Package ‘sparr’
August 27, 2017

Type Package
Title Spatial and Spatiotemporal Relative Risk
Version 2.1-12
Date 2017-08-27
Author T.M. Davies and J.C. Marshall
Maintainer Tilman M. Davies <tdavies@maths.otago.ac.nz>
Description Provides functions to estimate kernel-smoothed spatial and spatio-
temporal densities and relative risk functions, and perform subsequent inference.
Depends R (>= 2.10.1), spatstat
Imports spatstat.utils, doParallel, parallel, foreach, misc3d
Suggests fftwtools (>= 0.9.8)
License GPL (>= 2)
LazyLoad yes
NeedsCompilation no
RoxygenNote 6.0.1
Repository CRAN
Date/Publication 2017-08-27 16:21:32 UTC

R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>sparr-package</td>
<td>2</td>
</tr>
<tr>
<td>available.h0</td>
<td>7</td>
</tr>
<tr>
<td>bivariate.density</td>
<td>8</td>
</tr>
<tr>
<td>BOOT.density</td>
<td>12</td>
</tr>
<tr>
<td>BOOT.spattemp</td>
<td>15</td>
</tr>
<tr>
<td>burk</td>
<td>18</td>
</tr>
<tr>
<td>fft2d</td>
<td>19</td>
</tr>
<tr>
<td>fmd</td>
<td>20</td>
</tr>
<tr>
<td>LIK.density</td>
<td>21</td>
</tr>
<tr>
<td>LIK.spattemp</td>
<td>24</td>
</tr>
<tr>
<td>LSCV.risk</td>
<td>26</td>
</tr>
</tbody>
</table>
The sparr Package: Spatial and Spatiotemporal Relative Risk

Description

Provides functions to estimate fixed and adaptive kernel-smoothed spatial relative risk surfaces via the density-ratio method and perform subsequent inference. Fixed-bandwidth spatiotemporal density and relative risk estimation is also supported.

Details

Package: sparr
Date: 2017-08-27
Version: 2.1-12
License: GPL (>= 2)

Kernel smoothing, and the flexibility afforded by this methodology, provides an attractive approach to estimating complex probability density functions.

The spatial relative risk function, constructed as a ratio of estimated case to control densities (Bithell, 1990; 1991; Kelsall and Diggle, 1995a,b), describes the variation in the ‘risk’ of the disease, given the underlying at-risk population. This is a technique that has been applied successfully for mainly exploratory purposes in a number of different analyses (see for example Sabel et al., 2000; Prince et al., 2001; Wheeler, 2007). It has also grown in popularity in very different fields that pose similarly styled research questions, such as ecology (e.g. Campos and Fedigan, 2014); physiology (Davies et al., 2013); and archaeology (e.g. Bevan, 2012; Smith et al. 2015).

This package provides functions for spatial (i.e. bivariate/planar/2D) kernel density estimation (KDE), implementing both fixed and ‘variable’ or ‘adaptive’ (Abramson, 1982) smoothing parameter options. A selection of bandwidth calculators for bivariate KDE and the relative risk function
are provided, including one based on the maximal smoothing principle (Terrell, 1990), and others involving a leave-one-out cross-validation (see below). In addition, the ability to construct both Monte-Carlo and asymptotic p-value surfaces (‘tolerance’ contours of which signal statistically significant sub-regions of extremity in a risk surface - Hazelton and Davies, 2009; Davies and Hazelton, 2010) as well as some visualisation tools are provided.

Spatiotemporal estimation is also supported, largely following developments in Fernando and Hazelton (2014). This includes their fixed-bandwidth kernel estimator of spatiotemporal densities, relative risk, and asymptotic tolerance contours.

Key content of \texttt{sparr} can be broken up as follows:

**DATASETS**

\texttt{pbc} a case/control planar point pattern (\texttt{ppp.object}) concerning liver disease in northern England.

\texttt{fmd} an anonymised (jittered) case/control spatiotemporal point pattern of the 2001 outbreak of veterinary foot-and-mouth disease in Cumbria (courtesy of the Animal and Plant Health Agency, UK).

\texttt{burk} a spatiotemporal point pattern of Burkitt’s lymphoma in Uganda; artificially simulated control data are also provided for experimentation.

Also available are a number of relevant additional spatial datasets built-in to the \texttt{spatstat} package (Baddeley and Turner, 2005; Baddeley et al., 2015), such as \texttt{chorley}, which concerns the distribution of laryngeal cancer in an area of Lancashire, UK.

**SPATIAL**

*Bandwidth calculators*

\texttt{os} estimation of an isotropic smoothing parameter for fixed-bandwidth bivariate KDE, based on the oversmoothing principle introduced by Terrell (1990).

\texttt{ns} estimation of an isotropic smoothing parameter for fixed-bandwidth bivariate KDE, based on the asymptotically optimal value for a normal density (bivariate normal scale rule - see e.g. Wand and Jones, 1995).

\texttt{LSCV.density} a least-squares cross-validated (LSCV) estimate of an isotropic fixed bandwidth for bivariate, edge-corrected KDE (see e.g. Bowman and Azzalini, 1997).

\texttt{LIK.density} a likelihood cross-validated (LIK) estimate of an isotropic fixed bandwidth for bivariate, edge-corrected KDE (see e.g. Silverman, 1986).

\texttt{BOOT.density} a bootstrap approach to optimisation of an isotropic fixed bandwidth for bivariate, edge-corrected KDE (see e.g. Taylor, 1989).

\texttt{LSCV.risk} Estimation of a jointly optimal, common isotropic case-control fixed bandwidth for the kernel-smoothed risk function based on the mean integrated squared error (MISE), a weighted MISE, or the asymptotic MISE (see respectively Kelsall and Diggle, 1995a; Hazelton, 2008; Davies, 2013).

*Density and relative risk estimation*

\texttt{bivariate.density} kernel density estimate of bivariate data; fixed or adaptive smoothing.

\texttt{multiscale.density} multi-scale adaptive kernel density estimates for multiple global bandwidths as per Davies and Baddeley (2017).
**multiscale.slice** a single adaptive kernel estimate based on taking a slice from a multi-scale estimate.

**risk** estimation of a (log) spatial relative risk function, either from data or pre-existing bivariate density estimates; fixed (Kelsall and Diggle, 1995a) or both asymmetric (Davies and Hazelton, 2010) and symmetric (Davies et al., 2016) adaptive estimates are possible.

**tolerance** calculation of asymptotic or Monte-Carlo \( p \)-value surfaces.

**Visualisation**

S3 methods of the plot function; see **plot.bivden** for visualising a single bivariate density estimate from **bivariate.density, plot.rrs** for visualisation of a spatial relative risk function from **risk**, or **plot.msden** for viewing animations of multi-scale density estimates from **multiscale.density**.

**tol.contour** provides more flexibility for plotting and superimposing tolerance contours upon an existing plot of spatial relative risk (i.e. given output from **tolerance**).

**Printing and summarising**

S3 methods (**print.bivden, print.rrs, print.msden, summary.bivden, summary.rrs, and summary.msden**) are available for the bivariate density, spatial relative risk, and multi-scale adaptive density objects.

**SPATIOTEMPORAL**

**Bandwidth calculators**

**OS.spattemp** estimation of an isotropic smoothing parameter for the spatial margin and another for the temporal margin for spatiotemporal densities, based on the 2D and 1D versions, respectively, of the oversmoothing principle introduced by Terrell (1990).

**NS.spattemp** as above, based on the 2D and 1D versions of the normal scale rule (Silverman, 1986).

**LSCV.spattemp** least-squares cross-validated (LSCV) estimates of scalar spatial and temporal bandwidths for edge-corrected spatiotemporal KDE.

**LIK.spattemp** as above, based on likelihood cross-validation.

**BOOT.spattemp** bootstrap bandwidth selection for the spatial and temporal margins; for spatiotemporal, edge-corrected KDE (Taylor, 1989).

**Density and relative risk estimation**

**spattemp.density** fixed-bandwidth kernel density estimate of spatiotemporal data.

**spattemp.risk** fixed-bandwidth kernel density estimate of spatiotemporal relative risk, either with a time-static or time-varying control density (Fernando and Hazelton, 2014).

**spattemp.slice** extraction function of the spatial density/relative risk at prespecified time(s).

**Visualisation**

S3 methods of the plot function; see **plot.stden** for various options (including animation) for visualisation of a spatiotemporal density, and **plot.rrst** for viewing spatiotemporal relative risk surfaces (including animation and tolerance contour superimposition).

**Printing and summarising objects**

S3 methods (**print.stden, print.rrst, summary.stden, and summary.rrst**) are available for the spatiotemporal density and spatiotemporal relative risk objects respectively.
Dependencies

The sparr package depends upon spatstat. In particular, the user should familiarise themselves with ppp objects and im objects, which are used throughout. For spatiotemporal density estimation, sparr is assisted by importing from the misc3d package, and for the experimental capabilities involving parallel processing, sparr also currently imports doParallel, parallel, and foreach.

Citation

To cite use of current versions of sparr in publications or research projects please use:


Old versions of sparr (<= 2.1-09) can be referenced by Davies et al. (2011) (see reference list).

Author(s)

T.M. Davies
Dept. of Mathematics & Statistics, University of Otago, Dunedin, New Zealand.

J.C. Marshall
Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand.

Maintainer: T.M.D. <tdavies@maths.otago.ac.nz>

References


Bithell, J.F. (1990), An application of density estimation to geographical epidemiology, Statistics in Medicine, 9, 691-701.


available.h0

Available global bandwidth range

Description

Gets universally available global bandwidths as represented by several multi-scale density estimate objects.

Usage

available.h0(...) 

Arguments

... Any number of objects of class msden; possibly named.

Details

This simple function merely accesses and returns the maximum lower limit and minimum upper limit of all hPrange components of the msden objects passed through ... Natural numeric error arising from any changes to the bandwidth-axis discretisation resolution in the creation of the msden objects (i.e. through the ‘dimz’ argument) means individual global bandwidth ranges can differ slightly between affected multi-scale estimates, even if they are all applied to the same data set. Can additionally be useful when, for example, creating asymmetric relative risk surfaces based on slices of multi-scale densities with respect to the case and control data sets, because the bandwidth factors differ.

Throws an error if one or more of the hPrange components is incompatible (i.e. all hPrange components must overlap).

Value

A numeric vector of length 2 providing the range of available global bandwidths compatible with all supplied multi-scale density estimates.

Author(s)

T.M. Davies

See Also

multiscale.density, multiscale.slice

Examples

# See ?multiscale.slice
bivariateNdensity

Bivariate kernel density/intensity estimation

Description
Provides an isotropic adaptive or fixed bandwidth kernel density/intensity estimate of bivariate/planar/2D data.

Usage
bivariateNdensity(pp, h0, hp = NULL, adapt = FALSE, resolution = 128,
gamma.scale = "geometric", edge = c("uniform", "diggle", "none"),
weights = NULL, intensity = FALSE, trim = 5, xy = NULL,
pilot.density = NULL, leaveoneout = FALSE, parallelise = NULL,
davies.baddeley = NULL, verbose = TRUE)

Arguments
pp An object of class ppp giving the observed 2D data set to be smoothed.
h0 Global bandwidth for adaptive smoothing or fixed bandwidth for constant smoothing. A numeric value > 0.
hp Pilot bandwidth (scalar, numeric > 0) to be used for fixed bandwidth estimation of a pilot density in the case of adaptive smoothing. If NULL (default), it will take on the value of h0. Ignored when adapt = FALSE or if pilot.density is supplied as a pre-defined pixel image.
adapt Logical value indicating whether to perform adaptive kernel estimation. See 'Details'.
resolution Numeric value > 0. Resolution of evaluation grid; the density/intensity will be returned on a [resolution × resolution] grid.
gamma.scale Scalar, numeric value > 0; controls rescaling of the variable bandwidths. Defaults to the geometric mean of the bandwidth factors given the pilot density (as per Silverman, 1986). See 'Details'.
edge Character string giving the type of edge correction to employ. "uniform" (default) corrects based on evaluation grid coordinate and "diggle" reweights each observation-specific kernel. Setting edge = "none" requests no edge correction. Further details can be found in the documentation for density.ppp.
weights Optional numeric vector of nonnegative weights corresponding to each observation in pp. Must have length equal to npoints(pp).
intensity Logical value indicating whether to return an intensity estimate (integrates to the sample size over the study region), or a density estimate (default, integrates to 1).
trim Numeric value > 0; controls bandwidth truncation for adaptive estimation. See 'Details'.
xy  Optional alternative specification of the evaluation grid; matches the argument of the same tag in as.mask. If supplied, resolution is ignored.

pilot.density An optional pixel image (class im) giving the pilot density to be used for calculation of the variable bandwidths in adaptive estimation, or a ppp.object giving the data upon which to base a fixed-bandwidth pilot estimate using hp[1]. If used, the pixel image must be defined over the same domain as the data given resolution or the supplied pre-set xy evaluation grid; or the planar point pattern data must be defined with respect to the same polygonal study region as in pp.

leaveoneout Logical value indicating whether to compute and return the value of the density/intensity at each data point for an adaptive estimate. See ‘Details’.

parallelise Numeric argument to invoke parallel processing, giving the number of CPU cores to use when leaveoneout = TRUE. Experimental. Test your system first using parallel::detectCores() to identify the number of cores available to you.

davies.baddeley An optional numeric vector of length 3 to control bandwidth partitioning for approximate adaptive estimation, giving the quantile step values for the variable bandwidths for density/intensity and edge correction surfaces and the resolution of the edge correction surface. May also be provided as a single numeric value. See ‘Details’.

verbose Logical value indicating whether to print a function progress bar to the console when adapt = TRUE.

