Statistik für Punktprozesse

Seminar „Stochastische Geometrie und ihre Anwendungen“

WS 2009/2010

Inhalt

Inhalt

I. **Fragestellung / Problematik**

II. Ansätze für…

   a) … die Schätzung der Intensität

   b) … ein Testverfahren auf Stationarität

   c) … das Testen von Prozess-Typen

   d) … Test der Poissonannahmen

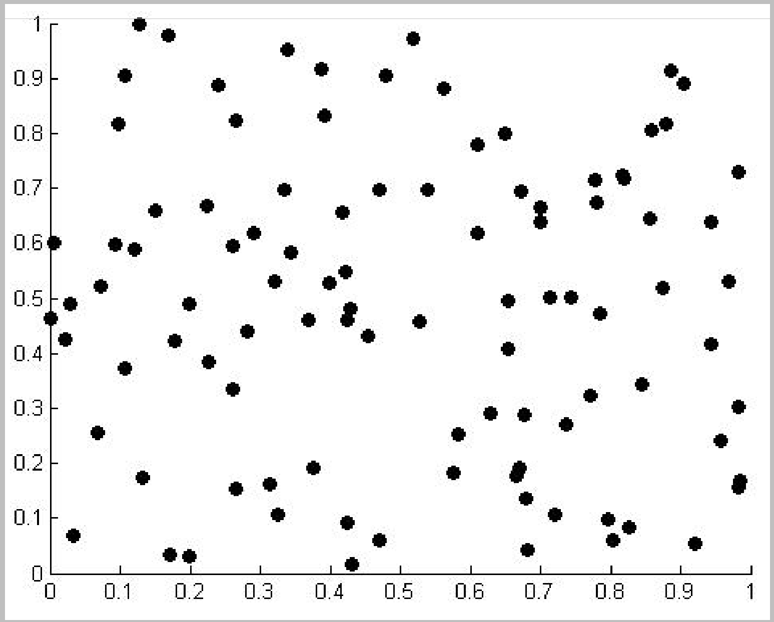
III. Zusammenfassung

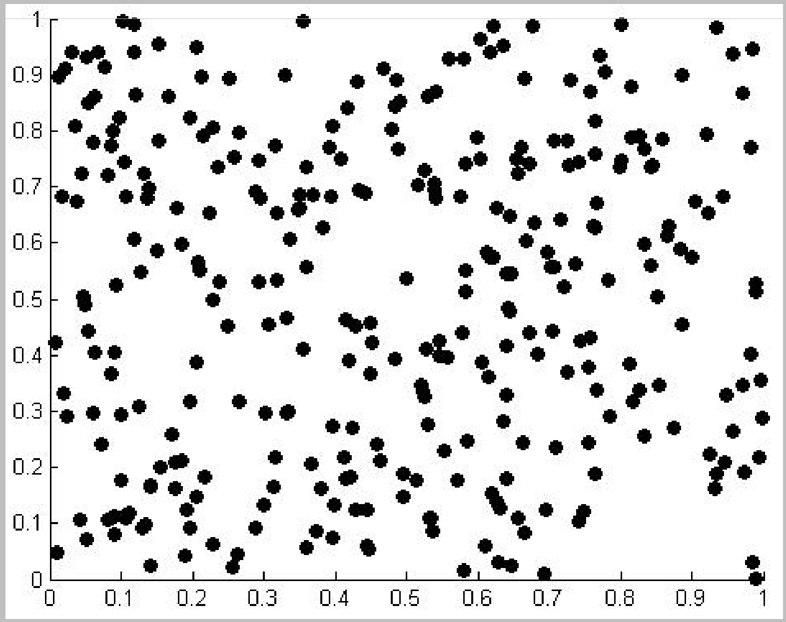
I. Fragestellung / Problematik

Gegeben: Punktmuster

Mögliche Interessen:

– Wie hoch ist die Intensität λ?



