Set Mean Estimation and Confidence Supersets using Oriented Distance Functions

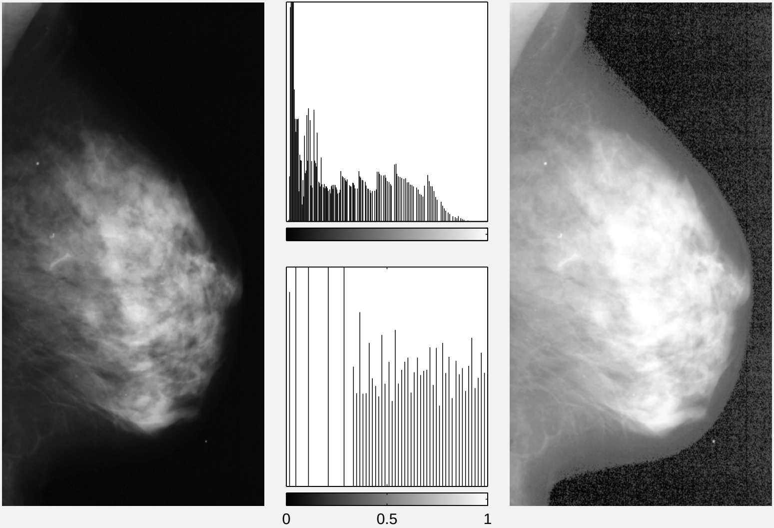
Hanna Jankowski

York University

joint work with Larissa Stanberry

Seattle Children's Research Institute

MOTIVATION



OUTLINE

1. distance functions and oriented distance functions (ODFs)
2. random closed sets (RCSs) and their expectation
   - selection expectation
   - Baddeley & Molchanov definition
   - ODF definition
3. properties of new definition
4. confidence regions/supersets
5. examples
   - sand grains
   - boundary reconstruction in a mammogram image

ORIENTED DISTANCE FUNCTION (ODF)

Fix $D \subset \mathbb{R}^d$, and let $d(x, y) = |x - y|$ denote Euclidean distance.

The distance function of $A \subset D$ such that $A \neq \emptyset$ is

$$d_A(x) = \inf_{y \in A} d(x, y) \quad \text{for } x \in D.$$

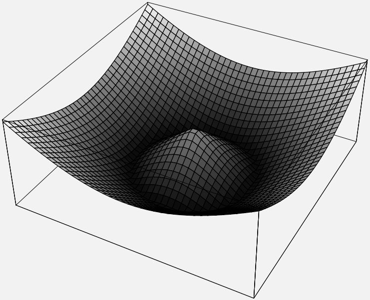
Note that $d_A(x) = d_C(x)$ iff $\overline{A} = \overline{C}$, and $A = \{x : d_A(x) = 0\}$.

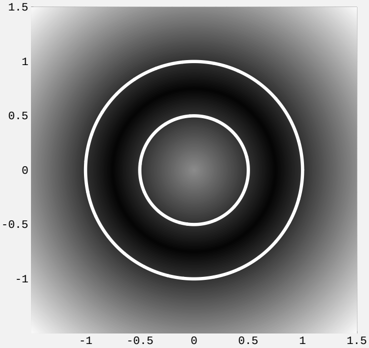
The oriented distance function of $A \subset D$ such that $\partial A \neq \emptyset$ is

$$b_A(x) = d_A(x) - d_{A^c}(x).$$

Here, $A = \{x : b_A(x) \leq 0\}$ and $\partial A = \{x : b_A(x) = 0\}$.

