

Das Boolesche Modell

Olaf Wied

November 30, 2009

1. Einführung und Definition
2. Eigenschaften
3. Modell mit konvexen Körnern
4. Statistiken und Anwendung in **R**

- ▶ Punkte im \mathbb{R}^d
- ▶ um jeden Punkt: beliebige , kompakte Mengen

- ▶ Punkte im \mathbb{R}^d
- ▶ um jeden Punkt: beliebige , kompakte Mengen
- ▶ Folge von Zufallsvektoren $\Phi = \{x_1, x_2, \dots\}$
- ▶ zufällige abgeschlossene Mengen Ξ_1, Ξ_2, \dots

- ▶ Punkte im \mathbb{R}^d
- ▶ um jeden Punkt: beliebige , kompakte Mengen
- ▶ Folge von Zufallsvektoren $\Phi = \{x_1, x_2, \dots\}$
- ▶ zufällige abgeschlossene Mengen Ξ_1, Ξ_2, \dots
- ▶ Das *Keim - Korn Modell* ist dann die Vereinigung:

$$\Xi = \bigcup_{n=1}^{\infty} (\Xi_n + x_n)$$

- ▶ mit $\Xi_n + x_n = \{x + x_n : x \in \Xi_n\}$

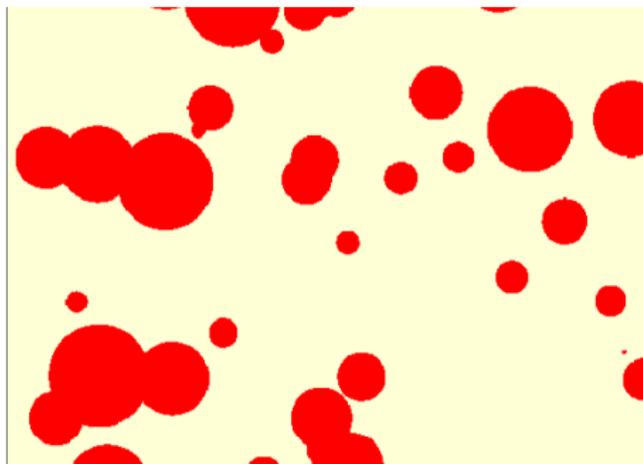


Figure: $\{x_n\}$ Poisson-Prozess und $\Xi_n = B(o, R_n)$ mit $R_n \sim U(u_1, u_2)$

- ▶ Sei Φ ein stationärer Poissonprozess mit Intensität λ

$$\Phi = \{x_1, x_2, \dots\}$$

- ▶ und $\{\Xi_n\}$ eine Folge u.i.v. zufälliger, kompakter Mengen $\subset \mathbb{R}^d$ unabhängig vom Poissonprozess Φ

$$\Xi_1, \Xi_2, \dots$$

- ▶ Dann ist das *Boolesche Modell* definiert als

$$\Xi = \bigcup_{n=1}^{\infty} (\Xi_n + x_n),$$

- ▶ wobei folgende Bedingung erfüllt sein soll

$$\mathbb{E}(\nu_d(\Xi_1 \oplus K)) < \infty, \forall K \text{ kompakt} \quad (1)$$

2. Eigenschaften

- ▶ aus (1) kann man folgern, dass das Boolesche Modell eine *zufällige, abgeschlossene Menge* (ZAM) ist
- ▶ (1) ist z.B. dann erfüllt, wenn $\mathbb{P}(\Xi_1 \subset B(o, R)) = 1$, wobei R eine (nicht negative) ZV ist mit $\mathbb{E}R^d < \infty$
- ▶ $\{(x_n, \Xi_n)\}$ bildet einen *markierten Poissonprozess* im \mathbb{R}^d
- ▶ die Verteilung der Ξ_n nennt man *Kornverteilung* (bzw. Markenverteilung)

