Halfspace Depth and Floating Body

Stanislav Nagy   Carsten Schütt   Elisabeth Werner

Mathematisches Seminar
CAU Kiel

Dept. Mathematics
CWRU

St. Petersburg, July 2019

Halfspace depth and floating body, STATISTICS SURVEYS *2019, Vol. 13, 52-118.*

Halfspace depth and floating body, STATISTICS SURVEYS *2019, Vol. 13, 52-118.*

Floating body is a concept from convex geometry/ differential geometry.

Halfspace depth and floating body, STATISTICS SURVEYS *2019, Vol. 13, 52-118.*

Floating body is a concept from convex geometry/ differential geometry.

Halfspace depth or Tukey depth is a concept from nonparametric statistics.

Halfspace depth and floating body, STATISTICS SURVEYS *2019, Vol. 13, 52-118.*

Floating body is a concept from convex geometry/ differential geometry.

Halfspace depth or Tukey depth is a concept from nonparametric statistics.

They are the same.

Notions and Definitions

Notions and Definitions

- A convex body K in $\mathbb{R}^n$ is a convex, compact subset with nonempty interior.

Notions and Definitions

- A convex body K in $\mathbb{R}^n$ is a convex, compact subset with nonempty interior.
- H denotes a hyperplane in $\mathbb{R}^n$ and H^- and H^+ its halfspaces.

Notions and Definitions

- A convex body K in $\mathbb{R}^n$ is a convex, compact subset with nonempty interior.
- H denotes a hyperplane in $\mathbb{R}^n$ and H^- and H^+ its halfspaces.
- The volume of a convex body K is denoted by $\text{vol}_n(K)$. It is the Lebesgue measure of K.

Floating body

Definition

Let K be a convex body in $\mathbb{R}^n$ and $\delta > 0$.

Floating body

Definition

Let K be a convex body in $\mathbb{R}^n$ and $\delta > 0$. A convex body $K_{[\delta]}$ is called a floating body of K if it is a subset of K such that every support hyperplane of $K_{[\delta]}$ cuts off a set of volume δ from K.

Floating body

Definition

Let K be a convex body in $\mathbb{R}^n$ and $\delta > 0$. A convex body $K_{[\delta]}$ is called a floating body of K if it is a subset of K such that every support hyperplane of $K_{[\delta]}$ cuts off a set of volume δ from K.

Let K_δ be the intersection of all halfspaces whose defining hyperplanes cut off a set of volume δ from K. i.e.

$$K_\delta = \bigcap_{\mathrm{vol}_n(K \cap H^-) = \delta} H^+$$

K_δ is called the convex floating body of K for index δ.

Floating body

In general, the floating body may not exist. The triangle in $\mathbb{R}^2$ does not have floating bodies for $0 < \delta$.

Floating body

In general, the floating body may not exist. The triangle in $\mathbb{R}^2$ does not have floating bodies for $0 < \delta$.

Theorem (Meyer-Reisner)

Every centrally symmetric convex body has floating bodies for all δ.

Applications of floating body

Affine surface area

Applications of floating body

Affine surface area

$$\lim_{\delta \to 0} \frac{\mathrm{vol}_n(K) - \mathrm{vol}_n(K_\delta)}{\delta^{\frac{2}{n+1}}} = c_n \int_{\partial K} \kappa^{\frac{1}{n+1}} d\mu_{\partial K}$$

