y_1|x,y_2) \).

Note

If variables other than factors or numerical vectors are involved in \( y_1 \), the normalizing constants can not be computed.

See Also

Fitting function sscden and dsscden.

Description

Evaluate conditional pdf, cdf, and quantiles for smoothing spline density estimates.

Usage

cdssden(object, x, cond, int=NULL)
cpssden(object, q, cond)
cqssden(object, p, cond)

Arguments

object Object of class "ssden".

x Data frame or vector of points on which conditional density is to be evaluated.

cond One row data frame of conditioning variables.

int Normalizing constant.

q Vector of points on which conditional cdf is to be evaluated.

p Vector of probabilities for which conditional quantiles are to be calculated.

Details

The argument x in cdssden is of the same form as the argument newdata in predict.lm, but can take a vector for 1-D conditional densities.

cpssden and cqssden naturally only work for 1-D conditional densities of a numerical variable.
Value

cdssden returns a list object with the following components.

pdf Vector of conditional pdf.
int Normalizing constant.

cpssden and cqssden return a vector of conditional cdf or quantiles.

Note

If variables other than factors or numerical vectors are involved in x, the normalizing constant can not be computed.

See Also

Fitting function ssden and dssden.

clim

Average Temperatures During December 1980 Through February 1981

Description

Average temperatures at 690 weather stations during December 1980 through February 1981.

Usage

data(clim)

Format

A data frame containing 690 observations on the following variables.


temp Average temperature, in Celsius.
geog Geographic location (latitude,longitude), in degrees, as a matrix.

Source

This is reformulated from the data frame climate in the R package assist by Yuedong Wang and Chunlei Ke.

ColoCan

Colorectal Cancer Mortality Rate in Indiana Counties

Description

County-wise death counts of colorectal cancer patients in Indiana during years 2000 through 2004.
Usage
data(ColoCan)

Format
A data frame containing 184 observations on the following variables.

- **event**: Death counts.
- **pop**: Population from Census 2000.
- **sex**: Gender of population.
- **wrt**: Proportion of Whites.
- **brt**: Proportion of Blacks.
- **ort**: Proportion of other minorities.
- **lat**: Latitude.
- **lon**: Longitude.
- **geog**: Geographic location, derived from **lat** and **lon**.
- **scrn**: Colorectal cancer screening rate.
- **name**: County name.

Details
geog was generated from **lat** and **lon** using the code given in the example section.

Source
Dr. Tonglin Zhang.

References

Examples
```r
# Converting latitude and longitude to x-y coordinates
# The 49th county is Marion, where Indianapolis is located.
# Not run: ltln2xy <- function(latlon, latlon0) {
# lat <- latlon[,1]*pi/180; lon <- latlon[,2]*pi/180
# lat0 <- latlon0[1]*pi/180; lon0 <- latlon0[2]*pi/180
# x <- cos(lat0)*sin(lon-lon0); y <- sin(lat-lat0)
# cbind(x,y)
# }
data(ColoCan)
latlon <- as.matrix(ColoCan[,c("lat","lon")])
ltn2xy(latlon, latlon[49,])
# Clean up
rmltln2xy,ColoCan, latlon)
# End(Not run)
```
Evaluating PDF, CDF, and Quantiles of Smoothing Spline Conditional Density Estimates

Description
Evaluate pdf, cdf, and quantiles for smoothing spline conditional density estimates.

Usage

\[
\begin{align*}
dsscden & (\text{object, } y, \ x) \\
psscden & (\text{object, } q, \ x) \\
qsscden & (\text{object, } p, \ x) \\
d.sscden & (\text{object, } x, \ y) \\
d.sscden1 & (\text{object, } x, \ y, \ \text{scale=TRUE})
\end{align*}
\]

Arguments

- **object**: Object of class "sscden" or "sscden1".
- **x**: Data frame of x values on which conditional density \(f(y|x)\) is to be evaluated.
- **y**: Data frame or vector of points on which conditional density \(f(y|x)\) is to be evaluated.
- **q**: Vector of points on which cdf is to be evaluated.
- **p**: Vector of probabilities for which quantiles are to be calculated.
- **scale**: Flag indicating whether to use approximate scaling without quadrature.

Details
The arguments \(x\) and \(y\) are of the same form as the argument \texttt{newdata} in \texttt{predict.lm}, but \(y\) in \texttt{dsscden} can take a vector for 1-D responses.

\texttt{psscden} and \texttt{qsscden} naturally only work for 1-D responses.

Value
A matrix or vector of pdf, cdf, or quantiles of \(f(y|x)\), with each column corresponding to a distinct \(x\) value.

See Also
Fitting function \texttt{sscden} and \texttt{cdsscden}.  

\[
\]
### Description

Evaluate pdf, cdf, and quantiles for smoothing spline density estimates.

### Usage

```r
  dssden(object, x)
  pssden(object, q)
  qssden(object, p)
  d.ssdn(object, x)
  d.ssdnl(object, x)
```

### Arguments

- `object`: Object of class "ssden".
- `x`: Data frame or vector of points on which density is to be evaluated.
- `q`: Vector of points on which cdf is to be evaluated.
- `p`: Vector of probabilities for which quantiles are to be calculated.

### Details

The argument `x` in `dssden` is of the same form as the argument `newdata` in `predict.lm`, but can take a vector for 1-D densities.

`pssden` and `qssden` naturally only work for 1-D densities.

### Value

A vector of pdf, cdf, or quantiles.

### See Also

Fitting function `ssden` and `cdssden`. 
**esc**

*Embryonic Stem Cell from Mouse*

**Description**

Data concerning mouse embryonic stem cell gene expression and transcription factor association strength.

**Usage**

`data(esc)`

**Format**

A data frame containing 1027 genes with the following variables.

- `y1`: Gene expression after 4 days.
- `y2`: Gene expression after 8 days.
- `y3`: Gene expression after 14 days.
- `klf4`: Score of TFAS with KLF4.
- `nanog`: Score of TFAS with NANOG.
- `oct4`: Score of TFAS with OCT4.
- `sox2`: Score of TFAS with SOX2.
- `clusterID`: Cluster identification.

**References**


---

**eyetrack**

*Eyesight Fixation in Eyetracking Experiments*

**Description**

Eyesight Fixation during some eyetracking experiments in linguistics studies.

**Usage**

`data(eyetrack)`
Format

A data frame containing 13891 observations on the following variables.

- **time**: Time, in ms.
- **color**: Binary indicator, 1 if eyesight fixed on target or color competitor, a factor.
- **object**: Binary indicator, 1 if eyesight fixed on target or object competitor, a factor.
- **id**: Identification of homogeneous sessions, a factor.
- **cnt**: Multiplicity count.

Source

Dr. Anouschka Foltz.

References


fitted.ssanova  

*Fitted Values and Residuals from Smoothing Spline ANOVA Fits*

Description

Methods for extracting fitted values and residuals from smoothing spline ANOVA fits.

Usage

```r
# S3 method for class 'ssanova'
fitted(object, ...)
# S3 method for class 'ssanova'
residuals(object, ...)

# S3 method for class 'gssanova'
fitted(object, ...)
# S3 method for class 'gssanova'
residuals(object, type="working", ...)
```

Arguments

- **object**: Object of class "ssanova" or "gssanova".
- **type**: Type of residuals desired, with two alternatives "working" (default) or "deviance".
- **...**: Ignored.

Details

The fitted values for "gssanova" objects are on the link scale, so are the "working" residuals.
gastric  

*Gastric Cancer Data*

**Description**

Survival of gastric cancer patients under chemotherapy and chemotherapy-radiotherapy combination.

**Usage**

data(gastric)

**Format**

A data frame containing 90 observations on the following variables.

- *futime*: Follow-up time, in days.
- *status*: Censoring status.
- *trt*: Factor indicating the treatments: 1 – chemotherapy, 2 – combination.

**Source**


---

gauss.quad  

*Generating Gauss-Legendre Quadrature*

**Description**

Generate Gauss-Legendre quadratures using the FORTRAN routine gaussq.f found on NETLIB.

**Usage**

gauss.quad(size, interval)

**Arguments**

- *size*: Size of quadrature.
- *interval*: Interval to be covered.

**Value**

*gauss.quad* returns a list object with the following components.

- *pt*: Quadrature nodes.
- *wt*: Quadrature weights.
Fitting Smoothing Spline ANOVA Models with Non-Gaussian Responses

Description
Fit smoothing spline ANOVA models in non-Gaussian regression. The symbolic model specification via formula follows the same rules as in \texttt{lm} and \texttt{glm}.

Usage
\begin{verbatim}
gssanova(formula, family, type=NULL, data=list(), weights, subset, offset, na.action=na.omit, partial=NULL, alpha=NULL, nu=NULL, id.basis=NULL, nbasis=NULL, seed=NULL, random=NULL, skip.iter=FALSE)
\end{verbatim}

Arguments
\begin{itemize}
\item **formula**: Symbolic description of the model to be fit.
\item **family**: Description of the error distribution. Supported are exponential families "binomial", "poisson", "Gamma", "inverse.gaussian", and "nbinomial". Also supported are accelerated life model families "weibull", "lognorm", and "loglogis".
\item **type**: List specifying the type of spline for each variable. See \texttt{mkterm} for details.
\item **data**: Optional data frame containing the variables in the model.
\item **weights**: Optional vector of weights to be used in the fitting process.
\item **subset**: Optional vector specifying a subset of observations to be used in the fitting process.
\item **offset**: Optional offset term with known parameter 1.
\item **na.action**: Function which indicates what should happen when the data contain NAs.
\item **partial**: Optional symbolic description of parametric terms in partial spline models.
\item **alpha**: Tuning parameter defining cross-validation; larger values yield smoother fits. Defaults are alpha=1 for family="binomial" and alpha=1.4 otherwise.
\item **nu**: Inverse scale parameter in accelerated life model families. Ignored for exponential families.
\item **id.basis**: Index designating selected "knots".
\item **nbasis**: Number of "knots" to be selected. Ignored when id.basis is supplied.
\item **seed**: Seed for reproducible random selection of "knots". Ignored when id.basis is supplied.
\item **random**: Input for parametric random effects in nonparametric mixed-effect models. See \texttt{mkran} for details.
\item **skip.iter**: Flag indicating whether to use initial values of theta and skip theta iteration. See \texttt{ssanova} for notes on skipping theta iteration.
\end{itemize}
Details

The model specification via formula is intuitive. For example, \( y \sim x_1 \times x_2 \) yields a model of the form

\[
y = C + f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2) + e
\]

with the terms denoted by "1", "x1", "x2", and "x1:x2".

The model terms are sums of unpenalized and penalized terms. Attached to every penalized term there is a smoothing parameter, and the model complexity is largely determined by the number of smoothing parameters.

Only one link is implemented for each family. It is the logit link for "binomial", and the log link for "poisson", and "Gamma". For "nbinomial", the working parameter is the log of the probability \( p \); see NegBinomial. For "weibull", "lognorm", and "loglogis", it is the location parameter for the log lifetime.

The selection of smoothing parameters is through direct cross-validation. The cross-validation score used for family="poisson" is taken from density estimation as in Gu and Wang (2003), and those used for other families are derived following the lines of Gu and Xiang (2001).