Details

Given a data set \(x_1, \ldots, x_n\) in 2D, the isotropic kernel estimate of its probability density function, \(\hat{f}(x)\), is given by

\[
\hat{f}(y) = \frac{n^{-1}}{n} \sum_{i=1}^{n} h(x_i)^{-2} K((y - x_i)/h(x_i))
\]

where \(h(x)\) is the bandwidth function, and \(K(.)\) is the bivariate standard normal smoothing kernel. Edge-correction factors (not shown above) are also implemented.

**Fixed** The classic fixed bandwidth kernel estimator is used when adapt = FALSE. This amounts to setting \(h(u) = h\theta\) for all \(u\). Further details can be found in the documentation for density.ppp.

**Adaptive** Setting adapt = TRUE requests computation of Abramson’s (1982) variable-bandwidth estimator. Under this framework, we have \(h(u) = h\theta \min[\hat{f}(u)^{-1/2}, G\text{trim}]\gamma\), where \(\hat{f}(u)\) is a fixed-bandwidth kernel density estimate computed using the pilot bandwidth hp.

- Global smoothing of the variable bandwidths is controlled with the global bandwidth \(h\theta\).
- In the above statement, \(G\) is the geometric mean of the “bandwidth factors” \(\hat{f}(x_i)^{-1/2}; i = 1, \ldots, n\). By default, the variable bandwidths are rescaled by \(\gamma = G\), which is set with gamma.scale = "geometric". This allows \(h\theta\) to be considered on the same scale as the smoothing parameter in a fixed-bandwidth estimate i.e. on the scale of the recorded data. You can use any other rescaling of \(h\theta\) by setting gamma.scale to be any scalar positive numeric value; though note this only affects \(\gamma\) – see the next bullet. When using a scale-invariant \(h\theta\), set gamma.scale = 1.
• The variable bandwidths must be trimmed to prevent excessive values (Hall and Marron, 1988). This is achieved through \( \text{trim} \), as can be seen in the equation for \( h(u) \) above. The trimming of the variable bandwidths is universally enforced by the geometric mean of the bandwidth factors \( G \) independent of the choice of \( \gamma \). By default, the function truncates bandwidth factors at five times their geometric mean. For stricter trimming, reduce \( \text{trim} \), for no trimming, set \( \text{trim} = \text{Inf} \).

• For even moderately sized data sets and evaluation grid resolution, adaptive kernel estimation can be rather computationally expensive. The argument \( \text{daviesNbaddeley} \) is used to approximate an adaptive kernel estimate by a sum of fixed bandwidth estimates operating on appropriate subsets of \( \text{pp} \). These subsets are defined by “bandwidth bins”, which themselves are delineated by a quantile step value \( 0 < \delta < 1 \). E.g. setting \( \delta = 0.05 \) will create 20 bandwidth bins based on the 0.05th quantiles of the Abramson variable bandwidths. Adaptive edge-correction also utilises the partitioning, with pixel-wise bandwidth bins defined using the value \( 0 < \beta < 1 \), and the option to decrease the resolution of the edge-correction surface for computation to a \([L \times L]\) grid, where \( 0 < L \leq \text{resolution} \). If \( \text{daviesNbaddeley} \) is supplied as a vector of length 3, then the values \([1, 2, 3]\) correspond to the parameters \( \delta, \beta, \text{ and } L_M = L_N \) in Davies and Baddeley (2017). If the argument is simply a single numeric value, it is used for both \( \delta \) and \( \beta \), with \( L_M = L_N = \text{resolution} \) (i.e. no edge-correction surface coarsening).

• Computation of leave-one-out values (when \( \text{leaveoneout} = \text{TRUE} \)) is done by brute force, and is therefore very computationally expensive for adaptive smoothing. This is because the leave-one-out mechanism is applied to both the pilot estimation and the final estimation stages. Experimental code to do this via parallel processing using the \text{foreach} routine is implemented. Fixed-bandwidth leave-one-out can be performed directly in \text{density.ppp}.

Value

If \( \text{leaveoneout} = \text{FALSE} \), an object of class "bivden". This is effectively a list with the following components:

- **z**: The resulting density/intensity estimate, a pixel image object of class \text{im}.
- **h0**: A copy of the value of \( h_0 \) used.
- **hp**: A copy of the value of \( h_p \) used.
- **h**: A numeric vector of length equal to the number of data points, giving the bandwidth used for the corresponding observation in \( \text{pp} \).
- **him**: A pixel image (class \text{im}), giving the ‘hypothetical’ Abramson bandwidth at each pixel coordinate conditional upon the observed data. NULL for fixed-bandwidth estimates.
- **q**: Edge-correction weights; a pixel image if edge = "uniform", a numeric vector if edge = "diggle", and NULL if edge = "none".
- **gamma**: The value of \( \gamma \) used in scaling the bandwidths. NA if a fixed bandwidth estimate is computed.
- **geometric**: The geometric mean \( G \) of the untrimmed bandwidth factors \( \tilde{f}(x_i)^{-1/2} \). NA if a fixed bandwidth estimate is computed.
- **pp**: A copy of the \text{ppp.object} initially passed to the \text{pp} argument, containing the data that were smoothed.
Else, if leaveoneout = TRUE, simply a numeric vector of length equal to the number of data points, giving the leave-one-out value of the function at the corresponding coordinate.

**Author(s)**

T.M. Davies and J.C. Marshall

**References**


**Examples**

data(chorley) # Chorley-Ribble data from package 'spatstat'

# Fixed bandwidth kernel density; uniform edge correction
chden1 %bivariate.density(chorley,h=1.5)

# Fixed bandwidth kernel density; diggle edge correction; coarser resolution
chden2 %bivariate.density(chorley,h=1.5,edge="diggle",resolution=64)

## Not run:

# Adaptive smoothing; uniform edge correction
chden3 %bivariate.density(chorley,h=1.5,hp=1,adapt=TRUE)

# Adaptive smoothing; uniform edge correction; partitioning approximation
chden4 %bivariate.density(chorley,h=1.5,hp=1,adapt=TRUE,davies.baddeley=0.025)

par(mfrow=c(2,2))
plot(chden1);plot(chden2);plot(chden3);plot(chden4)
BOOT.density

Description

Isotropic fixed or global (for adaptive) bandwidth selection for a standalone 2D density based on bootstrap estimation of the MISE.

Usage

BOOT.density(pp, hlim = NULL, eta = NULL, type = c("fixed", "adaptive"), hp = NULL, edge = c("uniform", "none"), ref.density = NULL, resolution = 64, rmdiag = TRUE, sim.adapt = list(N = 50, B = 100, dimz = 64, objective = FALSE), parallelise = NA, verbose = TRUE, ...)

Arguments

pp An object of class ppp giving the observed 2D data to be smoothed.

hlim An optional vector of length 2 giving the limits of the optimisation routine with respect to the bandwidth. If NULL, the function attempts to choose this automatically.

eta Fixed scalar bandwidth to use for the reference density estimate; if NULL it is calculated as the oversmoothing bandwidth of pp using os. Ignored if ref.density is supplied. See ‘Details’.

type A character string indicating selection type. Either "fixed" (default) for selection of a constant bandwidth for the fixed-bandwidth estimator based on theory extended from results in Taylor (1989); or "adaptive" for selection of the global bandwidth for an adaptive kernel density. See ‘Details’.

hp Pilot bandwidth used for adaptive estimates in the bootstrap; see the argument of the same tag in bivariate.density. Ignored when type = "fixed" or when ref.density is supplied.

edge Character string dictating edge correction for the bootstrapped estimates. "uniform" (default) corrects based on evaluation grid coordinate. Setting edge="none" requests no edge correction.

ref.density Optional. An object of class bivden giving the reference density from which data will be generated. Based on theory, this must be a fixed-bandwidth estimate if type = "fixed"; see ‘Details’. Must be edge-corrected if edge = "uniform".

resolution Spatial grid size; the optimisation will be based on a [resolution x resolution] density estimate.

rmdiag Logical control value for removal of mirrored evaluation points as suggested by Taylor (1989) in the theoretical expression of the fixed-bandwidth MISE estimate. See ‘Details’. Ignored when type = "adaptive"
**boot.density**

sim.adapt  List of control values for bootstrap simulation in the adaptive case; see ‘Details’. Ignored when type = ”fixed”.

parallelise  Optional numeric argument to reduce computation time by invoking parallel processing, by giving the number of CPU cores to use in either evaluation (fixed) or in the actual bootstrap replicate generation (adaptive). Experimental. Test your system first using parallel::detectCores() to identify the number of cores available to you.

verbose  Logical value indicating whether to print function progress during execution.

...  Optional arguments controlling scaling to be passed to multiscale.density for the adaptive bootstrap; ignored when type = ”fixed”.

**Details**

For a 2D kernel density estimate \( \hat{f} \) defined on \( W \in \mathbb{R}^2 \), the mean integrated squared error (MISE) is given by \( E[\int_W (\hat{f}(x) - f(x))^2 dx] \), where \( f \) is the corresponding true density. Given an observed data set \( X \) (argument pp) of \( n \) observations, this function finds the bandwidth \( h \) that minimises \( E^*[\int_W (\hat{f}^*(x) - \hat{f}(x))^2 dx] \),

where \( \hat{f}(x) \) is a density estimate of \( X \) constructed with ‘reference’ bandwidth \( \eta \) (argument eta or ref.density), and \( \hat{f}^*(x) \) is a density estimate using bandwidth \( h \) of \( n \) observations \( X^* \) generated from \( \hat{f}(x) \). The notation \( E^* \) denotes expectation with respect to the distribution of the \( X^* \).

**Fixed** When type = ”fixed”, the function assumes you want to select a constant bandwidth for use with the fixed-bandwith density estimator. This implementation is based on extending the remarkable results of Taylor (1989) (see also Sain et al., 1994), who demonstrates that when the Gaussian kernel is being used, we can find the optimal \( h \) with respect to the aforementioned bootstrap-estimated MISE without any actual resampling. This implementation extends these results to the bivariate setting, and allows for edge-correction of both the reference and bootstrap densities.

- Taylor (1989) does not distinguish between the reference bandwidth \( \eta \) and the target of optimisation, \( h \), thus allowing the reference bandwidth to vary alongside the target in the optimisation. This is not optimal, and this function always assumes a static reference bandwidth. Hall et al. (1992) indicate that a generous amount of smoothing is to be preferred in the reference density (hence the default eta set using OS).
- If ref.density is supplied, it must be a fixed-bandwidth density estimate as an object of class bivden for validity of the theory. Edge-correction must be present if edge = ”uniform”; and it must be evaluated on the same spatial domain as dictated by Window(pp) and resolution. If unsupplied, the function internally computes an appropriate fixed-bandwidth density estimate using eta as the reference bandwidth.
- Finally, Taylor (1989) argues it is preferable to avoid summation at identical evaluation grid points in the expression for the optimal bandwidth, which is performed when rmdiag = TRUE. Setting rmdiag = FALSE disables this correction.

**Adaptive** When type = ”adaptive”, the function assumes you want to select a global bandwidth (argument h0 in bivariate.density) for use in 2D adaptive kernel density estimation.
• An expression similar to Taylor (1989) is not possible for the adaptive estimator. Thus, in the adaptive setting, the optimal bootstrap bandwidth is calculated by brute force as was performed in Davies and Baddeley (2017) by taking advantage of the multiscale estimation theory implemented in `multiscale.density`. The value that minimises an interpolating cubic spline of the estimated MISE on bandwidth is identified as the optimal global bandwidth.

• The user can pass either a fixed or adaptive `bivden` object to `refNdensity`. If this is the case, `hp` is ignored and the pilot bandwidth for each iteration of the bootstrap in estimation of the \( f^*(x) \) uses `ref.density$hp` (if `ref.density` is adaptive) or `ref.density$h0` (if `ref.density` is fixed). When `ref.density` is unsupplied, the function uses a fixed-bandwidth kernel estimate with bandwidth `eta` as the reference density, and if additionally `hp` is unsupplied, the same value `eta` is used for the constant pilot bandwidth.

• Control over the bootstrap is achieved with four optional named arguments passed as a list to `simNadapt`. `N` controls the number of bootstrap iterates per bandwidth; `B` controls the resolution of the sequence of bandwidths trialled (i.e. between `hlim[1]` and `hlim[2]`); `dimz` specifies the resolution of the bandwidth axis in the trivariate convolution evaluated by `multiscale.density`; and `objective` specifies whether to return the set of estimated MISEs for all bandwidths (nice to plot), or merely the optimal bandwidth (see ‘Value’).

• The ... are intended for any relevant optional arguments to be passed to the internal call to `multiscale.density`, such as `gamma.scale` or `trim`.

Value

The optimal fixed or global (for adaptive) scalar bandwidth. If `simargs$objective = TRUE` for the adaptive bootstrap, the return object is instead a `simargs$x2` matrix, with the first column giving the trialled bandwidth and the second giving the corresponding value of the estimated bootstrap MISE.

Warning

Even with the implemented computational tricks, bootstrapping for bandwidth selection for spatial data is still computationally demanding, especially for adaptive kernel estimates. The user can reduce this time by keeping the evaluation grid at modest resolutions, and experimenting with parallelising the internal loops via `parallelise`. The ‘Examples’ section offers some rough indications of evaluation times on this author’s local machine.

