Statistische Methoden für planare Punktfelder

Franz Király
Universität Ulm
Seminar Simulation und Bildanalyse mit Java

27. Januar 2004

Gliederung

1. Einführung und Grundlagen

2. Intensitätsschätzung
   - Intensität
   - Schätzer für die Intensität

3. Schätzung von Charakteristiken zweiter Ordnung
   - K- und $\mathcal{K}$-Funktion
   - L-Funktion

4. Schätzung weiterer Charakteristiken
   - Paarkorrelationsfunktion
   - sphärische Kontaktverteilungsfunktion
   - Nächste-Nachbar-Abstands-Verteilungsfunktion
   - J-Funktion

1.1 Motivation

- statistische Untersuchung von planaren Punktmustern, z.B.:
  - Untersuchung der Verteilung von Pflanzen
  - Untersuchung von Versicherungsdaten
  - Untersuchung von porösen Materialien
- Statistische Aussagen über den das vorliegende Punktmuster am besten modellierenden Punktprozess an Hand der geschätzten Charakteristiken
- Statistische Beschreibung und Charakterisierung eines Punktmusters

1.2 Grundlagen

- Betrachtung eines **Beobachtungsfensters** $W \in \mathbb{R}^2$ kompakt, beschränkt (für planare Punktprozesse)

- Beobachtungsfenster im allgemeinen Fall abgeschlossenes Rechteck

- Verwendung der Punkte $x \in W$ als Input zur Schätzung bestimmter definierter Charakteristiken

Poisson-Prozess Cluster-Prozess HardCore-Prozess

1.3 Beispiele von Punktmustern

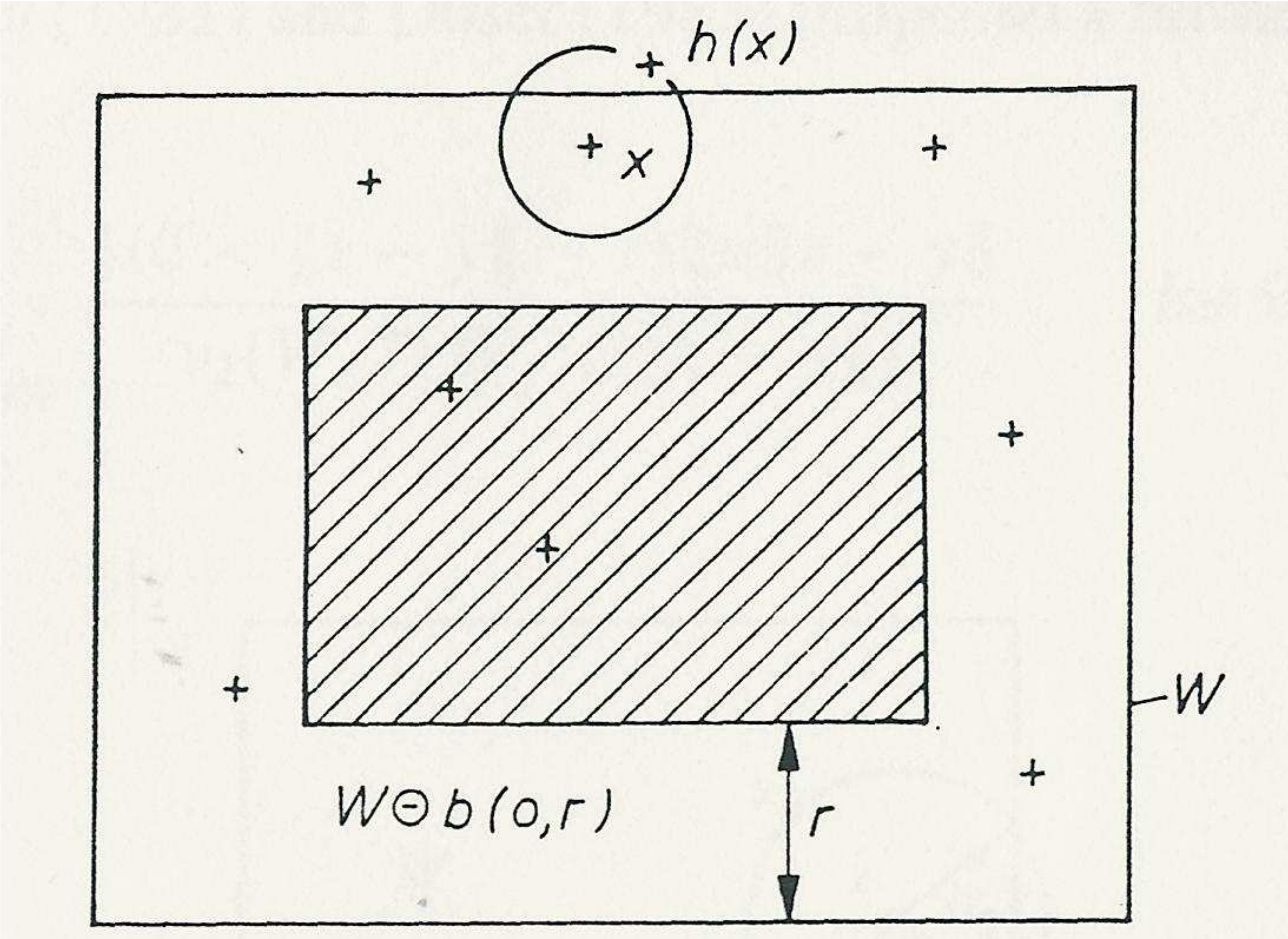
1.4 Minus-Sampling

- Schätzung mancher Charakteristika erfordert das Wissen über Punkte in der Umgebung (z.B. Kreis mit Radius r) von Punkten $x \in W$