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

► **Verteilung einer ZAM:**

Das Wahrscheinlichkeitsmaß $\mathbb{P}_{\Xi} : \sigma_{\mathcal{F}} \rightarrow [0, 1]$ mit $\mathbb{P}_{\Xi}(B) = \mathbb{P}(\Xi \in B)$, $\forall B \in \sigma_{\mathcal{F}}$

► **Kapazitätsfunktional:**

$T_{\Xi} : \mathcal{C} \rightarrow [0, 1]$ mit $T_{\Xi}(C) = \mathbb{P}(\Xi \cap C \neq \emptyset)$, $\forall C \in \mathcal{C}$

► **Stationarität:**

Eine ZAM heißt stationär, wenn $\mathbb{P}_{\Xi}(B) = \mathbb{P}_{\Xi+x}(B)$, $\forall x \in \mathbb{R}^d$.
Das ist äquivalent zu $T_{\Xi}(B) = T_{\Xi+x}(B)$.

► **Isotropie:**

Eine ZAM heißt isotrop, wenn $\mathbb{P}_{\Xi} = \mathbb{P}_{\delta(\Xi)}$ für jede Drehung δ um den Ursprung.
Das ist äquivalent zu $T_{\Xi}(B) = T_{\delta(\Xi)}(B)$.

$$\Xi = \bigcup_{n=1}^{\infty} (\Xi_n + x_n) \text{ ist stationär.}$$

Beweis:

$$\begin{aligned} T_{\Xi}(C) &= \mathbb{P}(\Xi \cap C \neq \emptyset) \\ &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} (\Xi_n + x_n) \cap C \neq \emptyset\right) \\ &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} (\Xi_n + (x_n + x)) \cap C \neq \emptyset\right) \\ &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} ((\Xi_n + x_n) + x) \cap C \neq \emptyset\right) \\ &= \mathbb{P}((\Xi + x) \cap C \neq \emptyset) = T_{\Xi+x}(C) \end{aligned}$$

$$\Xi = \bigcup_{n=1}^{\infty} (\Xi_n + x_n) \text{ ist isotrop, wenn } \Xi_1 \text{ isotrop ist.}$$

Beweis:

- ▶ analog zur Stationarität

1. Kapazitätsfunktional
2. Volumenanteil
3. Kovarianz
4. Kontaktverteilungsfunktion

- ▶ Erinnerung: $T_{\Xi}(C) = \mathbb{P}(\Xi \cap C \neq \emptyset)$
- ▶ im Booleschen Modell gilt

$$T_{\Xi}(C) = 1 - \exp(-\lambda \mathbb{E} \nu_d(\Xi_1 \oplus \check{C}))$$

Beweisskizze: Betrachte $\Phi_- = \{x_n \in \Phi : (\Xi_n + x_n) \cap C \neq \emptyset\}$

- ▶ Φ_- ist wieder ein Poisson-Prozess
- ▶ mit Intensitätsmaß

$$\Lambda_-(B) = \lambda \int_B \rho(x) dx, \quad \forall B \text{ Borelmenge}$$

mit

$$\rho(x) = \mathbb{P}((\Xi_1 + x) \cap C \neq \emptyset)$$

- Damit ($\forall C \in \mathcal{C}$):

$$\begin{aligned} T_{\Xi}(C) &= \mathbb{P}(\Xi \cap C \neq \emptyset) = 1 - \mathbb{P}(\Xi \cap C = \emptyset) \\ &= 1 - \exp(-\Lambda_-(\mathbb{R}^d)) \\ &= 1 - \exp\left(-\lambda \int_{\mathbb{R}^d} \mathbb{P}((\Xi_1 + x) \cap C \neq \emptyset) dx\right) \\ &= 1 - \exp(-\lambda \mathbb{E} \nu_d(\Xi_1 \oplus \check{C})) \end{aligned}$$

- ▶ Definition: $p = \mathbb{E}(\nu_d(\Xi \cap B)), \nu_d(B) = 1$
- ▶ $p = \mathbb{P}(o \in \Xi) = \mathbb{P}(x \in \Xi), \forall x \in \mathbb{R}^d$
- ▶ Damit: $p = 1 - \exp(-\lambda \mathbb{E}\nu_d(\Xi_1))$