Author(s)

T.M. Davies

References


**See Also**

*bivariate.density, OS, multiscale.density*

**Examples**

```r
## Not run:
data(pbc)

## Fixed bandwidth selection ##
BOOT.density(pbc) # ~20 secs
BOOT.density(pbc, eta=0.5/pbc/2) # halve default reference bandwidth
BOOT.density(pbc, eta=0.5/pbc)*2 # double default reference bandwidth

# supplying pre-defined reference density as fixed-bandwidth 'bivden' object
pbcfix <- bivariate.density(pbc, h=2.5, resolution=64)
system.time(hfix <- BOOT.density(pbc, ref.density=pbcfix, parallelise=4)) # parallelisation; 14 secs
hfix

## Global (for adaptive) bandwidth selection ##
# ~200 secs next line; use 'parallelise' for speedup
system.time(hada <- BOOT.density(pbc, type="adaptive")) # minimal usage for adaptive bootstrap
hada

# ~80 secs next line. Set custom h limits; increase reference bandwidth;
# set custom pilot bandwidth; return objective function
system.time(hada <- BOOT.density(pbc, hlim=c(0.9, 8), eta=3.5, type="adaptive",
                              hp=0.5/pbc/2, parallelise=6,
                              sim.adapt=list(objective=TRUE)))

hada[which.min(hada[,2]),1]
plot(hada); abline(v=hada[which.min(hada[,2]),1],col=2)

## End(Not run)
```

---

**Description**

Bootstrap bandwidths for a spatiotemporal kernel density estimate

Bandwidth selection for standalone spatiotemporal density/intensity based on bootstrap estimation of the MISE, providing an isotropic scalar spatial bandwidth and a scalar temporal bandwidth.
Usage

BOOT.spattemp(pp, tt = NULL, tlim = NULL, eta = NULL, nu = NULL, sedge = c("uniform", "none"), tedge = sedge, ref.density = NULL, sres = 64, tres = sres, start = NULL, verbose = TRUE)

Arguments

pp
An object of class ppp giving the spatial coordinates of the observations to be smoothed. Possibly marked with the time of each event; see argument tt.

tt
A numeric vector of equal length to the number of points in pp, giving the time corresponding to each spatial observation. If unsupplied, the function attempts to use the values in the marks attribute of the ppp.object in pp.

tlim
A numeric vector of length 2 giving the limits of the temporal domain over which to smooth. If supplied, all times in tt must fall within this interval (equality with limits allowed). If unsupplied, the function simply uses the range of the observed temporal values.

eta
Fixed scalar bandwidth to use for the spatial margin of the reference density estimate; if NULL it is calculated as the oversmoothing bandwidth of pp using OS. Ignored if ref.density is supplied. See ‘Details’.

nu
Fixed scalar bandwidth to use for the temporal margin of the reference density estimate; if NULL it is calculated from tt using the univariate version of Terrell’s (1990) oversmoothing principle. Ignored if ref.density is supplied. See ‘Details’.

sedge
Character string dictating spatial edge correction. "uniform" (default) corrects based on evaluation grid coordinate. Setting sedge="none" requests no edge correction.

tedge
As sedge, for temporal edge correction.

ref.density
Optional. An object of class stden giving the reference density from which data is assumed to originate in the bootstrap. Must be spatially edge-corrected if sedge = "uniform".

sres
Numeric value > 0. Resolution of the [sres x sres] evaluation grid in the spatial margin.

tres
Numeric value > 0. Resolution of the evaluation points in the temporal margin as defined by the tlim interval. If unsupplied, the density is evaluated at integer values between tlim[1] and tlim[2].

start
Optional positive numeric vector of length 2 giving starting values for the internal call to optim, in the order of (<spatial bandwidth>, <temporal bandwidth>).

verbose
Logical value indicating whether to print a function progress bar to the console during evaluation.

Details

For a spatiotemporal kernel density estimate \( \hat{f} \) defined on \( W \times T \in R^3 \), the mean integrated squared error (MISE) is given by \( E[ \int_W \int_T (\hat{f}(x,t) - f(x,t))^2 dt dx ] \), where \( f \) is the corresponding true
density. Given observed spatiotemporal locations $X$ (arguments pp and tt) of $n$ observations, this function finds the scalar spatial bandwidth $h$ and scalar temporal bandwidth $\lambda$ that jointly minimise

$$E^* \left[ \int_W \int_T (\hat{f}^*(x,t) - \hat{f}(x,t))^2 dt dx \right],$$

where $\hat{f}(x,t)$ is a density estimate of $X$ constructed with ‘reference’ bandwidths $\eta$ (spatial; argument eta) and $\nu$ (temporal; argument nu); $\hat{f}^*(x,t)$ is a density estimate using bandwidths $h$ and $\lambda$ of $n$ observations $X^*$ generated from $\hat{f}(x,t)$. The notation $E^*$ denotes expectation with respect to the distribution of the $X^*$. The user may optionally supply ref. density as an object of class stden, which must be evaluated on the same spatial and temporal domains $W$ and $T$ as the data (arguments pp, tt, and tlim). In this case, the reference bandwidths are extracted from this object, and eta and nu are ignored.

This function is based on an extension of the theory of Taylor (1989) to the spatiotemporal domain and to cope with the inclusion of edge-correction factors. No resampling is necessary due to the theoretical properties of the Gaussian kernel.

Value

A numeric vector of length 2 giving the jointly optimised spatial and temporal bandwidths (named h and lambda respectively).

Warning

Bootstrapping for spatiotemporal bandwidth selection for spatiotemporal data is very computationally demanding. Keeping verbose = TRUE offers an indication of the computational burden by printing each pair of bandwidths at each iteration of the optimisation routine. The ‘Examples’ section also offers some rough indications of evaluation times on this author’s local machine.

Author(s)

T. M. Davies

References


See Also

LSCV.spattemp, spattemp.density

Examples

```R
## Not run:

data(burk) # Burkitt's Uganda lymphoma data
burkcns <- burk$cases

#-85 secs
```
burk

Description

Data of the spatiotemporal locations of Burkitt’s lymphoma in the Western Nile district of Uganda from 1960 to 1975.

Format

burk is a named list with three members:

$cases An object of class ppp giving the spatial locations (eastings/northings) of the 188 cases of Burkitt’s lymphoma recorded in individuals of various ages (mostly children); the spatial study region as a polygonal owin; as well as the time (in days since 1/1/1960) of each observation stored as the marks of the points.

$cases.age A numeric vector of length 188 giving the age of each individual in $cases.

$controls An object of class ppp giving 500 artificially simulated spatial-only observations to pose as a ‘control’ data set representing the at-risk population. The data were generated from a smooth kernel estimate of the spatial margin of the cases. The similarity between the case point distribution and the true at-risk population dispersion can be seen in e.g. Figure 2 of Middleton and Greenland (1954).

Source

The case data were extracted from the burkitt object of the splancs R package; see


References


Examples

data(burk)
summary(burk$cases)

par(mfrow=c(1,3))
plot(burk$cases)
plot(burk$controls)
plot(density(marks(burk$cases)),xlim=range(marks(burk$cases)))

fft2d

- **2D fast-Fourier wrapper around 'fftwtools' or 'stats' package**

Description

Utilises the Fastest Fourier Transform in the West (FFTW) via the 'fftwtools' package if available, else reverts to built-in functionality

Usage

```r
fft2d(x, inverse = FALSE, fftw = sparr:::fftw_available())
```

Arguments

- **x**
  - A numeric matrix to be transformed.
- **inverse**
  - Whether it should compute the inverse transform (defaults to FALSE).
- **fftw**
  - Whether the `fftwtools` R package is available.

Details

This function is called wherever `sparr` seeks to perform a 2D fast-Fourier transform. Where available, computational expense is noticeably reduced by appealing to routines in the independent ‘FFTW’ toolbox. The user is encouraged to install the corresponding R package `fftwtools` from CRAN; this function will automatically detect and use the faster option, otherwise will defer to the built-in `fft`.

Value

The fast-Fourier (inverse) transform. A complex-valued matrix of the same size as `x`.

Author(s)

J.C. Marshall
Examples

```r
## Not run:

# System check
sparr:::fftw_available()

system.time(fft(matrix(1:2000^2,2000)))
system.time(fft2d(matrix(1:2000^2,2000)))

## End(Not run)
```

---

**fmd**

*Veterinary foot-and-mouth disease outbreak data*

---

**Description**

Data of the spatial locations and time of farms infected by veterinary foot-and-mouth disease in the county of Cumbria, UK, over a course of nearly 250 days between February and August in 2001. There are 410 infected farms (the cases), and 1866 uninfected farms (the controls). The data have been jittered and randomly thinned by an unspecified amount to preserve anonymity.

**Format**

`fmd` is a named list with two members:

- `$cases` An object of class `ppp` giving the spatial locations of the 410 infected farms within a polygonal study region representing the county of Cumbria. The `marks` component of this object contain the integer day of infection (from beginning of study period).
- `$controls` An object of class `ppp` defined over the same spatial study region with the locations of the 1866 uninfected farms.

**Acknowledgements**

The Animal and Plant Health Agency (APHA), UK, provided permission to use this dataset.

**References**


**LIK.density**

**Examples**

```r
data(fmd)
summary(fmd$cases)
summary(fmd$controls)

par(mfrow=c(1,2))
plot(fmd$cases)
plot(fmd$controls)
```

---

**LIK.density**

*Cross-validation bandwidths for spatial kernel density estimates*

**Description**

Isotropic fixed or global (for adaptive) bandwidth selection for standalone 2D density/intensity based on either unbiased least squares cross-validation (LSCV) or likelihood (LIK) cross-validation.

**Usage**

```r
LIK.density(pp, hlim = NULL, hseq = NULL, resolution = 64, edge = TRUE, 
auto.optim = TRUE, type = c("fixed", "adaptive"), seqres = 30, 
parallelise = NULL, verbose = TRUE, ...)
```

```r
LSCV.density(pp, hlim = NULL, hseq = NULL, resolution = 64, edge = TRUE, 
auto.optim = TRUE, type = c("fixed", "adaptive"), seqres = 30, 
parallelise = NULL, verbose = TRUE, ...)
```

**Arguments**

- **pp** An object of class `ppp` giving the observed 2D data to be smoothed.
- **hlim** An optional vector of length 2 giving the limits of the optimisation routine with respect to the bandwidth. If unspecified, the function attempts to choose this automatically.
- **hseq** An optional increasing sequence of bandwidth values at which to manually evaluate the optimisation criterion. Used only in the case (!auto.optim && is.null(hlim)).
- **resolution** Spatial grid size; the optimisation will be based on a `resolution × resolution` density estimate.
- **edge** Logical value indicating whether to edge-correct the density estimates used.
- **auto.optim** Logical value indicating whether to automate the numerical optimisation using `optimise`. If FALSE, the optimisation criterion is evaluated over hseq (if supplied), or over a sequence of values controlled by hlim and seqres.
type A character string; "fixed" (default) performs classical leave-one-out cross-validation for the fixed-bandwidth estimator. Alternatively, "adaptive" utilises multiscale adaptive kernel estimation (Davies & Baddeley, 2017) to run the cross-validation in an effort to find a suitable global bandwidth for the adaptive estimator. Note that data points are not 'left out' of the pilot density estimate when using this option. See also the entry for ....

seqres Optional resolution of an increasing sequence of bandwidth values. Only used if (!auto.optim && is.null(hseq)).

parallelise Numeric argument to invoke parallel processing, giving the number of CPU cores to use when !auto.optim and type = "fixed". Experimental. Test your system first using parallel::detectCores() to identify the number of cores available to you.

verbose Logical value indicating whether to provide function progress commentary.

Details

This function implements the bivariate, edge-corrected versions of fixed-bandwidth least squares cross-validation and likelihood cross-validation as outlined in Sections 3.4.3 and 3.4.4 of Silverman (1986) in order to select an optimal fixed smoothing bandwidth. With type = "adaptive" it may also be used to select the global bandwidth for adaptive kernel density estimates, making use of multi-scale estimation (Davies and Baddeley, 2017) via multiscale.density. Note that for computational reasons, the leave-one-out procedure is not performed on the pilot density in the adaptive setting; it is only performed on the final stage estimate. See also ‘Warning’ below.

Where LSCV.density is based on minimisation of an unbiased estimate of the mean integrated squared error (MISE) of the density, LIK.density is based on maximisation of the cross-validated leave-one-out average of the log-likelihood of the density estimate with respect to h.

Value

A single numeric value of the estimated bandwidth (if auto.optim = TRUE). Otherwise, a [seqres x 2] matrix giving the bandwidth sequence and corresponding CV function value.

Warning

Leave-one-out CV for bandwidth selection in kernel density estimation is notoriously unstable in practice and has a tendency to produce rather small bandwidths, particularly for spatial data. Satisfactory bandwidths are not guaranteed for every application. This method can also be computationally expensive for large data sets and fine evaluation grid resolutions. The user may need to experiment with adjusting hlim to find a suitable minimum.
**LIK.density**

**Author(s)**

T. M. Davies

**References**


**See Also**

Functions for bandwidth selection in package `spatstat`: `bw.diggle`; `bw.ppl`; `bw.scott`; `bw.frac`.