Applications of floating body

Affine surface area

$$\lim_{\delta \to 0} \frac{\mathrm{vol}_n(K) - \mathrm{vol}_n(K_\delta)}{\delta^{\frac{2}{n+1}}} = c_n \int_{\partial K} \kappa^{\frac{1}{n+1}} d\mu_{\partial K}$$

$$\mathrm{as}(K) = \int_{\partial K} \kappa^{\frac{1}{n+1}} d\mu_{\partial K}$$

Applications of floating body

Affine surface area

$$\lim_{\delta \to 0} \frac{\mathrm{vol}_n(K) - \mathrm{vol}_n(K_\delta)}{\delta^{\frac{2}{n+1}}} = c_n \int_{\partial K} \kappa^{\frac{1}{n+1}} d\mu_{\partial K}$$

$$\mathrm{as}(K) = \int_{\partial K} \kappa^{\frac{1}{n+1}} d\mu_{\partial K}$$

Affine isoperimetric inequality

Applications of floating body

Affine surface area

$$\lim_{\delta \to 0} \frac{\text{vol}_n(K) - \text{vol}_n(K_\delta)}{\delta^{\frac{2}{n+1}}} = c_n \int_{\partial K} \kappa^{\frac{1}{n+1}} d\mu_{\partial K}$$

$$\text{as}(K) = \int_{\partial K} \kappa^{\frac{1}{n+1}} d\mu_{\partial K}$$

Affine isoperimetric inequality

Blaschke-Santaló inequality

Halfspace depth

Order statistics are used to analyze univariate data/random variables.

Halfspace depth

Order statistics are used to analyze univariate data/random variables.

For multivariate data/ random variables there is no natural ordering.

Halfspace depth

Definition (Tukey, Donoho, Gasko)

Let $\mathbb{P}$ be a probability measure on $\mathbb{R}^n$. The halfspace depth or Tukey depth hD of $\mathbb{P}$ at x is

$$hD(x,\mathbb{P}) = \inf\{\mathbb{P}(H^-)|H \text{ is a hyperplane with } x \in H\}.$$

Halfspace depth

Definition (Tukey, Donoho, Gasko)

Let $\mathbb{P}$ be a probability measure on $\mathbb{R}^n$. The halfspace depth or Tukey depth hD of $\mathbb{P}$ at x is

$$hD(x,\mathbb{P}) = \inf\{\mathbb{P}(H^-) | H \text{ is a hyperplane with } x \in H\}.$$

The central region of $\mathbb{P}$ for index $\delta > 0$ is

$$\mathbb{P}_\delta = \{y \in \mathbb{R}^n | \; hD(y,\mathbb{P}) \geq \delta\}.$$

Halfspace depth

DEFINITION (TUKEY, DONOHO, GASKO)

Let $\mathbb{P}$ be a probability measure on $\mathbb{R}^n$. The halfspace depth or Tukey depth hD of $\mathbb{P}$ at x is

$$hD(x, \mathbb{P}) = \inf\{\mathbb{P}(H^-) | H \text{ is a hyperplane with } x \in H\}.$$

The central region of $\mathbb{P}$ for index $\delta > 0$ is

$$\mathbb{P}_\delta = \{y \in \mathbb{R}^n | \; hD(y, \mathbb{P}) \geq \delta\}.$$

A halfspace median $x_{\mathbb{P}}$ of $\mathbb{P}$ is a point with the maximal halfspace depth.

Halfspace depth

Definition (Tukey, Donoho, Gasko)

Let $\mathbb{P}$ be a probability measure on $\mathbb{R}^n$. The halfspace depth or Tukey depth hD of $\mathbb{P}$ at x is

$$hD(x,\mathbb{P}) = \inf\{\mathbb{P}(H^-) | H \text{ is a hyperplane with } x \in H\}.$$

The central region of $\mathbb{P}$ for index $\delta > 0$ is

$$\mathbb{P}_\delta = \{y \in \mathbb{R}^n | \ hD(y,\mathbb{P}) \geq \delta\}.$$

A halfspace median $x_{\mathbb{P}}$ of $\mathbb{P}$ is a point with the maximal halfspace depth.

Zuo, Serfling, Dyckerhoff, Mosler, Rousseeuw, Ruts

Halfspace depth

A halfspace median does not need to be unique.

