HDiR: An R Package for Nonparametric Plug-in Estimation of Directional Highest Density Regions

Paula Saavedra-Nieves* and Rosa M. Crujeiras*

*Universidade de Santiago de Compostela

VIII Xornada de Usuarios de R en Galicia
14th October 2021, Santiago de Compostela

1. Plug-in estimation of highest density regions in the Euclidean setting
   - Motivation from a leukaemia data set
   - Highest density regions estimation in the Euclidean space
   - Highest density regions estimation using R
   - Analysis of leukaemia data set

2. Plug-in estimation of highest density regions in the directional setting
   - Motivation from a circular and a spherical data sets
   - Highest density regions estimation in the directional space
   - Highest density regions estimation using the R package **HDiR**
   - Analysis of the circular and the spherical data sets using **HDiR**

A real problem: Leukaemia analysis

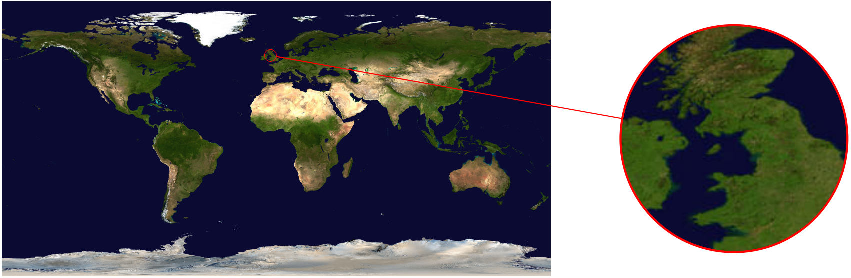


Figure: Geographical location of the North West of England.

NASA Terra satellite.

Leukaemia real example

Is there an excess of case intensity over that of population?

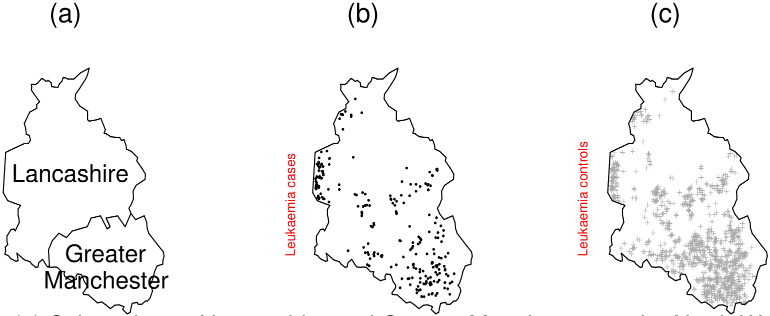


Figure: (a) Sub-regions of Lancashire and Greater Manchester on the North West of England. (b) distribution of 233 cases of diagnosed leukaemia between 1982 and 1998. (c) 988 controls on Lancashire and Greater Manchester.

Prof. Peter J. Diggle's website, Lancaster University.

Highest Density Regions estimation in $\mathbb{R}^d$

Given a random sample of points $\mathcal{X}_n = \{X_1, ..., X_n\}$ of a random vector X with values in $\mathbb{R}^d$, reconstructing the $t-$level set

$$G(t) = \{x \in \mathbb{R}^d : f(x) \geq t\}$$

where f denotes the density function of X and $t > 0$. Or, if the practitioner fixes a value $\tau \in (0, 1)$, estimating the Highest Density Region (HDR) with probability content $1 - \tau$

$$L(\tau) = \{x \in \mathbb{R}^d : f(x) \geq f_\tau\}$$

where f_τ can be seen as the largest constant such that

$$\mathbb{P}(X \in L(\tau)) \geq 1 - \tau$$

with respect to the distribution induced by f.

Highest Density Regions estimation in $\mathbb{R}^d$

Given a random sample of points $\mathcal{X}_n = \{X_1, ..., X_n\}$ of a random vector X with values in $\mathbb{R}^d$, reconstructing the $t-$level set

$$G(t) = \{x \in \mathbb{R}^d : f(x) \geq t\}$$

where f denotes the density function of X and $t > 0$. Or, if the practitioner fixes a value $\tau \in (0, 1)$, estimating the Highest Density Region (HDR) with probability content $1 - \tau$

$$L(\tau) = \{x \in \mathbb{R}^d : f(x) \geq f_\tau\}$$

where f_τ can be seen as the largest constant such that

$$\mathbb{P}(X \in L(\tau)) \geq 1 - \tau$$

with respect to the distribution induced by f.

Highest Density Regions estimation in $\mathbb{R}^d$

Given a random sample of points $\mathcal{X}_n = \{X_1, ..., X_n\}$ of a random vector X with values in $\mathbb{R}^d$, reconstructing the $t-$level set

$$G(t) = \{x \in \mathbb{R}^d : f(x) \geq t\}$$

where f denotes the density function of X and $t > 0$. Or, if the practitioner fixes a value $\tau \in (0, 1)$, estimating the Highest Density Region (HDR) with probability content $1 - \tau$

$$L(\tau) = \{x \in \mathbb{R}^d : f(x) \geq f_\tau\}$$

where f_τ can be seen as the largest constant such that

$$\mathbb{P}(X \in L(\tau)) \geq 1 - \tau$$

with respect to the distribution induced by f.