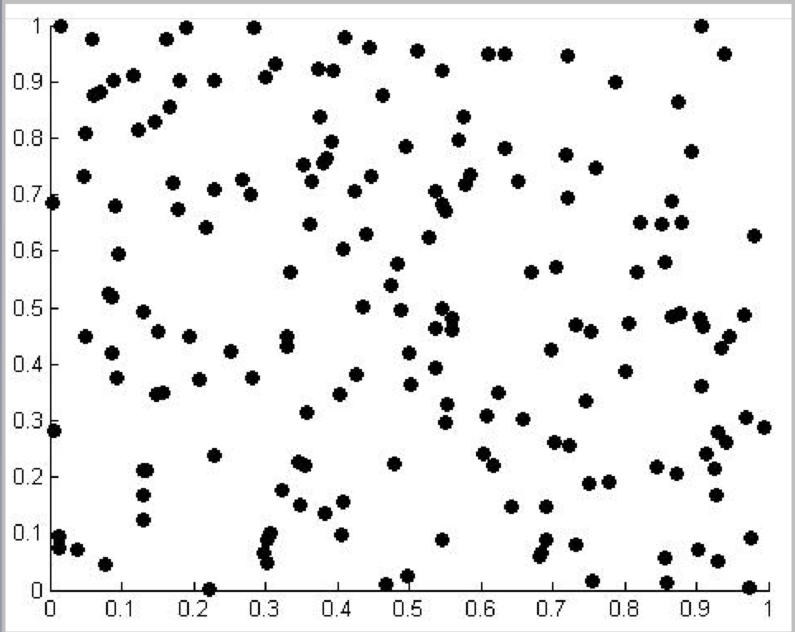


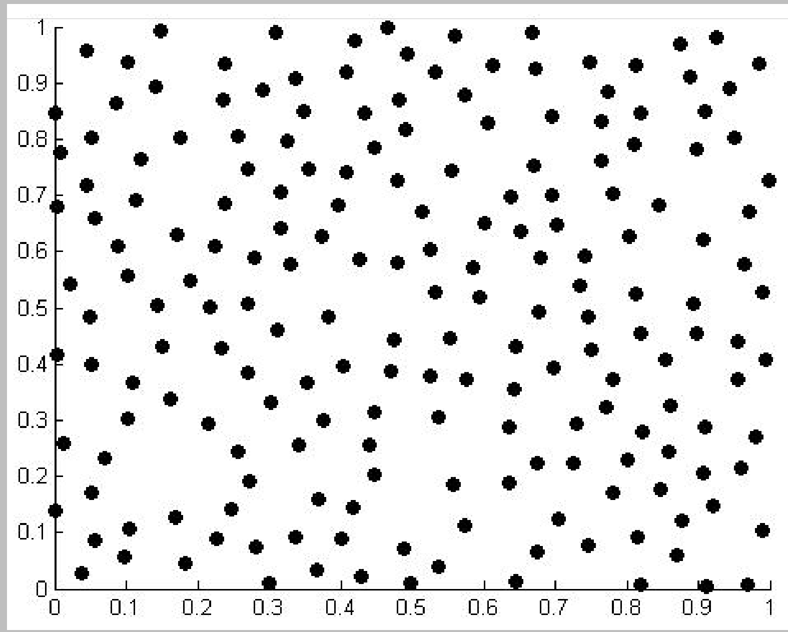
I. Fragestellung / Problematik

Gegeben: Punktmuster

Mögliche Interessen:

– Ist das Punktmuster vom Poisson Typ?