- **Problem** für Punkte am Rand des Beobachtungsfensters: gesamte Umgebung nicht im Beobachtungsfenster enthalten, d.h. **Randeffekte** können auftreten

- **Minus-Sampling**: betrachte das erodierte Fenster $\widetilde{W} := W \ominus b(o, r)$

- Mit Minus-Sampling werden die Randeffekte kompensiert.



1.4 Minus-Sampling: Beispiel

2. Intensitätsschätzung

- **Basisschritt** bei der Untersuchung von Punktmustern

- Intensität ist die erwartete Anzahl von Punkten des Prozesses pro Fläche

- Intensität λ: gewisse Analogie zum Erwartungswert einer ZV

- Erlaubt Aussagen über das erste Moment

- Intensitätsmaß im homogenen Fall uninteressant

2. Schätzer für λ

- Sei Φ Realisierung eines Punktprozesses mit $n = \Phi(W)$
- $\hat{\lambda} := \frac{n}{|W|}$ erwartungstreu für λ
- In der Praxis werden aber auch Schätzer für λ^2 benötigt.
- $\widehat{\lambda^2} := \frac{n(n-1)}{|W|^2}$ erwartungstreu für Poisson-Verteilung (Verwendung auch im allgemeinen Fall).

3. Charakteristika 2-ter Ordnung

- Erlauben Aussagen über Varianz und Kovarianz und zweite Momente

- $\lambda\mathcal{K}(B) := \mathbb{E}(\Phi(B \setminus \{0\}) \parallel 0)$

- $K(r) := \mathcal{K}(b(o,r)) \quad , r \geq 0$

- $\lambda K(r)$ ist die erwartete Anzahl von Punkten von Φ in $b(o,r)$ ohne den Punkt o, falls o als ein Punkt der Realisierung vorausgesetzt wird.

- $K(r)$ ist die entsprechede normierte Größe.

- $L(r) := \sqrt{\frac{K(r)}{\pi}}, \quad r \geq 0$

- $L(r)$ gibt über die Verteilung bezogen auf den Radius Auskunft.

3.1 Schätzer für $K(r)$

- Betrachten wir den Prozess Φ mit $\Phi(W) = n$ und $X_i \in W$, $1 \leq i \leq n$ den entsprechenden Punkten.

- Einfaches Verfahren zur Konstruktion von Schätzern: Ersetze erwartete Größen durch entsprechend gemessene.

- Schätze $\lambda K(r)$ mit
  $\kappa_1(r) = \frac{1}{n} \sum_{i=1}^{n} \Phi(b(X_i, r) \setminus \{X_i\}) \quad , r \geq 0$

- Problem: Ein solcher Schätzer ist im Allgemeinen nicht erwartungstreu.

3.1 Schätzer für $K(r)$

- Lösung: Schätze zunächst $\lambda^2 K(r)$ und dann λ^2 erwartungstreu.

- Berechne dann $\frac{\widehat{\lambda^2 K(r)}}{\widehat{\lambda^2}}$.

- Voraussetzungen wie zuvor (Φ, n)

- $\kappa_2(r) = \sum\limits_{i=1}^{n} \frac{\Phi(b(X_i,r)\setminus\{X_i\})}{|W|} \quad , r \geq 0$ ist erwartungstreu

- Problem: was passiert mit den Punkten $x \in W : d(x, \partial W) < r$?

3.1 Schätzer für $K(r)$

- Idee: betrachte nur die Punkte mit $x \in W : d(x, \partial W) \geq r$

- Dies sind genau die Punkte $x \in \widetilde{W} := W \ominus b(o, r)$.

- Die vorliegende Situation ist ein typischer Anwendungsbereich für Minus-Sampling.

