Stochastische Geometrie und ihre Anwendungen
Thema 3: Statistik für Punktprozesse

Henrike Häbel

Uni Ulm

12. November 2009

Gliederung

1. Wiederholung
2. Intensität λ
   - Schätzer $\hat{\lambda}$
   - Konfidenzintervall für λ
3. Test auf Stationarität
4. Test der Poisson-Hypothese
   - Plus-Sampling
   - Minus-Sampling
   - Distanzmethode

Gliederung

1. Wiederholung
2. Intensität λ
   - Schätzer $\hat{\lambda}$
   - Konfidenzintervall für λ
3. Test auf Stationarität
4. Test der Poisson-Hypothese
   - Plus-Sampling
   - Minus-Sampling
   - Distanzmethode

$\Phi(W)$

Für $W \in \mathcal{B}(\mathbb{R}^d)$ und $\Phi = \{x_n\}$ ein Punktprozess mit einer zufälligen Anzahl $n \in \mathbb{N}$ Punkten $x_i,\ 1 \leq i \leq n$ beschreibt

$$\Phi(W) = \#\{n \in \mathbb{N} | x_n \in W\}$$

die Anzahl der Punkte in W.

Definition (stationärer Poisson Prozess)

Φ heißt stationärer Poisson Prozess, wenn folgende Eigenschaften gelten

(i) $\Phi(W) \sim Poi(\mathbb{E}[\Phi(W)]) = Poi(\lambda\nu_d(W)) \quad \forall\ W \in \mathcal{B}(\mathbb{R}^d)$

(ii) $\Phi(W_1), \ldots, \Phi(W_n)$ sind unabhängig für $W_i \in \mathcal{B}(\mathbb{R}^d),\ W_i \cap W_j = \emptyset,\ 1 \leq, i, j \leq n,\ i \neq j$

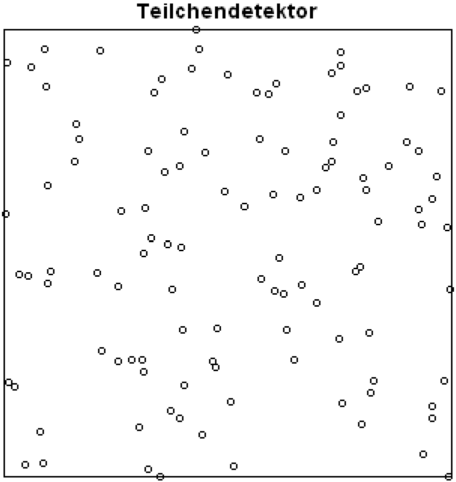
(iii) Φ ist stationär, das heißt es gilt die Translationsinvarianz $(\Phi(W_1),\ \ldots,\ \Phi(W_n)) \overset{D}{=} (\Phi(W_1 + v),\ \ldots,\ \Phi(W_n + v)),$ $\forall\ W_1,\ \ldots, W_n \in \mathcal{B}(\mathbb{R}^d),\ v \in \mathbb{R}^d$

Definition (Intensität für planare stationäre Punktprozesse)

Die erwartete Anzahl von Punkten x_j im Fenster W mit Fläche $\nu_d(W)$ wird mit der Intensität λ bezeichnet, wobei

$$\lambda = \frac{\mathbb{E}\left[\Phi(W)\right]}{\nu_2(W)}$$

117 Auftreffpunkte von α-Teilchen auf einer 60x60 Metallfolie
(1 EH = 2 μ m)



Gliederung

1 Wiederholung

2 Intensität λ
- Schätzer $\hat{\lambda}$
- Konfidenzintervall für λ

3 Test auf Stationarität

4 Test der Poisson-Hypothese
- Plus-Sampling
- Minus-Sampling
- Distanzmethode

Gliederung

1. Wiederholung
2. Intensität λ
   - Schätzer $\hat{\lambda}$
   - Konfidenzintervall für λ
3. Test auf Stationarität
4. Test der Poisson-Hypothese
   - Plus-Sampling
   - Minus-Sampling
   - Distanzmethode

Definition

Für einen stationären Poisson Prozess sei ein Schätzer für die Intensität λ wie folgt definiert:

$$\hat{\lambda} = \frac{\Phi(W)}{\nu_2(W)} \tag{1}$$

Für das Beipiel ergibt sich:

$$\hat{\lambda} = \frac{117}{60^2} = 0.0325$$

Theorem

$\hat{\lambda}$ ist ein erwartungstreuer Schätzer.

Theorem

Sei $(W_n)_{n\in\mathbb{N}}$ eine Folge mit $W \in \mathcal{B}^d$ und $\nu_d(W_n) \to \infty$ für $n \to \infty$. Dann ist $\hat{\lambda}$ schwach konsistent, das heißt, es gilt

$$\lim_{n\to\infty} \mathbb{P}(|\hat{\lambda}_{W_n} - \lambda| > \epsilon) = 0$$

Theorem

$\hat{\lambda}$ ist ein erwartungstreuer Schätzer.