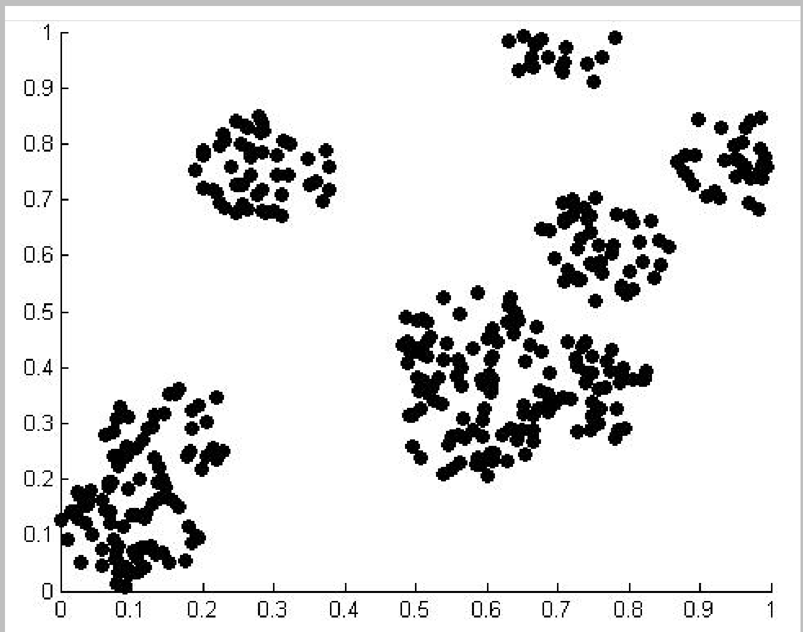


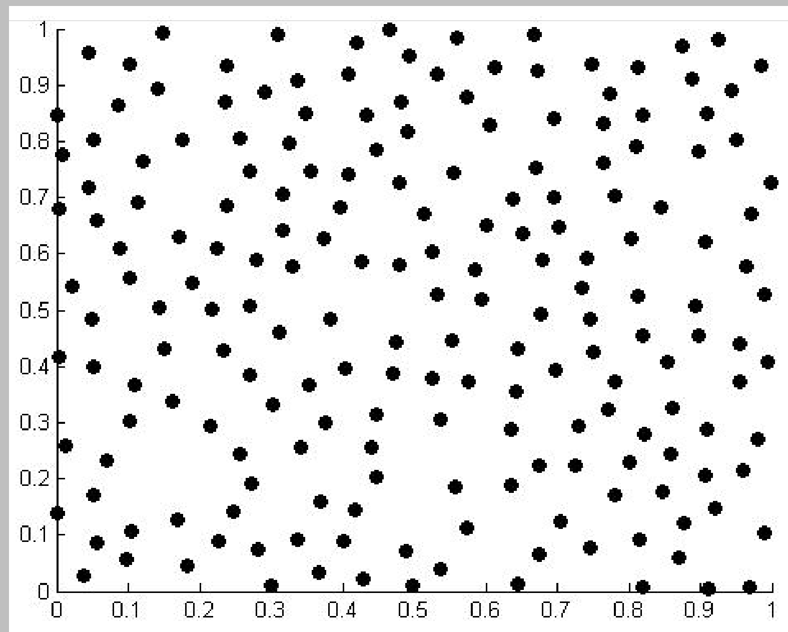
I. Fragestellung / Problematik

Gegeben: Punktmuster

Mögliche Interessen:

– Ist das Punktmuster von einem anderen Typ?



