



Über die statistische Analyse von Clusterpunktprozessen durch die Paarkorrelationsfunktion

Von der Fakultät für Mathematik und Informatik

der Technischen Universität Bergakademie Freiberg

genehmigte

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium

Dr. rer. nat.,

vorgelegt

von Dipl.-Math. Martin Snethlage

geboren am 21. Juli 1971 in Halle/Westfalen

Gutachter: Professor Dr. rer.nat. habil Dietrich Stoyan, Freiberg
Privatdozent Dr.-Ing. Joachim Ohser, Kaiserslautern
Professor Dr. rer.nat. Vicent J. Martínez, València

Tag der Verleihung: 8. September 2000

Inhaltsverzeichnis

| | |
|---|-----------|
| Zusammenfassung | 6 |
| Symbole | 8 |
| 1 Einführung | 10 |
| 2 Punktprozesse | 11 |
| 2.1 Grundlagen | 11 |
| 2.2 Clusterprozesse | 12 |
| 2.3 Der Zusammenhang zwischen $f(r)$ und $d(r)$ | 14 |
| 2.4 Vereinfachungen für die Rechnung mit Produktdichten | 15 |
| 2.5 Abstand zum nächsten Nachbarn | 18 |
| 2.6 Erzeugendes Funktional | 18 |
| 3 Schätzer der Paarkorrelationsfunktion | 20 |
| 3.1 $\rho^{(2)}$ -Schätzer | 20 |
| 3.1.1 Randkorrektur | 21 |
| 3.1.2 Optimale Kernfunktion | 22 |
| 3.1.3 Angepasste Bandweite | 23 |
| 3.1.4 Erwartungstreuer Schätzer | 26 |
| 3.1.5 Mehrere unabhängige Beobachtungsfenster | 27 |
| 3.2 λ^2 -Schätzer | 28 |
| 3.2.1 Bias von $\widehat{\lambda^2}$ | 28 |
| 3.2.2 Kenntnis des wahren λ | 29 |
| 3.2.3 Gewichteter Schätzer | 30 |
| 3.2.4 Erwartungstreuer λ^2 -Schätzer | 32 |
| 3.2.5 Mehrere unabhängige Beobachtungsfenster | 34 |
| 4 Variabilität der Schätzer der Paarkorrelationsfunktion | 35 |
| 4.1 Bootstrap | 35 |
| 4.1.1 Variance of estimators of the product density | 36 |
| 4.1.2 Confidence regions for the intensity function of an inhomogeneous Poisson process | 39 |

| | |
|---|-----------|
| Appendix | 41 |
| 4.2 Quotientenschätzer | 44 |
| 4.2.1 Reihenentwicklung | 45 |
| 4.2.2 Intensitätsschätzer | 47 |
| 4.2.3 Schätzer der Produktdichtefunktion | 48 |
| 4.2.4 Gemischte Größen | 51 |
| 4.2.5 Optimale $p(x, r)$ -Funktion | 52 |
| 4.2.6 Weitere Schätzer | 53 |
| 4.2.7 Schätzervergleich | 54 |
| 4.2.8 Simulationsergebnisse | 55 |
| Anhang | 57 |
| 4.3 Eine heuristische Näherung | 58 |
| 5 Bestimmung des Potenzgesetzes der Paarkorrelationsfunktion | 61 |
| 5.1 Grundlagen | 61 |
| 5.2 Mittelwert von $\hat{\gamma}$ | 62 |
| 5.3 Varianz von $\hat{\gamma}$ | 65 |
| 6 Modelle für Clusterpunktprozesse | 67 |
| 6.1 Punktfeldmodelle | 67 |
| 6.2 Störung | 68 |
| 6.2.1 Das Grundprinzip | 68 |
| 6.2.2 Asymptotisches Verhalten nach der Verwacklung | 69 |
| 6.2.3 Neyman-Scott-Prozesse höherer Ordnung | 70 |
| Anhang | 72 |
| Literaturverzeichnis | 75 |

Zusammenfassung

Diese Arbeit befasst sich mit der statistischen Analyse von Punktprozessen.

Punktprozesse sind zufällige Mengen im \mathbb{R}^d . In dieser Arbeit werden hauptsächlich Clusterprozesse zugrundegelegt, speziell die Klasse der Neyman-Scott-Prozesse. Bei diesen Punktprozessen ist die Wahrscheinlichkeit erhöht, dass sich in der Nähe eines Punktes ein weiterer Punkt des Punktprozesses befindet.

Ein wichtiges Werkzeug bei der Charakterisierung von Punktprozessen ist die Paarkorrelationsfunktion $g(r)$. Mit ihrer Hilfe können Aussagen über die relativen Häufigkeiten von bestimmten Zwischenpunktabständen gemacht werden.

Sei bei einem Punktprozess z. B. $g(r) > 1$. Im Mittel ist dann bei Realisierungen dieses Punktprozesses der Anteil an Punktepaaren mit einem Zwischenpunktabstand in der Größenordnung r größer als bei einem Poissonprozess.