- $\kappa_3(r) = \sum\limits_{x \in \widetilde{W}} \frac{\Phi(b(x,r) \setminus \{x\})}{|\widetilde{W}|}, \quad r \geq 0$ erwartungstreu

- Problem: Informationsverlust für große r bzw. kleine W

3.1 Schätzer für $K(r)$

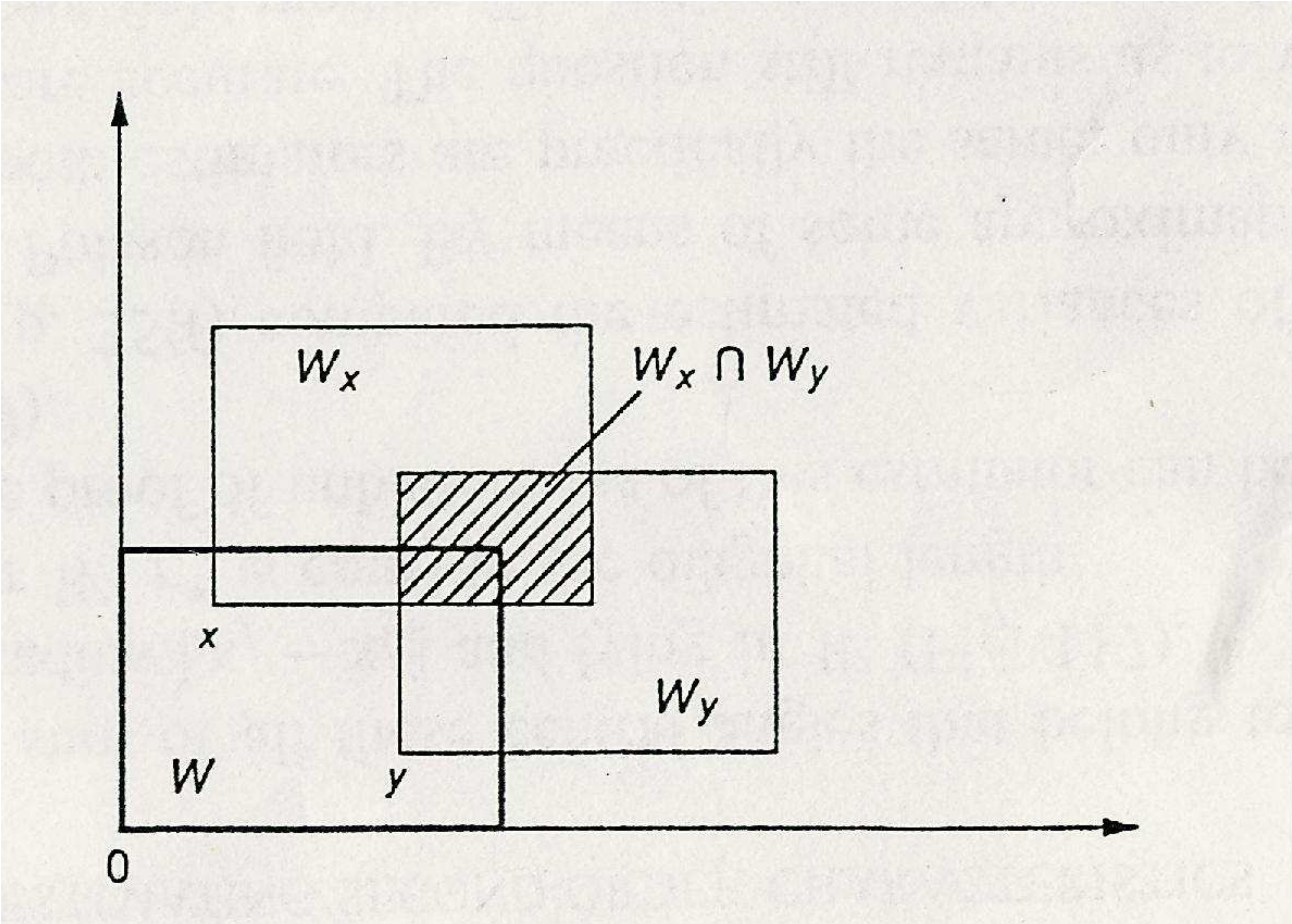
- Wir leiten einen weiteren Schätzer aus $\kappa_2(r) = \sum\limits_{i=1}^{n} \frac{\Phi(b(X_i,r)\setminus\{X_i\})}{|W|}$ her:

- $\kappa_2(r) = \sum\limits_{i=1}^{n} \sum\limits_{\substack{j=1 \\ j\neq i}}^{n} \frac{1_{b(x_i,r)}(x_j)}{|W|} = \sum\limits_{i=1}^{n} \sum\limits_{\substack{j=1 \\ j\neq i}}^{n} \frac{1_{b(o,r)}(x_j - x_i)}{|W|}$

- Ersetze W durch $W_{x_j} \cap W_{x_i}$:

- $\kappa(r) = \sum\limits_{i=1}^{n} \sum\limits_{\substack{j=1 \\ j\neq i}}^{n} \frac{1_{b(o,r)}(x_j - x_i)}{|W_{x_j} \cap W_{x_i}|}$

- Es gilt Erwartungstreue für $\widehat{\lambda^2 K(r)}$.



Verdeutlichung der Situation bei κ_3

3.2 Schätzer für $\mathcal{K}(B)$

- Es gilt offensichtlich $K(r) = \mathcal{K}(b(o,r))$.

- Konstruiere Schätzer in Analogie zu $K(r)$ für beliebige Borel-Mengen B.