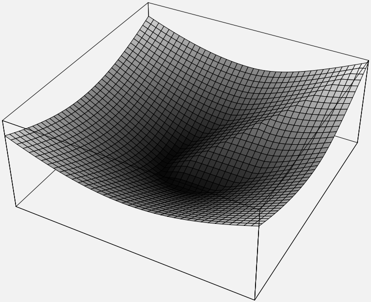
ODF OF A DONUT

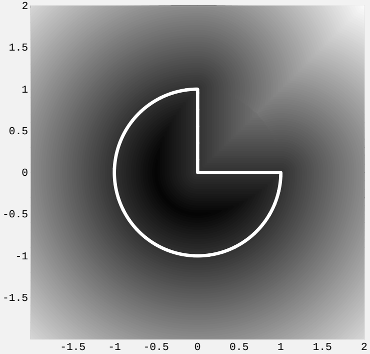




$$\text{donut} = \{x \in \mathbb{R}^2 : 0.5 \leq |x| \leq 1\}$$

ODF OF PACMAN





RANDOM CLOSED SET (RCS)

Let $\mathcal{F}$ be the family of closed subsets of $\mathbb{R}^d$ and $\mathcal{K}$ be the family of all compact subsets of $\mathbb{R}^d$. Consider a probability space given by the triple $(\Omega, \mathcal{A}, P)$.

Definition

A random closed set is the mapping $A : \Omega \mapsto \mathcal{F}$, such that for every compact set K

$$\{\omega : A(\omega) \cap K \neq \emptyset\} \in \mathcal{A}.$$

- Foundations laid by Choquet (1950s), Matheron (1975)
- Modern review by Molchanov (2005)
- As $\mathcal{F}$ is nonlinear, there is no natural way to define the expectation of a set.

SET EXPECTATION

Examples: Let $\mathbf{A} = \{x : |x| \leq \Theta\}$ and $\mathbf{B} = \{\xi\}$.

1. selection (Aumann) expectation
   - most studied
   - depends on structure of $(\Omega, \mathcal{A}, P)$
   - gives convex answer
   - $\mathrm{E}_A[\mathbf{A}] = \{x : |x| \leq \mathrm{E}[\Theta]\}$ and $\mathrm{E}_A[\mathbf{B}] = \{\mathrm{E}[\xi]\}$
2. Vorobe'ev expectation
   - most intuitive in terms of image analysis
   - $\mathrm{E}_V[\mathbf{A}] = \{x : |x| \leq \sqrt{\mathrm{E}[\Theta^2]}\}$ and $\mathrm{E}_V[\mathbf{B}] = \emptyset$
3. Baddeley & Molchanov definition
   - depends on significant user input (choice of two metrics)
   - complicated to calculate
   - $\mathrm{E}_{BM}[\mathbf{A}] = \{x : |x| \leq \mathrm{E}[\Theta]\}$ and $\mathrm{E}_{BM}[\mathbf{B}] = \{\mathrm{E}[\xi]\}$

OUR DEFINITION

Definition

Suppose that A is a random closed set such that $\partial A \neq \emptyset$ a.s. and $E|b_A(x_0)| < \infty$ for some $x_0 \in D$, then

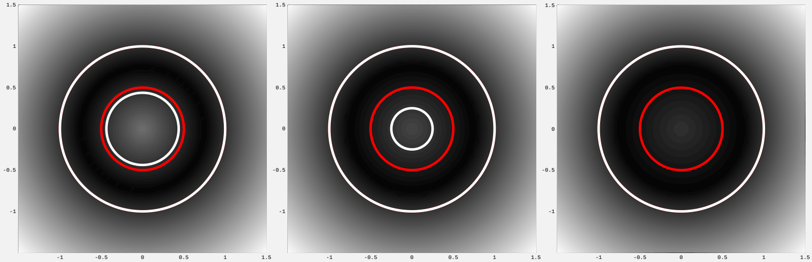
$$\mathrm{E}[\mathbf{A}] = \{x : \mathrm{E}[b_{\mathbf{A}}(x)] \leq 0\}$$
$$\Gamma[\mathbf{A}] = \{x : \mathrm{E}[b_{\mathbf{A}}(x)] = 0\}$$

- simple and intuitive, no user input
- algorithms for distance functions and level sets easily available (eg. MATLAB)
- includes definition for boundary
- $\mathrm{E}[\mathbf{A}] = \{x : |x| \leq \mathrm{E}[\Theta]\}$ and $\mathrm{E}[\mathbf{B}] = \emptyset$