Die Schätzung der Paarkorrelationsfunktion ist gerade im Fall von Clusterprozessen problematisch. Es wurden zahlreiche Artikel zu diesem Thema veröffentlicht, siehe z. B. [20], [26] oder [29]. Das zeigt, dass noch keine Einigkeit darüber herrscht, welches der optimale Schätzer ist. In dieser Arbeit wird versucht, diesen optimalen Schätzer zu finden.

In Kapitel 3 werden alle Bestandteile des üblichen Schätzers der Paarkorrelationsfunktion untersucht und für sie die jeweils geeignetste Repräsentation entwickelt. Dabei wird u. a. eine im gewissen Sinne optimale Kernfunktion wie auch ein optimaler Intensitätsschätzer entwickelt.

Kapitel 4 analysiert die Variabilität der Schätzer. Es gibt dazu mehrere Ansätze. Insbesondere wurden in letzter Zeit sogenannte Bootstrapmethoden entwickelt. Diese erzeugen auf der Grundlage des beobachteten Punktmusters Pseudodatensätze, die wie unabhängig erzeugte Daten behandelt werden. Mit ihrer Hilfe schätzt man dann bewertende Größen wie die Schätzvarianz, den Bias und Konfidenzintervalle. In dieser Arbeit werden zwei solcher Versuche, [2] und [4], genauer untersucht. Es stellt sich dabei heraus, dass sich Bootstrap-Methoden, zumindest in diesen Fällen, höchstens bedingt zur Berechnung statistischer Größen eignen.

Eine weitere Methode, Varianz und Erwartungswert des Schätzers der Paarkorrelationsfunktion zu berechnen, wurde von Landy und Szalay (1993) entwickelt. Mit Hilfe dieser Methode werden in dieser Arbeit verschiedene Schätzer verglichen. Dies führt zu einem in einem bestimmten Sinn besten Schätzer.

Die Kapitel 5 und 6 beschäftigen sich mit einer Anwendung aus der Astronomie. Hier wird das Punktmuster betrachtet, das aus den Positionen der Galaxien im All besteht. Formal wird davon ausgegangen, dass ein Punktprozess existiert, der dieses Punktmuster erzeugt hat.

Schätzungen legen nahe, dass die Paarkorrelationsfunktion dieses Galaxienpunktprozesses einen Pol der Ordnung 1,8 an der Stelle $r = 0$ besitzt. Kapitel 5 analysiert den Schätzer dieser Ordnung und stellt eine Verbesserung vor.

In Kapitel 6 werden einige Punktprozesse mit der Pol-Eigenschaft vorgestellt. Darunter befinden sich auch Neyman-Scott-Prozesse. Es wird dazu ein einfaches Verfahren entwickelt, das es ermöglicht, das asymptotische Verhalten der zugehörigen Paarkorrelationsfunktion und damit die Ordnung ihres Pols zu bestimmen.

Danksagung

Ich möchte mich ganz herzlich bei Herrn Professor Dr. Dietrich Stoyan bedanken für seine intensive Betreuung und für die vielen hilfreichen Anregungen und kritischen Diskussionen. Weiterhin unterstützten mich Markus Holtmann, Martin Kerscher und Michael Werner.

Ganz besonders möchte ich mich bei meiner Frau Jarmila dafür bedanken, dass sie soviel Geduld und Verständnis für mich aufbrachte.

Symbole

In der Tabelle sind viele der in dieser Arbeit verwendeten Symbole verzeichnet. Symbole, die nur an einer Stelle auftauchen, sind allerdings weggelassen worden. Die Seitenangabe bezieht sich auf das erste Vorkommen einer näheren Erklärung des Symbols.

| Symbol | Bedeutung | Seite |
|---------------------|---|-------|
| $\hat{\cdot}$ | Schätzer | 20 |
| $1_A(x)$ | Indikatorfunktion der Menge A | |
| $A(r)$ | Schätzer der Produktdichtefunktion | 48 |
| α | Potenz von $d^*(r) \asymp r^\alpha$ | 69 |
| α, β | | 45 |
| $A(x; z; r)$ | Oberflächeninhalt des Schnittes einer Kugel von Radius r mit einer Sphäre vom Radius z , deren Mittelpunkte den Abstand x haben | 69 |
| b_d | Volumeninhalt der d -dimensionalen Einheitskugel | 12 |
| $B(r)$ | Intensitätsschätzer | 47 |
| $b(x, r)$ | Kugel mit Radius r und Mittelpunkt x | 17 |
| $C_p(r)$ | Integral von $p(x, r)$ | 31 |
| $C_S(r)$ | Integral von $p_S(x, r)$ | 17 |
| d | Dimension des zugrundegelegten Raumes | 11 |
| $d(r)$ | Dichtefunktion des Abstandes zwischen einem Tochterpunkt und dem zugehörigen Elternpunkt | 13 |
| $d^*(r)$ | Dichtefunktion der absoluten Länge der Störung | 69 |
| ∂A | Rand der Menge A | 17 |
| E | Erwartungswert | |
| $F(r), f(r)$ | Dichtefunktion des Abstandes zwischen zwei Tochterpunkten desselben Clusters | 13 |
| $F^*(r), f^*(r)$ | Dichtefunktion des Abstandes zwischen zwei Tochterpunkten desselben Clusters nach einer Störung | 69 |
| $g(r)$ | Paarkorrelationsfunktion | 11 |
| $\hat{g}_S(r)$ | Oberflächengewichteter Schätzer der Paarkorrelationsfunktion | 31 |
| γ | Ordnung des Pols der Produktdichtefunktion | 61 |
| $\bar{\gamma}_W(r)$ | isotropisierte Mengenkovarianzfunktion von W | 18 |
| h | Bandweite | 20 |
| $h^*(r)$ | angepasste Bandweite | 25 |
| I_f | Integral über f | 47 |