One-dimensional example

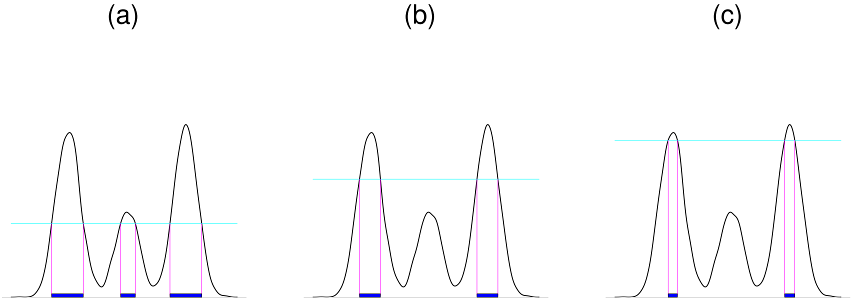


Figure: One-dimensional HDRs for a trimodal density taking τ equal to (a) 0.25, (b) 0.5 and (c) 0.75.

For low values of τ, HDR looks like the support of the distribution.

HDRs estimation

Two alternative routes for estimating a HDR:

- The only information we have comes from the sample.
  - Plug-in methodology.
- Some geometric properties of the HDR are known.
  - Excess mass methodology.
  - Hybrid methodology.

HDRs estimation

Two alternative routes for estimating a HDR:

- The only information we have comes from the sample.
  - Plug-in methodology.
- Some geometric properties of the HDR are known.
  - Excess mass methodology.
  - Hybrid methodology.

Plug-in estimation of HDRs

Plug-in methods propose

$$\hat{L}(\tau) = \{x \in \mathbb{R}^d : f_n(x) \geq \hat{f}_\tau\}$$

as an **estimator for $L(\tau)$** where

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_H(x - X_i)$$

where K is a symmetric density function with $K_H(z) = \mid H \mid^{-1/2} K(H^{1/2}z)$, H denotes the bandwidth matrix and $\hat{f}_\tau = f_\tau(f_n)$ denotes an estimator of the threshold f_τ.

Hyndman (1996) estimated f_τ as the quantile τ of the empirical distribution of $f_n(X_1), \cdots, f_n(X_n)$.

Hyndman, R.J. (1996). Computing and graphing highest density regions. The American Statistician, 50, 120–126.

Samworth, R.J. and Wand , M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. Annals of Statistics, 38, 1767–1792.

Plug-in estimation of HDRs

Plug-in methods propose

$$\hat{L}(\tau) = \{x \in \mathbb{R}^d : f_n(x) \geq \hat{f}_\tau\}$$

as an **estimator for $L(\tau)$** where

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_H(x - X_i)$$

where K is a symmetric density function with $K_H(z) = \mid H \mid^{-1/2} K(H^{1/2} z)$, H denotes the bandwidth matrix and $\hat{f}_\tau = f_\tau(f_n)$ denotes an estimator of the threshold f_τ.

Hyndman (1996) estimated f_τ as the quantile τ of the empirical distribution of $f_n(X_1), \cdots, f_n(X_n)$.

Hyndman, R.J. (1996). Computing and graphing highest density regions. The American Statistician, 50, 120–126.

Samworth, R.J. and Wand , M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. Annals of Statistics, 38, 1767–1792.

Plug-in estimation of HDRs

Plug-in methods propose

$$\hat{L}(\tau) = \{x \in \mathbb{R}^d : f_n(x) \geq \hat{f}_\tau\}$$

as an **estimator for $L(\tau)$** where

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_H(x - X_i)$$

where K is a symmetric density function with $K_H(z) = | H |^{-1/2} K(H^{1/2} z)$, H denotes the bandwidth matrix and $\hat{f}_\tau = f_\tau(f_n)$ denotes an estimator of the threshold f_τ.

Hyndman (1996) estimated f_τ as the quantile τ of the empirical distribution of $f_n(X_1), \cdots, f_n(X_n)$.

Hyndman, R.J. (1996). Computing and graphing highest density regions. The American Statistician, 50, 120–126.

Samworth, R.J. and Wand , M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. Annals of Statistics, 38, 1767–1792.

Highest density regions estimation with R

Plug-in estimation of Euclidean HDRs is already implemented in some packages:

`hdrcde`: It computes Euclidean HDRs in one and two dimensions. The specific HDR bandwidth selector proposed in Samworth and Wand (2010) is also implemented. Confidence regions for one-dimensional HDRs and bivariate HDRs scatterplots are also available.

`lsbs`: It implements the bandwidth selector for two-dimensional Euclidean HDRs proposed in Doss and Weng (2018).

Although not specifically designed for HDR or level set estimation, some packages such as **`sm`** and **`ks`** include tools for kernel density estimation allowing for graphical displays of density contours in two and three dimensions.

Doss, C. R. and Weng, G. (2018). Bandwidth selection for kernel density estimators of multivariate level sets and highest density regions. Electronic Journal of Statistics, 12(2), 4313-4376.

Samworth, R.J. and Wand , M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. Annals of Statistics, 38, 1767–1792.

Analysis of leukaemia data set

Is there an excess of case intensity over that of population?