A subset of the observations are selected as "knots." Unless specified via id.basis or nbasis, the number of "knots" \( q \) is determined by \( \max(30, 10n^{2/9}) \), which is appropriate for the default cubic splines for numerical vectors.

Value

gssanova returns a list object of class c("gssanova", "ssanova").

The method summary.gssanova can be used to obtain summaries of the fits. The method predict.gssanova can be used to evaluate the fits at arbitrary points along with standard errors, on the link scale. The method project.gssanova can be used to calculate the Kullback-Leibler projection for model selection. The methods residuals.gssanova and fitted.gssanova extract the respective traits from the fits.

Responses

For family="binomial", the response can be specified either as two columns of counts or as a column of sample proportions plus a column of total counts entered through the argument weights, as in glm.

For family="nbinomial", the response may be specified as two columns with the second being the known sizes, or simply as a single column with the common unknown size to be estimated through the maximum likelihood.

For family="weibull", "lognorm", or "loglogis", the response consists of three columns, with the first giving the follow-up time, the second the censoring status, and the third the left-truncation time. For data with no truncation, the third column can be omitted.

Note

For simpler models and moderate sample sizes, the exact solution of gssanova@ can be faster.

The results may vary from run to run. For consistency, specify id.basis or set seed.

In gss versions earlier than 1.0, gssanova was under the name gssanova1.
Author(s)

Chong Gu, <chong@stat.purdue.edu>

References


Examples

```r
## Fit a cubic smoothing spline logistic regression model
test <- function(x)
  {.3*(1e6*(x^11*(1-x)^6)+1e4*(x^3*(1-x)^10))-2}
  x <- (0:100)/100
  p <- 1-1/(1+exp(test(x)))
y <- rbinom(x,3,p)
logit.fit <- gssanova(cbind(y,3-y)-x,family="binomial")
## The same fit
logit.fit1 <- gssanova(y/3-x,"binomial",weights=rep(3,101),
  id.basis=logit.fit$id.basis)
## Obtain estimates and standard errors on a grid
est <- predict(logit.fit,data.frame(x=x),se=TRUE)
## Plot the fit and the Bayesian confidence intervals
plot(x,y/3,ylab="p")
lines(x,p,col=1)
lines(x,1-1/(1+exp(est$fit)),col=2)
lines(x,1-1/(1+exp(est$fit)+1.96*est$se),col=3)
lines(x,1-1/(1+exp(est$fit-1.96*est$se)),col=3)
## Fit a mixed-effect logistic model
data(bacteriuria)
bact.fit <- gssanova(infect~trt+time,family="binomial",data=bacteriuria,
  id.basis=(1:820)[bacteriuria$id%in%(3,38)],random=~1|id)
## Predict fixed effects
predict(bact.fit,data.frame(time=2:16,trt=as.factor(rep(1,15))),se=TRUE)
## Estimated random effects
bact.fit$bb
## Clean up
## Not run: rm(test,x,p,y,logit.fit,logit.fit1,est,bacteriuria,bact.fit)
dev.off()
## End(Not run)
```
Fitting Smoothing Spline ANOVA Models with Non-Gaussian Responses

Description

Fit smoothing spline ANOVA models in non-Gaussian regression. The symbolic model specification via formula follows the same rules as in `lm` and `glm`.

Usage

gssanova0(formula, family, type=NULL, data=list(), weights, subset, offset, na.action=na.omit, partial=NULL, method=NULL, varht=1, nu=NULL, prec=1e-7, maxiter=30)

gssanova1(formula, family, type=NULL, data=list(), weights, subset, offset, na.action=na.omit, partial=NULL, method=NULL, varht=1, alpha=1.4, nu=NULL, id.basis=NULL, nbasis=NULL, seed=NULL, random=NULL, skip.iter=FALSE)

Arguments

- **formula**: Symbolic description of the model to be fit.
- **family**: Description of the error distribution. Supported are exponential families "binomial", "poisson", "Gamma", "inverse.gaussian", and "nbinomial". Also supported are accelerated life model families "weibull", "lognorm", and "loglogis".
- **type**: List specifying the type of spline for each variable. See `mkterm` for details.
- **data**: Optional data frame containing the variables in the model.
- **weights**: Optional vector of weights to be used in the fitting process.
- **subset**: Optional vector specifying a subset of observations to be used in the fitting process.
- **offset**: Optional offset term with known parameter 1.
- **na.action**: Function which indicates what should happen when the data contain NAs.
- **partial**: Optional symbolic description of parametric terms in partial spline models.
- **method**: Score used to drive the performance-oriented iteration. Supported are method="v" for GCV, method="m" for GML, and method="u" for Mallows' CL.
- **varht**: Dispersion parameter needed for method="u". Ignored when method="v" or method="m" are specified.
- **nu**: Inverse scale parameter in accelerated life model families. Ignored for exponential families.
- **prec**: Precision requirement for the iterations.
- **maxiter**: Maximum number of iterations allowed for performance-oriented iteration, and for inner-loop multiple smoothing parameter selection when applicable.
alpha  Tuning parameter modifying GCV or Mallows’ CL.

idNbasis  Index designating selected “knots”.

nbasis  Number of “knots” to be selected. Ignored when idNbasis is supplied.

seed  Seed for reproducible random selection of “knots”. Ignored when idNbasis is supplied.

random  Input for parametric random effects in nonparametric mixed-effect models. See mkran for details.

skipNiter  Flag indicating whether to use initial values of theta and skip theta iteration. See ssanova for notes on skipping theta iteration.

Details

The model specification via formula is intuitive. For example, $y \sim x_1 \times x_2$ yields a model of the form

$$y = C + f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2) + e$$

with the terms denoted by "1", "x1", "x2", and "x1:x2".

The model terms are sums of unpenalized and penalized terms. Attached to every penalized term there is a smoothing parameter, and the model complexity is largely determined by the number of smoothing parameters.

Only one link is implemented for each family. It is the logit link for “binomial”, and the log link for “poisson”, “Gamma”, and “inverse.gaussian”. For “nbinomial”, the working parameter is the log of the probability $p$; see NegBinomial. For “weibull”, “lognorm”, and “loglogis”, it is the location parameter for the log lifetime.

The models are fitted by penalized likelihood method through the performance-oriented iteration as described in the reference. For family=“binomial”, “poisson”, “nbinomial”, “weibull”, “lognorm”, and “loglogis”, the score driving the performance-oriented iteration defaults to method=“u” with varht=1. For family=“Gamma” and “inverse.gaussian”, the default is method=“v”.

gssanova0 uses the algorithm of ssanova0 for the iterated penalized least squares problems, whereas gssanova1 uses the algorithm of ssanova.

In gssanova1, a subset of the observations are selected as “knots.” Unless specified via idNbasis or nbasis, the number of “knots” $q$ is determined by $\max(30, 10n^{2/9})$, which is appropriate for the default cubic splines for numerical vectors.

Value

gssanova0 returns a list object of class c(“gssanova0”, “ssanova0”, “gssanova”).
gssanova1 returns a list object of class c(“gssanova”, “ssanova”).

The method summary.gssanova0 or summary.gssanova can be used to obtain summaries of the fits. The method predict.gssanova0 or predict.gssanova can be used to evaluate the fits at arbitrary points along with standard errors, on the link scale. The methods residuals.gssanova and fitted.gssanova extract the respective traits from the fits.
Responses

For family="binomial", the response can be specified either as two columns of counts or as a column of sample proportions plus a column of total counts entered through the argument weights, as in \texttt{glm}.

For family="nbinomial", the response may be specified as two columns with the second being the known sizes, or simply as a single column with the common unknown size to be estimated through the maximum likelihood.

For family="weibull", "lognorm", or "loglogis", the response consists of three columns, with the first giving the follow-up time, the second the censoring status, and the third the left-truncation time. For data with no truncation, the third column can be omitted.

Note

The direct cross-validation of \texttt{gssanova} can be more effective, and more stable for complex models. For large sample sizes, the approximate solutions of \texttt{gssanova1} and \texttt{gssanova} can be faster than \texttt{gssanova@}.

The results from \texttt{gssanova1} may vary from run to run. For consistency, specify \texttt{id.basis} or set \texttt{seed}.

The method \texttt{project} is not implemented for \texttt{gssanova@}, nor is the mixed-effect model support through \texttt{mkran}.

In \texttt{gss} versions earlier than 1.0, \texttt{gssanova@} was under the name \texttt{gssanova}.

Author(s)

Chong Gu, <chong@stat.purdue.edu>

References


Examples

```r
## Fit a cubic smoothing spline logistic regression model
test <- function(x)
  .3*(1e6*(x^11*(1-x)^6)+1e4*(x^3*(1-x)^10))-2
x <- (0:100)/100
p <- 1-1/(1+exp(test(x)))
y <- rbinom(x,3,p)
logit.fit <- gssanova0(cbind(y,3-y)-x,family="binomial")
## The same fit
logit.fit1 <- gssanova0(y/3-x,"binomial",weights=rep(3,101))
## Obtain estimates and standard errors on a grid
est <- predict(logit.fit,data.frame(x=x),se=TRUE)
```
hzdrate.sshzd

## Plot the fit and the Bayesian confidence intervals
plot(x,y/3,ylab="p")
lines(x,p,col=1)
lines(x,1-1/(1+exp(est$fit)),col=2)
lines(x,1-1/(1+exp(est$fit+1.96*est$se)),col=3)
lines(x,1-1/(1+exp(est$fit-1.96*est$se)),col=3)
## Clean up
## Not run: rm(test,x,p,y,logit.fit,logit.fit1,est)
dev.off()
## End(Not run)

---

hzdrate.sshzd  
**Evaluating Smoothing Spline Hazard Estimates**

### Description
Evaluate smoothing spline hazard estimates by sshzd.

### Usage

```r
hzdrate.sshzd(object, x, se=FALSE, include=c(object$terms$labels,object$lab.p))
hzdcurve.sshzd(object, time, covariates=NULL, se=FALSE)
surveexp.sshzd(object, time, covariates=NULL, start=0)
```

### Arguments

- **object** 
  Object of class "sshzd".
- **x** 
  Data frame or vector of points on which hazard is to be evaluated.
- **se** 
  Flag indicating if standard errors are required.
- **include** 
  List of model terms to be included in the evaluation.
- **time** 
  Vector of time points.
- **covariates** 
  Vector of covariate values.
- **start** 
  Optional starting times of the intervals.

### Value

For se=FALSE, `hzdrate.sshzd` returns a vector of hazard evaluations, and `hzdcurve.sshzd` returns a vector or columns of hazard curve(s) evaluated on `time` points at the `covariates` values. For se=TRUE, `hzdrate.sshzd` and `hzdcurve.sshzd` return a list consisting of the following components.

- **fit** 
  Vector or columns of hazard.
- **se.fit** 
  Vector or columns of standard errors for log hazard.

`surveexp.sshzd` returns a vector or columns of expected survivals based on the cumulative hazards over (`start`, `time`) at the `covariates` values, which in fact are the (conditional) survival probabilities $S_{1}(\text{time})/S_{1}(\text{start})$. 
Note

For left-truncated data, `start` must be at or after the earliest truncation point.

See Also

Fitting function `sshzd`.

Description

Data extracted from the Eastern Lake Survey of 1984 conducted by the United States Environmental Protection Agency, concerning 112 lakes in the Blue Ridge.