Halfspace depth

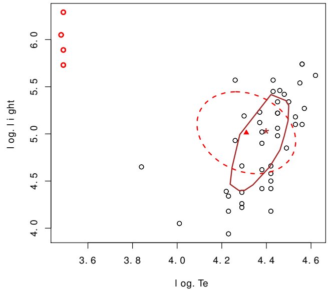
A halfspace median does not need to be unique.

Let $\mathbb{P}$ be the uniform measure on the vertices of a simplex in $\mathbb{R}^n$.

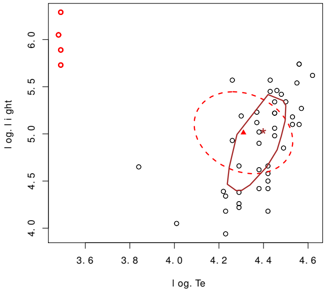
Halfspace depth

A halfspace median does not need to be unique.

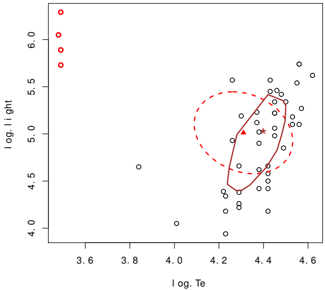
Let $\mathbb{P}$ be the uniform measure on the vertices of a simplex in $\mathbb{R}^n$.
Then all elements of the simplex are halfspace medians of $\mathbb{P}$.



The Hertzsprung-Russell diagram of the Star Cluster CYG OB1.



The Hertzsprung-Russell diagram of the Star Cluster CYG OB1. The four giant stars (red points) attract both the sample mean (red triangle), and the sample covariance (represented by the red ellipse).



The Hertzsprung-Russell diagram of the Star Cluster CYG OB1. The four giant stars (red points) attract both the sample mean (red triangle), and the sample covariance (represented by the red ellipse). The halfspace depth-based median (brown star), and the depth-based central region containing 25 % of the observations, provide a more appropriate representation of the location and the variability of the main data cloud.

Halfspace depth

Definition

We say that a measure $\mathbb{P}$ on $\mathbb{R}^n$ is smooth if for all hyperplanes H

$$\mathbb{P}(H) = 0.$$

Halfspace depth

Definition

We say that a measure $\mathbb{P}$ on $\mathbb{R}^n$ is smooth if for all hyperplanes H

$$\mathbb{P}(H) = 0.$$

Definition

We say that a measure $\mathbb{P}$ on $\mathbb{R}^n$ has contiguous support if there are no disjoint halfspaces H_1^- and H_2^- such that

Halfspace depth

Definition

We say that a measure $\mathbb{P}$ on $\mathbb{R}^n$ is smooth if for all hyperplanes H

$$\mathbb{P}(H) = 0.$$

Definition

We say that a measure $\mathbb{P}$ on $\mathbb{R}^n$ has contiguous support if there are no disjoint halfspaces H_1^- and H_2^- such that

$$\mathbb{P}(H_1^-) > 0 \qquad \mathbb{P}(H_2^-) > 0$$

Halfspace depth

Definition

We say that a measure $\mathbb{P}$ on $\mathbb{R}^n$ is smooth if for all hyperplanes H

$$\mathbb{P}(H) = 0.$$

Definition

We say that a measure $\mathbb{P}$ on $\mathbb{R}^n$ has contiguous support if there are no disjoint halfspaces H_1^- and H_2^- such that

$$\mathbb{P}(H_1^-) > 0 \qquad \mathbb{P}(H_2^-) > 0$$

and

$$\mathbb{P}(H_1^-) + \mathbb{P}(H_2^-) = 1.$$

Lemma

If $\mathbb{P}$ has contiguous support then the halfspace median is unique.