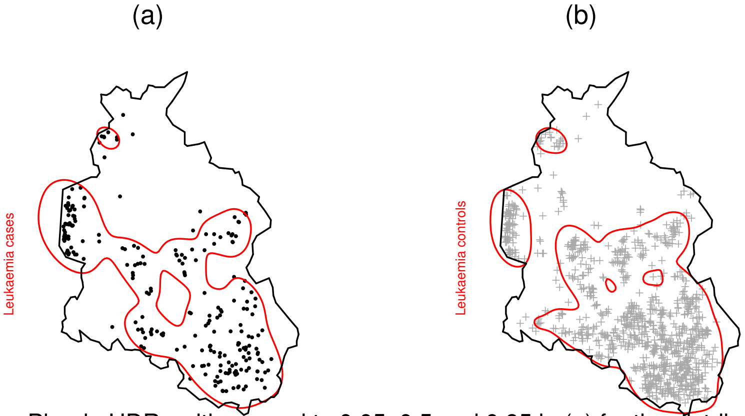


Figure: Plug-in HDRs with τ equal to 0.05, 0.5 and 0.95 in (a) for the distribution of 322 cases diagnosed of leukaemia and in (b) for the distribution of 988 controls.

Packages `ks` and `hdrcde`.

Analysis of leukaemia data set

Is there an excess of case intensity over that of population?

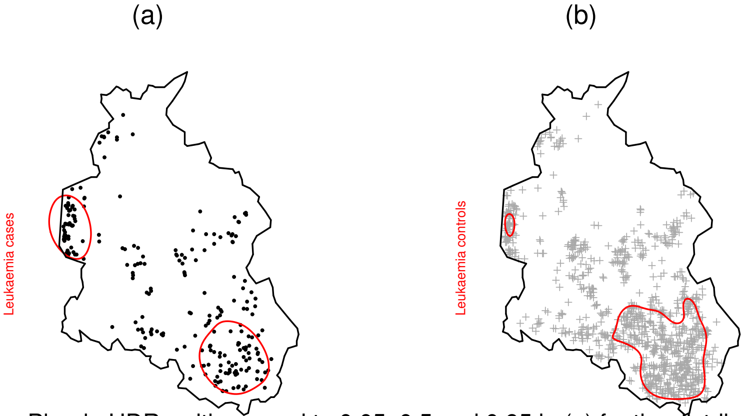


Figure: Plug-in HDRs with τ equal to 0.05, 0.5 and 0.95 in (a) for the distribution of 322 cases diagnosed of leukaemia and in (b) for the distribution of 988 controls.

Packages `ks` and `hdrcde`.

Analysis of leukaemia data set

Is there an excess of case intensity over that of population?

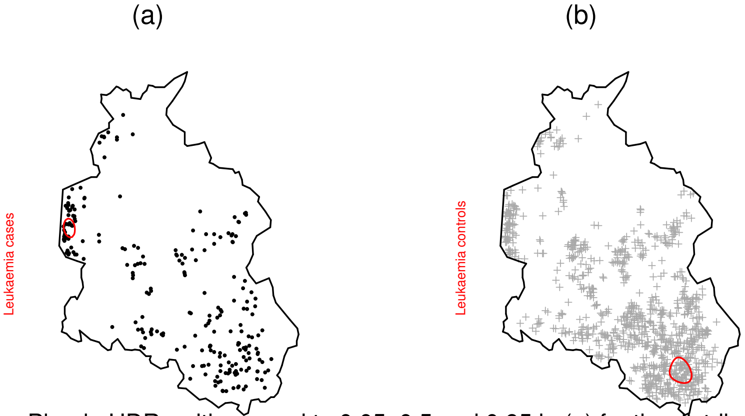


Figure: Plug-in HDRs with τ equal to 0.05, 0.5 and 0.95 in (a) for the distribution of 322 cases diagnosed of leukaemia and in (b) for the distribution of 988 controls.

Packages `ks` and `hdrcde`.

Other real problems (I): Sandhoppers orientation

Figure: Geographical location of Zouara beach in the Tunisian northwestern coast.

NASA Terra satellite.

Other real problems (I): Sandhoppers orientation

Does the moment of the day play a significant role in sandhoppers behavior?

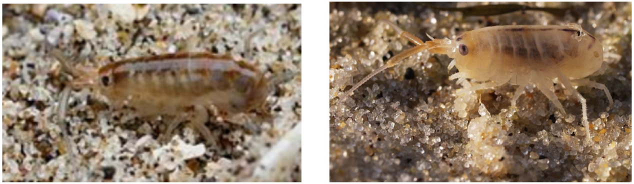


Figure: Talorchestia brito (left) and Talitrus saltator (right).

Scapini, F., Aloia, A., Bouslama, M. F., Chelazzi, L., Colombini, I., ElGtari, M., Fallaci, M. and Marchetti, G. M. (2002). Multiple regression analysis of the sources of variation in orientation of two sympatric sandhoppers, Talitrus saltator and Talorchestia brito, from an exposed Mediterranean beach, Behav. Ecol. Sociobiol., 51(5), 403-414.

Other real problems (I): Sandhoppers orientation

Does the moment of the day play a significant role in sandhoppers behavior?

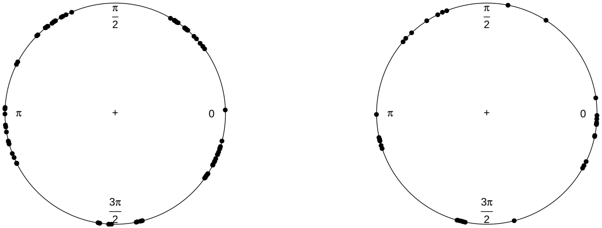


Figure: Orientation data (slightly jittered) corresponding to males of the specie Talitrus saltator registered in the morning (left) and in the afternoon (right) in April.

Scapini, F., Aloia, A., Bouslama, M. F., Chelazzi, L., Colombini, I., ElGtari, M., Fallaci, M. and Marchetti, G. M. (2002). Multiple regression analysis of the sources of variation in orientation of two sympatric sandhoppers, Talitrus saltator and Talorchestia brito, from an exposed Mediterranean beach, Behav. Ecol. Sociobiol., 51(5), 403-414.