Usage

```r
data(LakeAcidity)
```

Format

A data frame containing 112 observations on the following variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>ph</code></td>
<td>Surface ph.</td>
</tr>
<tr>
<td><code>cal</code></td>
<td>Calcium concentration.</td>
</tr>
<tr>
<td><code>lat</code></td>
<td>Latitude.</td>
</tr>
<tr>
<td><code>lon</code></td>
<td>Longitude.</td>
</tr>
<tr>
<td><code>geog</code></td>
<td>Geographic location, derived from <code>lat</code> and <code>lon</code></td>
</tr>
</tbody>
</table>

Details

`geog` was generated from `lat` and `lon` using the code given in the Example section.

Source


References


Examples

```r
## Converting latitude and longitude to x-y coordinates
## Not run: ltln2xy <- function(latlon,laton0) {
```
Minimizing Univariate Functions on Finite Intervals

Description

Minimize univariate functions on finite intervals using 3-point quadratic fit, with golden-section safeguard.

Usage

```r
nlm0(fun, range, prec=1e-7)
```

Arguments

- **fun**: Function to be minimized.
- **range**: Interval on which the function to be minimized.
- **prec**: Desired precision of the solution.

Value

`nlm0` returns a list object with the following components.

- **estimate**: Minimizer.
- **minimum**: Minimum.
- **evaluations**: Number of function evaluations.
Description

A subset of 500 hourly observations collected by the Norwegian Public Roads Administration at Alnabru in Oslo, Norway, between October 2001 and August 2003.

Usage

data(N02)

Format

A data frame containing 500 observations on the following variables.

- **noR**: Concentration of NO2, on log scale.
- **cars**: Traffic volume of the hour, on log scale.
- **temp**: Temperature 2 meters above ground, in Celsius.
- **wind**: Wind speed, meters/second.
- **tempR**: Temperature difference between 25 and 2 meters above ground, in Celsius.
- **windR**: Wind direction, in degrees between 0 and 360.

Source

Statlib Datasets Archive at http://lib.stat.cmu.edu/datasets, contributed by Magne Aldrin.

Description

Data from an experiment in which a single-cylinder engine was run with ethanol to see how the NOx concentration in the exhaust depended on the compression ratio and the equivalence ratio.

Usage

data(nox)

Format

A data frame containing 88 observations on the following variables.

- **nox**: NOx concentration in exhaust.
- **comp**: Compression ratio.
- **equi**: Equivalence ratio.
ozone

Source


References


---

**Ozone Concentration in Los Angeles Basin**

**Description**

Daily measurements of ozone concentration and eight meteorological quantities in the Los Angeles basin for 330 days of 1976.

**Usage**

data(ozone)

**Format**

A data frame containing 330 observations on the following variables.

- `upo3` Upland ozone concentration, in ppm.
- `vdht` Vandenberg 500 millibar height, in meters.
- `wdsp` Wind speed, in miles per hour.
- `hmtd` Humidity.
- `sbtp` Sandburg Air Base temperature, in Celsius.
- `ibht` Inversion base height, in foot.
- `dgpg` Dagget pressure gradient, in mmHg.
- `ibtp` Inversion base temperature, in Fahrenheit.
- `vsty` Visibility, in miles.
- `day` Calendar day, between 1 and 366.

**Source**

Unknown.

**References**


penny

*Thickness of US Lincoln Pennies*

**Description**


**Usage**

```r
data(nox)
```

**Format**

A data frame containing 90 observations on the following variables.

<table>
<thead>
<tr>
<th>year</th>
<th>Year minted.</th>
</tr>
</thead>
<tbody>
<tr>
<td>mil</td>
<td>Thickness in mils.</td>
</tr>
</tbody>
</table>

**Source**


**References**


---

**predict.sssanova**

*Predicting from Smoothing Spline ANOVA Fits*

**Description**

Evaluate terms in a smoothing spline ANOVA fit at arbitrary points. Standard errors of the terms can be requested for use in constructing Bayesian confidence intervals.

**Usage**

```r
## S3 method for class 'ssanova'
predict(object, newdata, se.fit=FALSE,
         include=c(object$terms$labels,object$lab.p), ...)
## S3 method for class 'ssanova0'
predict(object, newdata, se.fit=FALSE,
```
predict.ssanova

#include=c(object$terms$labels,object$lab.p), ...

## S3 method for class 'ssanova'
predict1(object, contr=c(1,-1), newdata, se.fit=TRUE,
include=c(object$terms$labels,object$lab.p), ...)

Arguments

- **object**: Object of class inheriting from "ssanova".
- **newdata**: Data frame or model frame in which to predict.
- **se.fit**: Flag indicating if standard errors are required.
- **include**: List of model terms to be included in the prediction. The offset term, if present, is to be specified by "offset".
- **contr**: Contrast coefficients.
- **...**: Ignored.

Value

For se.fit=FALSE, predict.ssanova returns a vector of the evaluated fit.
For se.fit=TRUE, predict.ssanova returns a list consisting of the following components.

- **fit**: Vector of evaluated fit.
- **se.fit**: Vector of standard errors.

Note

For mixed-effect models through ssanova or gssanova, the Z matrix is set to 0 if not supplied. To supply the Z matrix, add a component random=I(...) in newdata, where the as-is function I(...) preserves the integrity of the Z matrix in data frame.

predict1.ssanova takes a list of data frames in newdata representing x1, x2, etc. By default, it calculates f(x1)-f(x2) along with standard errors. While pairwise contrast is the targeted application, all linear combinations can be computed.

Author(s)

Chong Gu, <chong@stat.purdue.edu>

References


See Also

Fitting functions ssanova, ssanova0, gssanova, gssanova0 and methods summary.ssanova, summary.gssanova, summary.gssanova0, project.ssanova, fitted.ssanova.
Examples

```r
## THE FOLLOWING EXAMPLE IS TIME-CONSUMING
## Not run:
## Fit a model with cubic and thin-plate marginals, where geog is 2-D
data(LakeAcidity)
fit <- ssanova(ph~log(cal)+geog,,LakeAcidity)
## Obtain estimates and standard errors on a grid
new <- data.frame(cal=1,geog=1(matrix(0,1,2))
new <- model.frame(~log(cal)+geog,new)
predict(fit,new,se=TRUE)
## Evaluate the geog main effect
predict(fit,new,se=TRUE,inc="geog")
## Evaluate the sum of the geog main effect and the interaction
predict(fit,new,se=TRUE,inc=c("geog","log(cal):geog"))
## Evaluate the geog main effect on a grid
grid <- seq(-.04,.04,len=21)
new <- model.frame(~geog,list(geog=cbind(rep(grid,21),rep(grid,rep(21,21))))
est <- predict(fit,new,se=TRUE,inc="geog")
## Plot the fit and standard error
par(pty="s")
contour(grid,grid,matrix(est$fit,21,21),col=1)
contour(grid,grid,matrix(est$se,21,21),add=TRUE,col=2)
## Clean up
rm(LakeAcidity,fit,new,grid,est)
dev.off()
## End(Not run)
```

predict.sscox  Evaluating Smoothing Spline ANOVA Estimate of Relative Risk

Description

Evaluate terms in a smoothing spline ANOVA estimate of relative risk at arbitrary points. Standard errors of the terms can be requested for use in constructing Bayesian confidence intervals.

Usage

```r
## S3 method for class 'sscox'
predict(object, newdata, se.fit=FALSE,
         include=c(object$terms$labels,object$lab.p), ...)
```

Arguments

- `object` Object of class "sscox".
- `newdata` Data frame or model frame in which to predict.
- `se.fit` Flag indicating if standard errors are required.
- `include` List of model terms to be included in the prediction.
- `...` Ignored.
**Value**

For `se.fit=FALSE`, `predict.sscox` returns a vector of the evaluated relative risk. For `se.fit=TRUE`, `predict.sscox` returns a list consisting of the following components.

- **fit**: Vector of evaluated relative risk.
- **se.fit**: Vector of standard errors for log relative risk.

**Note**

For mixed-effect models through `sscox`, the Z matrix is set to 0 if not supplied. To supply the Z matrix, add a component `random=I(...)` in newdata, where the as-is function `I(...)` preserves the integrity of the Z matrix in data frame.

**Author(s)**

Chong Gu, <chong@stat.purdue.edu>

**See Also**

Fitting functions `sscox` and method `project.sscox`.

---

**predict.ssllrm**  
**Evaluating Log-Linear Regression Model Fits**

**Description**

Evaluate conditional density in a log-linear regression model fit at arbitrary x, or contrast of log conditional density possibly with standard errors for constructing Bayesian confidence intervals.

**Usage**

```r
## S3 method for class 'ssllrm'
predict(object, x, y=object$qd.pt, odds=NULL, se.odds=FALSE, ...)```

**Arguments**

- **object**: Object of class "ssllrm".
- **x**: Data frame of x values.
- **y**: Data frame of y values; y-variables must be factors.
- **odds**: Optional coefficients of contrast.
- **se.odds**: Flag indicating if standard errors are required. Ignored when `odds=NULL`.
- **...**: Ignored.
Value

For odds=NULL, predict.ssanova returns a vector/matrix of the estimated f(y|x).

When odds is given, it should match y in length and the coefficients must add to zero; predict.ssanova then returns a vector of estimated "odds ratios" if se.odds=FALSE or a list consisting of the following components if se.odds=TRUE.

- fit Vector of evaluated fit.
- se.fit Vector of standard errors.

Author(s)

Chong Gu, <chong@stat.purdue.edu>

See Also

Fitting function ssllrm.
Arguments

- \( x \) Object of class `ssanova`, `summary.ssanova`, `summary.gssanova`, or `ssden`.
- `digits` Number of significant digits to be printed in values.
- `...` Ignored.

See Also

`ssanova`, `ssanova0`, `gssanova`, `gssanova0`, `ssden`, `ssllrm`, `sshzd`, `summary.ssanova`, `summary.gssanova`, `summary.gssanova0`.

Description

Calculate Kullback-Leibler projection of smoothing spline ANOVA fits for model diagnostics.

Usage