| Symbol | Bedeutung | Seite |
|---|--|-------|
| $K(r)$ | Ripleysche K -Funktion | 11 |
| $k_h(v)$ | Kernfunktion mit Bandweite h | 20 |
| λ | Intensität des Punktprozesses | 11 |
| λ_d | mittlere Poisson-verteilte Anzahl der Punkte je Cluster | 13 |
| λ_p | Intensität des Elternprozesses | 12 |
| $\lambda_S(r)$ | Oberflächengewichteter Intensitätsschätzer | 31 |
| MSE | mittlerer quadratischer Fehler | |
| N | Anzahl der beobachteten Punkte im Beobachtungsfenster | 28 |
| ν_d | d -dimensionales Lebesgue-Maß | 17 |
| Φ | Punktprozess | 11 |
| $p(\ln r)$ | doppelt logarithmierte Produktdichtefunktion | 61 |
| $\hat{p}(\ln r)$ | Schätzer von $p(\ln r)$ | 62 |
| $p_S(x, r)$ | Oberflächen-Gewichtsfunktion $S(W \cap \partial b(x, r))$ | 17 |
| $p(x, r)$ | Gewichtsfunktion | 31 |
| q | Quotientenschätzer $\frac{A(r)}{B^2(r)}$ | 44 |
| R | Cluster-Radius beim Matérn-Cluster-Prozess | 13 |
| r_0 | | 61 |
| $\varrho^{(2)}(r), \varrho^{(2)}(x_1, x_2)$ | Produktdichtefunktion zweiter Ordnung | 11 |
| $\varrho^{(n)}(x_1, \dots, x_n)$ | Produktdichtefunktion n -ter Ordnung | 16 |
| $S(X)$ | Oberflächeninhalt der Menge X | 17 |
| $\sum_{x,y}^{\neq}$ | Summe über alle Punktepaare (x, y) mit unterschiedlichen Punkten | 20 |
| var | Varianz | |
| W | Beobachtungsfenster | 20 |
| $w^{(3)}(r)$ | Durchschnittswert von $w^{(3)}(x_1, x_2, x_3)$ | 50 |
| $w_{k_1 \dots k_n}^{(n)}$ | $w^{(n)}(x_{k_1}, \dots, x_{k_n})$ | 16 |
| $w(r)$ | $w^{(2)}(x_1, x_2)$ mit $ x_1 - x_2 = r$ | 16 |
| W_x | Translation von W um x | 18 |
| x_+ | Maximum von x und 0 | 24 |

1 Einführung

In vielen Bereichen der Wissenschaft wird mit Strukturen gearbeitet, die sich durch Punktmuster repräsentieren lassen. In der Forstwissenschaft werden zum Beispiel Wälder durch die Positionen ihrer Bäume dargestellt. Auf diese Weise kann man z. B. Aufschluß erhalten über Strukturen, die sich im Wald gebildet haben. In der Astronomie untersucht man das Punktmuster, das aus den Positionen der Galaxien im Weltall gebildet wird. Von seiner Analyse verspricht man sich, den Wert einiger zentraler physikalischer Konstanten zu bestimmen, die mit der Kosmogonese zusammenhängen.

Häufig sind die Punkte in solchen Punktmustern nicht gleichförmig verteilt. Sie können regelmäßige Strukturen bilden oder in Haufen konzentriert vorliegen. So wird bei vielen Wäldern beobachtet, dass die Bäume Cluster bilden, siehe z. B. [5], [8] oder [37]. Auch die Verteilung der Galaxien im All zeigt starke Unregelmäßigkeiten, siehe z. B. [29]. Hier werden Cluster mit unterschiedlichen Geometrien beobachtet. Neben eher konzentrisch angeordneten Galaxienhaufen existieren auch zylinder- und flächenartige Cluster.

Bei der Analyse der Punktmuster geht man formal davon aus, dass ein Punktprozess existiert, der das zu untersuchende Punktmuster erzeugt hat. Dieser Punktprozess ist dann das eigentliche Objekt der Untersuchung. Er wird anhand der vorliegenden Realisierung statistisch analysiert.

Im Fall von Clusterprozessen führt die Analyse zu großen Schwierigkeiten, da diese Punktprozesse ein hohes Maß an Variabilität besitzen. So sind z. B. bei der Paar-korrelationsfunktion $g(r)$ solche Werte interessant, die zu kleinen Argumenten r gehören. Gerade an diesen Stellen jedoch ist der Bias und die Varianz der üblichen Schätzer für $g(r)$ sehr hoch.

Um hier zu Verbesserungen zu kommen, werden im Folgenden systematisch alle Bestandteile des Schätzers untersucht. Die jeweils optimalen Repräsentationen dieser Bestandteile setzen sich dann zu einem Schätzer zusammen, der den bisher bekannten überlegen ist.