EXAMPLE: MISSING TIMBIT

The RCS **A** is equal to

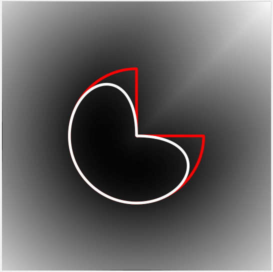
$$\begin{array}{lll} \text{a circle:} & \{x : |x| \leq 1\} & \text{with probability } 1 - p \\ \text{a donut:} & \{x : 0.5 \leq |x| \leq 1\} & \text{with probability } p \end{array}$$



$p = 0.9$ $p = 0.75$ $p < 2/3$

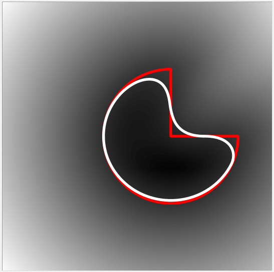
$\Gamma[\mathbf{A}]$ is shown in white.

EXAMPLE: PACMAN



- (white) mean pacman with uniform radius
- (red) pacman with mean radius

- (white) mean pacman with uniform NE shift
- (red) pacman with mean NE shift



- (white) mean pacman with uniform E shift
- (red) pacman with mean E shift

Properties

- some basics:
  - $\mathrm{E}[\mathbf{A}]$ is closed
  - $\partial\,\mathrm{E}[\mathbf{A}] \subset \Gamma[\mathbf{A}]$
  - if $\mathbf{A} \subset \mathbf{B}$ a.s. then $\mathrm{E}[\mathbf{A}] \subset \mathrm{E}[\mathbf{B}]$
  - if $\mathbf{A} = A$ a.s. then $\mathrm{E}[\mathbf{A}] = A$ and $\Gamma[\mathbf{A}] = \partial A$
  - if $\partial\mathbf{A} = B$ a.s. then $B \subset \Gamma[\mathbf{A}]$
- preservation of shape:
  - (translation) $\mathrm{E}[a + \mathbf{A}] = a + \mathrm{E}[\mathbf{A}]$
  - (dilation) $\mathrm{E}[\alpha\mathbf{A}] = \alpha\,\mathrm{E}[\mathbf{A}]$ for $\alpha \neq 0$
  - (equivariant w.r.t. orthogonal transformations) $\mathrm{E}[g\mathbf{A}] = g\,\mathrm{E}[\mathbf{A}]$, for $g(x) = \Lambda x + a$ and $\Lambda \in O(d)$
  - If $\mathbf{A}$ is convex a.s. then $\mathrm{E}[\mathbf{A}]$ is convex.
- preservation of smoothness:
  - If $\mathrm{E}[b_{\mathbf{A}}(x)]$ is smooth then $\Gamma[\mathbf{A}]$ is smooth.

ESTIMATION

Suppose that we observe the random sets $A_1, \ldots, A_n$ under IID sampling. Define $\bar{b}_n(x) = \sum_{i=1}^n b_{A_i}(x)/n$, and

$$\overline{A}_n = \{x : \bar{b}_n(x) \leq 0\} \quad \text{and} \quad \overline{\Gamma}_n = \{x : \bar{b}_n(x) = 0\}$$

Theorem

Suppose that $\mathrm{E}[\mathbf{A}]$ *satisfies* $\partial\,\mathrm{E}[\mathbf{A}^c] = \Gamma[\mathbf{A}]$ *then*

$$\overline{A}_n \to \mathrm{E}[\mathbf{A}] \quad a.s..$$

If in addition we have that $\partial\,\mathrm{E}[\mathbf{A}] = \Gamma[\mathbf{A}]$ *then*