- $\kappa(B) = \sum\limits_{i=1}^{n} \sum\limits_{\substack{j=1 \\ j\neq i}}^{n} \frac{1_B(x_j - x_i)}{|W_{x_j} \cap W_{x_i}|}$

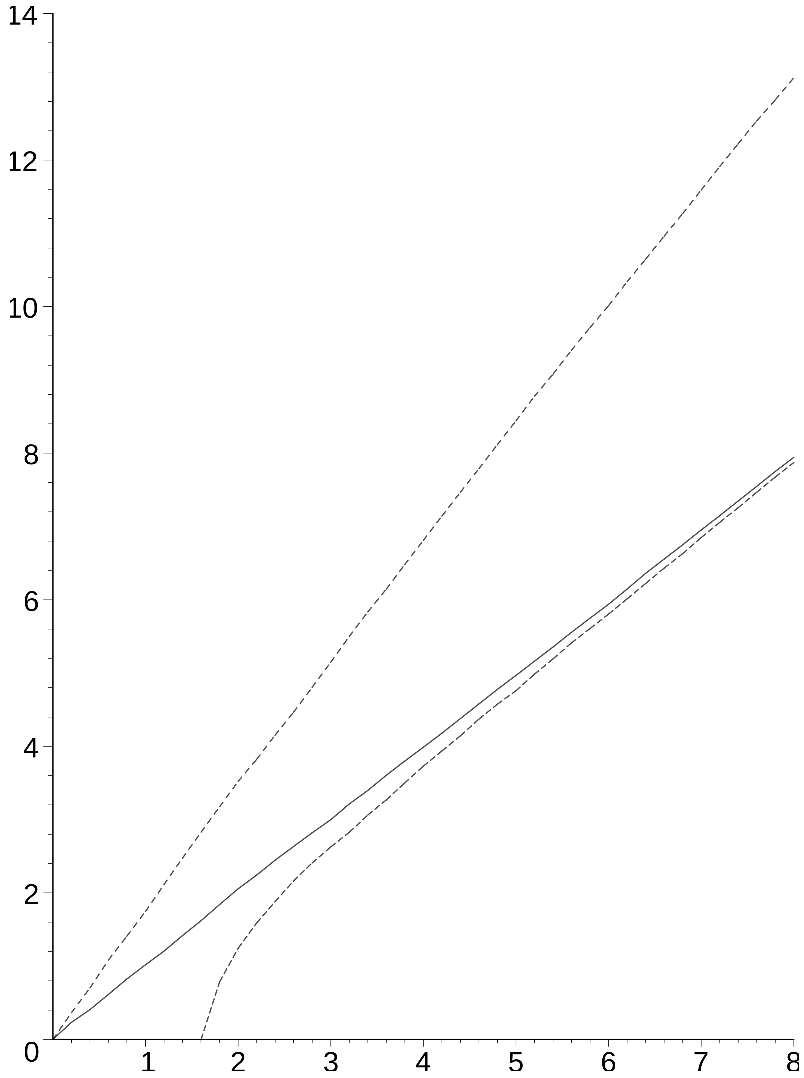
- Erwartungstreue für $\kappa(B)$ folgt analog wie bei κ_3.

3.3 Schätzer für $L(r)$

- Erinnerung:
  $L(r) := \sqrt{\frac{K(r)}{\pi}} \quad , r \geq 0.$

- Aus der Struktur folgt Verwendbarkeit der Schätzer für $K(r)$.

- Schätze also $\widehat{L}(r) = \sqrt{\frac{\widehat{K}(r)}{\pi}} \quad , r \geq 0.$



Anwendung des L-Schätzers $\widehat{L}$

4. weitere Charakteristika

- Paarkorrelationsfunktion $g(r)$
- Sphärische Kontaktverteilungsfunktion $H_s(r)$
- Nächste-Nachbar-Abstands-Verteilungsfunktion $D(r)$
- J-Funktion

4.1 Schätzverfahren für die Paarkorrelationsfunktion $g(r)$

- $g(r) := \frac{\rho^{(2)}(r)}{\lambda^2}$ ist normalisierte, infinitesimale Auftretenswahrscheinlichkeit von Punkten eines Punktprozesses an um r voneinander entfernten, festen Stellen.

- Berechne zunächst Schätzer für Produktdichte $\rho^{(2)}(r)$ und λ^2

- Teile dann durch Schätzer für λ^2

- $\widehat{g}(r) = \frac{\widehat{\rho}^{(2)}(r)}{\widehat{\lambda^2}}$

4.1.1 Schätzer für $\rho^{(2)}(r)$

- häufig angewandtes statistisches Verfahren: Kernschätzer
- im isotropen Fall: $\widehat{\rho}(r) = \frac{1}{2\pi r} \sum\limits_{i=1}^{n} \sum\limits_{\substack{j=1 \\ j\neq i}}^{n} \frac{k_h(r-d(x_j,x_i))}{|W_{x_j} \cap W_{x_i}|}$
- dabei ist $k_h(t)$ die sogenannte Kernfunktion

4.1.2 Kernfunktionen

- Verwendung verschiedener Kernfunktionen, z.B.:

- Epanetschnikoff-Kern: $e_h(t) := 1_{[-h,h]}(t)\frac{3}{4h}(1 - \frac{t^2}{h^2})$

