The logspline Package

September 28, 2007

Version 2.0.4

Date 2007-09-28

Title Logspline density estimation routines

Author Charles Kooperberg <clk@fhcrc.org>

Maintainer Charles Kooperberg <clk@fhcrc.org>

Description Routines for the logspline density estimation. oldlogspline uses the same algorithm as the logspline 1.0.x package - the Kooperberg and Stone (1992) algorithm (with an improved interface). The recommended routine logspline uses an algorithm from Stone et al (1997).

License GPL version 2 or newer

R topics documented:

dlogspline ......................................................... 1
doldlogspline ..................................................... 3
logspline ......................................................... 4
oldlogspline ....................................................... 6
plot.logspline .................................................. 8
plot.oldlogspline .............................................. 10
summary.logspline ............................................. 11
summary.oldlogspline .......................................... 12
unstrip ......................................................... 13

Index 15
Description

Density (dlogspline), cumulative probability (plogspline), quantiles (qlogspline), and random samples (rlogspline) from a logspline density that was fitted using the 1997 knot addition and deletion algorithm (logspline). The 1992 algorithm is available using the oldlogspline function.

Usage

dlogspline(q, fit)
plogspline(q, fit)
qlogspline(p, fit)
rlogspline(n, fit)

Arguments

q  vector of quantiles. Missing values (NAs) are allowed.
p  vector of probabilities. Missing values (NAs) are allowed.
n  sample size. If length(n) is larger than 1, then length(n) random values are returned.
fit  logspline object, typically the result of logspline.

Details

Elements of q or p that are missing will cause the corresponding elements of the result to be missing.

Value

Densities (dlogspline), probabilities (plogspline), quantiles (qlogspline), or a random sample (rlogspline) from a logspline density that was fitted using knot addition and deletion.

Author(s)

Charles Kooperberg (clk@fhcrc.org).

References


doldlogspline

See Also

logspline, plot.logspline, summary.logspline, oldlogspline.

Examples

x <- rnorm(100)
fit <- logspline(x)
qq <- qlogspline((1:99)/100, fit)
plot(qnorm((1:99)/100), qq) # qq plot of the fitted density
pp <- plogspline((-250:250)/100, fit)
plot((-250:250)/100, pp, type = "l")
lines((-250:250)/100, pnorm((-250:250)/100)) # asses the fit of the distribution
dd <- dlogspline((-250:250)/100, fit)
plot((-250:250)/100, dd, type = "l")
lines((-250:250)/100, dnorm((-250:250)/100)) # asses the fit of the density
rr <- rlogspline(100, fit) # random sample from fit

----------

doldlogspline  Logspline Density Estimation - 1992 version

Description

Probability density function (doldlogspline), distribution function (poldlogspline), quantiles (qoldlogspline), and random samples (roldlogspline) from a logspline density that was fitted using the 1992 knot deletion algorithm (oldlogspline). The 1997 algorithm using knot deletion and addition is available using the logspline function.

Usage

doldlogspline(q, fit)
poldlogspline(q, fit)
qoldlogspline(p, fit)
roldlogspline(n, fit)

Arguments

q  vector of quantiles. Missing values (NAs) are allowed.
p  vector of probabilities. Missing values (NAs) are allowed.
n  sample size. If length(n) is larger than 1, then length(n) random values are returned.
fit  oldlogspline object, typically the result of oldlogspline.

Details

Elements of q or p that are missing will cause the corresponding elements of the result to be missing.
Value

Densities (doldlogspline), probabilities (poldlogspline), quantiles (qoldlogspline), or a random sample (roldlogspline) from an oldlogspline density that was fitted using knot deletion.

Author(s)

Charles Kooperberg (clk@fhcrc.org).