Other real problems (II): Earthquakes distribution

Where are earthquakes likely to happen?

Play/Pause

Figure: Distribution of earthquakes around the world between October 2004 and April 2020.

European-Mediterranean Seismological Centre: `www.emsc-csem.org`.

HDRs estimation in S^{d-1}

Given a random sample of points $\mathcal{Y}_n = \{Y_1, ..., Y_n\}$ of a random vector Y with values in the unit sphere S^{d-1}, reconstructing the $t-$level set

$$G_g(t) = \{y \in S^{d-1} : g(y) \geq t\}$$

where g denotes the directional density function of Y and $t > 0$. Or, if the practitioner fixes a value $\tau \in (0, 1)$, estimating the Highest Density Region (HDR) with probability content $1 - \tau$

$$L_g(\tau) = \{y \in S^{d-1} : g(y) \geq g_\tau\}$$

where g_τ can be seen as the largest constant such that

$$\mathbb{P}(Y \in L_g(\tau)) \geq 1 - \tau$$

with respect to the distribution induced by g.

HDRs estimation in S^{d-1}

Given a random sample of points $\mathcal{Y}_n = \{Y_1, ..., Y_n\}$ of a random vector Y with values in the unit sphere S^{d-1}, reconstructing the $t-$level set

$$G_g(t) = \{y \in S^{d-1} : g(y) \geq t\}$$

where g denotes the directional density function of Y and $t > 0$. Or, if the practitioner fixes a value $\tau \in (0, 1)$, estimating the Highest Density Region (HDR) with probability content $1 - \tau$

$$L_g(\tau) = \{y \in S^{d-1} : g(y) \geq g_\tau\}$$

where g_τ can be seen as the largest constant such that

$$\mathbb{P}(Y \in L_g(\tau)) \geq 1 - \tau$$

with respect to the distribution induced by g.

HDRs estimation in S^{d-1}

Given a random sample of points $\mathcal{Y}_n = \{Y_1, ..., Y_n\}$ of a random vector Y with values in the unit sphere S^{d-1}, reconstructing the $t-$level set

$$G_g(t) = \{y \in S^{d-1} : g(y) \geq t\}$$

where g denotes the directional density function of Y and $t > 0$. Or, if the practitioner fixes a value $\tau \in (0, 1)$, estimating the Highest Density Region (HDR) with probability content $1 - \tau$

$$L_g(\tau) = \{y \in S^{d-1} : g(y) \geq g_\tau\}$$

where g_τ can be seen as the largest constant such that

$$\mathbb{P}(Y \in L_g(\tau)) \geq 1 - \tau$$

with respect to the distribution induced by g.

Circular examples

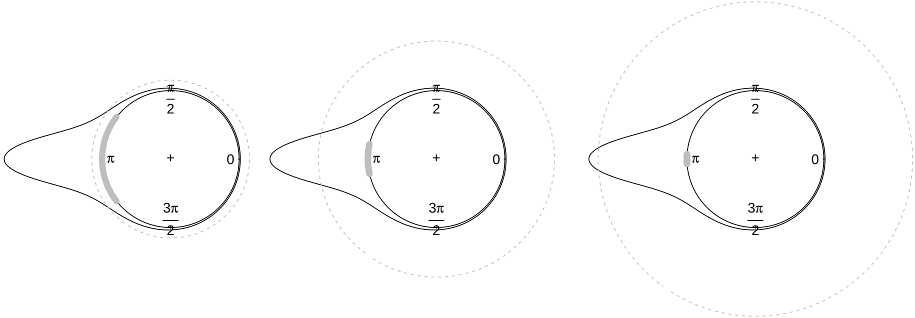


Figure: HDRs for $\tau = 0.2$ (first column), $\tau = 0.5$ (second column) and $\tau = 0.8$ (third column).

Function `circ.hdr` in package `HDiR`.

Circular examples

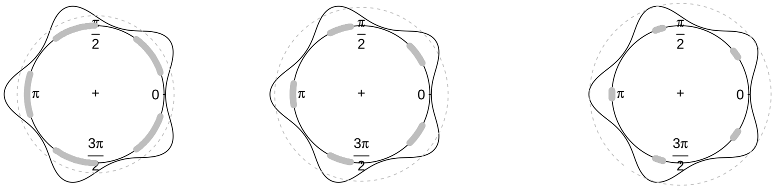


Figure: HDRs for $\tau = 0.2$ (first column), $\tau = 0.5$ (second column) and $\tau = 0.8$ (third column).

Function `circ.hdr` in package `HDiR`.

Circular examples

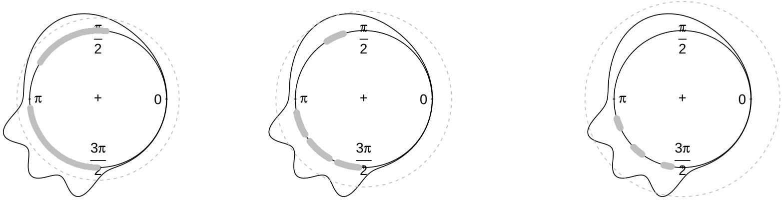


Figure: HDRs for $\tau = 0.2$ (first column), $\tau = 0.5$ (second column) and $\tau = 0.8$ (third column).

Function `circ.hdr` in package `HDiR`.