**Examples**

data(pbc)
pbccas <- split(pbc)$case

LIK.density(pbccas)
LSCV.density(pbccas)

```r
## Not run:
## FIXED

# custom limits
LIK.density(pbccas,hlim=c(0.01,4))
LSCV.density(pbccas,hlim=c(0.01,4))

# disable edge correction
LIK.density(pbccas,hlim=c(0.01,4),edge=FALSE)
LSCV.density(pbccas,hlim=c(0.01,4),edge=FALSE)

# obtain objective function
hcv <-LIK.density(pbccas,hlim=c(0.01,4),auto.optim=FALSE)
plot(hcv);abline(v=hcv[which.max(hcv[,2]),1],lty=2,col=2)

## ADAPTIVE
LIK.density(pbccas,type="adaptive")
LSCV.density(pbccas,type="adaptive")

# change pilot bandwidth used
LIK.density(pbccas,type="adaptive",hp=2)
LSCV.density(pbccas,type="adaptive",hp=2)

## End(Not run)
**LIK.spattemp**

Cross-validation bandwidths for spatiotemporal kernel density estimates

---

**Description**

Bandwidth selection for standalone spatiotemporal density/intensity based on either unbiased least squares cross-validation (LSCV) or likelihood (LIK) cross-validation, providing an isotropic scalar spatial bandwidth and a scalar temporal bandwidth.

**Usage**

```r
LIK.spattemp(pp, tt = NULL, tlim = NULL, sedge = c("uniform", "none"),
    tedge = sedge, parallelise = NA, start = NULL, verbose = TRUE)

LSCV.spattemp(pp, tt = NULL, tlim = NULL, sedge = c("uniform", "none"),
    tedge = sedge, sres = 64, tres = sres, parallelise = NA,
    start = NULL, verbose = TRUE)
```

**Arguments**

- `pp` An object of class `ppp` giving the spatial coordinates of the observations to be smoothed. Possibly marked with the time of each event; see argument `tt`.
- `tt` A numeric vector of equal length to the number of points in `pp`, giving the time corresponding to each spatial observation. If unsupplied, the function attempts to use the values in the `marks` attribute of the `ppp` object in `pp`.
- `tlim` A numeric vector of length 2 giving the limits of the temporal domain over which to smooth. If supplied, all times in `tt` must fall within this interval (equality with limits allowed). If unsupplied, the function simply uses the range of the observed temporal values.
- `sedge` Character string dictating spatial edge correction. "uniform" (default) corrects based on evaluation grid coordinate. Setting `sedge="none"` requests no edge correction.
- `tedge` As `sedge`, for temporal edge correction.
- `sres` Numeric value > 0. Resolution of the `[sres x sres]` evaluation grid in the spatial margin.
- `tres` Numeric value > 0. Resolution of the evaluation points in the temporal margin as defined by the `tlim` interval. If unsupplied, the density is evaluated at integer values between `tlim[1]` and `tlim[2]`.
- `parallelise` Optional numeric argument to invoke parallel processing, by giving the number of CPU cores to use optimisation. This is only useful for larger data sets of many thousand observations. Experimental. Test your system first using `parallel::detectCores()` to identify the number of cores available to you.
- `start` Optional positive numeric vector of length 2 giving starting values for the internal call to `optim`, in the order of (`<spatial bandwidth>`, `<temporal bandwidth>`).
Logical value indicating whether to print a function progress bar to the console during evaluation.

Value

A numeric vector of length 2 giving the jointly optimised spatial and temporal bandwidths (named h and lambda respectively).

Warning

Leave-one-out CV for bandwidth selection in kernel density estimation is notoriously unstable in practice and has a tendency to produce rather small bandwidths in the fixed bandwidth case. Satisfactory bandwidths are not guaranteed for every application. This method can also be computationally expensive for large data sets and fine evaluation grid resolutions.

Author(s)

T. M. Davies

References


See Also

`BOOT. spattemp`, `spattemp.density`

Examples

```r
## Not run:

data(burk) # Burkitt's Uganda lymphoma data
burkcas <- burk$cases

hlam1 <- LSCV. spattemp(burkcas) #-9 secs
hlam2 <- LSCV. spattemp(burkcas, tlim=c(400,5800))
hlam3 <- LSCV. spattemp(burkcas, start=c(7,400))
rbind(hlam1, hlam2, hlam3)

hlam1 <- LIK. spattemp(burkcas) #-3 secs
hlam2 <- LIK. spattemp(burkcas, tlim=c(400,5800))
hlam3 <- LIK. spattemp(burkcas, start=c(7,400))
rbind(hlam1, hlam2, hlam3)

## End(Not run)
```
LSCV-risk

Jointly optimal bandwidth selection for the spatial relative risk function

Description

Methods to find a jointly optimal, common case-control isotropic bandwidth for use in estimation of the fixed or adaptive kernel-smoothed relative risk function.

Usage

LSCV.risk(f, g = NULL, hlim = NULL, hseq = NULL, type = c("fixed", "adaptive"), method = c("kelsall-diggle", "hazelton", "davies"), resolution = 64, edge = TRUE, hp = NULL, pilot.symmetry = c("none", "f", "g", "pooled"), auto.optim = TRUE, seqlres = 30, parallelise = NA, verbose = TRUE, ...)

Arguments

f

Either a pre-calculated object of class bivden representing the ‘case’ (numerator) density estimate, or an object of class ppp giving the observed case data. Alternatively, if f is ppp object with dichotomous factor-valued marks, the function treats the first level as the case data, and the second as the control data, obviating the need to supply g.

g

As for f, for the ‘control’ (denominator) density; this object must be of the same class as f. Ignored if, as stated above, f contains both case and control observations.

hlim

An optional vector of length 2 giving the limits of the optimisation routine with respect to the bandwidth. If unspecified, the function attempts to choose this automatically.

hseq

An optional increasing sequence of bandwidth values at which to manually evaluate the optimisation criterion. Used only in the case (!auto.optim && is.null(hlim)).

type

A character string; "fixed" (default) performs classical leave-one-out cross-validation for a jointly optimal fixed bandwidth. Alternatively, "adaptive" utilises multiscale adaptive kernel estimation (Davies & Baddeley, 2017) to run the cross-validation in an effort to find a suitable jointly optimal, common global bandwidth for the adaptive relative risk function. See ‘Details’ and the entry for pilot.args.

method

A character string controlling the selector to use. There are three types, based on either the mean integrated squared error (MISE) (Kelsall and Diggle, 1995; default – method = "kelsall-diggle"); a weighted MISE (Hazelton, 2008 – method = "hazelton"); or an approximation to the asymptotic MISE (Davies, 2013 – method = "davies"). See ‘Details’.

resolution

Spatial grid size; the optimisation will be based on a [resolution × resolution] density estimate.
edge Logical value indicating whether to edge-correct the density estimates used.

hp A single numeric value or a vector of length 2 giving the pilot bandwidth(s) to be used for estimation of the pilot densities for adaptive risk surfaces. Ignored if type = "fixed".

pilot.symmetry A character string used to control the type of symmetry, if any, to use for the bandwidth factors when computing an adaptive relative risk surface. See ‘Details’. Ignored if type = "fixed".

auto.optim Logical value indicating whether to automate the numerical optimisation using optimise. If FALSE, the optimisation criterion is evaluated over hseq (if supplied), or over a sequence of values controlled by hlim and seqres.

seqres Optional resolution of an increasing sequence of bandwidth values. Only used if (!auto.optim && is.null(hseq)).

parallelise Numeric argument to invoke parallel processing, giving the number of CPU cores to use when !auto.optim. Experimental. Test your system first using parallel::detectCores() to identify the number of cores available to you.

verbose Logical value indicating whether to provide function progress commentary.

... Additional arguments such as dimz and trim to be passed to the internal calls to multiscaleNdensity.

Details

Given the established preference of using a common bandwidth for both case and control density estimates when constructing a relative risk surface, this function calculates a 'jointly optimal', common isotropic LSCV bandwidth for the (Gaussian) kernel-smoothed relative risk function (case-control density-ratio). It can be shown that choosing a bandwidth that is equal for both case and control density estimates is preferable to computing 'separately optimal' bandwidths (Kelsall and Diggle, 1995). The user can choose to either calculate a common smoothing parameter for a fixed-bandwidth relative risk surface (type = "fixed"; default), or a common global bandwidth for an adaptive risk surface (type = "adaptive"). See further comments below.

- method = "kelsall-diggle": the function computes the common bandwidth which minimises the approximate mean integrated squared error (MISE) of the log-transformed risk surface (Kelsall and Diggle, 1995).
- method = "hazelton": the function minimises a weighted-by-control MISE of the (raw) relative risk function (Hazelton, 2008).
- method = "davies": the optimal bandwidth is one that minimises a crude plug-in approximation to the asymptotic MISE (Davies, 2013). Only possible for type = "fixed".

For jointly optimal, common global bandwidth selection when type = "adaptive", the optimisation routine utilises multiscaleNdensity. Like LSCV.density, the leave-one-out procedure does not affect the pilot density, for which additional control is offered via the hp and pilot.symmetry arguments. The user has the option of obtaining a so-called symmetric estimate (Davies et al. 2016) via pilot.symmetry. This amounts to choosing the same pilot density for both case and control densities. By choosing "none" (default), the result uses the case and control data separately for the fixed-bandwidth pilots, providing the original asymmetric density-ratio of Davies and Hazelton (2010). By selecting either of "f", "g", or "pooled", the pilot density is calculated based on the...
case, control, or pooled case/control data respectively (using hp[1] as the fixed bandwidth). Davies et al. (2016) noted some beneficial practical behaviour of the symmetric adaptive surface over the asymmetric. (The pilot bandwidth(s), if not supplied in hp, are calculated internally via default use of LSCV.density, using the requested symmetric-based data set, or separately with respect to the case and control datasets f and g if pilot.symmetry = "none").

Value

A single numeric value of the estimated bandwidth (if auto.optim = TRUE). Otherwise, a list of two numeric vectors of equal length giving the bandwidth sequence (as hs) and corresponding CV function value (as CV).

Warning

The jointly optimal bandwidth selector can be computationally expensive for large data sets and fine evaluation grid resolutions. The user may need to experiment with adjusting hlim to find a suitable minimum.

Author(s)

T. M. Davies

References


See Also

bivariate.density
Examples

```r
## Not run:

data(pbc)
pbccas <- split(pbc)$case
pbccon <- split(pbc)$control

# FIXED (for common h)

LSCV.risk(pbccas,pbccon)
LSCV.risk(pbccas,pbccon,method="hazelton")
hcv <- LSCV.risk(pbccas,pbccon,method="davies",auto.optim=FALSE)
plot(hcv[,1],log(hcv[,2]));abline(v=hcv[which.min(hcv[,2]),1],col=2,lty=2)

# ADAPTIVE (for common h0)

LSCV.risk(pbccas,pbccon,type="adaptive")

# change pilot bandwidths used
LSCV.risk(pbccas,pbccon,type="adaptive",hp=c(OS(pbccas)/2,OS(pbccon)/2))

# specify pooled-data symmetric relative risk estimator
LSCV.risk(pbccas,pbccon,type="adaptive",hp=OS(pbc),pilot.symmetry="pooled")

# as above, for Hazelton selector
LSCV.risk(pbccas,pbccon,type="adaptive",method="hazelton")
LSCV.risk(pbccas,pbccon,type="adaptive",method="hazelton",hp=c(OS(pbccas)/2,OS(pbccon)/2))
LSCV.risk(pbccas,pbccon,type="adaptive",method="hazelton",hp=OS(pbc),pilot.symmetry="pooled")

## End(Not run)
```

---

**multiscale.density**  
*Multi-scale adaptive kernel density/intensity estimation*  

**Description**  
Computes adaptive kernel estimates of spatial density/intensity using a 3D FFT for multiple global bandwidth scales.

**Usage**

```r
multiscale.density(pp, h0, hp = NULL, h0fac = c(0.25, 1.5),
    edge = c("uniform", "none"), resolution = 128, dimz = 64,
    gamma.scale = "geometric", trim = 5, intensity = FALSE,
    pilot.density = NULL, xy = NULL, taper = TRUE, verbose = TRUE)
```
Arguments

**pp**
An object of class `ppp` giving the observed 2D data set to be smoothed.

**h0**
Reference global bandwidth for adaptive smoothing; numeric value > 0. Multiscale estimates will be computed by rescaling this value as per \( h0 \cdot hPfac \).

**hp**
Pilot bandwidth (scalar, numeric > 0) to be used for fixed bandwidth estimation of the pilot density. If `NULL` (default), it will take on the value of \( h0 \). Ignored when `pilot.density` is supplied as a pre-defined pixel image.

**h0fac**
A numeric vector of length 2 stipulating the span of the global bandwidths in the multiscale estimates. Interpreted as a multiplicative factor on \( h0 \). See ‘Details’.

**edge**
Character string dictating edge correction. "uniform" (default) corrects based on evaluation grid coordinate. Setting `edge="none"` requests no edge correction.

**resolution**
Numeric value > 0. Resolution of evaluation grid in the spatial domain; the densities/intensities will be returned on a \([\text{resolution} \times \text{resolution}]\) grid.

**dimz**
Resolution of z- (rescaled bandwidth)-axis in the trivariate convolution. Higher values increase precision of the multiscale estimates at a computational cost. See ‘Details’.

**gamma.scale**
Scalar, numeric value > 0; controls rescaling of the variable bandwidths. Defaults to the geometric mean of the bandwidth factors given the pilot density (as per Silverman, 1986). See the documentation for `bivariate.density`.

**trim**
Numeric value > 0; controls bandwidth truncation for adaptive estimation. See the documentation for `bivariate.density`.

**intensity**
Logical value indicating whether to return an intensity estimate (integrates to the sample size over the study region), or a density estimate (default, integrates to 1).

**pilot.density**
An optional pixel image (class `im`) giving the pilot density to be used for calculation of the variable bandwidths in adaptive estimation, or a `ppp` object giving the data upon which to base a fixed-bandwidth pilot estimate using \( hp \). See the documentation for `bivariate.density`.

**xy**
Optional alternative specification of the spatial evaluation grid; matches the argument of the same tag in `as.mask`. If supplied, `resolution` is ignored.

**taper**
Logical value indicating whether to taper off the trivariate kernel outside the range of \( h0 \cdot h0fac \) in the scale space; see Davies & Baddeley (2017). Keep at the default `TRUE` if you don’t know what this means.

**verbose**
Logical value indicating whether to print function progress.

Details

Davies & Baddeley (2017) investigated computational aspects of Abramson’s (1982) adaptive kernel smoother for spatial (2D) data. This function is the implementation of the 3D convolution via a fast-Fourier transform (FFT) which allows simultaneous calculation of an adaptive kernel estimate at multiple global bandwidth scales.

These ‘multiple global bandwidth scales’ are computed with respect to rescaling a reference value of the global bandwidth passed to the `h0` argument. This rescaling is defined by the range provided
multiscale.density

... to the argument h0fac. For example, by default, the function will compute the adaptive kernel estimate for a range of global bandwidths between 0.25*h0 and 1.5*h0. The exact numeric limits are subject to discretisation, and so the returned valid range of global bandwidths will differ slightly. The exact resulting range following function execution is returned as the h0range element of the result, see ‘Value’ below.

The distinct values of global bandwidth used (which define the aforementioned h0range) and hence the total number of pixel images returned depend on both the width of the span h0fac and the discretisation applied to the bandwidth axis through dimz. Increasing this z-resolution will provide more pixel images and hence greater numeric precision, but increases computational cost. The returned pixel images that represent the multiscale estimates are stored in a named list (see ‘Value’), whose names reflect the corresponding distinct global bandwidth. See ‘Examples’ for the easy way to extract these distinct global bandwidths.

The user can request an interpolated density/intensity estimate for any global bandwidth value within h0range by using the multiscale.slice function, which returns an object of class bivden.

Value

An object of class "msden". This is very similar to a bivden object, with lists of pixel images in the z, him, and q components (instead of standalone images).

z A list of the resulting density/intensity estimates; each member being a pixel image object of class im. They are placed in increasing order of the discretised values of h0.

h0 A copy of the reference value of h0 used.

h0range A vector of length 2 giving the actual range of global bandwidth values available (inclusive).

hp A copy of the value of hp used.

h A numeric vector of length equal to the number of data points, giving the bandwidth used for the corresponding observation in pp with respect to the reference global bandwidth h0.

him A list of pixel images (class im), corresponding to z, giving the ‘hypothetical’ Abramson bandwidth at each pixel coordinate conditional upon the observed data and the global bandwidth used.

q Edge-correction weights; list of pixel images corresponding to z if edge = "uniform", and NULL if edge = "none".

gamma The numeric value of gamma.scale used in scaling the bandwidths.

geometric The geometric mean of the untrimmed variable bandwidth factors. This will be identical to gamma if gamma.scale = "geometric" as per default.

pp A copy of the ppp.object initially passed to the pp argument, containing the data that were smoothed.

Author(s)

T.M. Davies and A. Baddeley
References


See Also

*bivariate.density, multiscale.slice*

Examples