$$\overline{\Gamma}_n \to \Gamma[\mathbf{A}] \quad a.s..$$

CONFIDENCE REGIONS/SUPERSETS

- Let $\mathbb{Z}_n(x) = \sqrt{n}\left(\bar{b}_n(x) - \mathrm{E}[b_{\mathbf{A}}(x)]\right)$ and assume that $\mathrm{E}[b_{\mathbf{A}}(x_0)^2] < \infty$ for some $x_0 \in D$ (compact).
- Then $\mathbb{Z}_n \Rightarrow \mathbb{Z}$, where $\mathbb{Z}$ is a smooth Gaussian field.
- Let q_1 and q_2 denote numbers such that $P(\sup_{x\in D} \mathbb{Z}(x) \leq q_1) = 0.95$ and $P(\sup_{x\in D} |\mathbb{Z}(x)| \leq q_2) = 0.95$.

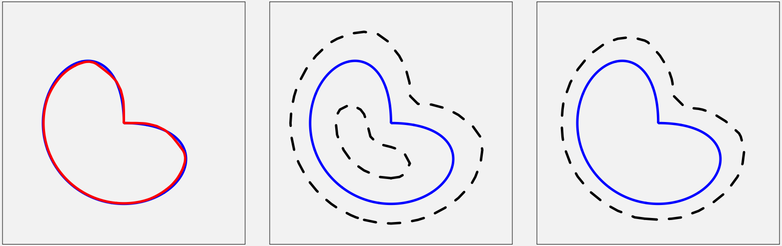
Definition

$$\left\{x : \bar{b}_m(x) \leq \frac{1}{\sqrt{m}}\, q_1\right\} \quad \text{and} \quad \left\{x : |\bar{b}_m(x)| \leq \frac{1}{\sqrt{m}}\, q_2\right\}$$

are 95% confidence regions for $\mathrm{E}[\mathbf{A}]$ and $\Gamma[\mathbf{A}]$.

CONFIDENCE REGIONS: PACMAN

Suppose the random model is pacman of random radius Θ where $\Theta \sim$Uniform$[0, 1]$. We observe 25 IID sets from this model.



LEFT: Estimated set (red) and the expected set (blue).

MID: Expected set boundary with 95% bootstrapped confidence interval.

RIGHT: Expected set with 95% bootstrapped confidence interval.

CONFIDENCE REGIONS: PROPERTIES

- Easy, visual way of describing variability around the mean
- CRs are conservative with a probability of at least 95% of capturing the expected set
- Quantiles are often intractable, but can be easily estimated via bootstrapping
- Immune to consistency conditions
- Allows for local changes in variability

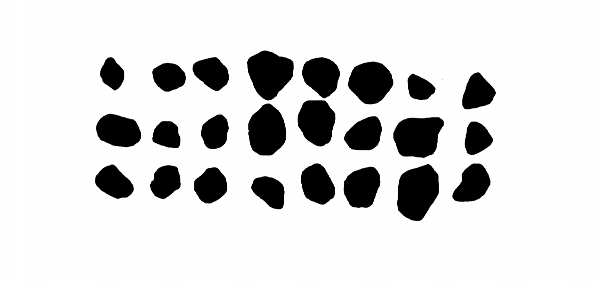
- EQUIVARIANCE PROPERTIES: let **C** denote the confidence region for E[**A**]. Then
  1. The confidence region for E[α**A**] is α**C**, for $\alpha \neq 0$.
  2. The confidence region for E[g**A**] is g**C**, where g is a rigid motion.

EMPIRICAL COVERAGE PROBABILITIES

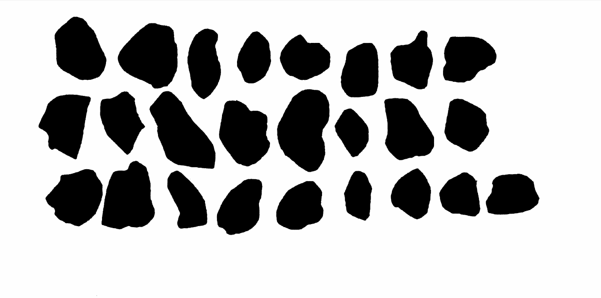
	$n = 25$		$n = 100$	
$100(1-\alpha)\%$	90%	95%	90%	95%
(A)	88.4/89.7	94.8/95.6	90.4/91.2	95.7/94.1
(B)	90.2/89.9	94.6/95.1	90.2/90.4	95.1/95.0
(C1)	90.1/91.2	94.4/95.1	91.6/91.1	95.3/95.3
(C2)	91.4/93.5	96.5/97.4	92.1/93.3	96.9/97.1
(D1)	92.0/91.0	96.3/95.7	91.8/91.7	96.2/95.7
(D2)	90.7/88.6	94.5/94.7	91.0/88.9	94.9/95.0