Spherical examples

Figure: Finite mixtures of von Mises-Fisher spherical models for simulations. HDRs are represented for $\tau = 0.2$, $\tau = 0.5$ and $\tau = 0.8$.

Function `sphere.hdr` in package `HDiR`.

Plug-in estimation of directional HDRs

Plug-in methods propose

$$\hat{L}_g(\tau) = \{y \in S^{d-1} : g_n(y) \geq \hat{g}_\tau\}$$

as an **estimator for $L_g(\tau)$** where

$$g_n(y) = \frac{1}{n}\sum_{i=1}^{n} K_{vM}(y; Y_i; 1/h^2),$$

where $1/h^2 > 0$ is concentration parameter and K_{vM} denotes the von Mises-Fisher kernel density.

Hyndman, R.J. (1996). Computing and graphing highest density regions. The American Statistician, 50, 120–126.

Saavedra-Nieves, P. and Crujeiras, R. M. (To appear). Nonparametric estimation of directional highest density regions. Advances in Data Analysis and Classification.

Plug-in estimation of directional HDRs

Plug-in methods propose

$$\hat{L}_g(\tau) = \{y \in S^{d-1} : g_n(y) \geq \hat{g}_\tau\}$$

as an **estimator for $L_g(\tau)$** where

$$g_n(y) = \frac{1}{n}\sum_{i=1}^{n} K_{vM}(y; Y_i; , 1/h^2),$$

where $1/h^2 > 0$ is concentration parameter and K_{vM} denotes the von Mises-Fisher kernel density.

Hyndman, R.J. (1996). Computing and graphing highest density regions. The American Statistician, 50, 120–126.

Saavedra-Nieves, P. and Crujeiras, R. M. (To appear). Nonparametric estimation of directional highest density regions. Advances in Data Analysis and Classification.

Plug-in estimation of directional HDRs

Plug-in methods propose

$$\hat{L}_g(\tau) = \{y \in S^{d-1} : g_n(y) \geq \hat{g}_\tau\}$$

as an **estimator for $L_g(\tau)$** where

$$g_n(y) = \frac{1}{n}\sum_{i=1}^{n} K_{vM}(y; Y_i; , 1/h^2),$$

where $1/h^2 > 0$ is concentration parameter and K_{vM} denotes the von Mises-Fisher kernel density.

Hyndman, R.J. (1996). Computing and graphing highest density regions. The American Statistician, 50, 120–126.

Saavedra-Nieves, P. and Crujeiras, R. M. (To appear). Nonparametric estimation of directional highest density regions. Advances in Data Analysis and Classification.

Circular examples

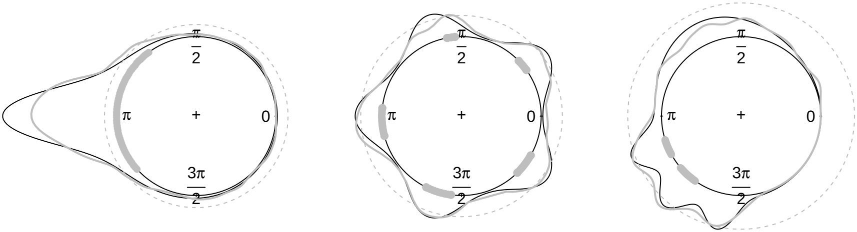


Figure: Plug-in HDRs from $\mathcal{Y}_{250}$ for three different circular densities with τ_1 (first column), τ_2 (second column) and τ_3 (third column) verifying $0 < \tau_1 < \tau_2 < \tau_3$.

Function `circ.plugin.hdr` in package `HDiR`.

Spherical examples

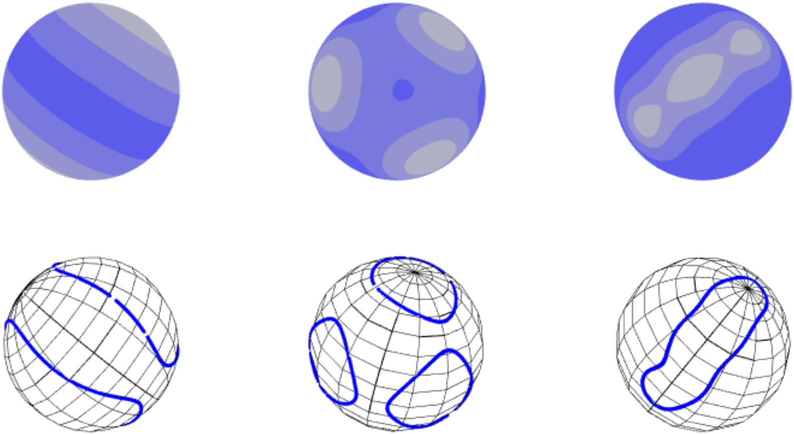


Figure: Plug-in HDRs from $\mathcal{Y}_{2500}$ for three different spherical densities with $\tau = 0.5$.

Function `sphere.plugin.hdr` in package `HDiR`.

Specific bandwidth selector for HDRs estimation

Since the closed expression of $d_H(\partial L_g(\tau), \partial \hat{L}_g(\tau))$ is not known, a new bandwidth selector is established as

$$h_1 = \arg\min_{h>0} \mathbb{E}_B \left[d_H(\partial L_g^*(\tau^*), \partial \hat{L}_g(\tau)) \right] \tag{1}$$

where $\mathbb{E}_B$ denotes the bootstrap expectation with respect to random samples $\mathcal{Y}_n^* = \{Y_1^*, \cdots, Y_n^*\}$ generated from the directional kernel g_n.

