David Scharenberg

Zufällige abgeschlossene Mengen und Statistik

Gliederung

I. Grundlegende Definitionen und Eigenschaften

II. Kenngrößen zufällig abgeschlossener Mengen

III. Schätzer für die Kenngrößen

Sei $\mathcal{F}$ die Familie der abgeschlossenen Teilmengen des $\mathbb{R}^d$ sowie $\mathcal{C}$ die Menge der kompakten Teilmengen des $\mathbb{R}^d$.

Dann ist die **Schnitt-σ-Algebra $\sigma_{\mathcal{F}}$** die kleinste σ-Algebra von Teilmengen von $\mathcal{F}$, so dass gilt:

$$\{ B \in \mathcal{F} : B \cap C \neq \emptyset \} \in \sigma_{\mathcal{F}} \qquad \forall\, C \in \mathcal{C}.$$

Alternative Darstellung:

$\sigma_{\mathcal{F}} = \sigma(\{ \{B \in \mathcal{F} : B \cap C \neq \emptyset\} : C \in \mathcal{C}\})$

Sei ein Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$ gegeben.

Die Abbildung $\Xi : \Omega \longrightarrow \mathcal{F}$ heißt **zufällige abgeschlossene Menge**, falls
$\{ \omega \in \Omega : \Xi(\omega) \cap C \neq \emptyset \} \in \mathcal{A} \quad \forall\, C \in \mathcal{C}$.

Das heißt, Ξ ist $(\mathcal{A}, \sigma_{\mathcal{F}})$-messbar:
$\{ \omega \in \Omega : \Xi(\omega) \in B \} \in \mathcal{A} \quad \forall\, B \in \sigma_{\mathcal{F}}$.

Mengenoperationen

Seien Ξ und $\Xi^*: \Omega \longrightarrow \mathcal{F}$ beliebige ZAM.

Dann sind auch

1. $\Xi \cup \Xi^*$,
2. $\Xi \cap \Xi^*$,
3. $\Xi \oplus \Xi^*$ und
4. $\partial\Xi$

ZAM.

Beachte: Unendliche Vereinigungen abgeschlossener Mengen sind i.a. nicht abgeschlossen!

Translation und Rotation

Sei Ξ: $\Omega \longrightarrow \mathcal{F}$ eine beliebige ZAM. Dann gilt

1. Ξ + x ist eine ZAM für alle x $\in \mathbb{R}^d$.
2. $\delta(\Xi)$ ist eine ZAM für jede Drehung δ: $\mathbb{R}^d \longrightarrow \mathbb{R}^d$ um den Nullpunkt.

Beispiele für ZAM:

Zufällige Punktprozesse im $\mathbb{R}^d$

Ein abgeschlossener Kreis im $\mathbb{R}^2$, wobei der Mittelpunkt in $[0,1]^2$ gleichverteilt ist und der Radius eine positive eindimensionale Zufallsvariable ist

Ebenso können bestimmte Faserprozesse als ZAM aufgefasst werden (-> Vortrag 13)

Gegeben: markierter Punktprozess $\{(S_n),(L_n)\}$.