```r
project(object, ...)  # S3 method for class 'ssanova'
project(object, include, ...)  # S3 method for class 'ssanova9'
project(object, include, ...)  # S3 method for class 'gssanova'
project(object, include, ...)  # S3 method for class 'ssden'
project(object, include, mesh=FALSE, ...)  # S3 method for class 'ssden1'
project(object, include, drop1=FALSE, ...)  # S3 method for class 'sscden'
project(object, include, ...)  # S3 method for class 'sscden1'
project(object, include, ...)  # S3 method for class 'sshzd'
project(object, include, mesh=FALSE, ...)  # S3 method for class 'sshzd1'
project(object, include, ...)  # S3 method for class 'ssccox'
project(object, include, ...)  # S3 method for class 'ssllrm'
project(object, include, ...)  # S3 method for class 'ssllrm1'
```
project

Arguments

object  Object of class "ssanova", "gssanova", "ssden", "ssden1", "sscden", "sscden1", "sshd", "sshd1", or "sslrm".
...  Additional arguments. Ignored in project.x.
include  List of model terms to be included in the reduced model space. The partial and offset terms, if present, are to be specified by "partial" and "offset", respectively.
mesh  Flag indicating whether to return evaluations of the projection.
drop1  If TRUE, calculate \[ p < - \text{length(include)} \text{projections with include[-i], i=1,...,p.} \]

Details

The entropy \( KL(\text{fit0},\text{null}) \) can be decomposed as the sum of \( KL(\text{fit0},\text{fit1}) \) and \( KL(\text{fit1},\text{null}) \), where \( \text{fit0} \) is the fit to be projected, \( \text{fit1} \) is the projection in the reduced model space, and null is the constant fit. The ratio \( KL(\text{fit0},\text{fit1})/KL(\text{fit0},\text{null}) \) serves as a diagnostic of the feasibility of the reduced model.

For regression fits, smoothness safe-guard is used to prevent interpolation, and \( KL(\text{fit0},\text{fit1})+KL(\text{fit1},\text{null}) \) may not match \( KL(\text{fit0},\text{null}) \) perfectly.

For mixed-effect models from ssanova and gssanova, the estimated random effects are treated as offset.

Value

The functions return a list consisting of the following components.

ratio  \[ KL(\text{fit0},\text{fit1})/KL(\text{fit0},\text{null}); \text{ the smaller the value, the more feasible the reduced model is.} \]
kl  \[ KL(\text{fit0},\text{fit1}). \]

For regression fits, the list also contains the following component.

check  \[ KL(\text{fit0},\text{fit1})/KL(\text{fit0},\text{null})+KL(\text{fit1},\text{null})/KL(\text{fit0},\text{null}); \text{ a value closer to 1 is preferred.} \]

For density and hazard fits, the list may contain the following optional component.

mesh  The evaluations of the projection.

Note

project.ssd1, project.scd1, and project.shzd1 calculates square error projections.

Author(s)

Chong Gu, <chong@stat.purdue.edu>

References

See Also

Fitting functions ssanova, gssanova, ssden, sshzd, and sshzd1.

---

**Sachs**

*Protein Expression in Human Immune System Cells*

---

**Description**

Data concerning protein expression levels in human immune system cells under stimulations.

**Usage**

```r
data(Sachs)
```

**Format**

A data frame containing 7466 cells, with flow cytometry measurements of 11 phosphorylated proteins and phospholipids, on the log10 scale of the original.

- praf: Raf phosphorylated at S259.
- pemek: Mek1/mek2 phosphorylated at S217/S221.
- plcg: Phosphorylation of phospholipase C - γ on Y783.
- pip2: Phosphatidylinositol 4,5-biphosphate.
- pip3: Phosphatidylinositol 3,4,5-triphosphate.
- p44.42: Erk1/erl2 phosphorylated at T202/Y204.
- pakts473: AKT phosphorylated at S473.
- pka: Phosphorylation of protein kinase A substrates on 3 sites.
- pkc: Phosphorylation of protein kinase C substrates on S660.
- p38: Erk1/erl2 phosphorylated at T180/Y182.
- pjnk: Erk1/erl2 phosphorylated at T183/Y185.

**Source**


---

**smolyak**

*Generating Smolyak Cubature*

---

**Description**

Generate delayed Smolyak cubatures using C routines modified from smolyak.c found in Knut Petras’ SMOLPACK.
Usage

smolyak.quad(d, k)

smolyak.size(d, k)

Arguments

d  Dimension of unit cube.

k  Depth of algorithm.

Value

smolyak.quad returns a list object with the following components.

pt  Quadrature nodes in rows of matrix.

wt  Quadrature weights.

smolyak.size returns an integer.

ssanova  Fitting Smoothing Spline ANOVA Models

Description

Fit smoothing spline ANOVA models in Gaussian regression. The symbolic model specification via
formula follows the same rules as in \texttt{lm}.

Usage

ssanova(formula, type=NULL, data=list(), weights, subset, offset,
         na.action=na.omit, partial=NULL, method="v", alpha=1.4,
         varht=1, id.basis=NULL, nbasis=NULL, seed=NULL, random=NULL,
         skip.iter=FALSE)

Arguments

formula  Symbolic description of the model to be fit.

type  List specifying the type of spline for each variable. See \texttt{mkterm} for details.

data  Optional data frame containing the variables in the model.

weights  Optional vector of weights to be used in the fitting process.

subset  Optional vector specifying a subset of observations to be used in the fitting pro-

offset  Optional offset term with known parameter 1.

na.action  Function which indicates what should happen when the data contain NAs.

partial  Optional symbolic description of parametric terms in partial spline models.
**Method for smoothing parameter selection.** Supported are method="v" for GCV, method="m" for GML (REML), and method="u" for Mallows’ CL.

**Alpha**
Parameter modifying GCV or Mallows’ CL; larger absolute values yield smoother fits; negative value invokes a stable and more accurate GCV/CL evaluation algorithm but may take two to five times as long. Ignored when method="m" are specified.

**Varht**
External variance estimate needed for method="u". Ignored when method="v" or method="m" are specified.

**Id.basis**
Index designating selected "knots".

**Nbasis**
Number of "knots" to be selected. Ignored when id.basis is supplied.

**Seed**
Seed to be used for the random generation of "knots". Ignored when id.basis is supplied.

**Random**
Input for parametric random effects in nonparametric mixed-effect models. See mkran for details.

**Skip.iter**
Flag indicating whether to use initial values of theta and skip theta iteration. See notes on skipping theta iteration.

**Details**

The model specification via formula is intuitive. For example, y~x1*x2 yields a model of the form

\[ y = C + f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2) + e \]

with the terms denoted by "1", "x1", "x2", and "x1:x2".

The model terms are sums of unpenalized and penalized terms. Attached to every penalized term there is a smoothing parameter, and the model complexity is largely determined by the number of smoothing parameters.

A subset of the observations are selected as "knots." Unless specified via id.basis or nbasis, the number of "knots" q is determined by \( \max(30, 10n^2/9) \), which is appropriate for the default cubic splines for numerical vectors.

Using q "knots," ssanova calculates an approximate solution to the penalized least squares problem using algorithms of the order \( O(nq^2) \), which for \( q \ll n \) scale better than the \( O(n^3) \) algorithms of ssanova0. For the exact solution, one may set \( q = n \) in ssanova, but ssanova0 would be much faster.

**Value**

ssanova returns a list object of class "ssanova".

The method summary.ssanova can be used to obtain summaries of the fits. The method predict.ssanova can be used to evaluate the fits at arbitrary points along with standard errors. The method project.ssanova can be used to calculate the Kullback-Leibler projection for model selection. The methods residuals.ssanova and fitted.ssanova extract the respective traits from the fits.
**Skipping Theta Iteration**

For the selection of multiple smoothing parameters, `nlm` is used to minimize the selection criterion such as the GCV score. When the number of smoothing parameters is large, the process can be time-consuming due to the great amount of function evaluations involved.

The starting values for the `nlm` iteration are obtained using Algorithm 3.2 in Gu and Wahba (1991). These starting values usually yield good estimates themselves, leaving the subsequent quasi-Newton iteration to pick up the “last 10%” performance with extra effort many times of the initial one. Thus, it is often a good idea to skip the iteration by specifying `skip.iter=TRUE`, especially in high-dimensions and/or with multi-way interactions.

`skip.iter=TRUE` could be made the default in future releases.

**Note**

To use GCV and Mallows’ CL unmodified, set `alpha=1`.

For simpler models and moderate sample sizes, the exact solution of `ssanova0` can be faster.

The results may vary from run to run. For consistency, specify `id.basis` or set `seed`.

In `gss` versions earlier than 1.0, `ssanova` was under the name `ssanova1`.

**Author(s)**

Chong Gu, <chong@stat.purdue.edu>

**References**


**Examples**

```r
## Fit a cubic spline
x <- runif(100); y <- 5 + 3*sin(2*pi*x) + rnorm(x)
cubic.fit <- ssanova(y~x)
## Obtain estimates and standard errors on a grid
new <- data.frame(x=seq(min(x),max(x),len=50))
est <- predict(cubic.fit,new,se=TRUE)
## Plot the fit and the Bayesian confidence intervals
plot(x,y,col=1); lines(new$x,est$fit,col=2)
lines(new$x,est$fit+1.96*est$se,col=3)
lines(new$x,est$fit-1.96*est$se,col=3)
## Clean up
## Not run: rm(x,y,cubic.fit,new,est)
```
ssanova0

Fitting Smoothing Spline ANOVA Models

Description

Fit smoothing spline ANOVA models in Gaussian regression. The symbolic model specification via formula follows the same rules as in lm.

Usage

ssanova0(formula, type=NULL, data=list(), weights, subset, 
          offset, na.action=na.omit, partial=NULL, method="v", 
          varht=1, prec=1e-7, maxiter=30)

Arguments

formula Symbolic description of the model to be fit.
type List specifying the type of spline for each variable. See mkterm for details.
data Optional data frame containing the variables in the model.
weights Optional vector of weights to be used in the fitting process.
subset Optional vector specifying a subset of observations to be used in the fitting process.
offset Optional offset term with known parameter 1.
na.action Function which indicates what should happen when the data contain NAs.
partial Optional symbolic description of parametric terms in partial spline models.
method Method for smoothing parameter selection. Supported are method="v" for GCV, method="m" for GML (REML), and method="u" for Mallow’s CL.
varht External variance estimate needed for method="u". Ignored when method="v" or method="m" are specified.
prec Precision requirement in the iteration for multiple smoothing parameter selection. Ignored when only one smoothing parameter is involved.
maxiter Maximum number of iterations allowed for multiple smoothing parameter selection. Ignored when only one smoothing parameter is involved.
Details

The model specification via formula is intuitive. For example, \( y \sim x_1 \times x_2 \) yields a model of the form

\[
y = C + f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2) + e
\]

with the terms denoted by "1", "x1", "x2", and "x1:x2".

The model terms are sums of unpenalized and penalized terms. Attached to every penalized term there is a smoothing parameter, and the model complexity is largely determined by the number of smoothing parameters.

ssanova0 and the affiliated methods provide a front end to RKPACK, a collection of RATFOR routines for nonparametric regression via the penalized least squares. The algorithms implemented in RKPACK are of the order \( O(n^3) \).

Value

ssanova0 returns a list object of class c("ssanova0", "ssanova").

The method summary.sanova0 can be used to obtain summaries of the fits. The method predict.sanova0 can be used to evaluate the fits at arbitrary points along with standard errors. The methods residuals.sanova and fitted.sanova extract the respective traits from the fits.

Note

For complex models and large sample sizes, the approximate solution of ssanova can be faster.

The method project is not implemented for ssanova0, nor is the mixed-effect model support through mkran.

In gss versions earlier than 1.0, ssanova0 was under the name ssanova.

Author(s)

Chong Gu, <chong@stat.purdue.edu>

References

Wahba, G. (1990), Spline Models for Observational Data. Philadelphia: SIAM.


Examples

```r
## Fit a cubic spline
x <- runif(100); y <- 5 + 3*sin(2*pi*x) + rnorm(x)
cubic.fit <- ssanova(y~x, method="m")
## Obtain estimates and standard errors on a grid
new <- data.frame(x=seq(min(x), max(x), len=50))
est <- predict(cubic.fit, new, se=TRUE)
## Plot the fit and the Bayesian confidence intervals
plot(x,y,col=1); lines(new$x, est$fit, col=2)
```
## ssanova9

### Fitting Smoothing Spline ANOVA Models with Correlated Data

#### Description

Fit smoothing spline ANOVA models with correlated Gaussian data. The symbolic model specification via `formula` follows the same rules as in `lm`.

#### Usage