This error criterion quantifies the differences between HDRs and their reconstructions instead of measuring the accuracy of kernel estimators!

Performance of h_1 was analyzed through an extensive simulation study in Saavedra - Nieves and Crujeiras (2020).

In general, h_1 is a competitive selector!

Saavedra-Nieves, P. and Crujeiras, R. M. (To appear). Nonparametric estimation of directional highest density regions. Advances in Data Analysis and Classification.

Spherical bootstrap bandwidth for HDRs estimation

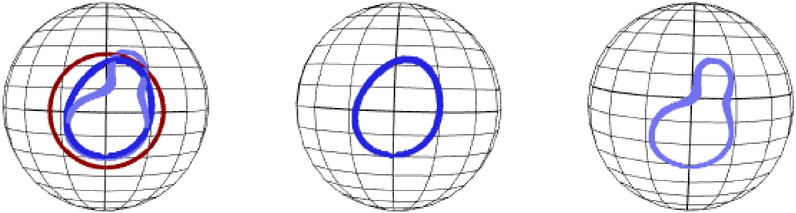


Figure: Theoretical (dark red colour) and estimated HDRs (left, bluish colours) from a random sample of size 500 when $\tau = 0.8$. Estimated HDRs from a random sample of size 500 using the specific bandwidth h_1 for spherical HDR reconstruction (center) and a cross-validation bandwidth (right) when $\tau = 0.8$.

Function `sphere.boot.bw` in package `HDiR`.

Highest density regions estimation with `HDiR` package

There are many libraries in the CRAN for directional data but none of them solves the problem of directional HDR reconstruction:

- **`circular`**: It provides functions for the statistical analysis (descriptive statistics, circular models, hypothesis tests), graphical representation and some classical circular datasets.
- **`Directional`**: Apart from hypothesis testing, discriminant and regression analysis, it allows to compute the kernel density estimation for hyper-spherical data using a von Mises-Fisher kernel.
- **`DirStats`**: This library computes a kernel density estimator and it also implements several bandwidth selectors.
- **`NPCirc`**: It is focused on nonparametric density and regression estimation methods for circular data. It also provides alternatives for choosing the smoothing parameter and a SiZer technique (CircSiZer) is developed for circular data.

Highest density regions estimation with `HDiR` package

Function	Description
`circ.boot.bw`	Circular bootstrap bandwidth for HDRs estimation
`circ.distances`	Euclidean and Hausdorff distances between two sets of points on the unit circle
`circ.hdr`	Computation of HDRs and level sets for a circular density
`circ.plugin.hdr`	Circular plug-in estimation of HDRs and level sets and confidence regions
`circ.scatterplot`	Circular scatterplot for plug-in HDRs
`dspheremix`	Densities for mixtures of spherical von Mises-Fisher
`rspheremix`	Random generation functions for mixtures of spherical von Mises-Fisher
`sphere.boot.bw`	Spherical bootstrap bandwidth for HDRs estimation
`sphere.distances`	Euclidean and Hausdorff distances between two sets of points on the unit sphere
`sphere.hdr`	Computation of HDRs for a spherical density
`sphere.plugin.hdr`	Spherical plug-in estimation of HDRs and level sets
`sphere.scatterplot`	Spherical scatterplot for plug-in HDRs

Table: Summary of `HDiR` package functions.

Highest density regions estimation with `HDiR` package

Function	Description
`circ.boot.bw`	Circular bootstrap bandwidth for HDRs estimation
`circ.distances`	Euclidean and Hausdorff distances between two sets of points on the unit circle
`circ.hdr`	Computation of HDRs and level sets for a circular density
`circ.plugin.hdr`	Circular plug-in estimation of HDRs and level sets and confidence regions
`circ.scatterplot`	Circular scatterplot for plug-in HDRs
`dspheremix`	Densities for mixtures of spherical von Mises-Fisher
`rspheremix`	Random generation functions for mixtures of spherical von Mises-Fisher
`sphere.boot.bw`	Spherical bootstrap bandwidth for HDRs estimation
`sphere.distances`	Euclidean and Hausdorff distances between two sets of points on the unit sphere
`sphere.hdr`	Computation of HDRs for a spherical density
`sphere.plugin.hdr`	Spherical plug-in estimation of HDRs and level sets
`sphere.scatterplot`	Spherical scatterplot for plug-in HDRs

Table: Summary of `HDiR` package functions.

Highest density regions estimation with HDiR package

Function	Description
`circ.boot.bw`	Circular bootstrap bandwidth for HDRs estimation
`circ.distances`	Euclidean and Hausdorff distances between two sets of points on the unit circle
`circ.hdr`	Computation of HDRs and level sets for a circular density
`circ.plugin.hdr`	Circular plug-in estimation of HDRs and level sets and confidence regions
`circ.scatterplot`	Circular scatterplot for plug-in HDRs
`dspheremix`	Densities for mixtures of spherical von Mises-Fisher
`rspheremix`	Random generation functions for mixtures of spherical von Mises-Fisher
`sphere.boot.bw`	Spherical bootstrap bandwidth for HDRs estimation
`sphere.distances`	Euclidean and Hausdorff distances between two sets of points on the unit sphere
`sphere.hdr`	Computation of HDRs for a spherical density
`sphere.plugin.hdr`	Spherical plug-in estimation of HDRs and level sets
`sphere.scatterplot`	Spherical scatterplot for plug-in HDRs

Table: Summary of HDiR package functions.