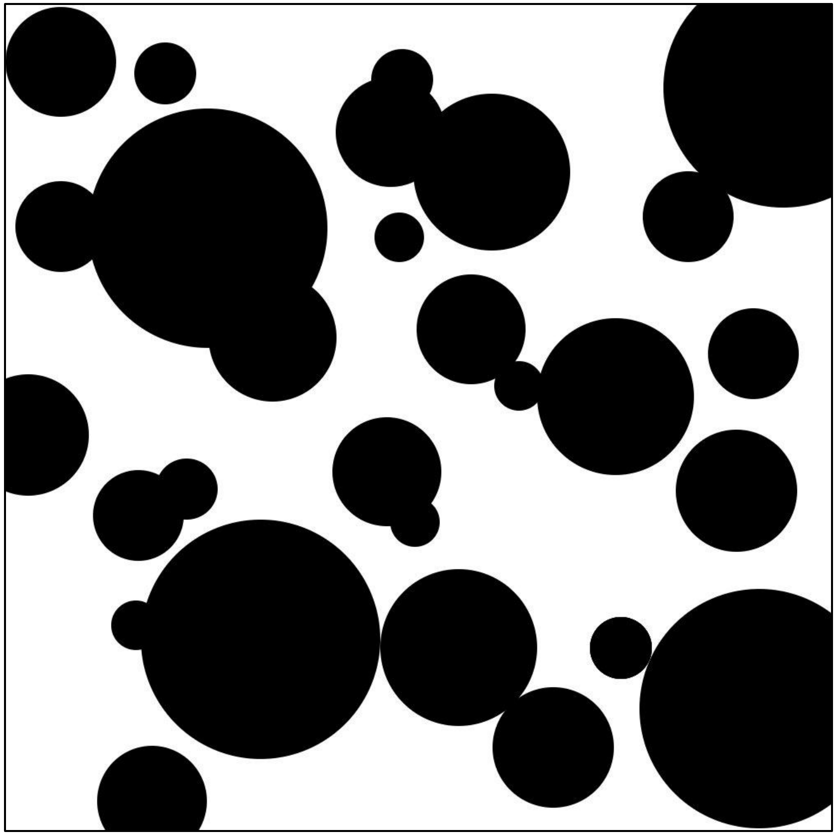
$\{S_n\}$ ist eine Folge von Zufallsvektoren im $\mathbb{R}^d$, die sog. Keime.

$L_n : \Omega \longrightarrow \mathcal{C}$ ein Zufallselement für alle $n \in \mathbb{N}$. Diese sog. Körner werden mit den Keimen assoziiert.

Hieraus geht dann das **Keim-Korn-Modell** hervor:

$$\Xi = \bigcup_{n=1}^{\infty} (L_n + S_n)$$

mit $(L_n + S_n) = \{x + S_n : x \in L_n\}$.



Realisierung eines Keim–Korn–Modells mit kreisförmigen Körnern

Sei Ξ eine beliebige ZAM.

Das Wahrscheinlichkeitsmaß $P_\Xi : \sigma_{\mathcal{F}} \longrightarrow [0,1]$ mit

$$P_\Xi(B)=P(\Xi \in B) \ \forall\, B \in \sigma_{\mathcal{F}}$$

heißt **Verteilung** von Ξ.

Sei Ξ eine beliebige ZAM.

Die Abbildung $T_\Xi : \mathcal{C} \longrightarrow [0,1]$ mit

$$T_\Xi(C)=P(\Xi \cap C \neq \emptyset) \;\forall\, C \in \mathcal{C}$$

heißt **Kapazitätsfunktional** von Ξ.

Anmerkung:
$T_\Xi(C)= P_\Xi(\{B \in \mathcal{F} : B \cap C \neq \emptyset\}) \;\forall\, C \in \mathcal{C}$

Für das Kapazitätsfunktional T_Ξ einer ZAM gilt, dass:

1. $T_\Xi(\emptyset) = 0$ und $0 \leqslant T_\Xi(C) \leqslant 1 \;\forall\; C \in \mathcal{C}$.

2. Für jede Folge $C_1, C_2, \ldots \in \mathcal{C}$ mit $C_1 \supseteq C_2 \supseteq \ldots$ gilt:

$$\lim_{n\to\infty} T_\Xi(C_n) = T_\Xi\Big(\bigcap_{n=1}^{\infty} C_n\Big)$$

3. Die Größen $S_0(C)$, $S_1(C;C_1)$, $S_2(C;C_1,C_2)$,… seien rekursiv definiert durch $S_0(C)=1-T_\Xi(C)$ und $S_n(C;C_1,\dots,C_n)=S_{n-1}(C;C_1,\dots,C_{n-1})-S_{n-1}(C\cup C_n;C_1,\dots,C_{n-1})$ für jedes $n\geqslant 1$ und für beliebige $C,C_1,\dots,C_n \in \mathcal{C}$.

   Dann gilt:
   $S_n(C;C_1,\dots,C_n)\geqslant 0$.

Eine Umkehrung liefert der sogenannte **Satz von Choquet**:

Sei T ein Funktional, das die Eigenschaften 1-3 der letzten Aussage erfüllt.