```r
glmmss(formula, type=NULL, data=list(), subset, offset, na.action=na.omit, partial=NULL, method="v", alpha=1.4, varht=1, id.basis=NULL, nbasis=NULL, seed=NULL, cov, skip.iter=FALSE)
```

#### Arguments

- **formula**: Symbolic description of the model to be fit.
- **type**: List specifying the type of spline for each variable. See `mkterm` for details.
- **data**: Optional data frame containing the variables in the model.
- **subset**: Optional vector specifying a subset of observations to be used in the fitting process.
- **offset**: Optional offset term with known parameter 1.
- **na.action**: Function which indicates what should happen when the data contain NAs.
- **partial**: Optional symbolic description of parametric terms in partial spline models.
method

Method for smoothing parameter selection. Supported are method="v" for V, method="m" for M, and method="u" for U; see the reference for definitions of U, V, and M.

alpha

Parameter modifying V or U; larger absolute values yield smoother fits. Ignored when method="m" are specified.

varht

External variance estimate needed for method="u". Ignored when method="v" or method="m" are specified.

idNbasis

Index designating selected "knots".

nbasis

Number of "knots" to be selected. Ignored when idNbasis is supplied.

seed

Seed to be used for the random generation of "knots". Ignored when id.basis is supplied.

cov

Input for covariance functions. See mkcov for details.

skipNiter

Flag indicating whether to use initial values of theta and skip theta iteration. See notes on skipping theta iteration.

fit

ssanova9 fit with ARMA error.

Details

The model specification via formula is intuitive. For example, y~x1*x2 yields a model of the form

\[ y = C + f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2) + e \]

with the terms denoted by "1", "x1", "x2", and "x1:x2".

The model terms are sums of unpenalized and penalized terms. Attached to every penalized term there is a smoothing parameter, and the model complexity is largely determined by the number of smoothing parameters.

A subset of the observations are selected as "knots." Unless specified via id.basis or nbasis, the number of "knots" q is determined by max(30, 10n^2/9), which is appropriate for the default cubic splines for numerical vectors.

Using q "knots," ssanova calculates an approximate solution to the penalized least squares problem using algorithms of the order O(nq^2), which for q << n scale better than the O(n^3) algorithms of ssanovaP. For the exact solution, one may set q = n in ssanova, but ssanovaP would be much faster.

Value

ssanova9 returns a list object of class c("ssanova9", "ssanova").

The method summary.ssanova can be used to obtain summaries of the fits. The method predict.ssanova can be used to evaluate the fits at arbitrary points along with standard errors. The method project.ssanova can be used to calculate the Kullback-Leibler projection for model selection. The methods residuals.ssanova and fitted.ssanova extract the respective traits from the fits.

para.arma returns the fitted ARMA coefficients for cov=list("arma",c(p,q)) in the call to ssanova9.
Skipping Theta Iteration

For the selection of multiple smoothing parameters, `nlm` is used to minimize the selection criterion such as the GCV score. When the number of smoothing parameters is large, the process can be time-consuming due to the great amount of function evaluations involved.

The starting values for the `nlm` iteration are obtained using Algorithm 3.2 in Gu and Wahba (1991). These starting values usually yield good estimates themselves, leaving the subsequent quasi-Newton iteration to pick up the "last 10%" performance with extra effort many times of the initial one. Thus, it is often a good idea to skip the iteration by specifying `skip.iter=TRUE`, especially in high-dimensions and/or with multi-way interactions.

`skip.iter=TRUE` could be made the default in future releases.

Note

The results may vary from run to run. For consistency, specify `id.basis` or set `seed`.

Author(s)

Chong Gu, <chong@stat.purdue.edu>

References


Examples

```r
x <- runif(100); y <- 5 + 3*sin(2*pi*x) + rnorm(x)
## independent fit
fit <- ssanova9(y~x,cov=list("known",diag(1,100)))
## AR(1) fit
fit <- ssanova9(y~x,cov=list("arma",c(1,0)))
para arma(fit)
## MA(1) fit
e <- rnorm(101); e <- e[-1]-.5*e[-101]
x <- runif(100); y <- 5 + 3*sin(2*pi*x) + e
fit <- ssanova9(y~x,cov=list("arma",c(0,1)))
para arma(fit)
## Clean up
## Not run: rm(x,y,e,fit)
```
Estimating Conditional Probability Density Using Smoothing Splines

Description

Estimate conditional probability densities using smoothing spline ANOVA models. The symbolic model specification via `formula` follows the same rules as in `lm`.

Usage

```r
sscden(formula, response, type=NULL, data=list(), weights, subset,
na.action=na.omit, alpha=1.4, id.basis=NULL, nbasis=NULL,
seed=NULL, ydomain=as.list(NULL), yquad=NULL, prec=1e-7,
maxiter=30, skip.iter=FALSE)

sscdn1(formula, response, type=NULL, data=list(), weights, subset,
na.action=na.omit, alpha=1.4, id.basis=NULL, nbasis=NULL,
seed=NULL, rho=list("xy"), ydomain=as.list(NULL), yquad=NULL,
prec=1e-7, maxiter=30, skip.iter=FALSE)
```

Arguments

- **formula**: Symbolic description of the model to be fit.
- **response**: Formula listing response variables.
- **type**: List specifying the type of spline for each variable. See `mkterm` for details.
- **data**: Optional data frame containing the variables in the model.
- **weights**: Optional vector of counts for duplicated data.
- **subset**: Optional vector specifying a subset of observations to be used in the fitting process.
- **na.action**: Function which indicates what should happen when the data contain NAs.
- **alpha**: Parameter defining cross-validation scores for smoothing parameter selection.
- **id.basis**: Index of observations to be used as "knots."
- **nbasis**: Number of "knots" to be used. Ignored when `id.basis` is specified.
- **seed**: Seed to be used for the random generation of "knots." Ignored when `id.basis` is specified.
- **ydomain**: Data frame specifying marginal support of conditional density.
- **yquad**: Quadrature for calculating integral on Y domain. Mandatory if response variables other than factors or numerical vectors are involved.
- **prec**: Precision requirement for internal iterations.
- **maxiter**: Maximum number of iterations allowed for internal iterations.
- **skip.iter**: Flag indicating whether to use initial values of theta and skip theta iteration. See `ssanova` for notes on skipping theta iteration.
- **rho**: rho function needed for `sscdn1`. 
Details

The model is specified via formula and response, where response lists the response variables. For example, `sscden(~ y * x, ~ y)` prescribe a model of the form

\[
\log f(y|x) = g_y(y) + g_{xy}(x, y) + C(x)
\]

with the terms denoted by "y", "y:x"; the term(s) not involving response(s) are removed and the constant C(x) is determined by the fact that a conditional density integrates to one on the y axis. `sscden1` does keep terms not involving response(s) during estimation, although those terms cancel out when one evaluates the estimated conditional density.

The model terms are sums of unpenalized and penalized terms. Attached to every penalized term there is a smoothing parameter, and the model complexity is largely determined by the number of smoothing parameters.

A subset of the observations are selected as "knots." Unless specified via `idNbasis` or `nbasis`, the number of "knots" q is determined by \( \max(30, 10n^2/9) \), which is appropriate for the default cubic splines for numerical vectors.

Value

`sscden` returns a list object of class "sscden". `sscden1` returns a list object of class c("sscden1", "sscden").

`dsscden` and `edsscden` can be used to evaluate the estimated conditional density \( f(y|x) \) and \( f(y1|x, y2) \);

`psscden`, `qsscden`, `cpsscden`, and `cqsscden` can be used to evaluate conditional cdf and quantiles.

The methods `project.sscden` or `project.sscden1` can be used to calculate the Kullback-Leibler or square-error projections for model selection.

Note

Default quadrature on the Y domain will be constructed for numerical vectors on a hyper cube, then outer product with factor levels will be taken if factors are involved. The sides of the hyper cube are specified by `ydomain`; for `ydomainDy` missing, the default is \( c(\min(y), \max(y)) + c(-1, 1) * (\max(y) - \min(y)) * .05 \).

On a 1-D interval, the quadrature is the 200-point Gauss-Legendre formula returned from `gauss.quad`.

For multiple numerical vectors, delayed Smolyak cubatures from `smolyak.quad` are used on cubes with the marginals properly transformed; see Gu and Wang (2003) for the marginal transformations.

The results may vary from run to run. For consistency, specify `idNbasis` or set `seed`.

For reasonable execution time in high dimensions, set `skip.iter=TRUE`.

Author(s)

Chong Gu, <chong@stat.purdue.edu>

References


Examples

data(penny); set.seed(5732)
fit <- sscden(~year*mil, ~mil, data=penny,
       ydomain=data.frame(mil=c(49,61)))
yy <- 1944+(0:92)/2
quan <- qsscden(fit, c(.05,.25,.5,.75,.95),
       data.frame(year=yy))
plot(penny$year+.1*runif(90), penny$mil, ylim=c(49,61))
for (i in 1:5) lines(yy, quan[i,])
## Clean up
## Not run: rm(penny, yy, quan)

Description

Estimate relative risk using smoothing spline ANOVA models. The symbolic model specification via formula follows the same rules as in \texttt{lm}, but with the response of a special form.

Usage

\texttt{sscox(formula, type=NULL, data=list(), weights=NULL, subset,}
\texttt{ na.action=na.omit, partial=NULL, alpha=1.4, id.basis=NULL,}
\texttt{ nbasis=NULL, seed=NULL, random=NULL, prec=1e-7, maxiter=30,}
\texttt{ skip.iter=FALSE)}

Arguments

\begin{itemize}
\item \texttt{formula} Symbolic description of the model to be fit, where the response is of the form \texttt{Surv(futime,status,start=0)}.
\item \texttt{type} List specifying the type of spline for each variable. See \texttt{mkterm} for details.
\item \texttt{data} Optional data frame containing the variables in the model.
\item \texttt{weights} Optional vector of counts for duplicated data.
\item \texttt{subset} Optional vector specifying a subset of observations to be used in the fitting process.
\item \texttt{na.action} Function which indicates what should happen when the data contain NAs.
\item \texttt{partial} Optional symbolic description of parametric terms in partial spline models.
\item \texttt{alpha} Parameter defining cross-validation score for smoothing parameter selection.
\item \texttt{id.basis} Index of observations to be used as "knots."
\item \texttt{nbasis} Number of "knots" to be used. Ignored when \texttt{id.basis} is specified.
\item \texttt{seed} Seed to be used for the random generation of "knots." Ignored when \texttt{id.basis} is specified.
\end{itemize}
random: Input for parametric random effects (frailty) in nonparametric mixed-effect models. See mkran for details.

prec: Precision requirement for internal iterations.

maxiter: Maximum number of iterations allowed for internal iterations.

skip.iter: Flag indicating whether to use initial values of theta and skip theta iteration. See ssanova for notes on skipping theta iteration.

Details

A proportional hazard model is assumed, and the relative risk is estimated via penalized partial likelihood. The model specification via formula is for the log relative risk. For example, \( \text{Suve}(t,d) \sim u \times v \) prescribes a model of the form

\[
\log f(u,v) = g_u(u) + g_v(v) + g_{u,v}(u,v)
\]

with the terms denoted by "u", "v", and "u\times v"; relative risk is defined only up to a multiplicative constant, so the constant term is not included in the model.

sscox takes standard right-censored lifetime data, with possible left-truncation and covariates; in \( \text{Surv}(\text{futime}, \text{status}, \text{start}=0) \sim \ldots \), futime is the follow-up time, status is the censoring indicator, and start is the optional left-truncation time.

Parallel to those in a ssanova object, the model terms are sums of unpenalized and penalized terms. Attached to every penalized term there is a smoothing parameter, and the model complexity is largely determined by the number of smoothing parameters.

The selection of smoothing parameters is through a cross-validation mechanism designed for density estimation under biased sampling, with a fudge factor \( \alpha \); \( \alpha = 1 \) is "unbiased" for the minimization of Kullback-Leibler loss but may yield severe undersmoothing, whereas larger \( \alpha \) yields smoother estimates.

A subset of the observations are selected as "knots." Unless specified via id.basis or nbasis, the number of "knots" \( q \) is determined by \( \max(30, 10n^2/9) \), which is appropriate for the default cubic splines for numerical vectors.

Value

sscox returns a list object of class "sscox".

The method predict.sscx can be used to evaluate the fits at arbitrary points along with standard errors. The method project.sscx can be used to calculate the Kullback-Leibler projection for model selection.

Note

The function \( \text{Surv}(\text{futime}, \text{status}, \text{start}=0) \) is defined and parsed inside sscx, not quite the same as the one in the survival package. The estimation is invariant of monotone transformations of time.

The results may vary from run to run. For consistency, specify id.basis or set seed.

Author(s)

Chong Gu, <chong@stat.purdue.edu>
References


Examples