Example

Let $0 < \alpha \leq 2$ and $\phi : \mathbb{R} \to \mathbb{R}$ a continuous function. Let X be a random vector such that for all $t \in \mathbb{R}^n$

$$\mathbb{E}e^{i\langle t,X\rangle} = \phi(\|t\|_\alpha).$$

Example

Let $0 < \alpha \leq 2$ and $\phi : \mathbb{R} \to \mathbb{R}$ a continuous function. Let X be a random vector such that for all $t \in \mathbb{R}^n$

$$\mathbb{E}e^{i\langle t,X\rangle} = \phi(\|t\|_\alpha).$$

Then we say that X is α-symmetric. $\mathbb{P}_X$ is the joint distribution. In this case we can compute the central regions/ floating bodies of $\mathbb{P}_X$.

Let $X = (X_1, \ldots, X_n)$ be a α-symmetric random vector on a probability space $(\Omega, \mathbb{P})$. Then

Let $X=(X_1,\dots,X_n)$ be a α-symmetric random vector on a probability space $(\Omega,\mathbb{P})$. Then

$$\begin{aligned} hD(x,\mathbb{P}_X) &= \inf\{\mathbb{P}_X(H^-)|H \text{ is a hyperplane with } x\in H\} \\ &= \inf_{t\neq 0}\mathbb{P}(\langle t,X\rangle\leq\langle t,x\rangle)=\inf_{t\neq 0}\mathbb{P}(\|t\|_\alpha X_1\leq\langle t,x\rangle) \end{aligned}$$

Let $X = (X_1, \ldots, X_n)$ be a α-symmetric random vector on a probability space $(\Omega, \mathbb{P})$. Then

$$\begin{aligned} hD(x, \mathbb{P}_X) &= \inf\{\mathbb{P}_X(H^-) | H \text{ is a hyperplane with } x \in H\} \\ &= \inf_{t \neq 0} \mathbb{P}(\langle t, X \rangle \leq \langle t, x \rangle) = \inf_{t \neq 0} \mathbb{P}(\|t\|_\alpha X_1 \leq \langle t, x \rangle) \end{aligned}$$

We have

$$\inf_{t \neq 0} \frac{\langle t, x \rangle}{\|t\|_\alpha} = -\|x\|_{\alpha^*}$$

Let $X = (X_1, \dots, X_n)$ be a α-symmetric random vector on a probability space $(\Omega, \mathbb{P})$. Then

$$\begin{aligned} hD(x, \mathbb{P}_X) &= \inf\{\mathbb{P}_X(H^-) | H \text{ is a hyperplane with } x \in H\} \\ &= \inf_{t \neq 0} \mathbb{P}(\langle t, X \rangle \leq \langle t, x \rangle) = \inf_{t \neq 0} \mathbb{P}(\|t\|_\alpha X_1 \leq \langle t, x \rangle) \end{aligned}$$

We have

$$\inf_{t \neq 0} \frac{\langle t, x \rangle}{\|t\|_\alpha} = -\|x\|_{\alpha^*}$$

where

$$\alpha^* = \begin{cases} \frac{\alpha}{\alpha - 1} & \alpha > 1 \\ \infty & 0 < \alpha \leq 1 \end{cases}$$

Let $X=(X_1,\dots,X_n)$ be a α-symmetric random vector on a probability space $(\Omega,\mathbb{P})$. Then

$$\begin{aligned} hD(x,\mathbb{P}_X) &= \inf\{\mathbb{P}_X(H^-) | H \text{ is a hyperplane with } x\in H\} \\ &= \inf_{t\neq 0} \mathbb{P}(\langle t,X\rangle \leq \langle t,x\rangle) = \inf_{t\neq 0} \mathbb{P}(\|t\|_\alpha X_1 \leq \langle t,x\rangle) \end{aligned}$$

We have

$$\inf_{t\neq 0} \frac{\langle t,x\rangle}{\|t\|_\alpha} = -\|x\|_{\alpha^*}$$

where

$$\alpha^* = \begin{cases} \frac{\alpha}{\alpha-1} & \alpha>1 \\ \infty & 0<\alpha\leq 1 \end{cases}$$

Therefore

$$hD(x,\mathbb{P}_X) = F_{X_1}(-\|x\|_{\alpha*})$$

where F_{x_1} is the distribution function of X_1.