Dann gibt es ein eindeutig bestimmtes Wahrscheinlichkeitsmaß P: $\sigma_{\mathcal{F}} \longrightarrow [0,1]$ so dass

$$T(C)=P(\{B \in \mathcal{F} : B \cap C \neq \emptyset\}) \ \forall\ C \in \mathcal{C}.$$

Somit spielt das Kapazitätsfunktional die gleiche definierende Rolle für die Verteilung einer ZAM wie die Verteilungsfunktion für die Verteilung einer Zufallsvariable.

Es besteht ein eineindeutiger Zusammenhang zwischen Verteilung und Kapazitätsfunktional.

Seien Ξ_1 und $\Xi_2 : \Omega \longrightarrow \mathcal{F}$ zwei beliebige ZAM. Dann gilt:

$$P_{\Xi_1} = P_{\Xi_2} \iff T_{\Xi_1} = T_{\Xi_2} .$$

Eine ZAM Ξ heißt **stationär**, wenn gilt:

$$P_\Xi = P_{\Xi+x} \ \forall \ x \in \mathbb{R}^d.$$

Äquivalent:

$$T_\Xi(C) = T_\Xi(C_x) \ \forall \ C \in \mathcal{C} \ \forall \ x \in \mathbb{R}^d$$

$$T_\Xi(C) = T_{\Xi+x}(C) \ \forall \ C \in \mathcal{C} \ \forall \ x \in \mathbb{R}^d$$

Beweis:

$T_\Xi(C) = P(\Xi \cap C \neq \emptyset) = P((\Xi - x) \cap C \neq \emptyset) = P(\Xi \cap (C + x) \neq \emptyset)$
$= T_\Xi(C_x)$

Eine ZAM Ξ heißt **isotrop**, wenn gilt:

$$P_{\Xi} = P_{\delta(\Xi)}$$

für jede Drehung δ: $\mathbb{R}^d \longrightarrow \mathbb{R}^d$ um den Nullpunkt.

Äquivalent:

$T_{\Xi}(C) = T_{\Xi}(\delta(C)) \; \forall \; C \in \mathcal{C}$

Eine stationäre und isotrope ZAM heißt **bewegungsinvariant**.

Zwei ZAM Ξ_1 und Ξ_2 heißen unabhängig, wenn für beliebige $A_1, A_2 \in \sigma_{\mathcal{F}}$ gilt:

$$P(\Xi_1 \in A_1, \Xi_2 \in A_2) = P(\Xi_1 \in A_1)\, P(\Xi_2 \in A_2)$$

Äquivalent: Für beliebige $C_1, C_2 \in \mathcal{C}$ gilt:

$$P(\Xi_1 \cap C_1 \neq \emptyset, \Xi_2 \cap C_2 \neq \emptyset) = T_{\Xi_1}(C_1)\, T_{\Xi_2}(C_2).$$

Eine Folge von Mengen $W_n \subseteq \mathbb{R}^d$, $n \in \mathbb{N}$ mit

1. W_n ist konvex und kompakt $\forall\, n \in \mathbb{N}$
2. $W_n \subseteq W_{n+1} \;\forall\, n \in \mathbb{N}$
3. $\sup\{r \geqslant 0 \colon b(x,r) \subseteq W_n$ für ein $x \in \mathbb{R}^d\} \longrightarrow \infty \; (n \longrightarrow \infty)$

heißt **mittelnde Folge**.

Eine ZAM Ξ heißt **ergodisch**, falls $\forall$ A,B $\in \sigma_{\mathcal{F}}$

$$\lim_{n\to\infty} \frac{1}{\upsilon_d(W_n)} \int_{W_n} P(A_x \cap B)dx = P(A)P(B)$$

für eine mittelnde Folge $W_n \subseteq R^d$, n $\in \mathbb{N}$,wobei $A_x = \{ F_x : F \in A \}$ und $x \in \mathbb{R}^d$.

Eine ZAM Ξ heißt **metrisch transitiv**, falls $\forall$ A $\in \sigma_{\mathcal{F}}$ gilt:

$$\boxed{\begin{gathered} P_\Xi((A \setminus A_x) \cup (A_x \setminus A)) = 0 \ \forall\, x \in \mathbb{R}^d \\ \Longrightarrow \\ P_\Xi(A) = 1 \text{ oder } P_\Xi(A) = 0 \end{gathered}}$$

Man kann zeigen, dass eine ZAM genau dann ergodisch ist, wenn sie metrisch transitiv ist.