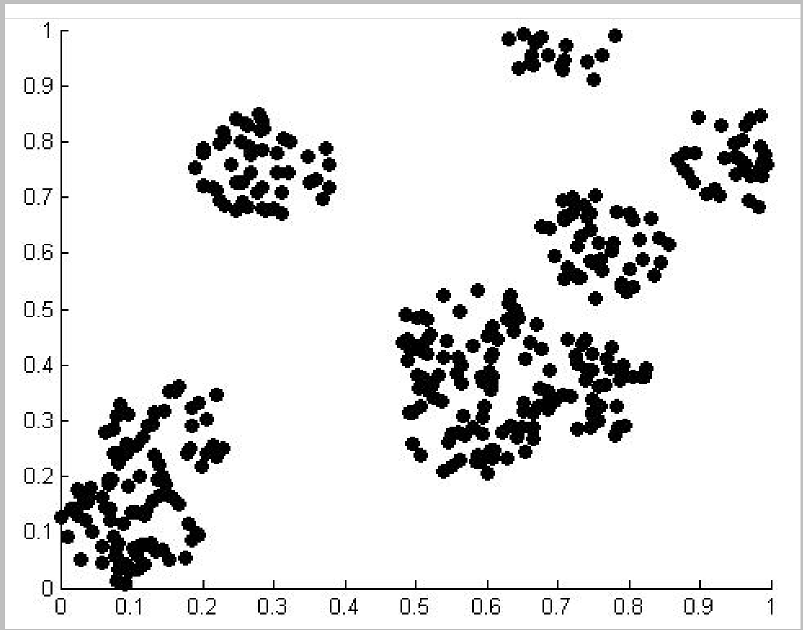


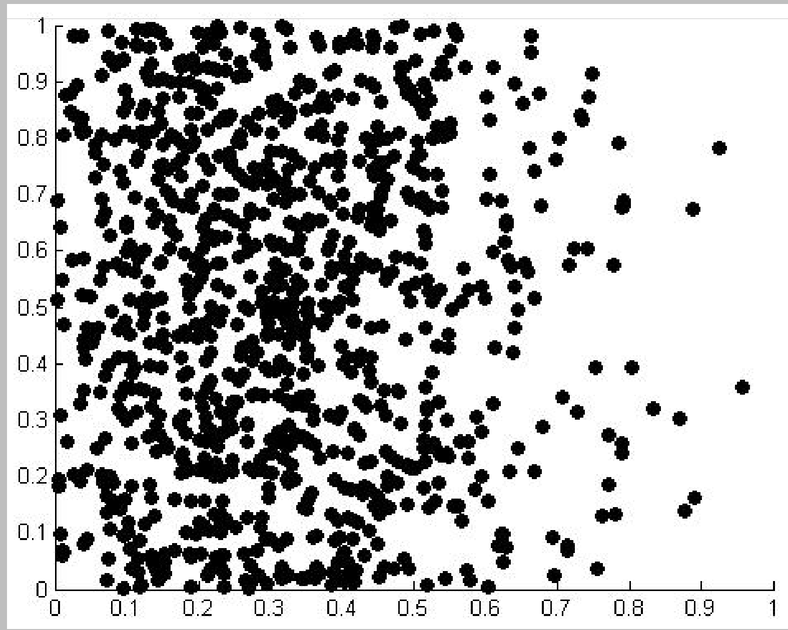
I. Fragestellung / Problematik

Gegeben: Punktmuster

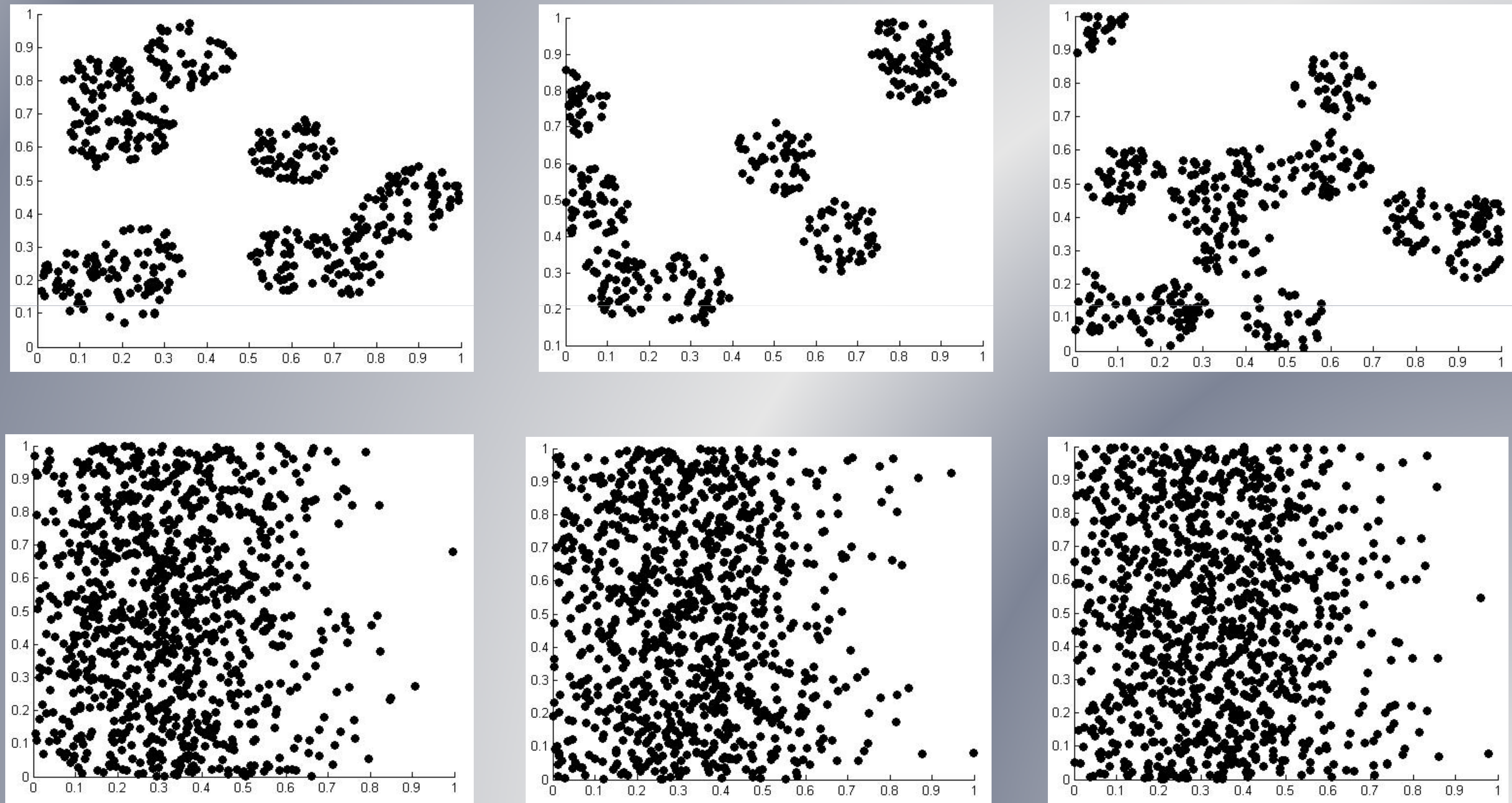
Mögliche Interessen:

– Ist das Punktmuster stationär?