- Rosenblatt-Kern: $r_h(t) := 1_{[-\sqrt{5}h,\sqrt{5}h]}(t)\frac{1}{h}(1 - \frac{t^2}{5h^2})\frac{3}{4\sqrt{5}}$

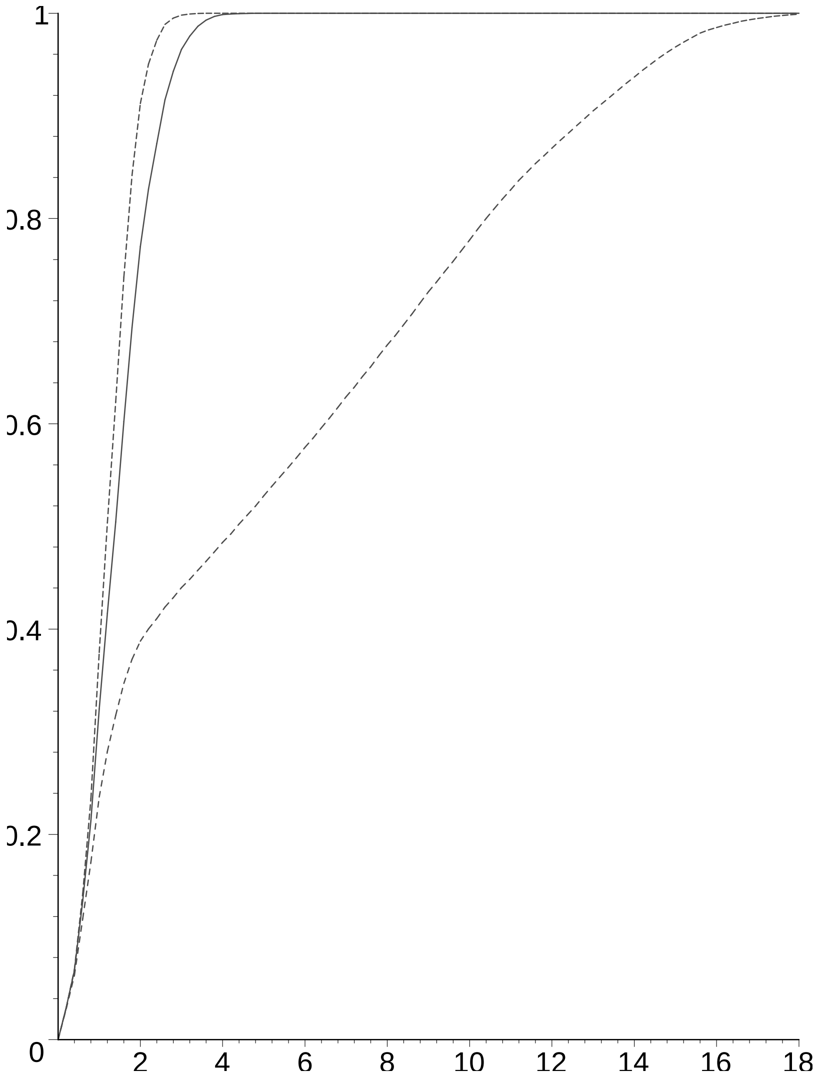
4.2 sphärische Kontaktverteilungsfunktion $H_s(r)$

- $H_s(r)$ ist die Verteilungsfunktion, die Auskunft über den Abstand des nächstgelegenen Punktes von Φ vom Ursprung gibt

- $H_s(r) := 1 - \mathbb{P}(\Phi(b(o,r)) = 0), \quad r \geq 0$

- Idee: betrachte den relativen Anteil der Fläche im Beobachtungsfenster, der dadurch entsteht, wenn man um jeden Punkt von Φ einen Kreis mit Radis r legt.

- $\widehat{H}_s(r) = \frac{|W \ominus b(o,r) \cap \bigcup_{x \in \Phi} b(x,r)|}{|W \ominus b(o,r)|}, \quad r \geq 0$ erwartungstreu



Anwendung des Schätzers $\widehat{H}_s$

4.3 Nächste-Nachbar-Abstands-Verteilungsfunktion

- Nächste Nachbar-Entfernung
  $\delta(x_i) := \min_{x \in \Phi} |x_i - x|$

- Sei Ψ_r der Teilpunktprozess, für den jeder Punkt einen nahen Nachbarn besitzt, d.h. $\delta(x) \leq r \wedge x \in \Phi$ gdw. $x \in \Psi_r$

- Idee: betrachte den Teil der Punkte im minus-gesampelten Beobachtungsfenster, die mindestens einen Nachbarn besitzen, der näher liegt als r.