References


See Also

logspline, oldlogspline, plot.oldlogspline, summary.oldlogspline

Examples

```r
x <- rnorm(100)
fit <- oldlogspline(x)
qq <- qoldlogspline((1:99)/100, fit)
plot(qnorm((1:99)/100), qq)  # qq plot of the fitted density
pp <- poldlogspline((-250:250)/100, fit)
plot((-250:250)/100, pp, type = "l")
lines((-250:250)/100, pnorm((-250:250)/100))  # assess the fit of the distribution
dd <- doldlogspline((-250:250)/100, fit)
plot((-250:250)/100, dd, type = "l")
lines((-250:250)/100, dnorm((-250:250)/100))  # assess the fit of the density
rr <- roldlogspline(100, fit)  # random sample from fit
```

---

# logspline

**Logspline Density Estimation**

**Description**

Fits a logspline density using splines to approximate the log-density using the 1997 knot addition and deletion algorithm (logspline). The 1992 algorithm is available using the oldlogspline function.

**Usage**

```r
logspline(x, lbound, ubound, maxknots = 0, knots, nknots = 0, penalty, silent = TRUE, mind = -1)
```
**Arguments**

x  
data vector. The data needs to be uncensored. oldlogspline can deal with right- left- and interval-censored data.

lboun, uboun  
lower/upper bound for the support of the density. For example, if there is a priori knowledge that the density equals zero to the left of 0, and has a discontinuity at 0, the user could specify lboun = 0. However, if the density is essentially zero near 0, one does not need to specify lboun.

maxknots  
the maximum number of knots. The routine stops adding knots when this number of knots is reached. The method has an automatic rule for selecting max-knots if this parameter is not specified.

knots  
ordered vector of values (that should cover the complete range of the observations), which forces the method to start with these knots. Overrules knots. If knots is not specified, a default knot-placement rule is employed.

nknots  
forces the method to start with nknots knots. The method has an automatic rule for selecting nknots if this parameter is not specified.

penalty  
the parameter to be used in the AIC criterion. The method chooses the number of knots that minimizes $-2 \times \log\text{likelihood} + \text{penalty} \times (\text{number of knots} - 1)$. The default is to use a penalty parameter of penalty = log(samplesize) as in BIC. The effect of this parameter is summarized in summary.logspline.

silent  
should diagnostic output be printed?

mind  
minimum distance, in order statistics, between knots.

**Value**

Object of the class logspline, that is intended as input for plot.logspline (summary plots), summary.logspline (fitting summary), dlogspline (densities), plogspline (probabilities), qlogspline (quantiles), rlogspline (random numbers from the fitted distribution).

The object has the following members:

call  
the command that was executed.

nknots  
the number of knots in the model that was selected.

coef.pol  
coefficients of the polynomial part of the spline. The first coefficient is the constant term and the second is the linear term.

coef.kts  
coefficients of the knots part of the spline. The k-th element is the coefficient of $(x - t(k))^3_+$ (where $x^3_+$ means the positive part of the third power of x, and $t(k)$ means knot k).

knots  
vector of the locations of the knots in the logspline model.

maxknots  
the largest number of knots minus one considered during fitting (i.e. with maxknots = 6 the maximum number of knots is 5).

penalty  
the penalty that was used.

bound  
first element: 0 - lboun was − inf 1 it was something else; second element: lboun, if specified; third element: 0 - uboun was inf, 1 it was something else; fourth element: uboun, if specified.
samples the sample size.
logl range range of the input data.
mind minimum distance in order statistics between knots required during fitting (the actual minimum distance may be much larger).

Author(s)
Charles Kooperberg (clk@fhcrc.org).

References


See Also
plot.logspline, summary.logspline, dlogspline, plogspline, qlogspline, rlogspline, oldlogspline.

Examples
```r
y <- rnorm(100)
fit <- logspline(y)
plot(fit)
# as (4 == length(-2, -1, 0, 1, 2) -1), this forces these initial knots,
# and does no knot selection
fit <- logspline(y, knots = c(-2, -1, 0, 1, 2), maxknots = 4, penalty = 0)
```

oldlogspline  Logspline Density Estimation - 1992 version

Description
Fits a logspline density using splines to approximate the log-density using the 1992 knot deletion algorithm (oldlogspline). The 1997 algorithm using knot deletion and addition is available using the logspline function.