Erinnerung

- ▶ möglicher Schätzer:

$$\hat{p}_W = \frac{\nu_d(W \cap \Xi)}{\nu_d(W)}$$

- ▶ mit $0 < \nu_d(W) < \infty$

- ▶ Definition: $C(x) = \mathbb{P}(o \in \Xi, x \in \Xi)$
- ▶ Es gilt: $C(x) = C(-x)$
- ▶ $C(x) = 2p - T_{\Xi}(\{o, x\})$

Kontaktverteilungsfunktion H_B

- ▶ Definition:

$$H_B(r) = 1 - \frac{\mathbb{P}(\Xi \cap rB = \emptyset)}{1 - p} \quad \forall r \geq 0$$

- ▶ mit konvexem und kompaktem "Strukturelement"
 $B \subset \mathbb{R}^d$, wobei $o \in B$
- ▶ $H_B(r)$ kann als Verteilung der ZV R aufgefasst werden (unter der Bedingung $R > 0$), mit :

$$R = \inf \{s : \Xi \cap sB \neq \emptyset\}$$

- ▶ Aus dem Resultat über das Kapazitätsfunktional und den Volumenanteil folgt:

$$H_B(r) = 1 - \exp(-\lambda(\mathbb{E}\nu_d(\Xi_1 \oplus r\check{B}) - \mathbb{E}\nu_d(\Xi_1)))$$

$$\hat{H}_{B,W}(r) = 1 - \frac{1 - \hat{p}_{B,W}(r)}{1 - \hat{p}_W}$$

mit

$$\hat{p}_{B,W}(r) = \frac{\nu_d(W \ominus r\check{B} \cap (\Xi \oplus r\check{B}))}{\nu_d(W \ominus r\check{B})}$$

und \hat{p}_W der eingeführte Schätzer des Volumenanteils p

3. Modell mit konvexen Körnern

Vorteile der Annahme isotroper und konvexer Körner:

- ▶ Vereinfachung vieler Formeln
- ▶ immer noch sehr flexibel

Verallgemeinerter Satz von Steiner

Sei Ξ eine isotrope, konvexe, zufällige, abgeschlossene Menge, mit $\mathbb{E}V_d(\Xi \oplus B(o, r)) < \infty, \forall r > 0$. Dann gilt

$$\mathbb{E}V_d(\Xi \oplus B) = \sum_{k=0}^d \alpha_{dk} \mathbb{E}V_{d-k}(\Xi) V_k(B), \quad \forall B \in \mathcal{K}$$

wobei $\alpha_{dk} = \frac{k! \kappa_k (d-k)! \kappa_{d-k}}{d! \kappa_d}$ und $\kappa_i = i$ -dimensionales Volumen der Einheitskugel in \mathbb{R}^i .

- ▶ insbesondere gilt für V_{d-1} und V_1 :

$$2V_{d-1}(B) = \text{"Oberfläche von } B\text{"}$$

$$V_1(B) = \frac{d\kappa_d}{2\kappa_{d-1}} \beta(B),$$

mit der mittleren Breite $\beta(B)$

Vereinfachung mit inneren Volumina

Gelte $\Xi_1, \Xi_2, \dots : \rightarrow \mathcal{K} \setminus \{\emptyset\}$ mit

$$\mathbb{E}V_d(\Xi_1 \oplus B(o, R)) < \infty, \forall R > 0$$

Dann gilt für:

- ▶ das Kapazitätsfunktional

$$T_{\Xi}(C) = 1 - \exp\left(-\lambda \sum_{k=0}^d \alpha_{dk} \mathbb{E}V_{d-k}(\Xi_1) V_k(C)\right), \forall C \in \mathcal{K}$$

- ▶ die Kontaktverteilungsfunktion

$$H_B(r) = 1 - \exp\left(-\lambda \sum_{k=1}^d \alpha_{dk} r^k \mathbb{E}V_{d-k}(\Xi_1) V_k(B)\right), \forall r \geq 0$$