Theorem

Sei $(W_n)_{n\in\mathbb{N}}$ eine Folge mit $W \in \mathcal{B}^d$ und $\nu_d(W_n) \to \infty$ für $n \to \infty$. Dann ist $\hat{\lambda}$ schwach konsistent, das heißt, es gilt

$$\lim_{n\to\infty} \mathbb{P}(|\hat{\lambda}_{W_n} - \lambda| > \epsilon) = 0$$

Gliederung

1. Wiederholung
2. Intensität λ
   - Schätzer $\hat{\lambda}$
   - Konfidenzintervall für λ
3. Test auf Stationarität
4. Test der Poisson-Hypothese
   - Plus-Sampling
   - Minus-Sampling
   - Distanzmethode

Theorem (asymptotische Normalverteiltheit)

Sei $(W_n)_{n\in\mathbb{N}}$ eine Folge in $W \in \mathcal{B}^d$ mit $\nu_d(W_n) \to \infty$ für $n \to \infty$. Dann ist $\hat{\lambda}$ asymptotisch normalverteilt, das heißt, es gilt

$$\lim_{n\to\infty} \mathbb{P}\left(-z_{\frac{\alpha}{2}} \leq \sqrt{\frac{\nu_2(W)}{\hat{\lambda}}}(\hat{\lambda} - \lambda) \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha,$$

Asympthotisches Konfidenzintervall zum Niveau $1 - \alpha$

$$z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{\nu_2(W)}} - \hat{\lambda} \leq \lambda \leq z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{\nu_2(W)}} + \hat{\lambda}$$

Theorem (asymptotische Normalverteiltheit)

Sei $(W_n)_{n\in\mathbb{N}}$ eine Folge in $W \in \mathcal{B}^d$ mit $\nu_d(W_n) \to \infty$ für $n \to \infty$. Dann ist $\hat{\lambda}$ asymptotisch normalverteilt, das heißt, es gilt

$$\lim_{n\to\infty} \mathbb{P}\left(-z_{\frac{\alpha}{2}} \leq \sqrt{\frac{\nu_2(W)}{\hat{\lambda}}}(\hat{\lambda} - \lambda) \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha,$$

Asympthotisches Konfidenzintervall zum Niveau $1 - \alpha$

$$z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}}{\nu_2(W)}} - \hat{\lambda} \leq \lambda \leq z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}}{\nu_2(W)}} + \hat{\lambda}$$

Gliederung

1. Wiederholung
2. Intensität λ
   - Schätzer $\hat{\lambda}$
   - Konfidenzintervall für λ
3. Test auf Stationarität
4. Test der Poisson-Hypothese
   - Plus-Sampling
   - Minus-Sampling
   - Distanzmethode

Φ stationärer Poisson Prozess, dann gilt [1] für

- $W_1, W_2 \subset W, W_1 \cap W_2 = \emptyset$
- $n_1,\ n_2$ zufällige Anzahl der Punkte

$$F = \frac{\nu_2(W_1)(2n_2 + 1)}{\nu_2(W_2)(2n_1 + 1)} \sim F_{2n_1+1,2n_2+1},$$

wobei die Indizes so gewählt werden, dass $F > 1$

Test

H_0 : Φ ist ein stationärer Poisson Prozess
H_0 wird zum Niveau α verworfen, falls

$$F < F_{2n_1+1,2n_2+1;\frac{\alpha}{2}} \text{ oder } F > F_{2n_1+1,2n_2+1;1-\frac{\alpha}{2}}$$

[1] heuritisch, gleiches gilt für alle folgenden Verteilungsangaben.

Gliederung

1. Wiederholung
2. Intensität λ
   - Schätzer $\hat{\lambda}$
   - Konfidenzintervall für λ
3. Test auf Stationarität
4. Test der Poisson-Hypothese
   - Plus-Sampling
   - Minus-Sampling
   - Distanzmethode

Wichtige Verteilungen für $r \geq 0$

- reduzierte Palmfunktion

$$P_o^!(Y) = \mathbb{P}(\Phi \setminus \{o\} \in Y \mid\mid o)$$

- Sphärische Kontaktverteilungsfunktion, wobei $b(x, r) \in \mathcal{B}(\mathbb{R}^d)$

$$H_S(r) = 1 - P(\varphi \in \mathbb{N} : \varphi(b(x, r)) = 1) = \mathbb{P}(o \in \bigcup_{x \in \Phi} b(x, r))$$

- Nächste-Nachbarfunktion, wobei $b(o, r) \in \mathcal{B}(\mathbb{R}^d)$

$$\begin{aligned} D(r) &= 1 - P_o^!(\varphi \in \mathbb{N} : \varphi(b(o, r)) = 0) \\ &= 1 - P_o(\varphi \in \mathbb{N} : \varphi(b(o, r)) = 1) \end{aligned}$$

Gliederung

1. Wiederholung
2. Intensität λ
   - Schätzer $\hat{\lambda}$
   - Konfidenzintervall für λ
3. Test auf Stationarität
4. Test der Poisson-Hypothese
   - Plus-Sampling
   - Minus-Sampling
   - Distanzmethode