Usage
```
oldlogspline(uncensored, right, left, interval, lbound, ubound, nknots, knots, penalty, delete = TRUE)
```
Arguments

uncensored vector of uncensored observations from the distribution whose density is to be estimated. If there are no uncensored observations, this argument can be omitted. However, either uncensored or interval must be specified.

right vector of right censored observations from the distribution whose density is to be estimated. If there are no right censored observations, this argument can be omitted.

left vector of left censored observations from the distribution whose density is to be estimated. If there are no left censored observations, this argument can be omitted.

interval two column matrix of lower and upper bounds of observations that are interval censored from the distribution whose density is to be estimated. If there are no interval censored observations, this argument can be omitted.

lbound, ubound lower/upper bound for the support of the density. For example, if there is a priori knowledge that the density equals zero to the left of 0, and has a discontinuity at 0, the user could specify lbound = 0. However, if the density is essentially zero near 0, one does not need to specify lbound. The default for lbound is \(-\infty\) and the default for ubound is \(\infty\).

nknots forces the method to start with nknots knots (delete = TRUE) or to fit a density with nknots knots (delete = FALSE). The method has an automatic rule for selecting nknots if this parameter is not specified.

knots ordered vector of values (that should cover the complete range of the observations), which forces the method to start with these knots (delete = TRUE) or to fit a density with these knots (delete = FALSE). Overrules nknots. If knots is not specified, a default knot-placement rule is employed.

penalty the parameter to be used in the AIC criterion. The method chooses the number of knots that minimizes \(-2 \times \log\text{likelihood} + \text{penalty} \times (\text{number of knots} - 1)\). The default is to use a penalty parameter of penalty = log(samplesize) as in BIC. The effect of this parameter is summarized in summary.oldlogspline.

delete should stepwise knot deletion be employed?

Value

Object of the class oldlogspline, that is intended as input for plot.oldlogspline, summary.oldlogspline, oldlogspline (densities), poldlogspline (probabilities), qoldlogspline (quantiles), roldlogspline (random numbers from the fitted distribution).

The object has the following members:

call the command that was executed.

knots vector of the locations of the knots in the oldlogspline model. old

coef coefficients of the spline. The first coefficient is the constant term, the second is the linear term and the k-th \((k > 2)\) is the coefficient of \((x - t(k - 2))^{3}\) (where \(x^{3}_{+}\) means the positive part of the third power of \(x\), and \(t(k - 2)\) means...
knot \( k - 2 \). If a coefficient is zero the corresponding knot was deleted from the model.

- **bound**: first element: 0 - lbound was \(-\infty\) it was something else; second element: lbound, if specified; third element: 0 - ubound was \(\infty\), 1 it was something else; fourth element: ubound, if specified.

- **logl**: the \( k \)-th element is the log-likelihood of the fit with \( k + 2 \) knots.

- **penalty**: the penalty that was used.

- **sample**: the sample size that was used.

- **delete**: was stepwise knot deletion employed?

**Author(s)**

Charles Kooperberg (clk@fhcrc.org).

**References**


**See Also**

`logspline`, `oldlogspline`, `plot.oldlogspline`, `summary.oldlogspline`, `doldlogspline`, `poldlogspline`, `qoldlogspline`, `roldlogspline`.

**Examples**

```r
# A simple example
y <- rnorm(100)
fit <- oldlogspline(y)
plot(fit)
# An example involving censoring and a lower bound
y <- rlnorm(1000)
censoring <- rexp(1000) * 4
delta <- 1 * (y <= censoring)
y[delta == 0] <- censoring[delta == 0]
fit <- oldlogspline(y[delta == 1], y[delta == 0], lbound = 0)
```
plot.logspline

Logspline Density Estimation

Description

Plots a \texttt{logspline} density, distribution function, hazard function or survival function from a \texttt{logspline} density that was fitted using the 1997 knot addition and deletion algorithm (\texttt{logspline}). The 1992 algorithm is available using the \texttt{oldlogspline} function.