Einfache Anwendung: Die Lineare Kontakverteilungsfunktion H_l

- ▶ $d = 2$
- ▶ $B =$ Strecke der Länge 1

$$\Rightarrow H_l(r) = 1 - \exp\left(-\lambda \frac{\mathbb{E}2V_1(\Xi_1)}{\pi} r\right), r \geq 0$$

- ▶ Mögliches Modell: Sichtweite im Wald
- ▶ Bäume mit erwartetem Radius $\mathbb{E}(R)$
- ▶ $\Rightarrow H_l(r) = 1 - \exp(-2\lambda r \mathbb{E}(R))$
- ▶ Bsp: $\mathbb{E}(R) = 0.20m$, $\lambda = 0.01m^{-2}$
- ▶ $\Rightarrow H_l(500) = 1 - 0.135$
- ▶ Interpretation

Fragestellung: Wie gut können Realdaten durch ein Boolesches Modell beschrieben werden?

- ▶ Gilt für die geschätzte Kontaktverteilungsfunktion (sphärisch,quadratisch)

$$\log(1 - \hat{H}_b(r)) = -a(B)r - b(B)r^2 ?$$

↔ Simulation und Anwendung in **R**

Erzeugen eines einfachen Booleschen Modells

```
library(spatstat)
```

```
n = rpois(1,500)
```

```
W = owin(c(0,1),c(0,1))
```

```
P = runifpoint(n,win=W)
```

```
plot(P)
```

Paket für Punktprozesse im \mathbb{R}^2

Poissonprozess generieren

Anzahl der Punkte

erzeugt das Fenster $[0,1] \times [0,1]$

erzeugt den Poissonprozess

als spezielles Objekt

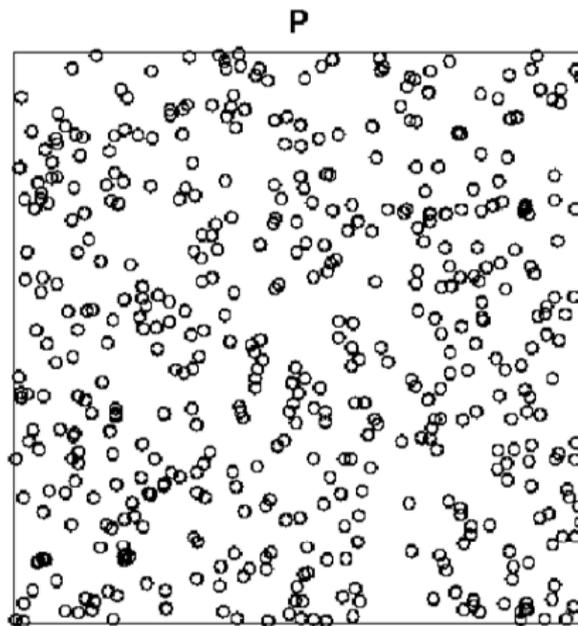


Figure: der erzeugte Poisson-Prozess

Erzeugen eines einfachen Booleschen Modells

```
R = runif(n,0.01,0.02)
```

```
B = owin(c(0,1),c(0,1),mask = matrix(TRUE,600,600))
```

```
X = raster.x(B)
```

```
Y = raster.y(B)
```

```
m = (X-P$x[1])^2 + (Y-P$y[1])^2 ≤ R[1]^2
```

```
for ( i in 2:length(P$x) ) {
```

```
mask - help = ( (X-P$x[i])^ 2 + (Y-P$y[i])^ 2 ≤ R[i]^ 2 )
```

```
m = m + mask - help
```

```
m = m ≥ 1
```

```
B = owin(c(0,1),c(0,1), mask = m)
```

```
plot(B)
```

gleichverteilte Radien

das Boolesche Modell

teilen das Fenster in 600*600

Pixel ein

welche Pixel liegen

in einem Kreis?