```r
## Relative Risk
data(stan)
fit.rr <- sscox(Surv(futime,status)-age,data=stan)
est.rr <- predict(fit.rr,data.frame(age=c(35,40)),se=TRUE)
## Base Hazard
risk <- predict(fit.rr,stan)
fit.bh <- sshzd(Surv(futime,status)-futime,data=stan,offset=log(risk))
tt <- seq(0,max(stan$futime),length=51)
est.bh <- hzdcurve.shzd(fit.bh,tt,se=TRUE)
## Clean up
## Not run: rm(stan,fit.rr,est.rr,risk,fit.bh,tt,est.bh)
```

---

**ssden**

*Estimating Probability Density Using Smoothing Splines*

**Description**

Estimate probability densities using smoothing spline ANOVA models. The symbolic model specification via `formula` follows the same rules as in `lm`, but with the response missing.

**Usage**

```r
ssden(formula, type=NULL, data=list(), alpha=1.4, weights=NULL,
   subset, na.action=na.omit, id.basis=NULL, nbasis=NULL, seed=NULL,
   domain=as.list(NULL), quad=NULL, qdsz.depth=NULL, bias=NULL,
   prec=1e-7, maxiter=30, skip.iter=FALSE)
```

**Arguments**

- **formula**
  - Symbolic description of the model to be fit.
- **type**
  - List specifying the type of spline for each variable. See `mkterm` for details.
- **data**
  - Optional data frame containing the variables in the model.
- **alpha**
  - Parameter defining cross-validation score for smoothing parameter selection.
- **weights**
  - Optional vector of bin-counts for histogram data.
subset: Optional vector specifying a subset of observations to be used in the fitting process.

na.action: Function which indicates what should happen when the data contain NAs.

id.basis: Index of observations to be used as "knots."

nbasis: Number of "knots" to be used. Ignored when id.basis is specified.

seed: Seed to be used for the random generation of "knots." Ignored when id.basis is specified.

domain: Data frame specifying marginal support of density.

quad: Quadrature for calculating integral. Mandatory if variables other than factors or numerical vectors are involved.

qdsz.depth: Depth to be used in smolyak_quad for the generation of quadrature.

bias: Input for sampling bias.

prec: Precision requirement for internal iterations.

maxiter: Maximum number of iterations allowed for internal iterations.

skip.iter: Flag indicating whether to use initial values of theta and skip theta iteration. See ssanova for notes on skipping theta iteration.

Details

The model specification via formula is for the log density. For example, \( \sim x_1 \times x_2 \) prescribes a model of the form

\[
\log f(x_1, x_2) = g_1(x_1) + g_2(x_2) + g_{12}(x_1, x_2) + C
\]

with the terms denoted by "x1", "x2", and "x1:x2"; the constant is determined by the fact that a density integrates to one.

The selective term elimination may characterize (conditional) independence structures between variables. For example, \( \sim x_1 \times x_2+x_1 \times x_3 \) yields the conditional independence of \( x_2 \) and \( x_3 \) given \( x_1 \).

Parallel to those in a ssanova object, the model terms are sums of unpenalized and penalized terms. Attached to every penalized term there is a smoothing parameter, and the model complexity is largely determined by the number of smoothing parameters.

The selection of smoothing parameters is through a cross-validation mechanism described in the references, with a parameter alpha; alpha=1 is "unbiased" for the minimization of Kullback-Leibler loss but may yield severe undersmoothing, whereas larger alpha yields smoother estimates.

A subset of the observations are selected as "knots." Unless specified via id.basis or nbasis, the number of "knots" q is determined by \( \max(30, 10n^{2/9}) \), which is appropriate for the default cubic splines for numerical vectors.

Value

ssden returns a list object of class "ssden". ssden1 returns a list object of class c("ssden1","ssden"). dssden and cdssden can be used to evaluate the estimated joint density and conditional density; pssden, qssden, cpssden, and cqssden can be used to evaluate (conditional) cdf and quantiles.

The method project.ssden can be used to calculate the Kullback-Leibler projection of "ssden" objects for model selection; project.ssden1 can be used to calculate the square error projection of "ssden1" objects.
ssden

Note
In ssden, default quadrature will be constructed for numerical vectors on a hyper cube, then outer product with factor levels will be taken if factors are involved. The sides of the hyper cube are specified by domain; for domain$x$ missing, the default is $c(\min(x), \max(x)) c(-1,1) \ast (\max(x)-\min(x)) \ast .05$.
In 1-D, the quadrature is the 200-point Gauss-Legendre formula returned from \texttt{gauss.quad}. In multi-D, delayed Smolyak cubatures from \texttt{smolyak.quad} are used on cubes with the marginals properly transformed; see Gu and Wang (2003) for the marginal transformations.

For reasonable execution time in higher dimensions, set \texttt{skip.iter=TRUE} in call to ssden.

If you get an error message from ssden stating "Newton iteration diverges", try to use a larger \texttt{qdsz.depth} which will execute slower, or switch to \texttt{ssden1}. The default values of \texttt{qdsz.depth} for dimensions 4, 5, 6+ are 12, 11, 10.

\texttt{ssden1} does not involve multi-D quadrature but does not perform as well as \texttt{ssden}. It can be used in very high dimensions where \texttt{ssden} is infeasible.

The results may vary from run to run. For consistency, specify \texttt{id.basis} or set seed.

Author(s)
Chong Gu, <chong@stat.purdue.edu>

References


Examples
```r
## 1-D estimate: Buffalo snowfall
data(buffalo)
buff.fit <- ssden(~buffalo, domain=data.frame(buffalo=c(0,150)))
plot(xx=seq(0,150,len=101),dssden(buff.fit,xx),type="l")
plot(xx,pssden(buff.fit,xx),type="l")
plot(qq=seq(0,1,len=51),qssden(buff.fit,qq),type="l")
## Clean up
## Not run: rm(buffalo,buff.fit,xx,qq)
dev.off()
## End(Not run)

## 2-D with triangular domain: AIDS incubation
data(aids)
## rectangular quadrature
quad.pt <- expand.grid(incu=((1:40)-.5)/40*100,infu=((1:40)-.5)/40*100)
quad.pt <- quad.pt[quad.pt$incu<quad.pt$infu,]
quad.wt <- rep(1,nrow(quad.pt))
quad.wt[quad.pt$incu>=quad.pt$infu] <- .5
quad.wt <- quad.wt/sum(quad.wt)*5e3
```
## additive model (pre-truncation independence)

```r
aids.fit <- ssden(~incu+infe,data=aids,subset=age>=60,
    domain=data.frame(incu=c(0,100),infe=c(0,100)),
    quad=list(pt=quad.pt,wt=quad.wt))
```

## conditional (marginal) density of infe

```r
jk <- cddssden(aids.fit,xx=seq(0,100,len=51),data.frame(infe=50))
plot(xx,jk$pdf,type="l")
```

## conditional (marginal) quantiles of infe (TIME-CONSUMING)

```r
# Not run:
cqssden(aids.fit,c(.05,.25,.5,.75,.95),data.frame(infe=50))
```

## End(Not run)

## Clean up

```r
# Not run: rm(aids,quad.pt,quad.wt,aids.fit,jk,xx)
dev.off()
```

## One factor plus one vector

```r
data(gastric)
gastric$trt
fit <- ssden(~futime*trt,data=gastric)
```

## conditional density

```r
cdssden(fit,c("1","2"),cond=data.frame(futime=150))
```

## conditional quantiles

```r
cqssden(fit,c(.05,.25,.5,.75,.95),data.frame(trt="1"))
```

## Clean up

```r
# Not run: rm(gastric,fit)
```

## Sampling bias

```r
# (X,T) is truncated to T<X<1 for T~U(0,1), so X is length-biased
rbias <- function(n) {
    t <- runif(n)
    x <- rnorm(n,.5,.15)
    ok <- (x>t)&(x<1)
    while(m=sum(ok)) {
        t[ok] <- runif(m)
        x[ok] <- rnorm(m,.5,.15)
        ok <- (x>t)&(x<1)
    }
    cbind(x,t)
}
x <- rbias(100)
x <- xt[,1]; t <- xt[,2]
```

## length-biased

```r
bias1 <- list(t=1,wt=1,fun=function(t,x){x[,]})
fit1 <- ssden(~x, domain=list(x=c(0,1)), bias=bias1)
plot(xx=seq(0,1,len=100), dssden(fit1,xx), type="1")
```

## truncated

```r
bias2 <- list(t=t,wt=rep(1/100,100),fun=function(t,x){x[,]})
fit2 <- ssden(~x, domain=list(x=c(0,1)), bias=bias2)
plot(xx,dssden(fit2,xx), type="1")
```

## Clean up