(A) $\mathbf{A} = [0,1]$ or $\{0,1\}$ with equal probability
(B) pacman with random radius
(C1)/(C2) union of two discs
(D1)/(D2) random ellipse

EXAMPLE: SAND GRAINS

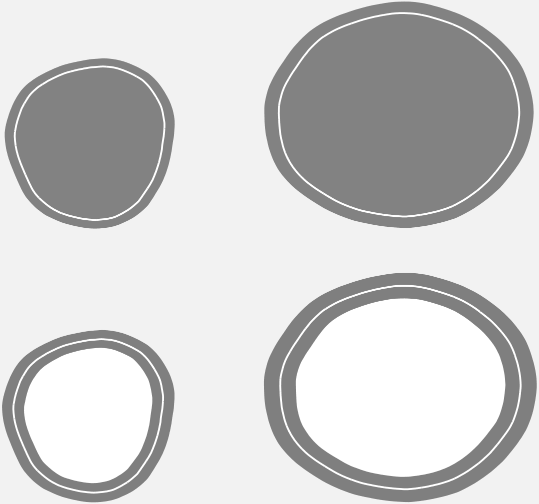


sand grains from the Baltic sea



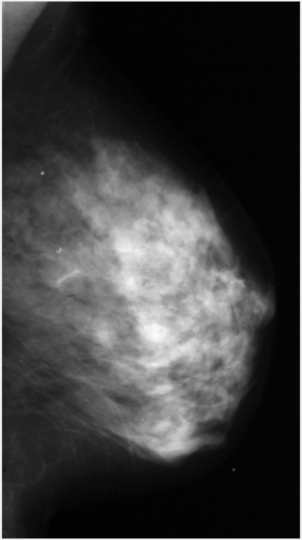
sand grains from the Zelenchuk river

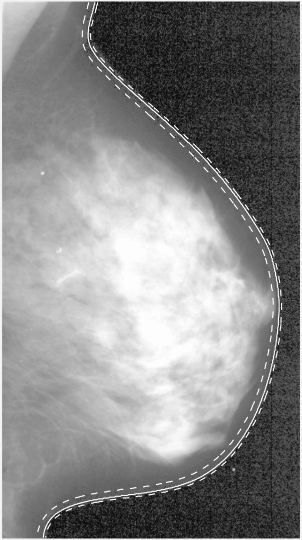
EXAMPLE: SAND GRAINS

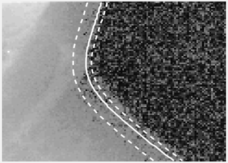


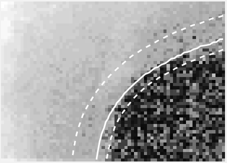
Zelenchuk river　　　　Baltic sea

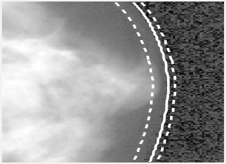
EXAMPLE: MAMMOGRAM











REFERENCES

- Stanberry, L. and Besag, J. (2009) Boundary reconstruction in binary images using splines. *Preprint.*
- Jankowski, H. and Stanberry, L. (2010) Expectations of random sets and their boundaries using oriented distance functions. *Journal of Mathematical Imaging and Vision* **36** 291–303.
- Jankowski, H. and Stanberry, L. (2011) Condence Regions for Means of Random Sets using Oriented Distance Functions *Scandinavian Journal of Statistics* To appear.