2 Punktprozesse

2.1 Grundlagen

Ein Punktprozess Φ ist eine zufällige lokal finite Menge des \mathbb{R}^d . Er heißt einfach, wenn alle Punkte mit Wahrscheinlichkeit 1 unterschiedlich sind. Er heißt stationär oder auch homogen, wenn Φ und seine Translation $\Phi_x = \{\omega + x : \omega \in \Phi\}$ für alle $x \in \mathbb{R}^d$ dieselbe Verteilung haben. In diesem Fall wird die mittlere Anzahl der Punkte im Einheitsquader mit der Intensität λ bezeichnet.

In dieser Arbeit wird immer von einem einfachen und stationären Punktprozess ausgegangen. Weitere Details zu Punktprozessen kann man z. B. in Daley und Vere-Jones (1988) oder König und Schmidt (1992) nachlesen.

Ein Punktprozess wird häufig durch seine Intensität und durch Größen zweiter Ordnung beschrieben. Letztere leiten sich vom faktoriellen Momentenmaß zweiter Ordnung $\alpha^{(2)}$ ab. Hier wird immer vorausgesetzt, dass $\alpha^{(2)}$ eine Dichte besitzt, die Produktdichtefunktion zweiter Ordnung $\varrho^{(2)}(x_1, x_2)$.

Intuitiv kann $\varrho^{(2)}(x_1, x_2)$ wie folgt verstanden werden: Es seien C_1 und C_2 zwei infinitesimal kleine Kreise mit den Mittelpunkten x_1 und x_2 und den Volumina dF_1 und dF_2 . (Wegen der Kleinheit der Kreise kann man annehmen, dass sie durchschnittsfremd sind.) Dann ist die Wahrscheinlichkeit dafür, dass in jedem der Kreise ein Punkt des Punktfeldes liegt, näherungsweise gleich

$$\varrho^{(2)}(x_1, x_2)dF_1dF_2.$$

Liegt ein stationärer und isotroper Punktprozess vor, so hängt $\varrho^{(2)}(x_1, x_2)$ nur vom Abstand zwischen x_1 und x_2 ab. In diesem Fall schreibt man auch $\varrho^{(2)}(r)$ mit $r = |x_1 - x_2|$. Neben der Produktdichtefunktion zweiter Ordnung gibt es auch noch Produktdichtefunktionen höherer Ordnung $\varrho^{(n)}(x_1, \dots, x_n)$. Sie werden analog zu $\varrho^{(2)}(x_1, x_2)$ definiert. Für genauere Erklärungen und Herleitungen vgl. König und Schmidt (1992).

Die wohl bekannteste Größe zweiter Ordnung ist Ripleys K -Funktion. Ist λ die Intensität eines Punktprozesses, so kann man die Größe $\lambda K(r)$ interpretieren als die durchschnittliche Anzahl von Punkten, die in einen Kreis mit Radius r fallen, dessen Mittelpunkt ein 'typischer' Punkt des Punktfeldes ist. Der Begriff des 'typischen' Punktes hängt mit der Palm'schen Verteilung zusammen. Eine genaue Beschreibung kann man z. B. in König und Schmidt (1992) oder in Stoyan, Kendall und Mecke (1995) nachlesen.

Für großes r wächst $K(r)$ asymptotisch wie r^d (im d -dimensionalen Raum). Das macht es schwer, den Verlauf dieser Funktion zu überblicken, und eine erste visuelle Interpretation ist nur schwer möglich. Deshalb wurden einige Derivate von $K(r)$ mit geeigneteren Eigenschaften entwickelt. Die prägnanteste unter ihnen ist die Paar-korrelationsfunktion $g(r)$. Sie geht im Wesentlichen durch Differentiation aus $K(r)$

hervor

$$g(r) = \frac{1}{\lambda^2} g^{(2)}(r) = \frac{\partial K(r)}{\partial r} / (db_d r^{d-1}), \quad r \geq 0.$$

Hierbei ist b_d das Volumen der d -dimensionalen Einheitskugel.

Die Paarkorrelationsfunktion eines Poissonprozesses ist konstant eins. Ist $g(r)$ für einen Punktprozess größer als eins, so deutet dies darauf hin, dass in Realisierungen dieses Punktprozesses Zwischenpunktabstände von etwa r häufiger vorzufinden sind als bei einem Poissonprozess gleicher Intensität. Entsprechendes gilt, wenn $g(r)$ kleiner als eins ist. Clusterprozesse haben Paarkorrelationsfunktionen, die mit kleiner werdendem r wachsen. Je stärker dieses Wachstum ist, umso größer ist der Grad der Clusterung.

Eine weitere populäre Charakteristik ist die $L(r)$ -Funktion

$$L(r) = \sqrt[d]{\frac{K(r)}{b_d}}, \quad r \geq 0.$$

Im Fall eines homogenen Poissonprozesses ist $L(r) = r$. Abweichungen von dieser Ideallinie nach oben bzw. unten deuten, verglichen mit einem Poissonprozess gleicher Intensität, auf gehäufte bzw. geordnetere Punkte hin. Die L -Funktion ist ein geeignetes Werkzeug bei statistischen Tests und wird auch bei der Parameterschätzung eingesetzt. Für die explorative Datenanalyse ist sie allerdings nicht so sehr geeignet. Der Unterschied zur Paarkorrelationsfunktion entspricht in etwa dem zwischen einer Verteilungsfunktion und ihrer Dichtefunktion, wobei letztere wesentlich aussagekräftiger ist. Ein weiterer Nachteil ist die schwierige Berechnung von Verteilungsparametern des zugehörigen $L(r)$ -Schätzers.