```r
## Not run:
data(chorley) # Chorley-Ribble data (package 'spatstat')
ch.multi <- multiscale.density(chorley,h0=1)
plot(ch.multi)

ch.pilot <- bivariate.density(chorley,h0=0.75) # with pre-defined pilot density
ch.multi2 <- multiscale.density(chorley,h0=1,pilot.density=ch.pilot$z)
plot(ch.multi2)

data(pbc)
# widen h0 scale, increase z-axis resolution
pbc.multi <- multiscale.density(pbc,h0=2,hp=1,h0fac=c(0.25,2.5),dimz=128)
plot(pbc.multi)

## End(Not run)
```

### multiscale.slice

Slicing a multi-scale density/intensity object

Description

Takes slices of a multi-scale density/intensity estimate at desired global bandwidths

Usage

`multiscale.slice(msob, h0, checkargs = TRUE)`

Arguments

- `msob` An object of class `msden` giving the multi-scale estimate from which to take slices.
- `h0` Desired global bandwidth(s); the density/intensity estimate corresponding to which will be returned. A numeric vector. All values must be in the available range provided by `msob$h0range`; see ‘Details’.
checkargs Logical value indicating whether to check validity of msob and h\( \theta \). Disable only if you know this check will be unnecessary.

Details

Davies & Baddeley (2017) demonstrate that once a multi-scale density/intensity estimate has been computed, we may take slices parallel to the spatial domain of the trivariate convolution to return the estimate at any desired global bandwidth. This function is the implementation thereof based on a multi-scale estimate resulting from a call to \texttt{multiscale.density}.

The function returns an error if the requested slices at h\( \theta \) are not all within the available range of pre-computed global bandwidth scalings as defined by the h\( \theta \)range component of msob.

Because the contents of the msob argument, an object of class msden, are returned based on a discretised set of global bandwidth scalings, the function internally computes the desired surface as a pixel-by-pixel linear interpolation using the two discretised global bandwidth rescalings that bound each requested h\( \theta \). (Thus, numeric accuracy of the slices is improved with an increase to the dimz argument of the preceding call to \texttt{multiscale.density} at the cost of additional computing time.)

Value

If h\( \theta \) is scalar, an object of class bivden with components corresponding to the requested slice at h\( \theta \). If h\( \theta \) is a vector, a list of objects of class bivden.

Author(s)

T.M. Davies

References


See Also

\texttt{multiscale.density, bivariate.density}

Examples

```r
## Not run:
data(chorley) # Chorley-Ribble data (package 'spatstat')
ch.multi <- multiscale.density(chorley,h\( \theta \)=1,h\( \theta \)fac=c(0.5,2))

available.h\( \theta \)(ch.multi)
ch.slices <- multiscale.slice(ch.multi,h\( \theta \)=c(0.7,1.1,1.6))

par(mfcol=c(2,3)) # plot each density and edge-correction surface
for(i in 1:3) { plot(ch.slices[[i]]$z); plot(ch.slices[[i]]$q) }

## End(Not run)
```
Normal scale (NS) bandwidth selector

Description

Provides the asymptotically optimal fixed bandwidths for spatial or spatiotemporal normal densities based on a simple expression.

Usage

\[
\text{NS}(\text{pp, nstar = c("npoints", "geometric"), scaler = c("silverman", "IQR", "sd", "var"))}
\]

\[
\text{NS.spattemp}(\text{pp, tt = NULL, nstar = "npoints", scaler = c("silverman", "IQR", "sd", "var"))}
\]

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pp</td>
<td>An object of class \text{pp} giving the observed 2D data to be smoothed.</td>
</tr>
<tr>
<td>nstar</td>
<td>Optional. Controls the value to use in place of the number of observations (n) in the normal scale formula. Either a character string, &quot;npoints&quot; (default) or &quot;geometric&quot; (only possible for NS), or a positive numeric value. See ‘Details’.</td>
</tr>
<tr>
<td>scaler</td>
<td>Optional. Controls the value for a scalar representation of the spatial (and temporal for \text{NS.spattemp}) scale of the data. Either a character string, &quot;silverman&quot; (default), &quot;IQR&quot;, &quot;sd&quot;, or &quot;var&quot;; or a positive numeric value. See ‘Details’.</td>
</tr>
<tr>
<td>tt</td>
<td>A numeric vector of equal length to the number of points in pp, giving the time corresponding to each spatial observation. If unsupplied, the function attempts to use the values in the \text{marks} attribute of the \text{pp.object} in pp.</td>
</tr>
</tbody>
</table>

Details

These functions calculate scalar smoothing bandwidths for kernel density estimates of spatial or spatiotemporal data: the optimal values would minimise the asymptotic mean integrated squared error assuming normally distributed data; see pp. 46-48 of Silverman (1986). The \text{NS} function returns a single bandwidth for isotropic smoothing of spatial (2D) data. The \text{NS.spattemp} function returns two values – one for the spatial margin and another for the temporal margin, based on independently applying the normal scale rule (in 2D and 1D) to the spatial and temporal margins of the supplied data.

Effective sample size

The formula requires a sample size, and this can be minimally tailored via \text{nstar}. By default, the function simply uses the number of observations in \text{pp}: \text{nstar = "npoints"}. Alternatively, the user can specify their own value by simply supplying a single positive numeric value to \text{nstar}. For \text{NS} (not applicable to \text{NS.spattemp}), if \text{pp} is a \text{pp.object} with factor-valued \text{marks}, then the user has the option of using \text{nstar = "geometric"}, which sets the sample size used in the formula to the geometric mean of the counts of observations of each mark. This can be useful for e.g. relative risk calculations, see Davies and Hazelton (2010).
Spatial (and temporal) scale  The `scaler` argument is used to specify spatial (as well as temporal, in use of `NS.spattemp`) scale. For isotropic smoothing in the spatial margin, one may use the ‘robust’ estimate of standard deviation found by a weighted mean of the interquartile ranges of the \( x \)- and \( y \)-coordinates of the data respectively (\`scaler = "IQR"\)). Two other options are the raw mean of the coordinate-wise standard deviations (\`scaler = "sd"\), or the square root of the mean of the two variances (\`scaler = "var"\)). A fourth option, \`scaler = "silverman"\) (default), sets the scaling constant to be the minimum of the “IQR” and “sd” options; see Silverman (1986), p. 47. In use of `NS.spattemp` the univariate version of the elected scale statistic is applied to the recorded times of the data for the temporal bandwidth. Alternatively, like `nstar`, the user can specify their own value by simply supplying a single positive numeric value to `scaler` for `NS`, or a numeric vector of length 2 (in the order of \[ \langle \text{spatial scale}, \text{temporal scale} \rangle \]) for `NS.spattemp`.

Value
A single numeric value of the estimated spatial bandwidth for `NS`, or a named numeric vector of length 2 giving the spatial bandwidth (as \( h \)) and the temporal bandwidth (as \( \lambda \)) for `NS.spattemp`.

Warning
The `NS` bandwidth is an approximation, and assumes that the target density is normal. This is considered rare in most real-world applications. Nevertheless, it remains a quick and easy ‘rule-of-thumb’ method with which one may obtain a smoothing parameter. Note that a similar expression for the adaptive kernel estimator is not possible (Davies et al., 2017).

Author(s)
T.M. Davies

References

Examples
```
data(pbc)
NS(pbc)
NS(pbc,nstar="geometric") # uses case-control marks to replace sample size
NS(pbc,scaler="var") # set different scalar measure of spread

data(burk)
```
**Description**

Provides fixed bandwidths for spatial or spatiotemporal data based on the maximal smoothing (oversmoothing) principle of Terrell (1990).

**Usage**

```r
OS(pp, nstar = c("npoints", "geometric"), scaler = c("silverman", "IQR", "sd", "var"))
```