- $\widehat{D}_1(r) = \frac{\Psi_r(W \ominus b(o,r))}{\Phi(W \ominus b(o,r))} \quad , r \geq 0$

- $\widehat{D}_1(r)$ ist lediglich asymptotisch erwartungstreu

4.3 Schätzer für $D(r)$

- Idee: betrachte zunächst den Schätzer $\widehat{\mathcal{D}}_2(r)$:

- $\widehat{\mathcal{D}}_2(r) := \sum\limits_{x \in \Psi_r} \frac{1_{W \ominus b(o,r)}(x)}{|W \ominus b(o,r)|} \quad , r \geq 0$

- $\widehat{\mathcal{D}}_2(r)$ ist ein erwartungstreuer Schätzer für $\lambda D(r)$.

- Konstruiere also $\widehat{D}_2(r)$ mit $\widehat{\mathcal{D}}_2(r) = \underbrace{\dfrac{\Psi_r(W \ominus b(o,r))}{|W \ominus b(o,r)|}}_{=:\widehat{\lambda}} \widehat{D}_2(r)$.

- Schätze also $D(r)$ mit $\widehat{D}_2(r) := \frac{\widehat{\mathcal{D}}_2(r)}{\widehat{\lambda}}$

- Problem: $\widehat{D}_2(r)$ nicht monoton, nicht strikt < 1

4.3 Schätzer für $D(r)$

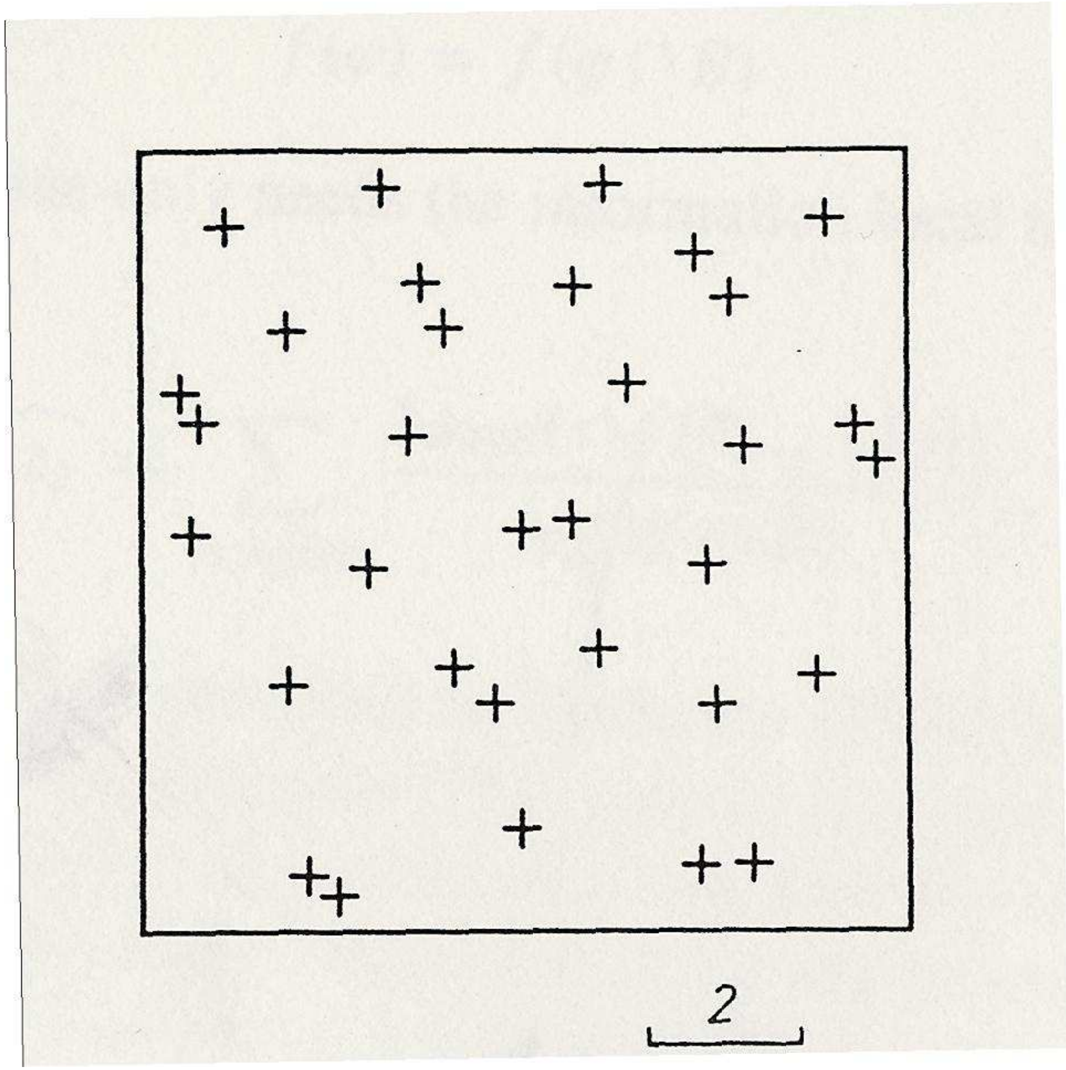
- Guter bekannter Schätzer: Hanisch-Schätzer

- Betrachte dazu
  $\widehat{\mathcal{D}}_H := \sum_{x \in \Psi_r} \frac{1_{W \ominus b(o,\delta(x))}(x)}{|W \ominus b(o,\delta(x))|}$ als Schätzer für $\lambda D(r)$.