nur die Pixel, die in (mind.)

einem Kreis liegen

B ist unser Boolesches Modell

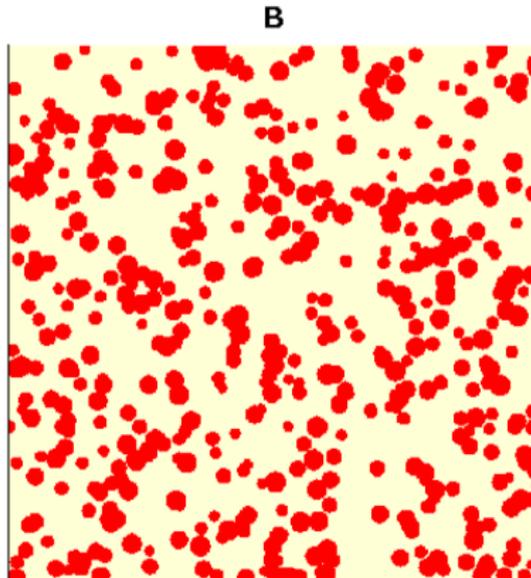


Figure: das erzeugte Boolesche Modell

Schätzen der sphärischen Kontaktverteilungsfunktion

$H = \text{Hest}(B)$

$r = H(r)$

$\lambda = \text{summary}(P)\$intensity$

$U = 2 \cdot \pi \cdot \text{mean}(R)$

```
h = function(x){  
  return(1-exp(-lambda*(U*x+pi*x^2)))  
}
```

sphärische Kontaktverteilungsfunktion

wollen nun $H(r)$ selber schätzen:

Schätzer für λ

erwarteter Umfang der Körner

Vergleich der geschätzten Kontaktverteilungsfunktionen

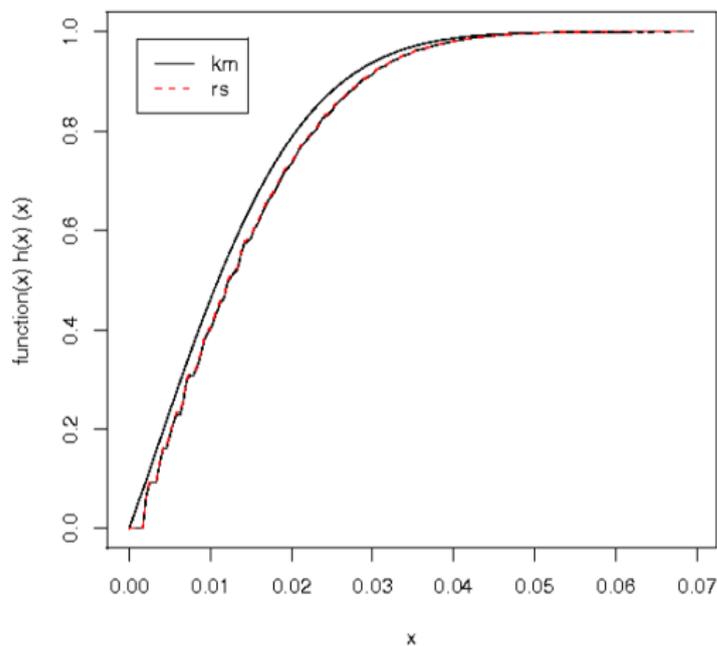


Figure: ähnlicher Verlauf

Testen der Annahmen über logarithmieren

```
L = -1/r*log(1-H$raw)
plot(r,L,type=l)
```

```
l = function(x){
return(lambda*(U+pi*x))
```

```
plot(function(x) l(x), 0,max(r), add=T)
```

nach Definition eine Gerade:

$$y = \lambda * (U + \pi * x)$$

logarithmierte Funktionen

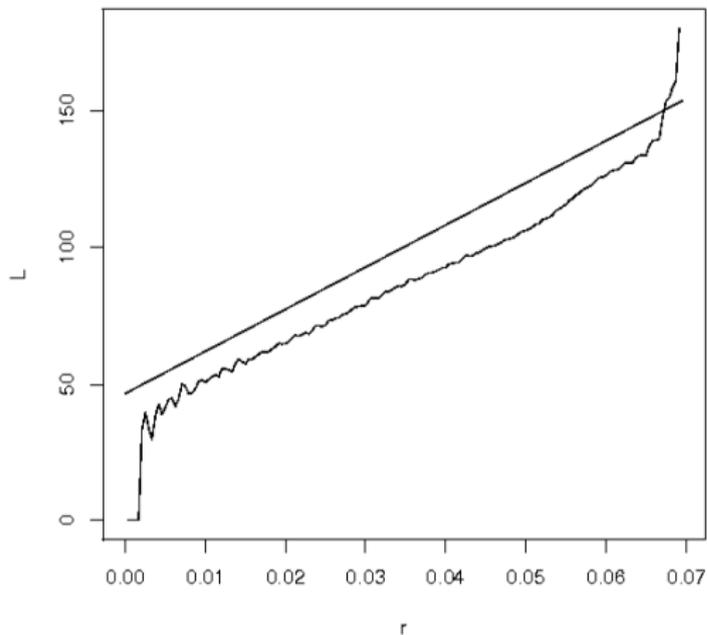


Figure: linear?

Ziel: Intensität λ und (erwartete) geometrische Charakteristika $\mathbb{E}V_j(\Xi_1)$ schätzen

- ▶ als Minimierungsproblem: $\int_{r_1}^{r_2} (F(r; \xi_1, \xi_2) - \hat{F}(r))^2 dr$
- ▶ mit spärlicher Kontaktverteilungsfunktion:
- ▶ $H_B(r) = 1 - \exp(-\lambda \sum_{j=0}^{d-1} \kappa_{d-j} \mathbb{E}V_j(\Xi_1) r^{d-j})$

$$\sum_{i=1}^m (-\log(1 - \hat{H}_{B,W}(r_i, \xi)) - \lambda(\kappa_d r_i^d + \sum_{j=1}^{d-1} \kappa_{d-j} \delta_j r_j^{d-j}))^2$$

- ▶ $m \geq d, r_i > 0$ und geeignete Schätzer \hat{H}
- ▶ ergibt Schätzer für $\lambda, \delta_1, \dots, \delta_{d-1}$ und damit
- ▶ $\hat{\delta}_d = \log(1 - \hat{p}_W) / \hat{\lambda}_W$ (Plug-in Schätzer)

$$d = m = 2$$

$$\sum_{i=1}^2 (-\log(1 - \hat{H}_{B,W}(r_i, \xi)) - 2r_i\lambda\delta_1 - \pi r_i^2\lambda)^2$$

- ▶ 2 Gleichungen mit 2 Unbekannten
- ▶ eindeutig nach λ auflösbar
- ▶ ergibt den Schätzer:

$$\lambda_{\hat{W}} = \frac{-\log(1 - \hat{H}_{B,W}(r_2, \xi)) + \frac{r_2}{r_1} \log(1 - \hat{H}_{B,W}(r_1, \xi))}{-r_2 r_1 \pi + \pi r_2^2}$$

- ▶ Ist $m = 2$ eine gute Wahl?

↔ Anwendung in **R**

Ergebnis für unser Beispiel

```
v = length(r)/4
r1 = H$r[v]
r2 = H$r[3*v]
schaetzer_lambda =
(-log(1-H$km[3*v])+r2/r1*log(1-H$km[v]))/(-pi*r2*r1+pi*r2^2)
schaetzer_lambda
[1] 557.3094
```

Wahl von r_1, r_2

Ergebnis

exakter Wert: 489

- ▶ Stoyan, D., Kendall, W.S., Mecke, J. (1995): Stochastic Geometry and its Applications, Wiley
- ▶ Schmidt, V. (2008): Räumliche Statistik (Vorlesungsskript)
- ▶ <http://www.spatstat.org/>