```r
OS.spattemp(pp, tt = NULL, nstar = "npoints", scaler = c("silverman", "IQR", "sd", "var"))
```

**Arguments**

- `pp` An object of class `ppp` giving the observed 2D data to be smoothed.
- `nstar` Optional. Controls the value to use in place of the number of observations $n$ in the oversmoothing formula. Either a character string, "npoints" (default) or "geometric" (only possible for `OS`), or a positive numeric value. See ‘Details’.
- `scaler` Optional. Controls the value for a scalar representation of the spatial (and temporal for `OS.spattemp`) scale of the data. Either a character string, "silverman" (default), "IQR", "sd", or "var"; or positive numeric value(s). See ‘Details’.
- `tt` A numeric vector of equal length to the number of points in `pp`, giving the time corresponding to each spatial observation. If unsupplied, the function attempts to use the values in the `marks` attribute of the `ppp.object` in `pp`.

**Details**

These functions calculate scalar smoothing bandwidths for kernel density estimates of spatial or spatiotemporal data: the “maximal amount of smoothing compatible with the estimated scale of the observed data”. See Terrell (1990). The `OS` function returns a single bandwidth for isotropic smoothing of spatial (2D) data. The `OS.spattemp` function returns two values – one for the spatial margin and another for the temporal margin, based on independently applying Terrell’s (1990) rule (in 2D and 1D) to the spatial and temporal margins of the supplied data.

**Effective sample size** The formula requires a sample size, and this can be minimally tailored via `nstar`. By default, the function simply uses the number of observations in `pp`: `nstar = "npoints"`. Alternatively, the user can specify their own value by simply supplying a single positive numeric value to `nstar`. For `OS` (not applicable to `OS.spattemp`), if `pp` is a `ppp.object` with factor-valued `marks`, then the user has the option of using `nstar = "geometric"`, which sets
the sample size used in the formula to the geometric mean of the counts of observations of each mark. This can be useful for e.g. relative risk calculations, see Davies and Hazelton (2010).

**Spatial (and temporal) scale** The scaler argument is used to specify spatial (as well as temporal, in use of `OS.spattemp`) scale. For isotropic smoothing in the spatial margin, one may use the ‘robust’ estimate of standard deviation found by a weighted mean of the interquartile ranges of the \(x\)- and \(y\)-coordinates of the data respectively (\texttt{scaler = \textquoteleft IQR\textquoteleft}). Two other options are the raw mean of the coordinate-wise standard deviations (\texttt{scaler = \textquoteleft sd\textquoteleft}), or the square root of the mean of the two variances (\texttt{scaler = \textquoteleft var\textquoteleft}). A fourth option, \texttt{scaler = \textquoteleft silverman\textquoteleft} (default), sets the scaling constant to be the minimum of the "IQR" and "sd" options; see Silverman (1986), p. 47. In use of \texttt{OS.spattemp} the univariate version of the elected scale statistic is applied to the recorded times of the data for the temporal bandwidth. Alternatively, like nstar, the user can specify their own value by simply supplying a single positive numeric value to \texttt{scaler} for \texttt{OS}, or a numeric vector of length 2 (in the order of \texttt{[\textless spatial scale\textgreater, \textless temporal scale\textgreater]}) for \texttt{OS.spattemp}.

**Value**

A single numeric value of the estimated spatial bandwidth for \texttt{OS}, or a named numeric vector of length 2 giving the spatial bandwidth (as \texttt{h}) and the temporal bandwidth (as \texttt{lambda}) for \texttt{OS.spattemp}.

**Author(s)**

T.M. Davies

**References**


**Examples**

data(pbc)

\texttt{OS(pbc)}

\texttt{OS(pbc,nstar=\textbf{\textquoteleft geometric\textquoteleft}) \# uses case-control marks to replace sample size}

\texttt{OS(pbc,scaler=\textbf{\textquoteleft var\textquoteleft}) \# set different scalar measure of spread}

data(burk)

\texttt{OS.spattemp(burk\$cases)}

\texttt{OS.spattemp(burk\$cases,scaler=\textbf{\textquoteleft sd\textquoteleft})}
Description

Data of the locations of 761 cases of primary biliary cirrhosis in several adjacent health regions of north-eastern England, along with 3020 controls representing the at-risk population, collected between 1987 and 1994. These data were first presented and analysed by Prince et al. (2001); subsequent analysis of these data in the spirit of sparr was performed in Davies and Hazelton (2010). Also included is the polygonal study region.

Format

pbc is a dichotomously marked ppp.object, with locations expressed in UK Ordnance Survey Coordinates (km).

Acknowledgements

The authors thank Prof. Peter Diggle at Lancaster University (http://www.lancs.ac.uk/staff/diggle/) for providing access to these data.

Source


References


Examples

data(pbc)
summary(pbc)
plot(pbc)
plot.bivden

Plotting sparr objects

Description

plot methods for classes bivden, stden, rrs, rrst and msden.

Usage

## S3 method for class 'bivden'
plot(x, what = c("z", "edge", "bw"), add.pts = FALSE,
 auto.axes = TRUE, override.par = TRUE, ...)

## S3 method for class 'msden'
plot(x, what = c("z", "edge", "bw"), sleep = 0.2,
 override.par = TRUE, ...)

## S3 method for class 'rrs'
plot(x, auto.axes = TRUE, tol.show = TRUE,
 tol.type = c("upper", "lower", "two.sided"), tol.args = list(levels = 0.05, lty = 1, drawlabels = TRUE), ...)

## S3 method for class 'rrst'
plot(x, tselect = NULL, type = c("joint", "conditional"),
 fix.range = FALSE, tol.show = TRUE, tol.type = c("upper", "lower",
 "two.sided"), tol.args = list(levels = 0.05, lty = 1, drawlabels = TRUE),
 sleep = 0.2, override.par = TRUE, ...)

## S3 method for class 'stden'
plot(x, tselect = NULL, type = c("joint", "conditional"),
 fix.range = FALSE, sleep = 0.2, override.par = TRUE, ...)

Arguments

x An object of class bivden, stden, rrs, rrst, or msden.

what A character string to select plotting of result ("z": default); edge-correction surface ("edge"); or variable bandwidth surface ("bw").

add.pts Logical value indicating whether to add the observations to the image plot using default points.

auto.axes Logical value indicating whether to display the plot with automatically added x-y axes and an 'L' box in default styles.

override.par Logical value indicating whether to override the existing graphics device parameters prior to plotting, resetting mfrom and mar. See 'Details' for when you might want to disable this.

... Additional graphical parameters to be passed to plot.im, or in one instance, to plot.ppp (see 'Details').
sleep

Single positive numeric value giving the amount of time (in seconds) to `Sys.sleep` before drawing the next image in the animation.

tol.show

Logical value indicating whether to show pre-computed tolerance contours on the plot(s). The object `x` must already have the relevant `p`-value surface(s) stored in order for this argument to have any effect.

tol.type

A character string used to control the type of tolerance contour displayed; a test for elevated risk ("upper"), decreased risk ("lower"), or a two-tailed test (two.sided).

tol.args

A named list of valid arguments to be passed directly to `contour` to control the appearance of plotted contours. Commonly used items are `levels`, `lty`, `lwd` and `drawlabels`.

tselect

Either a single numeric value giving the time at which to return the plot, or a vector of length 2 giving an interval of times over which to plot. This argument must respect the stored temporal bound in `xDtlim`, else an error will be thrown. By default, the full set of images (i.e. over the entire available time span) is plotted.

type

A character string to select plotting of joint/unconditional spatiotemporal estimate (default) or conditional spatial density given time.

fix.range

Logical value indicating whether use the same color scale limits for each plot in the sequence. Ignored if the user supplies a pre-defined `colormap` to the `col` argument, which is matched to ... above and passed to `plot.im`. See 'Examples'.

Details

In all instances, visualisation is deferred to `plot.im`, for which there are a variety of customisations available the user can access via .... The one exception is when plotting observation-specific "diggle" edge correction factors—in this instance, a plot of the spatial observations is returned with size proportional to the influence of each correction weight.

When plotting a `rrs` object, a pre-computed `p`-value surface (see argument tolerate in `risk`) will automatically be superimposed at a significance level of 0.05. Greater flexibility in visualisation is gained by using `tolerance` in conjunction with `contour`.

An `msden`, `stden`, or `rrst` object is plotted as an animation, one pixel image after another, separated by sleep seconds. If instead you intend the individual images to be plotted in an array of images, you should first set up your plot device layout, and ensure `override.par = FALSE` so that the function does not reset these device parameters itself. In such an instance, one might also want to set `sleep = 0`.

Value

Plots to the relevant graphics device.

Author(s)

T.M. Davies
print.bivden

Examples

## Not run:
data(pbc)
data(fmd)
data(burk)

# 'bivden' object
pbcden <- bivariate.density(split(pbc)$case,h0=3,hp=2,adapt=TRUE,davies.baddeley=0.05,verbose=FALSE)
plot(pbcden)
plot(pbcden,what="bw",main="PBC cases\n variable bandwidth surface",xlab="Easting",ylab="Northing")

# 'stden' object
burkden <- spattemp.density(burk$cases,tres=128) # observation times are stored in marks(burk$cases)
plot(burkden,fix.range=TRUE,sleep=0.1) # animation
plot(burkden,tselect=c(1000,3000),type="conditional") # spatial densities conditional on each time

# 'rrs' object
pbcr <- risk(pbc,h0=4,hp=3,adapt=TRUE,tolerate=TRUE,davies.baddeley=0.025,edge="diggle")
plot(pbcr) # default
plot(pbcr,tol.args=list(levels=c(0.05,0.01),lty=2:1,col="seagreen4"),auto.axes=FALSE)

# 'rrst' object
f <- spattemp.density(fmd$cases,h=6,lambda=8)
g <- bivariate.density(fmd$controls,h=6)
fmdrr <- spattemp.risk(f,g,tolerate=TRUE)
plot(fmdrr,sleep=0.1,fix.range=TRUE)
plot(fmdrr,type="conditional",sleep=0.1,tol.type="two.sided",
     tol.args=list(levels=0.05,drawlabels=FALSE))

# 'msden' object
pbcmult <- multiscale.density(split(pbc)$case,h0=4,h0fac=c(0.25,2.5))
plot(pbcmult) # densities
plot(pbcmult,what="edge") # edge correction surfaces
plot(pbcmult,what="bw") # bandwidth surfaces

## End(Not run)

print.bivden

Printing sparr objects

Description

print methods for classes bivden, stden, rrs, rrst and msden.

Usage

## S3 method for class 'bivden'
print(x, ...)
## S3 method for class 'msden'
print(x, ...)

## S3 method for class 'rrs'
print(x, ...)

## S3 method for class 'rrst'
print(x, ...)

## S3 method for class 'stden'
print(x, ...)

### Arguments

- **x**
  - An object of class `bivden`, `stden`, `rrs`, `rrst`, or `msden`
  - Ignored.

- **...**
  - Ignored.

### Author(s)

- T.M. Davies

---

### risk

**Spatial relative risk/density ratio**

### Description

Estimates a relative risk function based on the ratio of two 2D kernel density estimates.

### Usage

```r
risk(f, g = NULL, log = TRUE, h0 = NULL, hp = h0, adapt = FALSE,
    tolerate = FALSE, doplot = FALSE, pilot.symmetry = c("none", "f", "g",
    "pooled"), epsilon = 0, verbose = TRUE, ...)
```

### Arguments

- **f**
  - Either a pre-calculated object of class `bivden` representing the ‘case’ (numerator) density estimate, or an object of class `ppp` giving the observed case data. Alternatively, if `f` is `ppp` object with dichotomous factor-valued `marks`, the function treats the first level as the case data, and the second as the control data, obviating the need to supply `g`.

- **g**
  - As for `f`, for the ‘control’ (denominator) density; this object must be of the same class as `f`. Ignored if, as stated above, `f` contains both case and control observations.
log Logical value indicating whether to return the (natural) log-transformed relative risk function as recommended by Kelsall and Diggle (1995a). Defaults to TRUE, with the alternative being the raw density ratio.

h0 A single positive numeric value or a vector of length 2 giving the global bandwidth(s) to be used for case/control density estimates; defaulting to a common oversmoothing bandwidth computed via os on the pooled data using nstar = "geometric" if unsupplied. Ignored if f and g are already bivden objects.

hp A single numeric value or a vector of length 2 giving the pilot bandwidth(s) to be used for fixed-bandwidth estimation of the pilot densities for adaptive risk surfaces. Ignored if adapt = FALSE or if f and g are already bivden objects.

adapt A logical value indicating whether to employ adaptive smoothing for internally estimating the densities. Ignored if f and g are already bivden objects.

tolerate A logical value indicating whether to internally calculate a corresponding asymptotic p-value surface (for tolerance contours) for the estimated relative risk function. See ‘Details’.

doplot Logical. If TRUE, an image plot of the estimated relative risk function is produced using various visual presets. If additionally tolerate was TRUE, asymptotic tolerance contours are automatically added to the plot at a significance level of 0.05 for elevated risk (for more flexible options for calculating and plotting tolerance contours, see tolerance and tol.contour).

pilot.symmetry A character string used to control the type of symmetry, if any, to use for the bandwidth factors when computing an adaptive relative risk surface. See ‘Details’. Ignored if adapt = FALSE.

epsilon A single non-negative numeric value used for optional scaling to produce additive constant to each density in the raw ratio (see ‘Details’). A zero value requests no additive constant (default).

verbose Logical value indicating whether to print function progress during execution.

... Additional arguments passed to any internal calls of bivariate.density for estimation of the requisite densities. Ignored if f and g are already bivden objects.

Details

The relative risk function is defined here as the ratio of the ‘case’ density to the ‘control’ (Bithell, 1990; 1991). Using kernel density estimation to model these densities (Diggle, 1985), we obtain a workable estimate thereof. This function defines the risk function \( r \) in the following fashion:

\[
r = \frac{fd + \text{epsilon}\times\max(gd)}{gd + \text{epsilon}\times\max(gd)},
\]

where \( fd \) and \( gd \) denote the case and control density estimates respectively. Note the (optional) additive constants defined by \( \text{epsilon} \) times the maximum of each of the densities in the numerator and denominator respectively (see Bowman and Azzalini, 1997).

The log-risk function \( \rho \), given by \( \rho = \log[r] \), is argued to be preferable in practice as it imparts a sense of symmetry in the way the case and control densities are treated (Kelsall and Diggle, 1995a;b). The option of log-transforming the returned risk function is therefore selected by default.
When computing adaptive relative risk functions, the user has the option of obtaining a so-called symmetric estimate (Davies et al. 2016) via `pilot.symmetry`. This amounts to choosing the same pilot density for both case and control densities. By choosing "none" (default), the result uses the case and control data separately for the fixed-bandwidth pilots, providing the original asymmetric density-ratio of Davies and Hazelton (2010). By selecting either of "f", "g", or "pooled", the pilot density is calculated based on the case, control, or pooled case/control data respectively (using `hp[1]` as the fixed bandwidth). Davies et al. (2016) noted some beneficial practical behaviour of the symmetric adaptive surface over the asymmetric.

If the user selects `tolerate = TRUE`, the function internally computes asymptotic tolerance contours as per Hazelton and Davies (2009) and Davies and Hazelton (2010). When `adapt = FALSE`, the reference density estimate (argument `ref.density` in `tolerance`) is taken to be the estimated control density. The returned pixel image of `p`-values (see ‘Value’) is interpreted as an upper-tailed test i.e. smaller `p`-values represent greater evidence in favour of significantly increased risk. For greater control over calculation of tolerance contours, use `tolerance`.

**Value**

An object of class "rrs". This is a named list with the following components:

- `rr` A pixel image of the estimated risk surface.
- `f` An object of class `bivden` used as the numerator or 'case' density estimate.
- `g` An object of class `bivden` used as the denominator or 'control' density estimate.
- `p` Only included if `tolerate = TRUE`. A pixel image of the `p`-value surface for tolerance contours; NULL otherwise.

**Author(s)**

T.M. Davies

**References**


Examples

```r
data(pbc)
pbccas <- split(pbc)$case
pbcccon <- split(pbc)$control
h0 <- OS(pbc,nstar="geometric")

# Fixed
pbcr1 <- risk(pbccas,pbccon,h0=h0,tolrate=TRUE)

# Asymmetric adaptive
pbcr2 <- risk(pbccas,pbccon,h0=h0,adapt=TRUE,hp=c(OS(pbccas)/2,OS(pbccon)/2),
             tolerate=TRUE,davies.baddeley=0.05)

# Symmetric (pooled) adaptive
pbcr3 <- risk(pbccas,pbccon,h0=h0,adapt=TRUE,tolrate=TRUE,hp=OS(pbc)/2,
             pilot.symmetry="pooled",davies.baddeley=0.05)

# Symmetric (case) adaptive; from two existing 'bivden' objects
f <- bivariate.density(pbccas,h0=h0,hp=2,adapt=TRUE,pilot.density=pbccas,
            edge="diggle",davies.baddeley=0.05,verbose=FALSE)
g <- bivariate.density(pbccon,h0=h0,hp=2,adapt=TRUE,pilot.density=pbccas,
            edge="diggle",davies.baddeley=0.05,verbose=FALSE)
pbcr4 <- risk(f,g,tolrate=TRUE,verbose=FALSE)

par(mfrow=c(2,2))
plot(pbcr1,override.par=FALSE,main="Fixed")
plot(pbcr2,override.par=FALSE,main="Asymmetric adaptive")
plot(pbcr3,override.par=FALSE,main="Symmetric (pooled) adaptive")
plot(pbcr4,override.par=FALSE,main="Symmetric (case) adaptive")
```

**Description**

Provides a fixed-bandwidth kernel estimate of continuous spatiotemporal data.

**Usage**