```r
# Not run: rm(rbias,xt,x,t,bias1,fit1,bias2,fit2)
```
Estimating Hazard Function Using Smoothing Splines

Description

Estimate hazard function using smoothing spline ANOVA models. The symbolic model specification via formula follows the same rules as in \texttt{lm}, but with the response of a special form.

Usage

\begin{verbatim}
sshzd(formula, type=NULL, data=list(), alpha=1.4, weights=NULL, subset, offset, na.action=na.omit, partial=NULL, id.basis=NULL, nbasis=NULL, seed=NULL, random=NULL, prec=1e-7, maxiter=30, skip.iter=FALSE)

sshzd1(formula, type=NULL, data=list(), alpha=1.4, weights=NULL, subset, na.action=na.omit, rho="marginal", partial=NULL, id.basis=NULL, nbasis=NULL, seed=NULL, random=NULL, prec=1e-7, maxiter=30, skip.iter=FALSE)
\end{verbatim}

Arguments

- \texttt{formula}: Symbolic description of the model to be fit, where the response is of the form \texttt{Surv(futime,status,start=0)}.
- \texttt{type}: List specifying the type of spline for each variable. See \texttt{mkterm} for details.
- \texttt{data}: Optional data frame containing the variables in the model.
- \texttt{alpha}: Parameter defining cross-validation score for smoothing parameter selection.
- \texttt{weights}: Optional vector of counts for duplicated data.
- \texttt{subset}: Optional vector specifying a subset of observations to be used in the fitting process.
- \texttt{offset}: Optional offset term with known parameter 1.
- \texttt{na.action}: Function which indicates what should happen when the data contain NAs.
- \texttt{partial}: Optional symbolic description of parametric terms in partial spline models.
- \texttt{id.basis}: Index of observations to be used as "knots."
- \texttt{nbasis}: Number of "knots" to be used. Ignored when \texttt{id.basis} is specified.
- \texttt{seed}: Seed to be used for the random generation of "knots." Ignored when \texttt{id.basis} is specified.
- \texttt{random}: Input for parametric random effects (frailty) in nonparametric mixed-effect models. See \texttt{mkran} for details.
- \texttt{prec}: Precision requirement for internal iterations.
- \texttt{maxiter}: Maximum number of iterations allowed for internal iterations.
- \texttt{skip.iter}: Flag indicating whether to use initial values of theta and skip theta iteration. See \texttt{ssanova} for notes on skipping theta iteration.
- \texttt{rho}: Choice of rho function for \texttt{sshzd1}: "marginal" or "weibull".
Details

The model specification via formula is for the log hazard. For example, \( \text{Suve}(t,d) \sim t*u \) prescribes a model of the form

\[
\log(f(t, u)) = C + g_t(t) + g_u(u) + g_{t,u}(t, u)
\]

with the terms denoted by "1", "t", "u", and "t:u". Replacing \( t*u \) by \( t+u \) in the formula, one gets a proportional hazard model with \( g_{t,u} = 0 \).

\( \text{sshzd} \) takes standard right-censored lifetime data, with possible left-truncation and covariates; in \( \text{Surv}(\text{futime, status, start=0}) \sim \ldots \), \text{futime} is the follow-up time, \text{status} is the censoring indicator, and \text{start} is the optional left-truncation time. The main effect of \text{futime} must appear in the model terms specified via \ldots .

Parallel to those in a \( \text{ssanova} \) object, the model terms are sums of unpenalized and penalized terms. Attached to every penalized term there is a smoothing parameter, and the model complexity is largely determined by the number of smoothing parameters.

The selection of smoothing parameters is through a cross-validation mechanism described in Gu (2002, Sec. 7.2), with a parameter \( \text{alpha} \); \( \text{alpha}=1 \) is "unbiased" for the minimization of Kullback-Leibler loss but may yield severe undersmoothing, whereas larger \( \text{alpha} \) yields smoother estimates.

A subset of the observations are selected as "knots." Unless specified via \text{id.basis} or \text{nbasis}, the number of "knots" \( q \) is determined by \( \max(30, 10n^{2/9}) \), which is appropriate for the default cubic splines for numerical vectors.

Value

\( \text{sshzd} \) returns a list object of class "sshzd". \( \text{sshzd1} \) returns a list object of class c("sshzd1", "sshzd").

\( \text{hzdrate.sshzd} \) can be used to evaluate the estimated hazard function. \( \text{hzdcurve.sshzd} \) can be used to evaluate hazard curves with fixed covariates. \( \text{survexp.sshzd} \) can be used to calculated estimated expected survival.

The method \( \text{project.sshzd} \) can be used to calculate the Kullback-Leibler projection of "sshzd" objects for model selection; \( \text{project.sshzd1} \) can be used to calculate the square error projection of "sshzd1" objects.

Note

The function \( \text{Surv}(\text{futime, status, start=0}) \) is defined and parsed inside \( \text{sshzd} \), not quite the same as the one in the \text{survival} package.

Integration on the time axis is done by the 200-point Gauss-Legendre formula on \( c(\min(\text{start}), \max(\text{futime})) \), returned from \text{gauss.quad}.

\( \text{sshzd1} \) can be up to 50 times faster than \( \text{sshzd} \), at the cost of performance degradation.

The results may vary from run to run. For consistency, specify \text{id.basis} or set seed.

Author(s)

Chong Gu, <chong@stat.purdue.edu>
References


Examples

```r
## Model with interaction
data(gastric)
gastric.fit <- sshzd(Surv(futime,status)-futime*trt,data=gastric)
## exp(-Lambda(600)), exp(-(Lambda(1200)-Lambda(600))), and exp(-Lambda(1200))
surve.exp.sshzd(gastric.fit,c(600,1200,1200),data.frame(trt=as.factor(1)),c(0,600,0))
## Clean up
## Not run: rm(gastric,gastric.fit)
dev.off()
## End(Not run)

## THE FOLLOWING EXAMPLE IS TIME-CONSUMING
## Proportional hazard model
## Not run:
data(stan)
stan.fit <- sshzd(Surv(futime,status)-futime+age,data=stan)
## Evaluate fitted hazard
hzdrate.sshzd(stan.fit,data.frame(futime=c(10,20),age=c(20,30)))
## Plot lambda(t,age=20)
tt <- seq(0,60,leng=101)
hh <- hzdcurve.sshzd(stan.fit,tt,data.frame(age=20))
plot(tt,hh,type="l")
## Clean up
rm(stan,stan.fit,tt,hh)
dev.off()
## End(Not run)
```

---

### ssllrm

**Fitting Smoothing Spline Log-Linear Regression Models**

**Description**

Fit smoothing spline log-linear regression models. The symbolic model specification via formula follows the same rules as in `lm`. 
Usage

ssllrm(formula, response, type=NULL, data=list(), weights, subset,
na.action=na.omit, alpha=1, id.basis=NULL, nbasis=NULL,
seed=NULL, random=NULL, prec=1e-7, maxiter=30, skip.iter=FALSE)

Arguments

  formula   Symbolic description of the model to be fit.
  response  Formula listing response variables.
  type      List specifying the type of spline for each variable. See mkterm for details.
  data      Optional data frame containing the variables in the model.
  weights   Optional vector of weights to be used in the fitting process.
  subset    Optional vector specifying a subset of observations to be used in the fitting process.
  na.action Function which indicates what should happen when the data contain NAs.
  alpha     Parameter modifying GCV or Mallow's CL; larger absolute values yield smoother fits; negative value invokes a stable and more accurate GCV/CL evaluation algorithm but may take two to five times as long. Ignored when method="m" are specified.
  id.basis  Index designating selected "knots".
  nbasis    Number of "knots" to be selected. Ignored when id.basis is supplied.
  seed      Seed to be used for the random generation of "knots". Ignored when id.basis is supplied.
  random    Input for parametric random effects in nonparametric mixed-effect models. See mkran for details.
  prec      Precision requirement for internal iterations.
  maxiter   Maximum number of iterations allowed for internal iterations.
  skip.iter Flag indicating whether to use initial values of theta and skip theta iteration. See ssanova for notes on skipping theta iteration.

Details

The model is specified via formula and response, where response lists the response variables. For example, ssllrm(~y1*y2*x, -y1*y2) prescribe a model of the form

$$\log f(y_1, y_2|x) = g_1(y_1) + g_2(y_2) + g_{12}(y_1, y_2) + g_{x1}(x, y_1) + g_{x2}(x, y_2) + g_{x12}(x, y_1, y_2) + C(x)$$

with the terms denoted by "y1", "y2", "y1:y2", "y1:x", "y2:x", and "y1:y2:x"; the term(s) not involving response(s) are removed and the constant C(x) is determined by the fact that a conditional density integrates (adds) to one on the y axis.

The model terms are sums of unpenalized and penalized terms. Attached to every penalized term there is a smoothing parameter, and the model complexity is largely determined by the number of smoothing parameters.

A subset of the observations are selected as "knots." Unless specified via id.basis or nbasis, the number of "knots" q is determined by $max(30, 10n^{2/9})$, which is appropriate for the default cubic splines for numerical vectors.
Value

`ssllrm` returns a list object of class "ssllrm".

The method `predict.ssllrm` can be used to evaluate \( f(y|x) \) at arbitrary \( x \), or contrasts of \( \log(f(y|x)) \) such as the odds ratio along with standard errors. The method `project.ssllrm` can be used to calculate the Kullback-Leibler projection for model selection.

Note

The responses, or y-variables, must be factors, and there must be at least one numerical x’s. For response, there is no difference between \(~y_1+y_2\) and \(~y_1*y_2\).

The results may vary from run to run. For consistency, specify `id.basis` or set `seed`.

Author(s)

Chong Gu, <chong@stat.purdue.edu>

References


Examples

```r
## Simulate data

test <- function(x)
  {.3*(1e6*(x^11*(1-x)^6)+1e4*(x^3*(1-x)^10))-2}
x <- (0:100)/100
p <- 1-1/(1+exp(test(x)))
y <- rbinom(x,3,p)
y1 <- as.ordered(y)
y2 <- as.factor(rbinom(x,1,p))

## Fit model
fit <- ssllrm(~y1*y2*x, ~y1+K~yR)

## Evaluate \( f(y|x) \)
est <- predict(fit,data.frame(x=x),
  data.frame(y=as.factor(0:3),y2=as.factor(rep(0,4))))

## \( f(y|x) \) at all y values (fit$qd.pt)
est <- predict(fit,data.frame(x=x))

## Evaluate contrast of log \( f(y|x) \)
est <- predict(fit,data.frame(x=x),odds=c(-1,5,5,0),
  data.frame(y=as.factor(0:3),y2=as.factor(rep(0,4)),se=TRUE)

## Odds ratio \( \log(f(0,0|x)/f(3,0|x)) \)
est <- predict(fit,data.frame(x=x),odds=c(1,-1),
  data.frame(y=as.factor(c(0,3)),y2=as.factor(c(0,1))),se=TRUE)

## KL projection
```
stan  Stanford Heart Transplant Data

Description
Survival of patients from the Stanford heart transplant program.

Usage
data(stan)

Format
A data frame containing 184 observations on the following variables.

- **time**: Follow-up time after transplant, in days.
- **status**: Censoring status.
- **age**: Age at transplant.
- **futime**: Square root of time.

Source

summary.gssanova  Assessing Smoothing Spline ANOVA Fits with Non-Gaussian Responses

Description
Calculate various summaries of smoothing spline ANOVA fits with non-Gaussian responses.

Usage
```r
## S3 method for class 'gssanova'
summary(object, diagnostics=FALSE, ...)
```
Arguments

object Object of class "gssanova".
diagnostics Flag indicating if diagnostics are required.
... Ignored.

Details

Similar to the iterated weighted least squares fitting of glm, penalized likelihood regression fit can be calculated through iterated penalized weighted least squares.
The diagnostics are based on the "pseudo" Gaussian response model behind the weighted least squares problem at convergence.

Value

summary.gssanova returns a list object of class "summary.gssanova" consisting of the following components. The entries pi, kappa, cosines, and roughness are only calculated if diagnostics=TRUE.
call Fitting call.
family Error distribution.
alpha Parameter used to define cross-validation in model fitting.
fitted Fitted values on the link scale.
dispersion Assumed or estimated dispersion parameter.
residuals Working residuals on the link scale.
rss Residual sum of squares.
dev.resid Deviance residuals.
deviance Deviance of the fit.
dev.null Deviance of the null model.
penalty Roughness penalty associated with the fit.
pi "Percentage decomposition" of "explained variance" into model terms.
kappa Concurvity diagnostics for model terms. Virtually the square roots of variance inflation factors of a retrospective linear model.
cosines Cosine diagnostics for practical significance of model terms.
roughness Percentage decomposition of the roughness penalty penalty into model terms.

Author(s)

Chong Gu, <chong@stat.purdue.edu>

References


See Also

Fitting function gssanova and methods predict.gssanova, project.gssanova, fitted.gssanova.
Assessing Smoothing Spline ANOVA Fits with Non-Gaussian Responses

Description

Calculate various summaries of smoothing spline ANOVA fits with non-Gaussian responses.

Usage