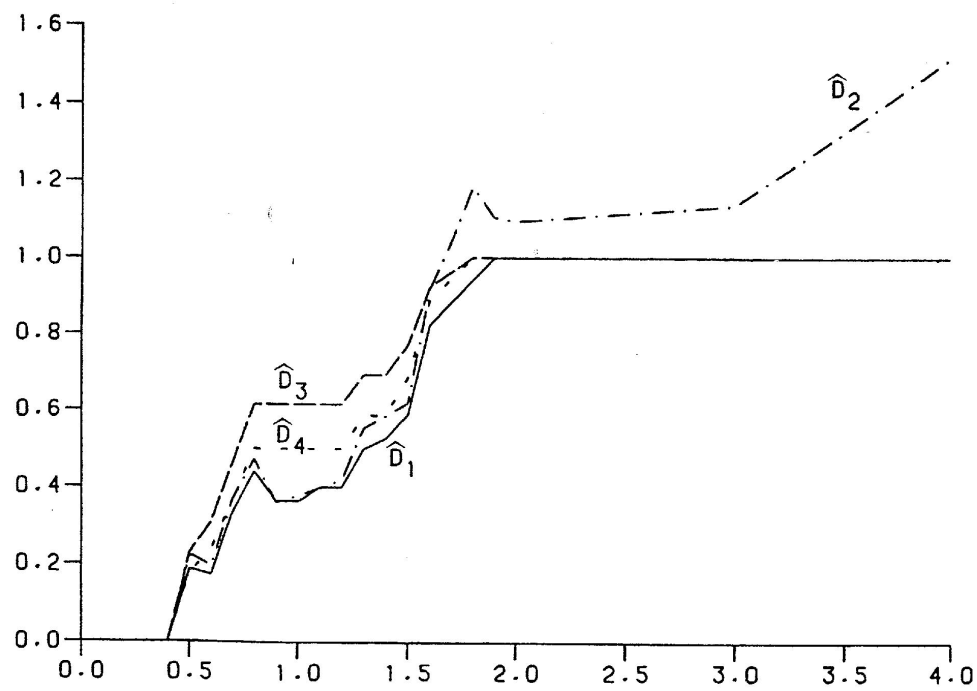
- Schätzung von λ mit den gleichen Daten:

- $\widehat{\lambda}_H := \sum_{x \in \Psi_r} \frac{1_{W \ominus b(o,\delta(x))}(x)}{|W \ominus b(o,\delta(x))|}$

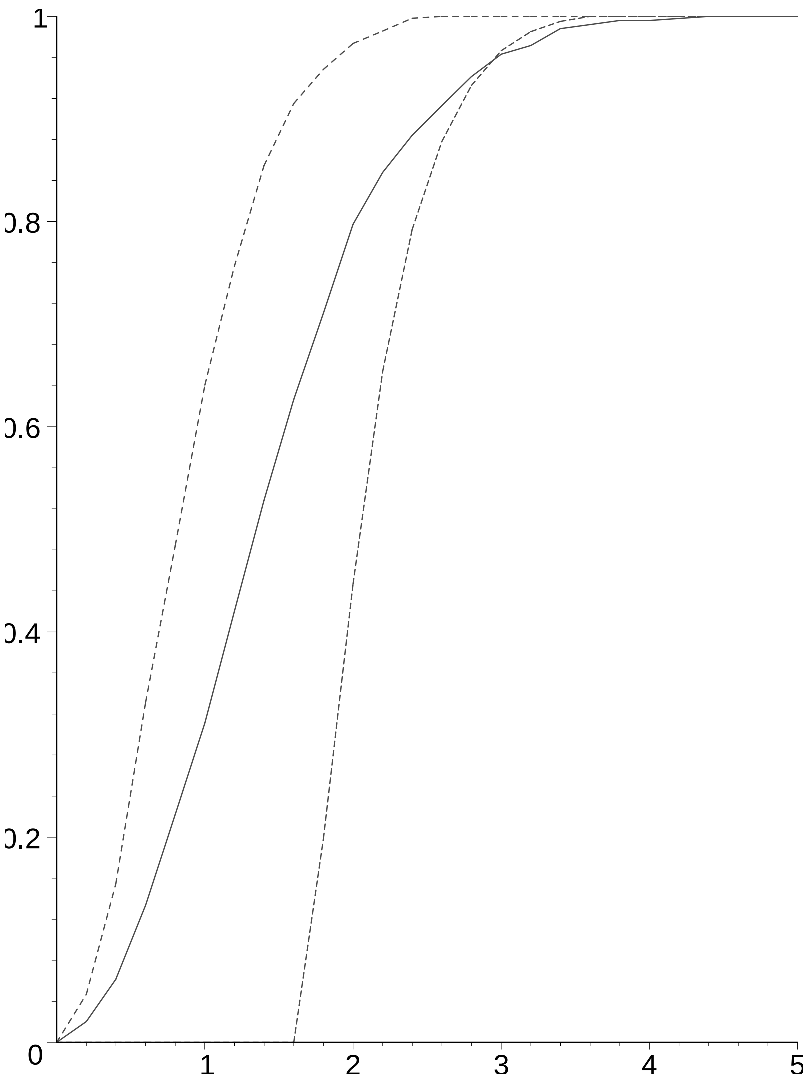
- $\frac{\widehat{\mathcal{D}}(r)}{\widehat{\lambda}}$ ist ein erwartungstreuer Schätzer für $D(r)$.