```r
spattemp.density(pp, h = NULL, tt = NULL, lambda = NULL,
                 tlim = NULL, sedge = c("uniform", "none"),
                 edge = sedge, sres = 128, tres = NULL, verbose = TRUE)
```
Arguments

**pp**
An object of class `ppp` giving the spatial coordinates of the observations to be smoothed. Possibly marked with the time of each event; see argument `tt`.

**h**
Fixed bandwidth to smooth the spatial margin. A numeric value > 0. If unsupplied, the oversmoothing bandwidth is used as per `os`.

**tt**
A numeric vector of equal length to the number of points in `pp`, giving the time corresponding to each spatial observation. If unsupplied, the function attempts to use the values in the `marks` attribute of the `pppNobject` in `pp`.

**lambda**
Fixed bandwidth to smooth the temporal margin; a numeric value > 0. If unsupplied, the function internally computes the Sheather-Jones bandwidth using `bw.SJ` (Sheather & Jones, 1991).

**tlim**
A numeric vector of length 2 giving the limits of the temporal domain over which to smooth. If supplied, all times in `tt` must fall within this interval (equality with limits allowed). If unsupplied, the function simply uses the range of the observed temporal values.

**sedge**
Character string dictating spatial edge correction. "uniform" (default) corrects based on evaluation grid coordinate. Setting `sedge="none"` requests no edge correction.

**tedge**
As `sedge`, for temporal edge correction.

**sres**
Numeric value > 0. Resolution of the `[sres x sres]` evaluation grid in the spatial margin.

**tres**
Numeric value > 0. Resolution of the evaluation points in the temporal margin as defined by the `tlim` interval. If unsupplied, the density is evaluated at integer values between `tlim[1]` and `tlim[2]`.

**verbose**
Logical value indicating whether to print a function progress bar to the console during evaluation.

Details

This function produces a fixed-bandwidth kernel estimate of a single spatiotemporal density, with isotropic smoothing in the spatial margin, as per Fernando & Hazelton (2014). Estimates may be edge-corrected for an irregular spatial study window and for the bounds on the temporal margin as per `tlim`; this edge-correction is performed in precisely the same way as the "uniform" option in `bivariate.density`.

Specifically, for $n$ trivariate points in space-time (`pp`, `tt`, `tlim`), we have

$$
\hat{f}(x, t) = n^{-1} \sum_{i=1}^{n} h^{-2} \lambda^{-1} K((x - x_i)/h) L((t - t_i)/\lambda)/(q(x)q(t)),
$$

where $x \in W \subset R^2$ and $t \in T \subset R$; $K$ and $L$ are the 2D and 1D Gaussian kernels controlled by fixed bandwidths $h$ (h) and $\lambda$ (lambda) respectively; and $q(x) = \int_{W} h^{-2} K((u - x)/h) du$ and $q(t) = \int_{T} \lambda^{-1} L((w - t)/\lambda) dw$ are optional edge-correction factors (sedge and tedge).

The above equation provides the **joint or unconditional** density at a given space-time location $(x, t)$. In addition to this, the function also yields the **conditional** density at each grid time, defined as

$$
\hat{f}(x|t) = \hat{f}(x, t)/\hat{f}(t),
$$
where $\hat{f}(t) = n^{-1} \sum_{i=1}^{n} \lambda^{-1} L((t - t_i)/\lambda)/q(t)$ is the univariate kernel estimate of the temporal margin. Normalisation of the two versions $\hat{f}(x,t)$ and $\hat{f}(x|t)$ is the only way they differ. Where in the unconditional setting we have $\int_W \int_T \hat{f}(x,t) dt dx = 1$, in the conditional setting we have $\int_W \hat{f}(x|t) dx = 1$ for all $t$. See Fernando & Hazelton (2014) for further details and practical reasons as to why we might prefer one over the other in certain situations.

The objects returned by this function (see ‘Value’ below) are necessary for kernel estimation of spatiotemporal relative risk surfaces, which is performed by `spattemp.risk`.

### Value

An object of class "stden". This is effectively a list with the following components:

- **z**: A named (by time-point) list of pixel images corresponding to the joint spatiotemporal density over space at each discretised time.
- **z.cond**: A named (by time-point) list of pixel images corresponding to the conditional spatial density given each discretised time.
- **h**: The scalar bandwidth used for spatial smoothing.
- **lambda**: The scalar bandwidth used for temporal smoothing.
- **tlim**: A numeric vector of length two giving the temporal bound of the density estimate.
- **spatial.z**: A pixel image giving the overall spatial margin as a single 2D density estimate (i.e. ignoring time).
- **temporal.z**: An object of class `density` giving the overall temporal margin as a single 1D density estimate (i.e. ignoring space).
- **qs**: A pixel image giving the edge-correction surface for the spatial margin. NULL if `sedge = "none"`.
- **qt**: A numeric vector giving the edge-correction weights for the temporal margin. NULL if `tedge = "none"`.
- **pp**: A `ppp` object of the spatial data passed to the argument of the same name in the initial function call, with `marks` of the observation times.
- **tgrid**: A numeric vector giving the discretised time grid at which the spatiotemporal density was evaluated (matches the names of `z` and `z.cond`).

### Author(s)

T.M. Davies

### References


New York.

See Also

*bivariate-density, spattemp-risk, spattemp-slice*

Examples

data(burk)
burkcas <- burk$cases

burkden1 <- spattemp-density(burkcas,tres=128)
summary(burkden1)

## Not run:

hlam <- LIK.spattemp(burkcas,tlim=c(400,5900),verbose=FALSE)
burkden2 <- spattemp-density(burkcas,h=hlam[1],lambda=hlam[2],tlim=c(400,5900),tres=256)
tims <- c(1000,2000,3500)
par(mfcol=c(2,3))
for(i in tims){
  plot(burkden2,i,override.par=FALSE,fix.range=TRUE,main=paste("joint",i))
  plot(burkden2,i,"conditional",override.par=FALSE,main=paste("cond.",i))
}

## End(Not run)

spattemp-risk  Spatiotemporal relative risk/density ratio

Description

Produces a spatiotemporal relative risk surface based on the ratio of two kernel estimates of spa-
tiotemporal densities.

Usage

spattemp-risk(f, g, log = TRUE, tolerate = FALSE, finiteness = TRUE, verbose = TRUE)

Arguments

f
An object of class *stden* representing the ‘case’ (numerator) density estimate.

g
Either an object of class *stden*, or an object of class *bivden* for the ‘control’
(denominator) density estimate. This object must match the spatial (and tempo-
ral, if *stden*) domain of f completely; see ‘Details’.
log
Logical value indicating whether to return the log relative risk (default) or the raw ratio.

tolerate
Logical value indicating whether to compute and return asymptotic $p$-value surfaces for elevated risk; see ‘Details’.

finiteness
Logical value indicating whether to internally correct infinite risk (on the log-scale) to the nearest finite value to avoid numerical problems. A small extra computational cost is required.

verbose
Logical value indicating whether to print function progress during execution.

Details
Fernando & Hazelton (2014) generalise the spatial relative risk function (e.g. Kelsall & Diggle, 1995) to the spatiotemporal domain. This is the implementation of their work, yielding the generalised log-relative risk function for $x \in W \subset \mathbb{R}^2$ and $t \in T \subset \mathbb{R}$. It produces

$$\hat{\rho}(x, t) = \log(\hat{f}(x, t)) - \log(\hat{g}(x, t)),$$

where $\hat{f}(x, t)$ is a fixed-bandwidth kernel estimate of the spatiotemporal density of the cases (argument $f$) and $\hat{g}(x, t)$ is the same for the controls (argument $g$).

- When argument $g$ is an object of class stden arising from a call to spattemp.density, the resolution, spatial domain, and temporal domain of this spatiotemporal estimate must match that of $f$ exactly, else an error will be thrown.

- When argument $g$ is an object of class bivden arising from a call to bivariate.density, it is assumed the ‘at-risk’ control density is static over time. In this instance, the above equation for the relative risk becomes $\hat{\rho} = \log(\hat{f}(x, t)) + \log |T| - \log(g(x))$. The spatial density estimate in $g$ must match the spatial domain of $f$ exactly, else an error will be thrown.

- The estimate $\hat{\rho}(x, t)$ represents the joint or unconditional spatiotemporal relative risk over $W \times T$. This means that the raw relative risk $\hat{r}(x, t) = \exp \hat{\rho}(x, t)$ integrates to 1 with respect to the control density over space and time: $\int_W \int_T r(x,t)g(x,t)dtdx = 1$. This function also computes the conditional spatiotemporal relative risk at each time point, namely

$$\hat{\rho}(x|t) = \log \hat{f}(x|t) - \log \hat{g}(x|t),$$

where $\hat{f}(x|t)$ and $\hat{g}(x|t)$ are the conditional densities over space of the cases and controls given a specific time point $t$ (see the documentation for spattemp.density). In terms of normalisation, we therefore have $\int_W r(x|t)g(x|t)dx = 1$. In the case where $\hat{g}$ is static over time, one may simply replace $\hat{g}(x|t)$ with $\hat{g}(x)$ in the above.

- Based on the asymptotic properties of the estimator, Fernando & Hazelton (2014) also define the calculation of tolerance contours for detecting statistically significant fluctuations in such spatiotemporal log-relative risk surfaces. This function can produce the required $p$-value surfaces by setting tolerate = TRUE; and if so, results are returned for both the unconditional (x,t) and conditional (x|t) surfaces. See the examples in the documentation for plot_rrst for details on how one may superimpose contours at specific $p$-values for given evaluation times $t$ on a plot of relative risk on the spatial margin.
Value

An object of class “rrst”. This is effectively a list with the following members:

- **rr**: A named (by time-point) list of pixel images corresponding to the joint spatiotemporal relative risk over space at each discretised time.
- **rr.cond**: A named list of pixel images corresponding to the conditional spatial relative risk given each discretised time.
- **P**: A named list of pixel images of the \( p \)-value surfaces testing for elevated risk for the joint estimate. If \( \text{tolerate} = \text{FALSE} \), this will be NULL.
- **P.cond**: As above, for the conditional relative risk surfaces.
- **f**: A copy of the object \( f \) used in the initial call.
- **g**: As above, for \( g \).
- **tlim**: A numeric vector of length two giving the temporal bound of the density estimate.

Author(s)

T.M. Davies

References


See Also

- spattemp.density, spattemp.slice, bivariate.density

Examples

```r
## Not run:
data(fmd)
fmdcas <- fmd$cases
defmdcon <- fmd$controls

f <- spattemp.density(fmdcas,h=6,lambda=8)  # stden object as time-varying case density
g <- bivariate.density(fmdcon,h0=6)  # bivden object as time-static control density
rho <- spattemp.risk(f,g,tolerate=TRUE)
print(rho)

par(mfrow=c(2,3))
plot(rho$f$spatial.z,main="Spatial margin (cases)")  # spatial margin of cases
plot(rho$f$temporal.z,main="Temporal margin (cases)")  # temporal margin of cases
plot(rho$g$z,main="Spatial margin (controls)")  # spatial margin of controls
plot(rho,tselect=50,type="conditional",tol.args=list(levels=c(0.05,0.0001),
    lty=2:1,lwd=1:2),override.par=FALSE)
plot(rho,tselect=100,type="conditional",tol.args=list(levels=c(0.05,0.0001),
```
spattemp.slice  
  
  lty=2:1, lwd=1:2, override.par=FALSE)  
  plot(rho, tselect=200, type="conditional", tol.args=list(levels=c(0.05, 0.0001)),  
       lty=2:1, lwd=1:2, override.par=FALSE)  
  
  ## End(Not run)  
  
spattemp.slice  

Slicing a spatiotemporal object

Description

Takes slices of the spatiotemporal kernel density or relative risk function estimate at desired times

Usage

spattemp.slice(stob, tt, checkargs = TRUE)

Arguments

- **stob**: An object of class `stden` or `rrst` giving the spatiotemporal estimate from which to take slices.
- **tt**: Desired time(s); the density/risk surface estimate corresponding to which will be returned. This value must be in the available range provided by `stob$tlim`; see ‘Details’.
- **checkargs**: Logical value indicating whether to check validity of `stob` and `tt`. Disable only if you know this check will be unnecessary.

Details

Contents of the `stob` argument are returned based on a discretised set of times. This function internally computes the desired surfaces as pixel-by-pixel linear interpolations using the two discretised times that bound each requested `tt`.

The function returns an error if any of the requested slices at `tt` are not within the available range of times as given by the `tlim` component of `stob`.

Value

A list of lists of pixel images, each of which corresponds to the requested times in `tt`, and are named as such.

If `stob` is an object of class `stden`:

- **z**: Pixel images of the joint spatiotemporal density corresponding to `tt`.
- **z.cond**: Pixel images of the conditional spatiotemporal density given each time in `tt`.

If `stob` is an object of class `rrst`:

- **rr**: Pixel images of the joint spatiotemporal relative risk corresponding to `tt`.


Pixel images of the conditional spatiotemporal relative risk given each time in \( t_t \).

**P**

Only present if `tolerate = TRUE` in the preceding call to `spattemp.risk`.

Pixel images of the \( p \)-value surfaces for the joint spatiotemporal relative risk.

**P.cond**

Only present if `tolerate = TRUE` in the preceding call to `spattemp.risk`.

Pixel images of the \( p \)-value surfaces for the conditional spatiotemporal relative risk.

### Author(s)

T.M. Davies

### References


### See Also

`spattemp-density, spattemp-risk, bivariate-density`

### Examples

```r
## Not run:
data(fmd)
fmdcas <- fmd$cases
fmdcon <- fmd$controls

f <- spattemp-density(fmdcas,h=6,lambda=8)
g <- bivariate-density(fmdcon,h0=6)
rho <- spattemp-risk(f,g,tolerate=TRUE)

f$tlim # requested slices must be in this range

# slicing 'stden' object
f.slice1 <- spattemp.slice(f,tt=50) # evaluation timestamp
f.slice2 <- spattemp.slice(f,tt=150.5) # interpolated timestamp
par(mfrow=c(2,2))
plot(f.slice1$z$'50')
plot(f.slice1$z.cond$'50')
plot(f.slice2$z$'150.5')
plot(f.slice2$z.cond$'150.5')

# slicing 'rrst' object
rho.slice <- spattemp.slice(rho,tt=c(50,150.5))
par(mfrow=c(2,2))
plot(rho.slice$r$'50');tol.contour(rho.slice$P$'50',levels=0.05,add=TRUE)
plot(rho.slice$r$'150.5');tol.contour(rho.slice$P$'150.5',levels=0.05,add=TRUE)
plot(rho.slice$rr.cond$'50');tol.contour(rho.slice$P.cond$'50',levels=0.05,add=TRUE)
plot(rho.slice$rr.cond$'150.5');tol.contour(rho.slice$P.cond$'150.5',levels=0.05,add=TRUE)
```
## Description

summary methods for classes `bivden`, `stden`, `rrs`, `rrst` and `msden`.

## Usage

```r
## S3 method for class 'bivden'
summary(object, ...)