2.2 Clusterprozesse

Clusterprozesse sind Punktprozesse, bei denen die Wahrscheinlichkeit erhöht ist, in der Nähe eines Punktes einen weiteren Punkt des Punktprozesses vorzufinden. Eine wichtige Klasse von Clusterprozessen ist die der Neyman-Scott-Prozesse. Diese Prozesse sind spezielle Fälle von Poisson-Clusterprozessen, die häufig in der Punktprozessstatistik benutzt werden. Sie entstehen durch unabhängige Clusterbildung, deren Zentren, auch Elternpunkte genannt, einen homogenen Poissonprozess der Intensität λ_p bilden. Eine zufällige Anzahl von Tochterpunkten wird unabhängig und, bezogen auf den jeweiligen Elternpunkt, identisch verteilt um jeden Elternpunkt gestreut. Die mittlere Anzahl der Tochterpunkte im typischen Cluster ist c . Der resultierende Punktprozess besteht nur aus den Tochterpunkten.

Ein solcher Neyman-Scott-Prozess ist nach Konstruktion stationär mit der Intensität

$$\lambda = \lambda_p \cdot \mathbf{E}c.$$

Er ist auch isotrop, wenn die Verteilung eines Tochterpunktes bezüglich seines Elternpunktes nur von dem Zwischenpunkt Abstand zwischen Tochter- und Elternpunkt abhängt. Die Dichtefunktion dieser Verteilung wird mit $d(r)$ bezeichnet.

Neben der Funktion $d(r)$ gibt es noch eine weitere Größe, die die Clustergeometrie beschreibt

$$F(r) = P \{ |x - y| \leq r : x, y \text{ gehören zu demselben Cluster} \}, \quad (1)$$

also die Verteilungsfunktion des Abstandes zweier unabhängiger, zufälliger Tochterpunkte desselben Clusters.

Die Dichtefunktion von $F(r)$ wird $f(r)$ genannt. In Kapitel 2.3 wird der Zusammenhang zwischen $d(r)$ und $f(r)$ genauer untersucht.

In dieser Arbeit wird hauptsächlich eine leichte Verallgemeinerung der Neyman-Scott-Prozesse betrachtet. Die Cluster müssen nicht mehr unbedingt globular sein, sie können statt dessen auch auf Geraden und im Fall dreidimensionaler Punktprozesse auch auf Flächen liegen. Die Tochterpunkte sind in jedem Fall wieder gemäß der Dichtefunktion $d(r)$ um das jeweilige Clusterzentrum (innerhalb der vorgegebenen Geometrie) verteilt. Im Folgenden ist mit Neyman-Scott-Prozessen immer diese verallgemeinerte Art gemeint. In jedem Fall gilt für die Produktdichtefunktion zweiter Ordnung

$$\rho^{(2)}(r) = \lambda^2 g(r) = \lambda^2 \left(1 + \frac{f(r)}{db_d r^{d-1} \lambda_p} \right).$$

Møller (1998) führte eine andere Klasse von Punktprozessen ein, die den Clusterfeldern ähnliche Punktfelder erzeugt, die sogenannten Log-Gaußschen-Cox-Prozesse. Sie haben den Vorteil, dass sie durch die Intensität und durch die Produktdichtefunktion zweiter Ordnung $\rho^{(2)}(r)$ vollständig beschrieben werden. Allerdings besitzen sie eine viel geringere Variabilität als Neyman-Scott-Prozesse. Das zeigt sich u. a. darin, dass $\rho^{(2)}(r)$ nie einen Pol an der Stelle $r = 0$ haben kann. Dieses ist aber in manchen Anwendungen erwünscht. Wie schon eingangs erwähnt, ist man in der Astronomie der Meinung, dass der Galaxien-Prozess eine Paarkorrelationsfunktion mit einem Pol der Ordnung 1,8 besitzt (siehe Kapitel 5). Um dieses Punktfeld befriedigend zu modellieren, reicht ein Log-Gaußscher-Cox-Prozess offenbar nicht aus.

Beispiele

Drei Beispiele für Neyman-Scott-Prozesse sind der Matérn-Cluster-, der Strecken- und der Thomasprozess. Bei allen ist die Anzahl der Tochterpunkte je Cluster Poisson-verteilt mit Parameter λ_d . Beim

- Matérn-Cluster-Prozess liegen die Tochterpunkte gleichverteilt in Kugeln mit festem Radius R ;

- Streckenprozess liegen die Tochterpunkte stattdessen gleichverteilt auf Strecken von fester Länge s ;
- Thomasprozess werden die Tochterpunkte gemäß einer d -dimensionalen Normalverteilung mit Streuungsparameter σ um das jeweilige Clusterzentrum verteilt;

Abbildung 1 zeigt jeweils eine Realisierung der drei Punktprozesse im Einheitsquadrat. Die dabei benutzten Prozessparameter sind $\lambda_p = 100$, $\lambda_d = 10$ und für den Matérn-Cluster-Prozess $R = 0.05$, für den Streckenprozess $s = 0.3$ und für den Thomasprozess $\sigma = 0.02$.