I. Fragestellung / Problematik



Inhalt

I. Fragestellung / Problematik

II. Ansätze für…

- **a) … die Schätzung der Intensität**
- b) … ein Testverfahren auf Stationarität
- c) … das Testen von Prozess-Typen
- d) … Test der Poissonannahmen

III. Zusammenfassung

II. Ansätze für…

a) … die Schätzung der Intensität

- „Natürlicher Ansatz“ $\hat{\lambda}_W = \dfrac{\Phi(W)}{\nu_d(W)}$
- „Leere Quadrate“ Methode

II. Ansätze für…

a) … die Schätzung der Intensität

- „Natürlicher Ansatz“ $\hat{\lambda}_W = \dfrac{\Phi(W)}{\nu_d(W)}$

Eigenschaften (im Poisson Fall):

- Erwartungstreu:

$$\mathrm{E}\left(\hat{\lambda}_W\right) = \mathrm{E}\left(\frac{\Phi(W)}{\nu_d(W)}\right) = \frac{\mathrm{E}(\Phi(W))}{\nu_d(W)} = \frac{\lambda \cdot \nu_d(W)}{\nu_d(W)} = \lambda$$

II. Ansätze für…

a) … die Schätzung der Intensität

Falls W_n Folge von Beobachtungsfenstern, $\lim_{n\to\infty} \nu_d(W_n) = \infty$

- Schwach konsistent:

$$\lim_{n\to\infty} \mathrm{P}\left(|\hat{\lambda}_{W_n} - \lambda| > \varepsilon\right) = 0 \quad \forall \varepsilon > 0$$

- Aysmptotisch normalverteilt:

$$\lim_{n\to\infty} \mathrm{P}\left(\sqrt{\frac{\nu_d(W_n)}{\lambda}} \cdot \left(\hat{\lambda}_{W_n} - \lambda\right) \le x\right) = \Phi(x)$$

Falls zusätzlich $W_1 \subset W_2 \subset \ldots$

- Stark konsistent:

$$\mathrm{P}(\lim_{n\to\infty} \hat{\lambda}_{W_n} = \lambda) = 1$$

II. Ansätze für…

a) … die Schätzung der Intensität

Konfidenzintervall (Approximation):

$$\hat{\lambda}_W = \frac{\Phi(W)}{\nu_d(W)} \Leftrightarrow \hat{\lambda}_W \cdot \nu_d(W) = \Phi(W) \sim Poi(\lambda \cdot \nu_d(W))$$

Damit gilt für große $\Phi(W)$ mit Wahrscheinlichkeit 1-α:

$$\left(\frac{z_{\alpha/2}}{2} - \sqrt{\Phi(W)} \right)^2 \leq \lambda \cdot \nu_d(W) \leq \left(\frac{z_{\alpha/2}}{2} + \sqrt{\Phi(W)+1} \right)^2$$

(siehe Crow&Gardner 1959, Sachs 1984)

II. Ansätze für…

a) … die Schätzung der Intensität

- „Leere Quadrate“ -Methode

Schritt1: Zerlege W in p gleichgroße Quadrate

Schritt2: Sei p* die Anzahl der leeren Quadrate in W und sei

$$p_0 := \frac{p^*}{p}$$

Schritt3: Dann ist ein Schätzer für λ gegeben durch:

$$\hat{\lambda} = -\frac{\log(p_0)}{a^2}$$

Inhalt

II. Ansätze für…
b) …ein Testverfahren auf Stationarität

- wähle 2 disjunkte Beobachtungsfenster W_1 und W_2
- seien n_1 bzw. n_2 die Anzahl der Atome in W_1 bzw. W_2

Testgröße: $$F = \frac{\nu_d(W_1)\cdot(2n_2+1)}{\nu_d(W_2)\cdot(2n_1+1)} \sim F_{2n_1+1,2n_2+1}$$

(o.B.d.A. sei F>1)

Test: H_o: Φ ist stationär vs. H_1: Φ ist nicht stationär

H_o ist abzulehnen mit Signifikanzlevel α, falls

$$F > F_{2n1+1,2n2+1,1-\alpha/2}$$

(siehe Cox 1953)

Inhalt

I. Fragestellung / Problematik

II. Ansätze für…

a) … die Schätzung der Intensität

b) … ein Testverfahren auf Stationarität

c) … das Testen von Prozess-Typen

d) … Test der Poissonannahmen

III. Zusammenfassung

II. Ansätze für…

c) … das Testen anderer Prozess-Typen

Kenngrößen von Punktprozessen sind z.B.:

– K-Funktion (und L-Funktion)