Punktprozess



Vergleich der Schätzer für $D(r)$



Anwendung des Hanisch-Schätzers $\widehat{D}_h(x)$

4.4 Schätzer für die J-Funktion

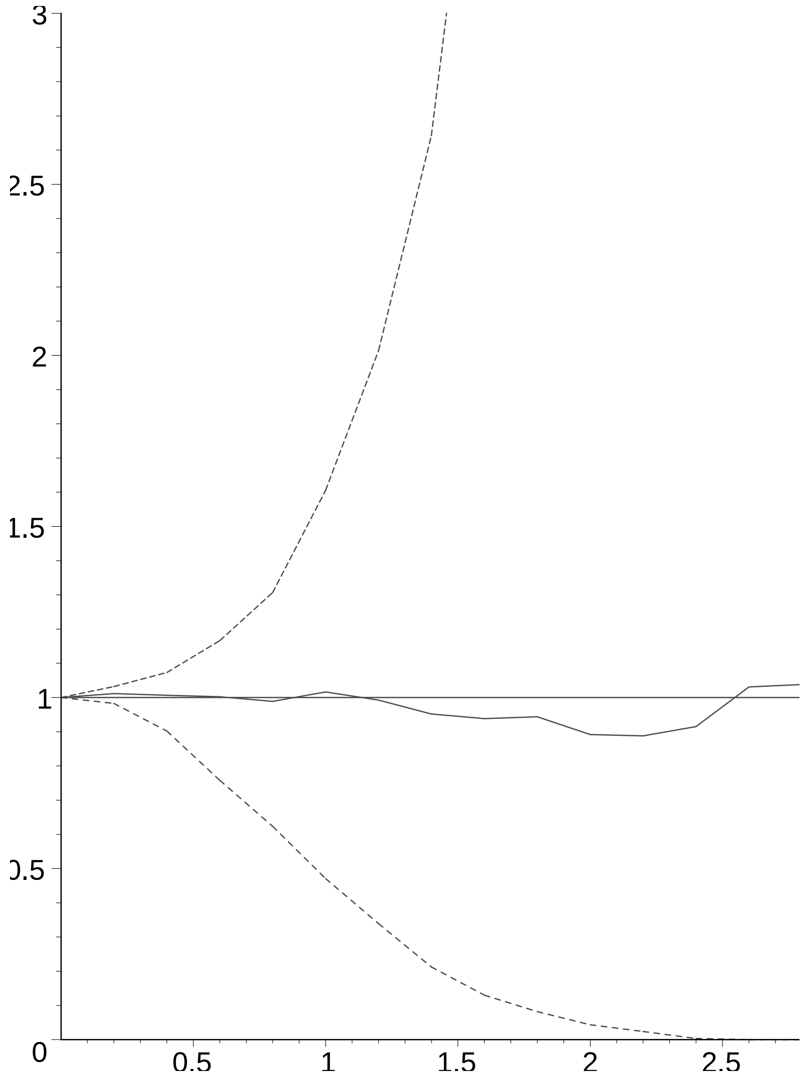
- Erinnerung: $J(r) := \frac{1-D(r)}{1-H_s(r)}$ mit $r \geq 0, H_s(r) < 1$

- J-Funktion gibt Aufschluss über Anordnung der Realisierung

- Verfahren wie bei L-Funktion: ersetze Größen durch Schätzer:

- $\hat{J}(r) := \frac{1-\hat{D}(r)}{1-\hat{H}_s(r)}$

- Güte der erfolgten Schätzung von $\hat{J}(r)$ hängt von $\hat{D}(r)$ und $\hat{H}_s(r)$ ab



Schätzung von J-Funnktionen

5. Zusammenfassung

- Schätzer liefern Information über das **vorliegende Punktmuster**

- Schätzer unterscheiden sich in ihrer Praktikabilität und ihren stochstischen Eigenschaften

- Meist kann ein Punktprozessmodell an Hand der Berechnung mehrerer Charakteristiken identifiziert werden, welches das gegebene Punktmuster “gut” beschreibt

Literatur

[1] Stoyan, Kendall, Mecke: Stochastic Geometry and its Applications, 2nd Ed.
Wiley & Sons, 1995

[2] Stoyan, Stoyan: Fractals, Random Shapes and Point Fields, Methods of Geometrical Statistics
Wiley & Sons, 1994

[3] Hanisch: Some remarks on estimators of the distance function of nearest-neighbour distance in stationary spatial point-patterns
Math. Operationsf. Statist., ss. statistics, ser. statistics 15, 409-412, 1994