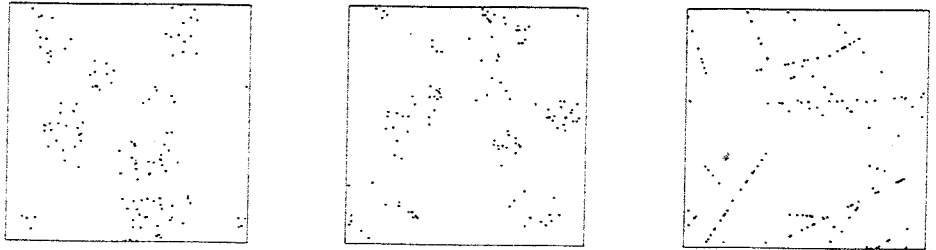


Abbildung 1: Realisierungen des Matérn-Cluster-, des Thomas- und des Streckenprozesses.

2.3 Der Zusammenhang zwischen $f(r)$ und $d(r)$

Bei Neyman-Scott-Prozessen wird die Geometrie der Cluster durch die beiden Funktionen $f(r)$ und $d(r)$ beschrieben. Sie hängen über die folgende Gleichung miteinander zusammen

$$f(r) = \frac{\partial}{\partial r} \left(2 \int_0^\infty \int_0^{r_2} F(r|r_1 \leq r_2) d(r_1) d(r_2) dr_1 dr_2 \right),$$

siehe Stoyan und Stoyan (1994). Hierbei ist $F(r|r_1 \leq r_2)$ die bedingte Verteilungsfunktion des Abstandes zwischen zwei Punkten, die gleichverteilt sind auf den Einheitskugeln der Radien r_1 bzw. r_2 (mit $r_1 \leq r_2$). Wegen

$$F(r|r_1 \leq r_2) = \begin{cases} 0 & \text{für } r < r_2 - r_1 \\ 1 & \text{für } r > r_2 + r_1 \end{cases}$$

ist

$$f(r) = 2 \frac{\partial}{\partial r} \left(\int_0^{r/2} \int_0^{r_2} d(r_1) d(r_2) dr_1 dr_2 + \right.$$

$$\begin{aligned}
& + \int_{r/2}^r \int_0^{r-r_2} d(r_1)d(r_2)dr_1dr_2 + \\
& + \int_{r/2}^r \int_{r-r_2}^{r_2} F(r|r_1 \leq r_2)d(r_1)d(r_2)dr_1dr_2 + \\
& + \int_r^\infty \int_{r_2-r}^{r_2} F(r|r_1 \leq r_2)d(r_1)d(r_2)dr_1dr_2 \Big).
\end{aligned}$$

Ist $F(r|r_1 \leq r_2)$ stetig in r_1 und r_2 und existiert

$$f(r|r_1 \leq r_2) = \frac{\partial}{\partial r} F(r|r_1 \leq r_2),$$

so kann man leicht zeigen, dass

$$\begin{aligned}
f(r) & = 2 \int_{r/2}^r \int_{r-r_2}^{r_2} f(r|r_1 \leq r_2)d(r_1)d(r_2)dr_1dr_2 + \\
& + 2 \int_r^\infty \int_{r_2-r}^{r_2} f(r|r_1 \leq r_2)d(r_1)d(r_2)dr_1dr_2.
\end{aligned}$$

Für $r_2 - r_1 \leq r \leq r_2 + r_1$ ist

$$F(r|r_1 \leq r_2) = \begin{cases} 1 & \text{für 1-dim. Cluster} \\ \frac{1}{\pi} \arccos\left(\frac{r_1^2 + r_2^2 - r^2}{2r_1r_2}\right) & \text{für 2-dim. Cluster} \\ \frac{2r_1r_2 - r_1^2 - r_2^2 + r^2}{4r_1r_2} & \text{für 3-dim. Cluster} \end{cases},$$

siehe u. a. [33], und so

$$f(r|r_1 \leq r_2) = \begin{cases} \frac{2r}{\pi} \sqrt{4r_1^2r_2^2 - (r_1^2 + r_2^2 - r^2)^2} & \text{für 2-dim. Cluster} \\ \frac{r}{2r_1r_2} & \text{für 3-dim. Cluster} \end{cases}$$

Im Fall eindimensionaler Cluster ist $F(r|r_1 \leq r_2)$ nicht stetig. Hier gilt

$$f(r) = \int_r^\infty d(r_1 - r)d(r_1)dr_1.$$

2.4 Vereinfachungen für die Rechnung mit Produktdichten

In den folgenden Kapiteln werden analytische Rechnungen durchgeführt, bei denen auch Produktdichtefunktionen höherer Ordnung vorkommen. Dabei werden einige Vereinfachungen gemacht. In diesem Kapitel werden die wichtigsten von ihnen näher erläutert.