Sei nun $C_0=\{(x_1,\dots,x_d) \in \mathbb{R}^d : -\frac{1}{2} \leqslant x_i < \frac{1}{2},\ i=1,\dots,d\}$,

Ξ eine stationäre und ergodische ZAM,

h eine translationsinvariante und additive Abbildung mit

$h\colon \mathcal{B}(\mathbb{R}^d) \longrightarrow \mathbb{R}$,

W_n eine mittelnde Folge

und $K \subseteq C_0$ eine beliebige konvexe und kompakte Menge.

1. $\exists$ c > 0 mit $\mathbb{E}(|h(K\cap\Xi)|) \leqslant c$. Dann gilt:

$$\lim_{n\to\infty} E\left(\left|\frac{h(W_n\cap\Xi)}{\upsilon_d(W_n)} - E(h(C_0\cap\Xi))\right|\right) = 0$$

2. $\exists$ Zufallsvariable $\xi \geqslant 0$ mit $\mathbb{E}\xi < \infty$ und $|h(K\cap\Xi)| \leqslant \xi$ P-fs. Dann gilt:

$$\lim_{n\to\infty} \frac{h(W_n\cap\Xi)}{\upsilon_d(W_n)} = E(h(C_0\cap\Xi)) \quad \text{P-fs.}$$

Sei $\Xi : \Omega \longrightarrow \mathcal{F}$ eine beliebige ZAM.

Wir betrachten das **zufällige Feld** $\{ \mathbb{1}_{x \in \Xi}, x \in \mathbb{R}^d \}$ gegeben durch

$$\mathbb{1}_{x \in \Xi}(\omega) = \begin{cases} 1, \text{ wenn } x \in \Xi(\omega) \\ 0, \text{ wenn } x \notin \Xi(\omega) \end{cases}$$

Beachte den eineindeutigen Zusammenhang zwischen der Verteilung von Ξ und $\{ \mathbb{1}_{x \in \Xi}, x \in \mathbb{R}^d \}$!

$\Longrightarrow$ Betrachte Funktionen des zufälligen Feldes

Die Funktion $m(x)$, $x \in \mathbb{R}^d$ des zufälliges Feldes $\{ \mathbb{1}_{x\in\Xi}, x \in \mathbb{R}^d \}$ mit $m(x)=\mathbb{E}\, \mathbb{1}_{x\in\Xi}\ \forall\, x \in \mathbb{R}^d$ heißt **Mittelwertfunktion**.

Diese lässt sich durch **Einpunkt-Überdeckungswahrscheinlichkeiten** $\{p_x, x \in \mathbb{R}^d\}$ der ZAM Ξ mit $p_x= P(x\in\Xi)$ beschreiben.

Es gilt: $m(x)= p_x$.

Im stationären Fall hängt p_x nicht von $x \in \mathbb{R}^d$ ab.

II. Kenngrößen zufällig abgeschlossener Mengen

1. Volumenanteil

Das heißt für alle $x \in \mathbb{R}^d$ gilt:
$p_x = P(x \in \Xi) = P(0 \in \Xi) = P(\Xi \cap 0 \neq \emptyset) = T_{\Xi}(\{0\})$.

Wir nennen $p = P(0 \in \Xi)$ den **Volumenanteil** von Ξ.

Für diesen gilt:
$p = \mathbb{E}(\upsilon_d(\Xi \cap C_0))$
wobei C_0 das Einheitsquadrat ist.

Der Volumenanteil lieferte eine Beschreibung der Struktur erster Ordnung der ZAM Ξ, ähnlich der Intensität eines Punktprozesses.

Eine Beschreibung der Struktur zweiter Ordnung liefert die Kovarianz der ZAM Ξ.

Wieder betrachten wir zunächst das zufällige Feld $\{ \mathbb{1}_{x\in\Xi}, x \in \mathbb{R}^d \}$:

Die Funktion $k(x,y)$, $x,y \in \mathbb{R}^d$ des zufälliges Feldes $\{ \mathbb{1}_{x\in\Xi}, x \in \mathbb{R}^d \}$ mit $k(x,y)=\mathbb{E}(\mathbb{1}_{x\in\Xi}\mathbb{1}_{y\in\Xi})-\mathbb{E}\mathbb{1}_{x\in\Xi}\mathbb{E}\mathbb{1}_{y\in\Xi} \forall x,y \in \mathbb{R}^d$ heißt **Kovarianzfunktion**.