## S3 method for class 'msden'
summary(object, ...)

## S3 method for class 'rrs'
summary(object, ...)

## S3 method for class 'rrst'
summary(object, ...)

## S3 method for class 'stden'
summary(object, ...)
```

## Arguments

- **object**: An object of class `bivden`, `stden`, `rrs`, `rrst`, or `msden`.
- **...**: Ignored.

## Author(s)

T.M. Davies
**tol.contour**

*Plot tolerance contours*

**Description**

Draw contours based on a $p$-value matrix.

**Usage**

```r
tol.contour(pim, test = c("upper", "lower", "two-sided"), ...)
```

**Arguments**

- `pim`: A pixel image of $p$-values, typically obtained from a call to `tolerance`, computed with respect to a test for elevated risk.
- `test`: An optional character string giving the type of manipulation to be applied to the $p$-values, corresponding to a test for significantly elevated risk ("upper"; default); for reduced risk ("lower"); or for both ("two-sided").
- `...`: Additional arguments to be passed to `contour`. Commonly used options include `add` (to superimpose the contours upon an existing plot); `levels` (to control the specific significance levels at which to delineate the $p$-values); and `lty` or `lwd` for aesthetics.

**Details**

Note that no checks on the numeric content of `pim` are made. The function assumes the pixel image of $p$-values in `pim` is supplied with respect to an upper-tailed test for elevated risk (this is exactly the way the $p$-value surface is returned when `tolerance` is used). This is important if one makes subsequent use of `test`, which manipulates the $p$-values to draw at desired significance levels.

**Value**

Opens a new graphics device and displays a `contour` plot if `add = FALSE`, otherwise adds the contours to the plot in the existing active graphics device.

**Author(s)**

T. M. Davies

**Examples**

```r
# See ?tolerance
```
Description

Calculates a p-value surface based on asymptotic theory or Monte-Carlo (MC) permutations describing the extremity of risk given a fixed or adaptive kernel-smoothed density-ratio, allowing the drawing of tolerance contours.

Usage

tolerance(rs, method = c("ASY", "MC"), ref.density = NULL, beta = 0.025, ITER = 100, parallelise = NULL, verbose = TRUE, ...)

Arguments

rs
An object of class rrs giving the estimated relative risk function for which to calculate the p-value surface.

method
A character string specifying the method of calculation. "ASY" (default) instructs the function to compute the p-values using asymptotic theory. "MC" computes the values by random permutations of the data. See ‘Details’.

ref.density
Required if rs is based on fixed-bandwidth estimates of the case and control densities and method = "ASY". Either a pixel image or an object of class bivden giving the reference density to use in asymptotic formulae. May be unnormalised. Ignored if rs is based on adaptive kernel estimates or if method = "MC".

beta
A numeric value 0 < beta < 1 giving the fineness of the adaptive bandwidth partitioning to use for calculation of the required quantities for asymptotic adaptive p-value surfaces. Smaller values provide more accurate bandwidth bins at the cost of additional computing time, see Davies and Baddeley (2017); the default is sensible in most cases. Ignored if rs is based on fixed-bandwidth kernel estimates.

ITER
Number of iterations for the Monte-Carlo permutations. Ignored if method = "ASY".

parallelise
Numeric argument to invoke parallel processing, giving the number of CPU cores to use when method = "MC". Experimental. Test your system first using parallel::detectCores() to identify the number of cores available to you.

verbose
Logical value indicating whether to print function progress during execution.

...
Additional arguments to be passed to risk when method = "MC". While most information needed for the MC repetitions is implicitly gleaned from the object passed to rs, this ellipsis is typically used to set the appropriate epsilon and pilot.symmetry values for the internal calls to risk.
Details

This function implements developments in Hazelton and Davies (2009) (fixed) and Davies and Hazelton (2010) (adaptive) to compute pointwise $p$-value surfaces based on asymptotic theory of kernel-smoothed relative risk surfaces. Alternatively, the user may elect to calculate the $p$-value surfaces using Monte-Carlo methods (see Kelsall and Diggle, 1995). Superimposition upon a plot of the risk surface contours of these $p$-values at given significance levels (i.e. “tolerance contours”) can be an informative way of exploring the statistical significance of the extremity of risk across the defined study region.

The Monte-Carlo method, while not dependent on asymptotic theory, is computationally expensive and it has been suggested might have some undesirable practical consequences in certain settings (Hazelton and Davies, 2009). When performing the MC simulations, the same global (and pilot, if necessary) bandwidths and edge-correction regimen is employed as were used in the initial density estimates of the observed data. With regard to arguments to be passed to internal calls of `risk`, the user should take care to use ... to set the epsilon value to match that which was used in creation of the object passed to `rs` (if this was set to a non-default value). Furthermore, if performing MC simulations for the adaptive relative risk function, the function borrows the value of the beta argument to speed things up via partitioning, and the user should additionally access ... to set the same pilot symmetry value as was used for creation of the object passed to `rs`, in the same way as for any non-default use of epsilon. This will ensure the simulations are all performed under the same conditions as the original risk function.

The asymptotic approach to the $p$-value calculation can be advantageous over a Monte-Carlo method, which can lead to excessive computation time for adaptive risk surfaces and large datasets. See the aforementioned references for further comments.

Choosing different options for the argument `test` simply manipulates the ‘direction’ of the $p$-values. That is, plotting tolerance contours at a significance level of 0.05 for a $p$-value surface calculated with `test = "double"` is equivalent to plotting tolerance contours at significance levels of 0.025 and 0.975 for `test = "upper"`.

Implementation of the Monte-Carlo contours for the fixed-bandwidth estimator simply involves random allocation of case/control marks and re-estimation of the risk surface ITER times, against which the original estimate is compared. The bandwidth(s) for case and control densities in the permuted estimates are controlled by `hpsim`. If your risk surface is adaptive, `hpsim` is used to control the pilot bandwidths, `hpsim` the global bandwidths. In particular, for the adaptive symmetric estimator (Davies et al. 201), it is assumed that the original estimate was itself calculated as a symmetric estimate via use of the `pdef` argument. The `symchoice` argument governs the specific permuted data set to use for the pilot estimate, and if `hpsim` is NULL, the pilot bandwidth thereof is taken from the stored `pdef` object in the original estimate. An error will occur if you attempt to set `symchoice` with an `rs` argument in this function that does not contain density estimates with present `pdef` objects of class “bivden”. See the help file for `bivariate.density` for details on using the `pdef` argument.

In addition to the usage noted above, you may define either `hpsim` and/or `hpsim` as functions which re-calculate the case and control pilot (or fixed) bandwidth(s) and the global bandwidth(s) at each iteration, based on the data set of the permuted case/control marks. If so, these must strictly be functions that take the case data as the first argument and the control data as the second argument, each as a two-column matrix of the x-y coordinates. The function must strictly return a numeric vector of length 1 or 2; these entries to be assigned to the relevant density estimates as per the usage notes on supply of a numeric vector for `hpsim`. Take care – warnings will be issued if, for
example, you specify a hpsim function that returns two numeric values, but your rs object is a symmetric-adaptive estimate (in which case it only makes sense to yield one pilot bandwidth!)

Value

A pixel image of the estimated $p$-value surface.

Note

The returned $p$-values are geared so that “smallness” corresponds to statistical significance of elevated risk, that is, an upper-tailed test. The complement of the $p$-values will yield significance of reduced risk; a lower-tailed test. When using tol.contour, the user can control what type of contours to display.

Author(s)

T. M. Davies

References

Kelsall, J.E. and Diggle, P.J. (1995), Kernel estimation of relative risk, Bernoulli, 1, 3-16.

Examples

```r
## Not run:
data(pbc)
h0 <- LSCV.risk(pbc,method="hazelton");h0
pbccas <- split(pbc)[[1]]
pbcon <- split(pbc)[[2]]

# ASY
riskfix <- risk(pbc,h0=h0)
fixtol1 <- tolerance(riskfix,ref.density=density(pbc,OS(pbc)))

riskada <- risk(pbc,h0=h0,adapt=TRUE,hp=NS(pbc),pilot.symmetry="pooled",davies.baddeley=0.025)
adatol1 <- tolerance(riskada)

par(mfrow=c(1,2))
plot(riskfix)
tol.contour(fixtol1,levels=c(0.1,0.05,0.01),lty=3:1,add=TRUE)
plot(riskada)
tol.contour(adatol1,levels=c(0.1,0.05,0.01),lty=3:1,add=TRUE)
```
# MC
fixtol2 <- tolerance(riskfix, method="MC", ITER=200)
adatol2 <- tolerance(riskada, method="MC", ITER=200, parallelise=4) # ~1 minute with parallelisation
par(mfrow=c(1,2))
plot(riskfix)
tol.contour(fixtol2, levels=c(0.1, 0.05, 0.01), lty=3:1, add=TRUE)
plot(riskada)
tol.contour(adatol2, levels=c(0.1, 0.05, 0.01), lty=3:1, add=TRUE)

## End(Not run)
Index

*Topic data
  burk, 18
  fmd, 20
  pbc, 38

*Topic package
  sparr-package, 2

as.mask, 9, 30
available.h0, 7

bivariate.density, 3, 4, 8, 12, 13, 15, 22, 28, 30, 32, 33, 43, 46, 48–50, 52, 56
bivden, 12, 13, 26, 31, 33, 39, 41–44, 48, 49, 53, 55
bivden (bivariate.density), 8
BOOT.density, 3, 12
BOOT.spattemp, 4, 15, 25
burk, 3, 18
burkitt, 18
bw.diggle, 23
bw.frac, 23
bw.ppl, 23
bw.scott, 23
bw.SJ, 46

chorley, 3
colourmap, 40
contour, 40, 54
density, 47
density.ppp, 8–10
doParallel, 5

fft, 19
fft2d, 19
fmd, 3, 20
foreach, 5, 10

im, 5, 9, 10, 30, 31, 44, 47, 50, 51, 54, 55, 57
LIK.density, 3, 21

LIK.spattemp, 4, 24
LSCV.density, 3, 27, 28
LSCV.density (LIK.density), 21
LSCV.risk, 3, 26
LSCV.spattemp, 4, 17
LSCV.spattemp (LIK.spattemp), 24

marks, 16, 20, 24, 26, 34, 36, 42, 46, 47
msden, 7, 32, 33, 39–42, 53
msden (multiscale.density), 29
multiscale.density, 3, 4, 7, 13–15, 22, 27, 29, 33
multiscale.slice, 4, 7, 31, 32, 32

NS, 3, 34
NS.spattemp, 4

optim, 16, 17, 24
optimise, 21, 27
OS, 3, 12, 13, 15, 16, 36, 43, 46
OS.spattemp, 4
owin, 18

parallel, 5
pbc, 3, 38
plot.bivden, 4, 39
plot.im, 39, 40
plot.msden, 4
plot.msden (plot.bivden), 39
plot.ppp, 39
plot.rrs, 4
plot.rrs (plot.bivden), 39
plot.rrst, 4, 49
plot.rrst (plot.bivden), 39
plot.std, 4
plot.std (plot.bivden), 39
points, 39
ppp, 5, 8, 12, 16, 18, 20, 21, 24, 26, 30, 34, 36, 42, 46
ppp.object, 3, 9, 10, 16, 24, 30, 31, 34, 36, 38, 46, 47
print.bivden, 4, 41
print.msd, 4
print.msd (print.bivden), 41
print.rrs, 4
print.rrs (print.bivden), 41
print.rrst, 4
print.rrst (print.bivden), 41
print.std, 4
print.std (print.bivden), 41

risk, 4, 40, 42, 55, 56
rrs, 39–42, 53, 55
rrs (risk), 42
rrst, 39–42, 51, 53
rrst (spattemp.risk), 48

sparr, 38
sparr (sparr-package), 2
sparr-package, 2
spatstat, 3, 5, 23
spattemp.density, 4, 17, 25, 45, 49, 50, 52
spattemp.risk, 4, 47, 48, 48, 52
spattemp.slice, 4, 48, 50, 51
std, 16, 17, 39–42, 48, 49, 51, 53
std (spattemp.density), 45
summary.bivden, 4, 53
summary.msd, 4
summary.msd (summary.bivden), 53
summary.rrs, 4
summary.rrs (summary.bivden), 53
summary.rrst, 4
summary.rrst (summary.bivden), 53
summary.std, 4
summary.std (summary.bivden), 53
Sys.sleep, 40

tol.contour, 4, 43, 54, 57
tolerance, 4, 40, 43, 44, 54, 55