Die Produktdichtefunktionen der ersten vier Ordnungen eines Neyman-Scott-Prozesses kann man wie folgt schreiben

$$\begin{aligned}
 \varrho^{(1)}(x) &= \lambda \\
 \varrho^{(2)}(x_1, x_2) &= \lambda^2 \left(1 + w_{12}^{(2)} \right) \\
 \varrho^{(3)}(x_1, x_2, x_3) &= \lambda^3 \left(1 + w_{12}^{(2)} + w_{13}^{(2)} + w_{23}^{(2)} + w_{123}^{(3)} \right) \\
 \varrho^{(4)}(x_1, x_2, x_3, x_4) &= \lambda^4 \left(1 + w_{12}^{(2)} + w_{13}^{(2)} + w_{14}^{(2)} \right. \\
 &\quad + w_{23}^{(2)} + w_{24}^{(2)} + w_{34}^{(2)} \\
 &\quad + w_{123}^{(3)} + w_{124}^{(3)} + w_{134}^{(3)} + w_{234}^{(3)} \\
 &\quad + w_{12}^{(2)} w_{34}^{(2)} + w_{13}^{(2)} w_{24}^{(2)} + w_{14}^{(2)} w_{23}^{(2)} \\
 &\quad \left. + w_{1234}^{(4)} \right),
 \end{aligned}$$

siehe [27]. Hierbei sind die $w_{k_1 \dots k_n}^{(n)} = w^{(n)}(x_{k_1}, \dots, x_{k_n})$ Funktionen, die nur von den Zwischenpunktabständen der Punkte $(x_{k_i})_{i \leq n}$ abhängen. Ist $n = 2$, so schreibt man wegen der Stationarität auch vereinfachend $g(r) = 1 + w(r)$, wobei $w(r) = w^{(2)}(x_1, x_2)$ mit $|x_1 - x_2| = r$.

Die $w_{k_1 \dots k_n}^{(n)}$ sind Null, wenn der Abstand zweier Punkte von $(x_{k_i})_{i \leq n}$ so groß ist, daß sie unabhängig voneinander sind. Im Falle eines Matérn-Cluster-Prozesses z. B. beträgt dieser Maximalabstand $2R$, da zwei Punkte, deren Abstand größer als $2R$ ist, nicht vom selben Cluster abstammen können und sie sich somit zueinander wie zwei Poisson-verteilte und damit unabhängige Punkte verhalten, siehe z. B. [17].

Der Maximalabstand muss nicht für jeden Neyman-Scott-Prozess existieren (das ist z. B. für den Thomas-Prozess der Fall). Allerdings kann man auch dann annehmen, dass zwei weit voneinander entfernt liegende Punkte angenähert unabhängig sind; man kann sich dies leicht anhand der Konstruktion der Neyman-Scott-Prozesse klar machen.

Liegen mindestens zwei der Punkte $(x_{k_i})_{i \leq n}$ rein zufällig, also gleichverteilt, in einem (verglichen mit dem Maximalabstand) relativ großen Beobachtungsfenster, so ist die Wahrscheinlichkeit groß, dass sie weit voneinander entfernt liegen. Deshalb nimmt man dann an, dass hier $w_{k_1 \dots k_n}^{(n)} \approx 0$ gilt. Es gilt also für zwei beliebige Funktionen $f(x)$ und $g(y)$ ungefähr

$$\begin{aligned}
 \int_W \varrho^{(2)}(x, y) f(x) g(y) dx dy &= \lambda^2 \int_W \left(1 + w^{(2)}(x, y) \right) f(x) g(y) dx dy \\
 &\approx \lambda^2 \int_W f(x) g(y) dx dy.
 \end{aligned}$$

Liegen allerdings alle Punkte mehr oder weniger nahe beieinander, so kann $w_{k_1 \dots k_n}^{(n)}$

nicht vernachlässigt werden. Das gilt z. B. für den Ausdruck

$$\int_{W^2} \varrho^{(2)}(x, y) \mathbf{1}_{[-h, h]}(r - |x - y|) dx dy.$$

Hier ist der Abstand zwischen x und y ungefähr r (wenn h klein ist gegenüber r). Es ist also $w^{(2)}(x, y) = w(|x - y|) \approx w(r)$ und so

$$\begin{aligned} & \int_{W^2} \varrho^{(2)}(x, y) \mathbf{1}_{[-h, h]}(r - |x - y|) dx dy \\ &= \lambda^2 \int_{W^2} (1 + w(r)) \mathbf{1}_{[-h, h]}(r - |x - y|) dx dy \\ &= \varrho^{(2)}(r) \int_{W^2} \mathbf{1}_{[-h, h]}(r - |x - y|) dx dy. \end{aligned} \quad (2)$$

Wie man in Kapitel 4.2 sehen wird, lassen sich auf diese Weise sehr komplexe Ausdrücke ganz erheblich vereinfachen.