Usage

\begin{verbatim}
plot.logspline(x, n = 100, what = "d", add = FALSE, xlim, xlab = ",
ylab = ", type = "l", ...
\end{verbatim}

Arguments

\begin{itemize}
\item \texttt{x} \texttt{logspline} object, typically the result of \texttt{logspline}.
\item \texttt{n} the number of equally spaced points at which to plot the density.
\item \texttt{what} what should be plotted: "d" (density), "p" (distribution function), "s" (survival function) or "h" (hazard function).
\item \texttt{add} should the plot be added to an existing plot.
\item \texttt{xlim} range of data on which to plot. Default is from the 1th to the 99th percentile of the density, extended by 10\% on each end.
\item \texttt{xlab,ylab} labels plotted on the axes.
\item \texttt{type} type of plot.
\item \texttt{...} other plotting options, as desired
\end{itemize}

Details

This function produces a plot of a \texttt{logspline} fit at \texttt{n} equally spaced points roughly covering the support of the density. (Use \texttt{xlim = c(from, to)} to change the range of these points.)

Author(s)

Charles Kooperberg (clk@fhcrc.org).

References


plot.oldlogspline

See Also

logspline, summary.logspline, dlogspline, plogspline, qlogspline, rlogspline, oldlogspline.

Examples

y <- rnorm(100)
fit <- logspline(y)
plot(fit)

plot.oldlogspline

Logspline Density Estimation - 1992 version

Description

Plots an oldlogspline density, distribution function, hazard function or survival function from a logspline density that was fitted using the 1992 knot deletion algorithm. The 1997 algorithm using knot deletion and addition is available using the logspline function.

Usage

plot.oldlogspline(x, n = 100, what = "d", xlim, xlab = "", ylab = "", type = "l", ...)

Arguments

x 
logspline object, typically the result of logspline.
n 
the number of equally spaced points at which to plot the density.
what 
what should be plotted: "d" (density), "p" (distribution function), "s" (survival function) or "h" (hazard function).
xlim 
range of data on which to plot. Default is from the 1th to the 99th percentile of the density, extended by 10% on each end.
xlab, ylab 
labels plotted on the axes.
type 
type of plot.
... other plotting options, as desired

Details

This function produces a plot of a oldlogspline fit at n equally spaced points roughly covering the support of the density. (Use xlim=c(from, to) to change the range of these points.)

Author(s)

Charles Kooperberg (clk@fhcrc.org).
References


See Also

logspline, oldlogspline, summary.oldlogspline, doldlogspline, poldlogspline, qoldlogspline, roldlogspline.

Examples

```r
y <- rnorm(100)
fit <- oldlogspline(y)
plot(fit)
```

summary.logspline  Logspline Density Estimation

Description

This function summarizes both the stepwise selection process of the model fitting by logspline, as well as the final model that was selected using AIC/BIC. A logspline object was fit using the 1997 knot addition and deletion algorithm. The 1992 algorithm is available using the oldlogspline function.

Usage

```r
summary.logspline(object, ...)
print.logspline(x, ...)
```

Arguments

- `object, x`  logspline object, typically the result of logspline
- `...`  other arguments are ignored.

Details

These function produce identical printed output. The main body is a table with five columns: the first column is a possible number of knots for the fitted model; the second column is the log-likelihood for the fit; the third column is $-2 \times \text{loglikelihood} + \text{penalty} \times (\text{number of knots} - 1)$, which is the AIC criterion; logspline selected the model with the smallest value of AIC;
the fourth and fifth columns give the endpoints of the interval of values of penalty that would yield the model with the indicated number of knots. (NAs imply that the model is not optimal for any choice of penalty.) At the bottom of the table the number of knots corresponding to the selected model is reported, as is the value of penalty that was used.

Author(s)

Charles Kooperberg (clk@fhcrc.org).

References


See Also

`logspline`, `plot.logspline`, `dlogspline`, `plogspline`, `qlogspline`, `rlogspline`, `oldlogspline`.