```r
## S3 method for class 'gssanova0'
summary(object, diagnostics=FALSE, ...)
```

Arguments

- `object` Object of class "gssanova".
- `diagnostics` Flag indicating if diagnostics are required.
- `...` Ignored.

Details

Similar to the iterated weighted least squares fitting of `glm`, penalized likelihood regression fit can be calculated through iterated penalized weighted least squares.

The diagnostics are based on the "pseudo" Gaussian response model behind the weighted least squares problem at convergence.

Value

`summary.gssanova0` returns a list object of class "summary.gssanova0" consisting of the following components. The entries `pi`, `kappa`, `cosines`, and `roughness` are only calculated if `diagnostics=TRUE`.

- `call` Fitting call.
- `family` Error distribution.
- `method` Method for smoothing parameter selection.
- `dispersion` Assumed or estimated dispersion parameter.
- `iter` Number of performance-oriented iterations performed.
- `fitted` Fitted values on the link scale.
- `residuals` Working residuals on the link scale.
- `rss` Residual sum of squares.
- `dev.resid` Deviance residuals.
- `deviance` Deviance of the fit.
- `dev.null` Deviance of the null model.
alpha Estimated size for family="nbinomial" with one column responses. Estimated inverse scale of log life time for family="nbinomial", "lognorm", or "loglogis".

penalty Roughness penalty associated with the fit.

pi "Percentage decomposition" of "explained variance" into model terms.

kappa Concurvity diagnostics for model terms. Virtually the square roots of variance inflation factors of a retrospective linear model.

cosines Cosine diagnostics for practical significance of model terms.

roughness Percentage decomposition of the roughness penalty penalty into model terms.

Author(s)

Chong Gu, <chong@stat.purdue.edu>

References


See Also

Fitting function gssanova and methods predict.ssanova, fitted.gssanova.

summary.ssanova Assessing Smoothing Spline ANOVA Fits

Description

Calculate various summaries of smoothing spline ANOVA fits.

Usage

```r
## S3 method for class 'ssanova'
summary(object, diagnostics=FALSE, ...)
## S3 method for class 'ssanova0'
summary(object, diagnostics=FALSE, ...)
## S3 method for class 'ssanova9'
summary(object, diagnostics=FALSE, ...)
```

Arguments

- `object` Object of class "ssanova".
- `diagnostics` Flag indicating if diagnostics are required.
- `...` Ignored.
Value

summary.ssanova returns a list object of \texttt{class} "summary.ssanova" consisting of the following components. The entries \texttt{pi}, \texttt{kappa}, \texttt{cosines}, and \texttt{roughness} are only calculated if \texttt{diagnostics=TRUE}; see the reference below for details concerning the diagnostics.

- \texttt{call} Fitting call.
- \texttt{method} Method for smoothing parameter selection.
- \texttt{fitted} Fitted values.
- \texttt{residuals} Residuals.
- \texttt{sigma} Assumed or estimated error standard deviation.
- \texttt{r.squared} Fraction of "explained variance" by the fitted model.
- \texttt{rss} Residual sum of squares.
- \texttt{penalty} Roughness penalty associated with the fit.
- \texttt{pi} "Percentage decomposition" of "explained variance" into model terms.
- \texttt{kappa} Concurvity diagnostics for model terms. Virtually the square roots of variance inflation factors of a retrospective linear model.
- \texttt{cosines} Cosine diagnostics for practical significance of model terms.
- \texttt{roughness} Percentage decomposition of the roughness penalty \texttt{penalty} into model terms.

Author(s)

Chong Gu, \texttt{<chong@stat.purdue.edu>}

References


See Also

Fitting functions \texttt{ssanova, ssanova@} and methods \texttt{predict.ssanova, project.ssanova, fitted.ssanova}.

---

wesdr Progression of Diabetic Retinopathy

Description

Data derived from the Wisconsin Epidemiological Study of Diabetic Retinopathy.

Usage

data(wesdr)

Format

A data frame containing 669 observations on the following variables.
dur  Duration of diabetes at baseline, in years.
gly  Percent of glycosylated hemoglobin at baseline.
bmi  Body mass index at baseline.
ret  Binary indicator of retinopathy progression at first follow-up.

Source


References


Index

*Topic datasets
  aids, 2
  bacteriuria, 3
  buffalo, 3
  clim, 6
  ColoCan, 6
  esc, 10
  eyetrack, 10
  gastric, 12
  LakeAcidity, 20
  NO2, 22
  nox, 22
  ozone, 23
  penny, 24
  Sachs, 31
  stan, 53
  wesdr, 57

*Topic distribution
  cdsscden, 4
  cdssden, 5
  dsscden, 8
  dssden, 9
  sscden, 40
  ssden, 44

*Topic htest
  project, 29

*Topic math
  gauss.quad, 12
  nlm0, 21
  smolyak, 31

*Topic models
  cdsscden, 4
  cdssden, 5
  dsscden, 8
  dssden, 9
  fitted.ssanova, 11
  gssanova, 13
  gssanova0, 16
  hzdrate.sshzd, 19
  predict.ssanova, 24
  predict.sscox, 26
  predict.ssllrm, 27
  print, 28
  project, 29
  sssanova, 32
  sssanova0, 35
  sssanova9, 37
  sscden, 40
  sscox, 42
  ssden, 44
  sshzd, 48
  ssllrm, 50
  summary.gssanova, 53
  summary.gssanova0, 55
  summary.sssanova, 56

*Topic regression
  fitted.ssanova, 11
  gssanova, 13
  gssanova0, 16
  predict.sssanova, 24
  predict.ssscox, 26
  predict.ssllrm, 27
  sssanova, 32
  sssanova0, 35
  sssanova9, 37
  ssllrm, 50
  summary.gssanova, 53
  summary.gssanova0, 55
  summary.sssanova, 56

*Topic smooth
  cdsscden, 4
  cdssden, 5
  dsscden, 8
  dssden, 9
  fitted.ssanova, 11
  gssanova, 13
  gssanova0, 16
  hzdrate.sshzd, 19
  predict.sssanova, 24
  predict.ssscox, 26
  predict.ssllrm, 27
predict.ssanova, 24  
predict.sscox, 26  
predict.ssllrm, 27  
print, 28  
project, 29  
ssanova, 32  
ssanova0, 35  
ssanova9, 37  
sscden, 40  
sscox, 42  
ssden, 44  
sshzd, 48  
ssllrm, 50  
summary.gssanova, 53  
summary.gssanova0, 55  
summary.ssanova, 56  

*Topic survival*  
hzdrate.sshzd, 19  
predict.sscox, 26  
sscox, 42  
sshzd, 48  
aids, 2  
bacteriuria, 3  
buffalo, 3  
cdsscden, 4, 8, 41  
cdssden, 5, 9, 45  
class, 54, 55, 57  
clim, 6  
ColoCan, 6  
cpsscden, 41  
cpsscden (cdsscden), 4  
cpssden, 45  
cpssden (cdssden), 5  
cqsscden, 41  
cqsscden (cdsscden), 4  
cqssden, 45  
cqssden (cdssden), 5  
d.s.scden (dsscden), 8  
d.s.scden1 (dsscden), 8  
d.s.sden (dssden), 9  
d.s.sden1 (dssden), 9  
dsscden, 5, 8, 41  
dssden, 6, 9, 43  
esc, 10  
eyetrack, 10  
fitted.gssanova, 14, 17, 54, 56  
fitted.gssanova (fitted.ssanova), 11  
fitted.ssanova, 11, 25, 33, 36, 38, 57  
gastric, 12  
gauss.quad, 12, 41, 46, 49  
glm, 13, 14, 16, 18, 54, 55  
gssanova, 13, 18, 25, 29, 31, 54  
gssanova0, 14, 16, 18, 25, 29, 56  
gssanova1, 18  
gssanova1 (gssanova0), 16  
hzcurve.sshzd, 49  
hzcurve.sshzd (hzdrate.sshzd), 19  
hzdrate.sshzd, 19, 49  
LakeAcidity, 20  
lm, 13, 16, 32, 35, 37, 40, 42, 44, 48, 50  
mkcov, 38  
mkran, 13, 17, 18, 33, 36, 43, 48, 51  
mkterm, 13, 16, 32, 35, 37, 40, 42, 44, 48, 51  
NegBinomial, 14, 17  
nlm, 34, 39  
nlm0, 21  
NO2, 22  
nox, 22  
ozone, 23  
para arma (ssanova9), 37  
penny, 24  
predict.lm, 4, 5, 8, 9  
predict.ssanova, 14, 17, 24, 33, 38, 54, 57  
predict.ssanova0, 17, 36, 56  
predict.ssanova0 (predict.ssanova), 24  
predict.sscox, 26, 43  
predict.ssllrm, 27, 52  
predict1 (predict.ssanova), 24  
print, 28  
project, 18, 29, 36  
project.gssanova, 14, 54  
project.ssanova, 25, 33, 57  
project.ssanova9, 38  
project.sscden, 41  
project.sscden1, 41  
project.sscox, 27, 43
INDEX

project.ssd, 45
project.ssd1, 45
project.sshzd, 49
project.sshzd1, 49
project.sslr, 52
pssc, 41
pssc (dssc), 8
pssd, 45
pssd (dssd), 9

qssc, 41
qssc (dssc), 8
qssd, 45
qssd (dssd), 9

residuals.gssanova, 14, 17
residuals.gssanova (fitted.ssanova), 11
residuals.ssanova, 33, 36, 38
residuals.ssanova (fitted.ssanova), 11

Sachs, 31
smolyak, 31
smolyak.quad, 41, 45, 46
ssanova, 13, 17, 25, 29, 31, 32, 36, 40, 43, 45, 48, 49, 51, 57
ssanova0, 17, 25, 29, 33, 34, 35, 38, 57
ssanova9, 37
ss, 5, 8, 40
ss (sscden), 40
sscox, 27, 42
ssd, 6, 9, 29, 31, 44
ssd1 (ssd), 44
sshd, 20, 29, 31, 48
sshd1, 31
sshd1 (sshd), 48
sslr, 28, 29, 50
stan, 53
summary.gssanova, 14, 17, 25, 29, 53
summary.gssanova0, 17, 25, 29, 55
summary.ssanova, 25, 29, 33, 56
summary.ssanova0, 36
summary.ssanova0 (summary.ssanova), 56
summary.ssanova9, 38
summary.ssanova9 (summary.ssanova), 56
survexp.sshzd, 49
survexp.sshzd (hzdrate.sshzd), 19

wesdr, 57