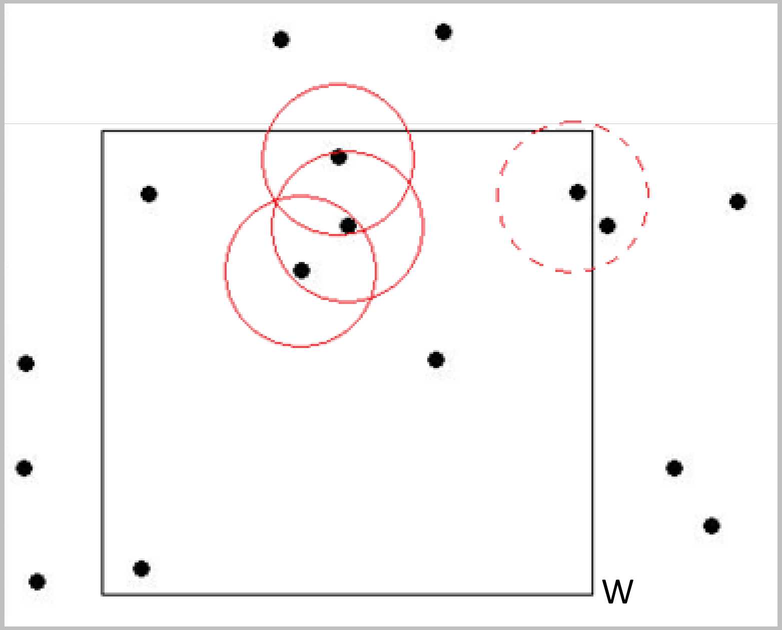
– Sphärische Kontaktverteilungsfunktion $H_s(r)$

– Nächster-Nachbar-Abstandverteilungsfunktion $D(r)$

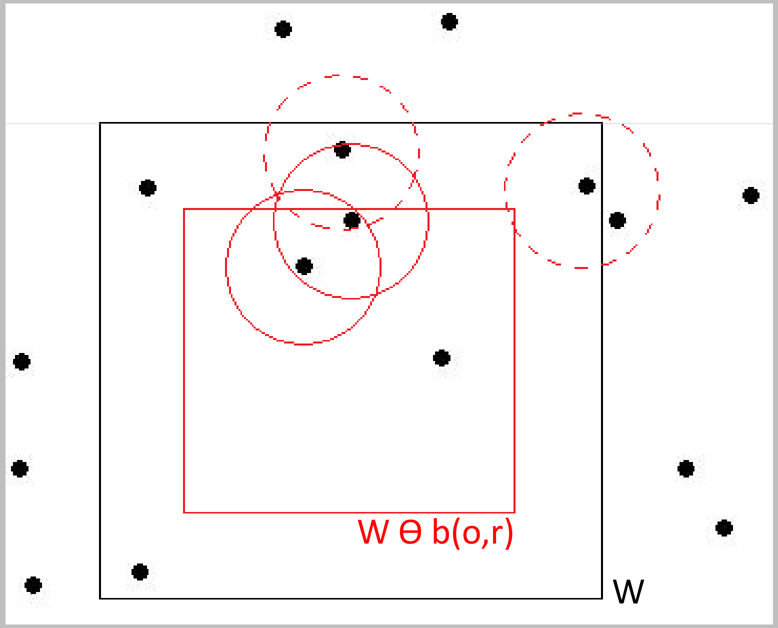
II. Ansätze für…

c) … das Testen anderer Prozess-Typen

Problem:



Minus-Sampling:



II. Ansätze für…

c) … das Testen anderer Prozess-Typen

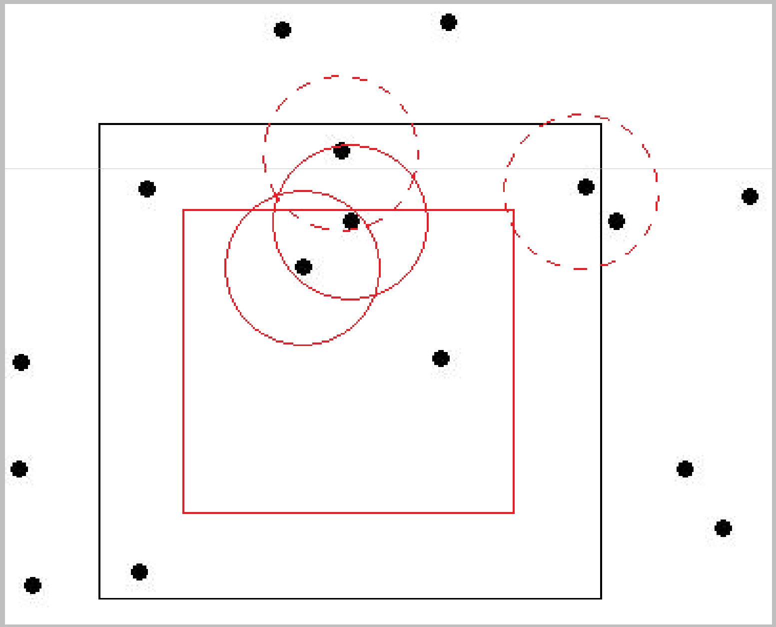
Plus-Sampling vs. Minus-Sampling:

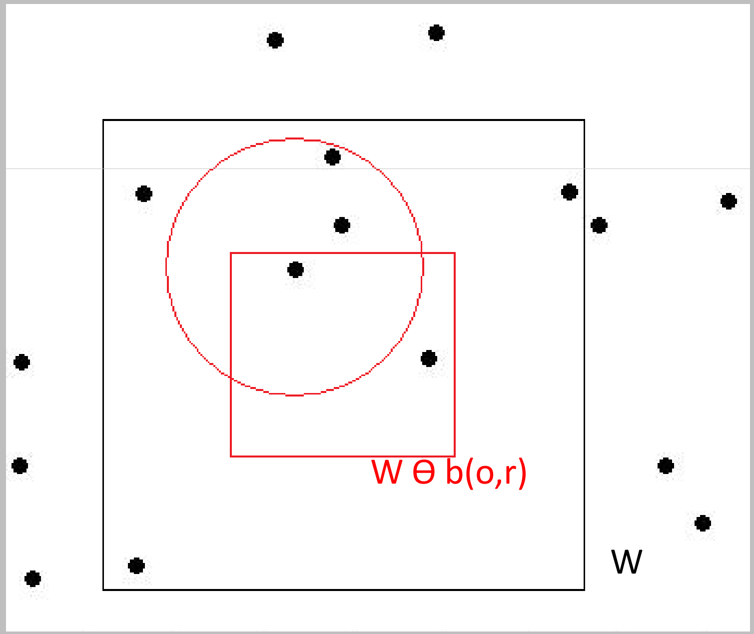
- Plus-Sampling vergrößert das Beobachtungsfenster, benötigt mehr Informationen; es entsteht eine verzerrte Sicht, wenn man die benötigten Zusatzinformationen nicht zur Verfügung hat
- Minus-Sampling verkleinert das Beobachtungsfenster, benötigt weniger Informationen (verwirft aber auch Informationen!)

II. Ansätze für...

c) ... das Testen anderer Prozess-Typen

$\hat{D}(r)$ bei Minus-Sampling nicht monoton in r!





II. Ansätze für…

c) … das Testen anderer Prozess-Typen

<u>K-Funktion</u> (zweites reduziertes Momentenmaß):

$\mathrm{K}(B) := \frac{\mu^!(B)}{\lambda}$ wobei $\mu^!(B)$ das Intensitätsmaß der reduzierten Palmschen Verteilung ist;

<u>Ripleysche K-Funktion</u>: $K(r) := \mathrm{K}(b(o,r))$

Und aus Slivnyak‘s Theorem folgt:

$$K(r) = \frac{\mu^!(b(o,r))}{\lambda} \overset{Slivniak}{=} \frac{\mu(b(o,r))}{\lambda} = \pi \cdot r^2$$

II. Ansätze für…

c) … das Testen anderer Prozess-Typen

Schätzer für die Ripleysche K-Funktion:

$$\lambda^2 \hat{K}(r) = \sum_{n_1, n_2}^{\neq} \frac{1_{b(o,r)}(S_{n_2} - S_{n_1}) \cdot 1_{W \times W}(S_{n_1}, S_{n_2})}{\nu_2(W \cap (W + (S_{n_2} - S_{n_1})))}$$

Beziehungsweise:

$$\hat{K}(r) = \frac{\lambda^2 \hat{K}(r)}{\hat{\lambda}^2} \quad \text{mit} \quad \hat{\lambda}^2 = \frac{\Phi(W) \cdot (\Phi(W) - 1)}{(\nu_2(W))^2}$$

II. Ansätze für…

c) … das Testen anderer Prozess-Typen

Sphärische Kontaktverteilungsfunktion:

$$H_s(r) := P(Z_x \le r) = P\left(\min\left\{r^0 \ge 0 : \Phi(b(x, r^0)) > 0\right\} \le r\right)$$

$$= P\left(o \in \bigcup_{x \in \Phi} b(x, r)\right) \overset{Poisson}{=} 1 - e^{-\lambda \cdot \pi \cdot r^2}$$

Somit ergibt sich als Schätzer:

$$\hat{H}_s(r) = \frac{\nu_2\left((W \ominus b(o, r)) \cap \bigcup_{x \in \Phi} b(x, r) \right)}{\nu_2(W \ominus b(o, r))}$$

II. Ansätze für…

c) … das Testen anderer Prozess-Typen

Nächster-Nachbar-Abstandverteilungsfunktion:

$$D(r) = 1 - \mathrm{P}_0^!\left(\varphi \in \mathrm{N} : \varphi(b(o,r)) = 0\right)$$

$$= \frac{1}{\lambda \cdot \nu_2(W)} \cdot \mathrm{E} \sum_{n=1}^{\infty} 1_{\left(S_n \in W, \min\limits_{n \neq m} |S_n - S_m| \leq r\right)} \overset{Poisson}{=} 1 - e^{-\lambda \cdot \pi \cdot r^2}$$