Definition

A nonempty convex set $\mathbb{P}_{[\delta]}$ is the floating body of $\mathbb{P}$ for index δ if for each supporting hyperplane H of $\mathbb{P}_{[\delta]}$ with $H^+ \supseteq \mathbb{P}_{[\delta]}$ we have $\mathbb{P}(H^-) = \delta$.

Definition

A nonempty convex set $\mathbb{P}_{[\delta]}$ is the floating body of $\mathbb{P}$ for index δ if for each supporting hyperplane H of $\mathbb{P}_{[\delta]}$ with $H^+ \supseteq \mathbb{P}_{[\delta]}$ we have $\mathbb{P}(H^-) = \delta$.

Definition

The convex floating body of the probability measure $\mathbb{P}$ on $\mathbb{R}^n$ with index δ is the intersection of all closed halfspaces whose defining hyperplanes cut off a set of probability content at most δ from $\mathbb{P}$, i.e.

$$\mathbb{P}_\delta^{FB} = \bigcap_{\mathbb{P}(H^-)\leq\delta} H^+.$$

Proposition

Assume that $\mathbb{P}_{[\delta]}$ exists. Then

$$\mathbb{P}_{\delta}^{FB} = \mathbb{P}_{[\delta]} = \mathbb{P}_{\delta}.$$

Proposition

Assume that $\mathbb{P}_{[\delta]}$ exists. Then

$$\mathbb{P}_{\delta}^{FB} = \mathbb{P}_{[\delta]} = \mathbb{P}_{\delta}.$$

Example

Let K be a convex body in $\mathbb{R}^n$ and $\mathbb{P}_K$ the uniform measure on K. Then

$$K_\delta = (\mathbb{P}_K)_\delta.$$

Question

Do the central regions/ floating bodies of a measure determine the measure?

Question

Do the central regions/ floating bodies of a measure determine the measure?

Partial results: Struyf, Rousseeuw, Koshevoy, Hassairi, Regaieg, Kong, Zuo

No! Counterexample by S. Nagy.

No! Counterexample by S. Nagy.
Consider the α-symmetric random vector $X = (X_1, \ldots, X_n)$ with $\alpha = 1$ and $\phi(\theta) = \sqrt{\theta}$

$$\psi_X(t) = \exp\left(-\|t\|_1^{\frac{1}{2}}\right) = \exp\left(-\sqrt{\sum_{j=1}^{n} |t_j|}\right)$$

No! Counterexample by S. Nagy.
Consider the α-symmetric random vector $X = (X_1, \ldots, X_n)$ with $\alpha = 1$ and $\phi(\theta) = \sqrt{\theta}$

$$\psi_X(t) = \exp\left(-\|t\|_1^{\frac{1}{2}}\right) = \exp\left(-\sqrt{\sum_{j=1}^{n} |t_j|}\right)$$

and $Y = (Y_1, \ldots, Y_n)$ with $\alpha = \frac{1}{2}$ and $\phi(\theta) = \sqrt{\theta}$

$$\psi_Y(t) = \exp\left(-\|t\|_{\frac{1}{2}}^{\frac{1}{2}}\right) = \exp\left(-\sum_{j=1}^{n} \sqrt{|t_j|}\right)$$

No! Counterexample by S. Nagy.
Consider the α-symmetric random vector $X = (X_1, \dots, X_n)$ with $\alpha = 1$ and $\phi(\theta) = \sqrt{\theta}$

$$\psi_X(t) = \exp\left(-\|t\|_1^{\frac{1}{2}}\right) = \exp\left(-\sqrt{\sum_{j=1}^{n} |t_j|}\right)$$

and $Y = (Y_1, \dots, Y_n)$ with $\alpha = \frac{1}{2}$ and $\phi(\theta) = \sqrt{\theta}$

$$\psi_Y(t) = \exp\left(-\|t\|_{\frac{1}{2}}^{\frac{1}{2}}\right) = \exp\left(-\sum_{j=1}^{n} \sqrt{|t_j|}\right)$$

Clearly, $\mathbb{P}_X$ and $\mathbb{P}_Y$ are different.