$D(r)$ wird geschätzt mittels der empirischen Verteilungsfunktion aller Distanzen

$$\|x - h(x)\| \,, \ h(x) \mathrel{\hat{=}} \text{nächster Nachbar}$$

+ Kein Informationsverlust
– mehr Informationen als in W enthalten benötigt
– $h(x)$ eventuell nicht ermittelbar $\Rightarrow$ Verzerrung

$D(r)$ wird geschätzt mittels der empirischen Verteilungsfunktion aller Distanzen

$$\|x - h(x)\|\ ,\ \ h(x) \mathrel{\hat{=}} \text{nächster Nachbar}$$

- $+$ Kein Informationsverlust
- $-$ mehr Informationen als in W enthalten benötigt
- $-$ $h(x)$ eventuell nicht ermittelbar $\Rightarrow$ Verzerrung

Gliederung

1. Wiederholung
2. Intensität λ
   - Schätzer $\hat{\lambda}$
   - Konfidenzintervall für λ
3. Test auf Stationarität
4. Test der Poisson-Hypothese
   - Plus-Sampling
   - Minus-Sampling
   - Distanzmethode

$D(r)$ wird geschätzt mittels der empirischen Verteilungsfunktion der Distanzen

$$\|x - h(x)\|\,,\ h(x) \mathrel{\hat{=}} \text{nächster Nachbar}$$

in $W \ominus b(o, r)$

+ keine Verzerrung
− Informationsverlust

$D(r)$ wird geschätzt mittels der empirischen Verteilungsfunktion der Distanzen

$$\|x - h(x)\|\,,\ h(x) \mathrel{\hat{=}} \text{nächster Nachbar}$$

in $W \ominus b(o, r)$

- $+$ keine Verzerrung
- $-$ Informationsverlust

Gliederung

1. Wiederholung
2. Intensität λ
   - Schätzer $\hat{\lambda}$
   - Konfidenzintervall für λ
3. Test auf Stationarität
4. Test der Poisson-Hypothese
   - Plus-Sampling
   - Minus-Sampling
   - Distanzmethode

Vorbereitung

H_0 : Φ ist ein stationärer Poisson Prozess
Unter H_0 gilt:

$$D(r) = H_S(r) = 1 - \exp^{-\lambda \pi r^2}$$

$$\rightarrow u_1^2, \ldots, u_m^2, \; v_1^2, \ldots, v_m^2 \sim Exp(\lambda \pi)$$

Tests

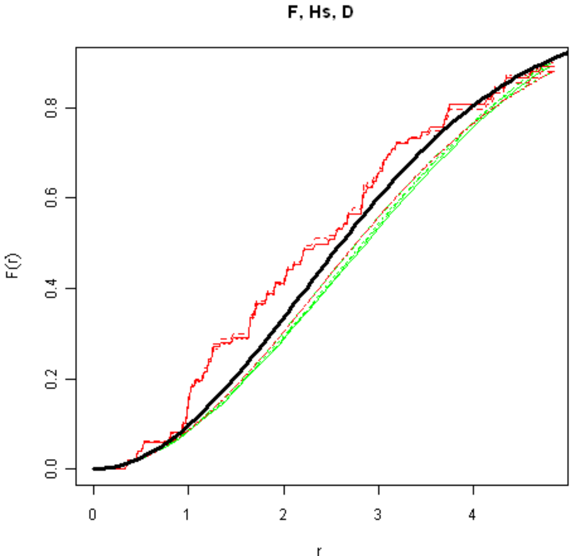
H_0 wird bei sehr kleinen oder großen Werten von h_F verworfen, wobei

$$h_F = \frac{\sum_{i=1}^m u_i^2}{\sum_{i=1}^m v_i^2} \sim F_{2m,2m}$$

Das heißt, H_0 wird verworfen, falls

$$h_F < F_{2m,2m,\frac{\alpha}{2}} \text{ oder } h_F > F_{2m,2m,1-\frac{\alpha}{2}}$$

Vergleich der Methoden, wobei $F = 1 - e^{-\hat{\lambda}\pi r^2}$, $\hat{H}_S$ grün, $\hat{D}$ rot



VIELEN DANK

FÜR DIE AUFMERKSAMKEIT

Quellenverzeichnis:

1. D. Stoyan, W.S. Kendall and J. Mecke,
   Stochastic Geometry and Its Applications, 2nd ed.,
   Wiley & Sons, Chichester, 1995.
2. V. Schmidt, *Räumliche Statistik für Punktprozesse und weitere Modelle der stochastischen Geometrie*, Ulm, 2008.
3. D.R. Cox, *Some Simple Approximate Tests for Poisson Variates*,
   Biometrika PDF-Version COX 40 (3.4):354.(1953), 1953.
4. K. Byth, B.D. Ripley,
   On Sampling Spatial Patterns by Distance Methods,
   Biometrics 36, 279-284, 1980.

Maßband argiope.spinnen-forum.de/mv.mitropa/mitropa.html

Institut http://www.uni-ulm.de/mawi/mawi-stochastik.html

Sterne http://depts.washington.edu/probab/research.php