Highest density regions estimation with `HDiR` package

Function	Description
`circ.boot.bw`	Circular bootstrap bandwidth for HDRs estimation
`circ.distances`	Euclidean and Hausdorff distances between two sets of points on the unit circle
`circ.hdr`	Computation of HDRs and level sets for a circular density
`circ.plugin.hdr`	Circular plug-in estimation of HDRs and level sets and confidence regions
`circ.scatterplot`	Circular scatterplot for plug-in HDRs
`dspheremix`	Densities for mixtures of spherical von Mises-Fisher
`rspheremix`	Random generation functions for mixtures of spherical von Mises-Fisher
`sphere.boot.bw`	Spherical bootstrap bandwidth for HDRs estimation
`sphere.distances`	Euclidean and Hausdorff distances between two sets of points on the unit sphere
`sphere.hdr`	Computation of HDRs for a spherical density
`sphere.plugin.hdr`	Spherical plug-in estimation of HDRs and level sets
`sphere.scatterplot`	Spherical scatterplot for plug-in HDRs

Table: Summary of `HDiR` package functions.

Confidence regions for estimated circular HDRs

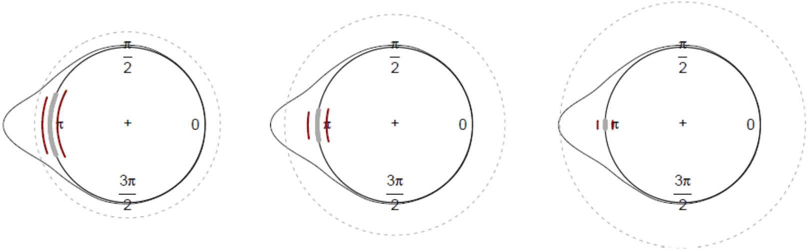


Figure: Confidence regions for an estimated circular HDR with τ_1 (first column), τ_2 (second column) and τ_3 (third column) verifying $\tau_1 < \tau_2 < \tau_3$.

Function `circ.plugin.hdr` in package `HDiR`.

Highest density regions estimation with `HDiR` package

Function	Description
`circ.boot.bw`	Circular bootstrap bandwidth for HDRs estimation
`circ.distances`	Euclidean and Hausdorff distances between two sets of points on the unit circle
`circ.hdr`	Computation of HDRs and level sets for a circular density
`circ.plugin.hdr`	Circular plug-in estimation of HDRs and level sets and confidence regions
`circ.scatterplot`	Circular scatterplot for plug-in HDRs
`dspheremix`	Densities for mixtures of spherical von Mises-Fisher
`rspheremix`	Random generation functions for mixtures of spherical von Mises-Fisher
`sphere.boot.bw`	Spherical bootstrap bandwidth for HDRs estimation
`sphere.distances`	Euclidean and Hausdorff distances between two sets of points on the unit sphere
`sphere.hdr`	Computation of HDRs for a spherical density
`sphere.plugin.hdr`	Spherical plug-in estimation of HDRs and level sets
`sphere.scatterplot`	Spherical scatterplot for plug-in HDRs

Table: Summary of `HDiR` package functions.

Spherical density models in `HDiR` package

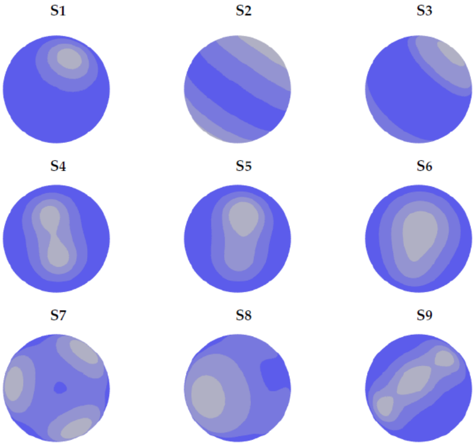


Figure: Finite mixtures of von Mises-Fisher spherical models. HDRs are represented for $\tau = 0.2$, $\tau = 0.5$ and $\tau = 0.8$.

Functions `dspheremix` and `rspheremix` in package `HDiR`.

Highest density regions estimation with HDiR package

Function	Description
`circ.boot.bw`	Circular bootstrap bandwidth for HDRs estimation
`circ.distances`	Euclidean and Hausdorff distances between two sets of points on the unit circle
`circ.hdr`	Computation of HDRs and level sets for a circular density
`circ.plugin.hdr`	Circular plug-in estimation of HDRs and level sets and confidence regions
`circ.scatterplot`	Circular scatterplot for plug-in HDRs
`dspheremix`	Densities for mixtures of spherical von Mises-Fisher
`rspheremix`	Random generation functions for mixtures of spherical von Mises-Fisher
`sphere.boot.bw`	Spherical bootstrap bandwidth for HDRs estimation
`sphere.distances`	Euclidean and Hausdorff distances between two sets of points on the unit sphere
`sphere.hdr`	Computation of HDRs for a spherical density
`sphere.plugin.hdr`	Spherical plug-in estimation of HDRs and level sets
`sphere.scatterplot`	Spherical scatterplot for plug-in HDRs

Table: Summary of HDiR package functions.