For every $x \in \mathbb{R}^n$ the halfspace depth are the same

$$hD(x, \mathbb{P}_X) = F_{X_1}(-\|x\|_\infty) \qquad\qquad hD(x, \mathbb{P}_Y) = F_{Y_1}(-\|x\|_\infty)$$

where F_{X_1} is the distribution function of X_1 and F_{Y_1} is the distribution function of Y_1.

For every $x \in \mathbb{R}^n$ the halfspace depth are the same

$$hD(x, \mathbb{P}_X) = F_{X_1}(-\|x\|_\infty) \qquad hD(x, \mathbb{P}_Y) = F_{Y_1}(-\|x\|_\infty)$$

where F_{X_1} is the distribution function of X_1 and F_{Y_1} is the distribution function of Y_1. But $F_{X_1} = F_{Y_1}$

$$\psi_{X_1}(t_1) = \exp(-\sqrt{|t_1|}) = \psi_{Y_1}(t_1)$$

Theorem

Let $\mathbb{P}$ have contiguous support and let $x_{\mathbb{P}}$ be the halfspace median of $\mathbb{P}$. Then the following are equivalent:

Theorem

Let $\mathbb{P}$ have contiguous support and let $x_{\mathbb{P}}$ be the halfspace median of $\mathbb{P}$. Then the following are equivalent:

(i) For each $\delta \in (0, \frac{1}{2})$ the floating body $\mathbb{P}_{[\delta]}$ of $\mathbb{P}$ exists.

Theorem

Let $\mathbb{P}$ have contiguous support and let $x_{\mathbb{P}}$ be the halfspace median of $\mathbb{P}$. Then the following are equivalent:

(i) For each $\delta \in (0, \frac{1}{2})$ the floating body $\mathbb{P}_{[\delta]}$ of $\mathbb{P}$ exists.

(ii) For all hyperplanes H we have $\mathbb{P}(H) = 0$ and

$$\mathbb{P}(H^-) = \begin{cases} \sup_{x \in H} hD(x, \mathbb{P}) & x_{\mathbb{P}} \notin H^- \\ 1 - \sup_{x \in H} hD(x, \mathbb{P}) & x_{\mathbb{P}} \in H^- \end{cases}$$

Theorem

Let $\mathbb{P}$ have contiguous support and let $x_{\mathbb{P}}$ be the halfspace median of $\mathbb{P}$. Then the following are equivalent:

(i) For each $\delta \in (0, \frac{1}{2})$ the floating body $\mathbb{P}_{[\delta]}$ of $\mathbb{P}$ exists.

(ii) For all hyperplanes H we have $\mathbb{P}(H) = 0$ and

$$\mathbb{P}(H^-) = \begin{cases} \sup_{x \in H} hD(x, \mathbb{P}) & x_{\mathbb{P}} \notin H^- \\ 1 - \sup_{x \in H} hD(x, \mathbb{P}) & x_{\mathbb{P}} \in H^- \end{cases}$$

Consequently, if (i) is true then $\mathbb{P}$ is characterized by its halfspace depth.

Theorem

Suppose that for all hyperplanes H we have $\mathbb{P}(H) = 0$ and that $\mathbb{P}$ has contiguous support. Then the map $\delta \to \mathbb{P}_\delta$ is continuous w.r.t. the Hausdorff distance.

Theorem

Suppose that for all hyperplanes H we have $\mathbb{P}(H) = 0$ and that $\mathbb{P}$ has contiguous support. Then the map $\delta \to \mathbb{P}_\delta$ is continuous w.r.t. the Hausdorff distance.

Question

For which $\mathbb{P}$ do the central regions/ floating bodies have a C^1- or C^2-boundary?