Somit ergibt sich der Schätzer:

$$\hat{D}(r) = \frac{1}{\#\{n : S_n \in (W \ominus b(o,r))\}} \cdot \sum_{n=1}^{\infty} 1_{\left(S_n \in (W \ominus b(o,r)), \min\limits_{n \neq m} |S_n - S_m| \leq r\right)}$$

II. Ansätze für…

c) … das Testen anderer Prozess-Typen

Konstruktion von Tests:

- Getestet werden soll (mit Hilfe der eingeführten Kenngrößen/Schätzer), ob ein Punktmuster zu einer bestimmten Art von Punktprozess gehört
- Problem: Tatsächliche Verteilung von Testgrößen zu diesen Kenngrößen ist schwierig zu ermitteln
- Lösung: Monte Carlo Simulation

II. Ansätze für…

c) … das Testen anderer Prozess-Typen

Monte Carlo Simulation:

Schritt1: Hypothetischer Punktprozess Φ_0 , Beobachtungsfenster W, empirische Testgröße T_0

Schritt2: Simuliere n mal die Testgröße T_i im Beobachtungsfenster W des hypothetischen Punktprozesses Φ_0

Schritt3: Ordne T_n einschließlich T_0 der Größe nach. Liegt T_0 in einem kritischen Bereich, also „zu weit außen“, ist die Hypothese (mit Signifikanzlevel 1-α) abzulehnen

Inhalt

II. Ansätze für…

d) … das Testen der Poissonannahmen

- Distanz-Methoden
- Methoden, welche die K-Funktion benutzen
- Quadrat-Methoden

II. Ansätze für…

d) … das Testen der Poissonannahmen

– Distanz-Methoden:

Im Poisson-Fall muss gelten: $H_s(r) = D(r)$

Intuitives Vorgehen:

- Distanzmessungen durchführen
- Daraus empirische Verteilungen berechnen
- Auf Gleichheit überprüfen

Problem:

- Unabhängigkeitsannahme nicht erfüllt

II. Ansätze für…

d) … das Testen der Poissonannahmen

Lösungsmethode (von Byth & Ripley):

- Wähle 2m Lokationen in W (ca. 10% der Punkte)
- Mit der ersten Hälfte berechnet man die Distanzen zum nächsten Punkt: $(v_1, \ldots, v_m)$
- Mit der zweiten Hälfte definiert man Regionen, aus deren unmittelbarer Umgebung (so gewählt, dass durchschnittlich 5 Punkte darin liegen) man zufällig einen Punkt auswählt und zu diesem den Nächsten-Nachbar-Abstand berechnet: $(u_1, \ldots, u_m)$

II. Ansätze für…

d) … das Testen der Poissonannahmen

- Damit ergeben sich folgende Testgrößen mit approximativen Verteilungen:

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

$$h_N = \frac{1}{m} \sum_{i=1}^{m} \frac{u_i^2}{\left(u_i^2 + v_i^2\right)} \quad \sim \quad N\left(\frac{1}{2}, \frac{1}{12m}\right)$$

(siehe Byth&Ripley 1980)

II. Ansätze für…

d) … das Testen der Poissonannahmen

– Methoden, welche die K-Funktion benutzen:

Im Poisson-Fall muss gelten: $K(r) = \pi \cdot r^2$

- Man betrachtet statt der K-Funktion die L-Funktion:

$$L(r) = \sqrt{\frac{K(r)}{\pi}} \overset{Poisson}{=} r$$

- Damit betrachtet man die Teststatistik:

$$\tau = \max_{r \leq r_{\max}} \left| \hat{L}(r) - r \right|$$

II. Ansätze für…

d) … das Testen der Poissonannahmen

– Quadrat-Methoden:

Zerlegt man das Beobachtungsfenster W in gleich große Quadrate der Fläche $\nu_2(Q)$, sollte gelten:

- Durchschnittlich $\lambda \cdot \nu_2(Q)$ Punkte pro Quadrat
- Anzahl der Punkte pro Quadrat sind unabhängig

Aufbauend auf diesen Verteilungseigenschaften kann man Testverfahren konstruieren

II. Ansätze für…

d) … das Testen der Poissonannahmen

Index-of-dispersion Test :

Mit k = Anzahl der Quadrate

$\bar{x}$ = Mittelwert der Anzahl der Punkte pro Quadrat

s² = Varianz der Anzahl der Punkte pro Quadrat

gilt approximativ:

$$I = \frac{(k-1)s^2}{\bar{x}} \sim \chi^2_{k-1}$$

(gute Annäherung, falls k>6 und $\lambda \cdot \nu_2(Q) > 1$)

Inhalt