Hausdorff distance between HDRs boundaries

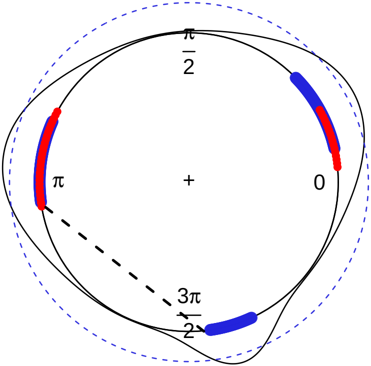


Figure: The black dotted line represents the Hausdorff distance between the boundary of the theoretical HDR (blue colour) and its plug-in estimator (red colour) for a circular model.

Function `circ.distances` in package `HDiR`.

Highest density regions estimation with `HDiR` package

Function	Description
`circ.boot.bw`	Circular bootstrap bandwidth for HDRs estimation
`circ.distances`	Euclidean and Hausdorff distances between two sets of points on the unit circle
`circ.hdr`	Computation of HDRs and level sets for a circular density
`circ.plugin.hdr`	Circular plug-in estimation of HDRs and level sets and confidence regions
`circ.scatterplot`	Circular scatterplot for plug-in HDRs
`dspheremix`	Densities for mixtures of spherical von Mises-Fisher
`rspheremix`	Random generation functions for mixtures of spherical von Mises-Fisher
`sphere.boot.bw`	Spherical bootstrap bandwidth for HDRs estimation
`sphere.distances`	Euclidean and Hausdorff distances between two sets of points on the unit sphere
`sphere.hdr`	Computation of HDRs for a spherical density
`sphere.plugin.hdr`	Spherical plug-in estimation of HDRs and level sets
`sphere.scatterplot`	Spherical scatterplot for plug-in HDRs

Table: Summary of `HDiR` package functions.

Circular and spherical scatterplots

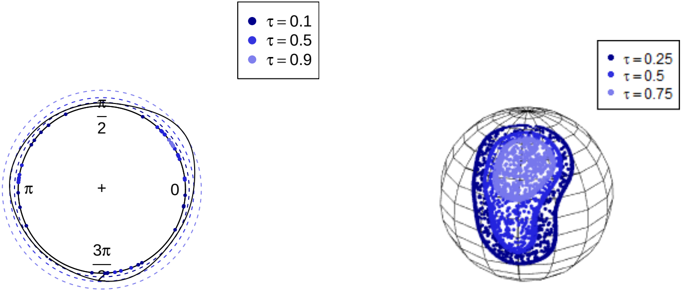


Figure: Circular and spherical scatterplots from $\mathcal{Y}_{50}$ and $\mathcal{Y}_{1500}$, respectively.

Functions `circ.scatterplot` and `sphere.scatterplot` in package `HDiR`.

Highest density regions estimation with `HDiR` package

Dataset	Description
`earthquakes`	Geographical coordinates (latitude and longitude) of earthquakes of magnitude greater than or equal to 2.5 degrees between October 2004 and April 2020
`sandhoppers`	Orientation of two sandhoppers species, Talitrus saltator and Talorchestia brito under different natural conditions

Table: Summary of `HDiR` datasets.

Analysis of sandhoppers orientation

Does the moment of the day play a significant role in sandhoppers behavior?

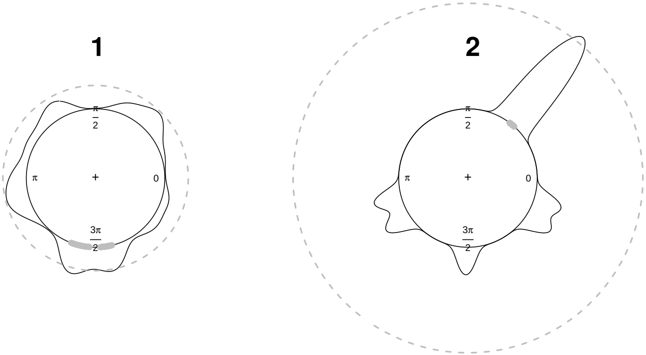
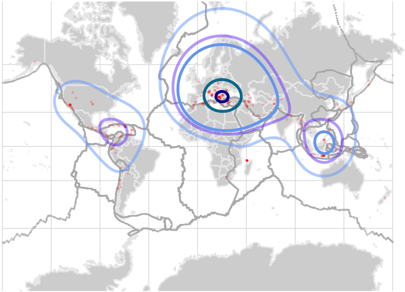


Figure: Plug-in HDRs when $\tau = 0.8$. The largest modes of these distributions can be observed.

Regions 1 and 2 correspond to the orientation for males of the specie Talorchestia Brito when the orientation is measure in noon and morning during April, respectively.

Analysis of earthquakes distribution

Play/Pause



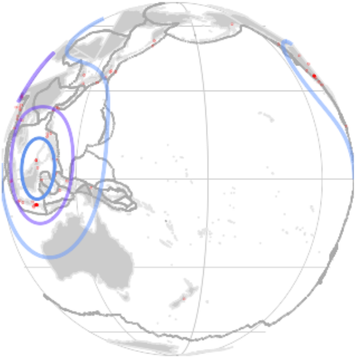


Figure: Distribution of earthquakes around the world between October 2004 and April 2020 (red color). Contours of plug-in HDRs for $\tau_1 = 0.1$, $\tau_2 = 0.3$, $\tau_3 = 0.5$, $\tau_4 = 0.7$ and $\tau_5 = 0.9$ (bluish colors).

Thanks!

HDiR: An R Package for Nonparametric Plug-in Estimation of Directional Highest Density Regions

Paula Saavedra-Nieves* and Rosa M. Crujeiras*

*Universidade de Santiago de Compostela

VIII Xornada de Usuarios de R en Galicia
14th October 2021, Santiago de Compostela