Examples

```r
y <- rnorm(100)
fit <- logspline(y)
summary(fit)
```

Description

This function summarizes both the stepwise selection process of the model fitting by `oldlogspline`, as well as the final model that was selected using AIC/BIC. A `logspline` object was fit using the 1992 knot deletion algorithm (`oldlogspline`). The 1997 algorithm using knot deletion and addition is available using the `logspline` function.

Usage

```r
summary.oldlogspline(object, ...)
print.oldlogspline(x, ...)
```

Arguments

- `object, x`: codeoldlogspline object, typically the result of `oldlogspline`
- `...`: other arguments are ignored.
Details

These function produces the same printed output. The main body is a table with five columns: the first column is a possible number of knots for the fitted model; the second column is the log-likelihood for the fit; the third column is $-2 \times \log\text{likelihood} + \text{penalty} \times (\text{number of knots} - 1)$, which is the AIC criterion; logspline selected the model with the smallest value of AIC; the fourth and fifth columns give the endpoints of the interval of values of penalty that would yield the model with the indicated number of knots. (NAs imply that the model is not optimal for any choice of penalty.) At the bottom of the table the number of knots corresponding to the selected model is reported, as is the value of penalty that was used.

Author(s)

Charles Kooperberg (clk@fhcrc.org).

References


See Also

logspline, oldlogspline, plot.oldlogspline, doldlogspline, poldlogspline, qoldlogspline, roldlogspline.

Examples

```r
y <- rnorm(100)
fit <- oldlogspline(y)
summary(fit)
```

unstrip

Reformat data as vector or matrix

Description

This function tries to convert a data.frame or a matrix to a no-frills matrix without labels, and a vector or time-series to a no-frills vector without labels.

Usage

unstrip(x)
unstrip

Arguments

x               one- or two-dimensional object.

Details

Many of the functions for `logspline, oldlogspline, lspec, polyclass, hare, heft,`
and `polymars` were written in the “before data.frame” era; `unstrip` attempts to keep all these
functions useful with more advanced input objects. In particular, many of these functions call `unstrip` before doing anything else.

Value

If `x` is two-dimensional a matrix without names, if `x` is one-dimensional a numerical vector

Author(s)

Charles Kooperberg (clk@fhcrc.org).

Examples

data(co2)
unstrip(co2)
data(iris)
unstrip(iris)
Index

*Topic classes
unstrip, 13

*Topic distribution
  dlogspline, 1
  doldlogspline, 3
  logspline, 4
  oldlogspline, 6
  plot.logspline, 8
  plot.oldlogspline, 10
  summary.logspline, 11
  summary.oldlogspline, 12

*Topic smooth
  dlogspline, 1
  doldlogspline, 3
  logspline, 4
  oldlogspline, 6
  plot.logspline, 8
  plot.oldlogspline, 10
  summary.logspline, 11
  summary.oldlogspline, 12

*Topic utilities
unstrip, 13

dlogspline, 1, 5, 6, 9, 12
oldlogspline, 3, 7, 8, 11, 13
hare, 14
heft, 14
logspline, 1–3, 4, 4, 6, 8–14
lspec, 14
oldlogspline, 1–4, 6, 6, 8–14
plogspline, 5, 6, 9, 12
plogspline(dlogspline), 1
plot.logspline, 2, 5, 6, 8, 12
plot.oldlogspline, 4, 7, 8, 10, 13
poldlogspline, 7, 8, 11, 13
poldlogspline(doldlogspline), 3
polyclass, 14
polymars, 14
print.logspline
  (summary.logspline), 11
print.oldlogspline
  (summary.oldlogspline), 12
qlogspline, 5, 6, 9, 12
qlogspline(dlogspline), 1
qoldlogspline, 7, 8, 11, 13
qoldlogspline(doldlogspline), 3
rlogspline, 5, 6, 9, 12
rlogspline(dlogspline), 1
roldlogspline, 7, 8, 11, 13
roldlogspline(doldlogspline), 3
summary.logspline, 2, 5, 6, 9, 11
summary.oldlogspline, 4, 7, 8, 11, 12
unstrip, 13