Diese lässt sich durch **Zweipunkt-Überdeckungswahrscheinlichkeiten** $\{p_{x,y}, x,y \in \mathbb{R}^d \}$ der ZAM Ξ mit $p_{x,y}= P(x\in\Xi, y\in\Xi)$ beschreiben.

Es gilt:

$k(x,y)=p_{x,y}-p_x p_y \ \forall x,y \in \mathbb{R}^d$

Im stationären Fall gilt:

$p_{x,y}=p_{0,y-x} \; \forall \; x,y \in \mathbb{R}^d$

$k(x,y)=p_{0,y-x}-p^2$

Dann nennen wir die Funktion $q: \mathbb{R}^d \longrightarrow [0,1]$ mit $q(x)=p_{0,x}$ **Kovarianz** von Ξ.

Beachte: Die Kovarianz ist nicht zentriert und nicht-negativ.

1. Für die Kovarianz gilt:
$q(x)= p_{0,x}= P(0 \in \Xi, x \in \Xi) = P(-x \in \Xi, 0 \in \Xi) = P(0 \in \Xi, -x \in \Xi)$
$=p_{0,-x}=q(-x) \; \forall \, x \in \mathbb{R}^d$

2. Falls Ξ isotrop ist, gilt:
$q(x)=q^*(||x||) \; \forall \, x \in \mathbb{R}^d$
Der Wert der Kovarianz hängt lediglich von der Norm eines Punktes des $\mathbb{R}^d$ ab.
Oft schreibt man dann $r=||x||$ statt x.
Beachte die unterschiedlichen Definitionsbereiche!

Sei Ξ eine stationäre ZAM und B $\in \mathcal{C}$ eine bzgl. des Ursprungs 0 sternförmige Menge.

Dann nennt man die Funktion $H_B:[0,\infty) \longrightarrow [0,1]$ mit $H_B(r)=$ $P(rB \cap \Xi \neq \emptyset \mid 0 \notin \Xi) \; \forall \, r \geqslant 0$ **Kontaktverteilungsfunktion** von Ξ.
B heißt strukturierendes Element von H_B.

Wichtige Spezialfälle sind:

1. Die lineare Kontaktverteilungsfunktion $H_\ell(r)$

Sei $u \in \mathbb{R}^d$ mit $\|u\|=1$ und $B = \{ cu \mid c \in [0,1] \}$.

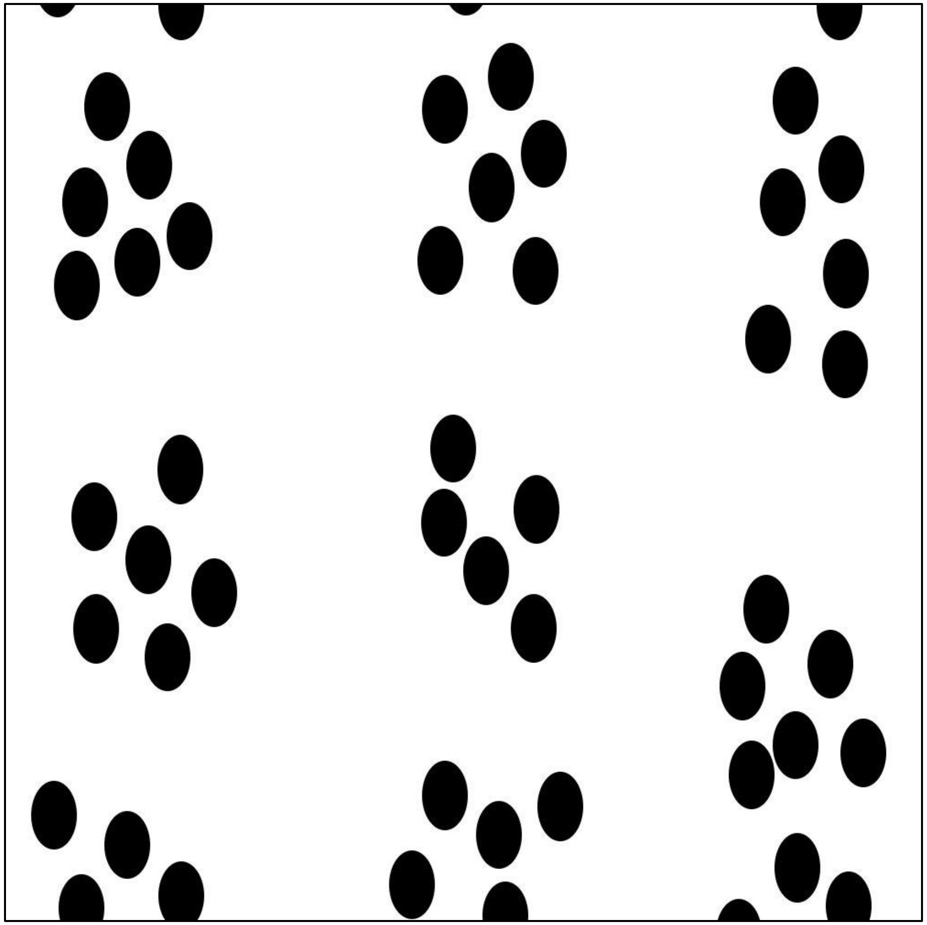
Beachte:

Falls Ξ isotrop ist, dann ist die genaue Wahl von u irrelevant.

Im anisotropen Fall gilt das nicht.

II. Kenngrößen zufällig abgeschlossener Mengen

3. Kontaktverteilungsfunktion



2. Die sphärische Kontaktverteilungsfunktion $H_s(r)$

Wähle $B=\{ x \in \mathbb{R}^d : \|x\| \leqslant 1 \}$.

Manchmal nennt man $H_s(r)$ das „Gesetz des ersten Kontakts“.

$1- H_s(r) = P(\{B(m,r) \cap \Xi \neq \emptyset \mid m \notin \Xi\}) \; \forall \, m \in \mathbb{R}^d$

Sei Ξ eine stationäre und isotrope ZAM im $\mathbb{R}^2$.

Die Punktzählmethode

Wähle ein Gitter von Punkten $x_1,\ldots,x_n$ (deterministisch oder zufällig)
Ein Schätzer für den Volumenanteil ist:

$$\hat{p}_p = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(x_i \in \Xi)$$

Dieser Schätzer ist unverzerrt:

$$E(\hat{p}_p) = \frac{1}{n} E \sum_{i=1}^{n} \mathrm{I}(x_i \in \Xi) = \frac{1}{n} np = p$$

Für seine Varianz gilt: (k Kovarianzfunktion)

$$\sigma_p^2 = \frac{1}{n^2} E((np - \sum_{i=1}^{n} \mathrm{I}(x_i \in \Xi))^2)$$
$$= \frac{1}{n^2} (np(1-p) + 2 \sum_{i>j} q^*(\| x_i - x_j \|)$$

Falls n „groß“ ist und k „schnell gegen 0 geht“, dann kann man die Varianz approximieren durch:

$$\sigma_p^2 \approx \frac{p(1-p)}{n}$$

Im Falle unabhängiger Indikatorfunktionen $\mathbb{1}_\Xi(x_i)$ gilt insbesondere:

$$\sigma_p^2 = \frac{p(1-p)}{n}$$

Beispiele für weitere Schätzer:
-Lineare Methode
-Flächenmethode

Anmerkungen:
-Oft: k(x,y) unbekannt $\Longrightarrow$ Vergleich der Varianzen der Schätzer nicht möglich

-Numerischer Aufwand zu beachten:
Geringere Varianz oft (aber nicht immer!) durch höheren Aufwand „erkauft“

Erinnerung: $q(x)= P(0 \in \Xi, x \in \Xi)$

Außerdem gilt:

$0 \in \Xi, x \in \Xi \Leftrightarrow 0 \in \Xi^* = \{ \Xi \ominus \{0,-x\} \}$

Ein Schätzer für $q(x)$ ist nun gegeben durch einen Schätzer des Volumenanteils der ZAM Ξ^*.

D. Stoyan, W.S. Kendall, J. Mecke: Stochastic Geometry and its Applications

Volker Schmidt: Vorlesungsskript Räumliche Statistik