Das letzte Integral in (2) lässt sich vereinfachen zu

$$\begin{aligned} & \int_W \int_W \mathbf{1}_{[-h, h]}(r - |x - y|) dy dx \\ &= \int_W \nu_d(W \cap (b(x, r + h) - b(x, r - h))) dx \\ &\approx 2h \int_W S(W \cap \partial b(x, r)) dx \\ &= 2h \int_W p_S(x, r) dx \\ &= 2h C_S(r), \end{aligned} \quad (3)$$

wobei $b(x, r)$ bzw. $\partial b(x, r)$ die Kugel bzw. Sphäre mit Radius r und Mittelpunkt x ist, $\nu_d(X)$ das d -dimensionale Lebesgue-Maß und $S(X)$ der Oberflächeninhalt der Menge X . Die Funktionen $p_S(x, r)$ und $C_S(r)$ sind entsprechend dem obigen Gleichungssystem definiert,

$$p_S(x, r) = S(W \cap \partial b(x, r))$$

und

$$C_S(r) = \int_W p_S(x, r) dx.$$

Sie werden erst an späterer Stelle gebraucht. Es ist

$$\begin{aligned} C_S(r) &= \int_W S(W \cap \partial b(x, r)) dx \\ &= \int_W \int_{\partial b(x, r)} \mathbf{1}_W(y) dy dx \end{aligned}$$

$$\begin{aligned}
&= \int_{\partial b(0,r)} \int_W \mathbf{1}_{W-x}(y) dx dy \\
&= \int_{\partial b(0,r)} \nu_d(W \cap W_{-y}) dy \\
&= db_d r^{d-1} \bar{\gamma}_W(r).
\end{aligned}$$

Hierbei ist $W_x = \{w + x : w \in W\}$ die Translation von W um x und $\bar{\gamma}_W(r)$ die isotropisierte Mengenkovarianzfunktion, siehe z. B. [33].

2.5 Abstand zum nächsten Nachbarn

Die D -Funktion ist die Verteilungsfunktion des Abstandes vom 'typischen' Punkt des Punktprozesses zu seinem nächsten Nachbarn. Die sphärische Kontaktverteilungsfunktion $H_s(r)$ ist die Verteilungsfunktion des Abstandes vom Ursprung zum nächsten Nachbarn im Punktprozess, siehe [35]. Für beide Funktionen existieren befriedigende Schätzer. Allerdings sind die Funktionen im Fall von Clusterprozessen analytisch nur schwer berechenbar. Schon für die $H_s(r)$ -Funktion des Matérn-Cluster-Prozess im \mathbb{R}^2 ergeben sich sehr komplizierte Integrale:

$$H_s(r) = 1 - \exp(-2\pi\lambda_p(I_1(r) + I_2(r))), \quad r \geq 0,$$

mit

$$I_1(r) = \int_0^{|R-r|} t \left(1 - \exp \left(\frac{2\lambda_c}{\pi R^2} \left(\int_{x_3}^{x_2} \sqrt{R^2 - x^2} dx + \int_{x_1}^{x_3} \sqrt{r^2 - (x-t)^2} dx \right) \right) \right) dt$$

und

$$I_2(r) = \int_{|R-r|}^{R+r} t \left(1 - \exp \left(-\frac{\lambda_c}{R^2} (\min(R, r))^2 \right) \right) dt,$$

siehe Last und Holtmann (1999). Die Formel für die entsprechende $D(r)$ -Funktion ist noch wesentlich komplizierter. Gleiches gilt auch für andere Standardpunktprozesse, weshalb diese Größen nur wenig geeignet für analytische Berechnungen sind. Nichtsdestoweniger haben sie sehr wohl ihren Sinn bei Gütetests.

2.6 Erzeugendes Funktional

Das erzeugende Funktional ist in der mathematischen Analyse von Clusterprozessen ein hilfreiches Werkzeug. Deshalb liegt es nahe, seine Nützlichkeit auch für die explorative Datenanalyse zu untersuchen.

Das erzeugende Funktional G_Φ eines Punktprozesses Φ ist wie folgt definiert. Sei U die Familie aller nichtnegativen, beschränkten messbaren Funktionen u auf \mathbb{R}^d , deren Träger $\{x \in \mathbb{R}^d : u(x) > 0\}$ ebenfalls beschränkt ist. Weiterhin sei V die Familie aller Funktionen $v = 1 - u$ für $u \in U$ mit $0 \leq u \leq 1$. Dann ist G_Φ definiert als

$$\begin{aligned} G_\Phi(v) &= \mathbf{E} \left(\prod_{x \in \Phi} v(x) \right) \\ &= \int \prod_{x \in \varphi} v(x) P(d\varphi), \quad v \in V. \end{aligned}$$

$G_\Phi(v)$ ist von seiner Natur her der $H_s(r)$ -Funktion verwandt. Ist nämlich $v(x) = v_r^0(x) = 1 - 1_{b(o,r)}(x)$, dann ist $G(v_r^0)$ sogar äquivalent zu $H_s(r)$

$$H_s(r) = 1 - G(v_r^0).$$

Man kann das erzeugende Funktional also als eine Art Verallgemeinerung von $H_s(r)$ ansehen. Deshalb ist es nicht überraschend, dass beide Größen ähnliche Eigenschaften haben: Wie schon für $H_s(r)$ existiert auch für das erzeugende Funktional ein befriedigender Schätzer. Leider ist es aber auch ebenso ungeeignet für analytische Berechnungen, da die entsprechenden Formeln sehr kompliziert sind, (siehe z. B. [35]).

Obwohl ein Punktprozess eindeutig durch das erzeugende Funktional definiert ist, scheint es, wie auch $H_s(r)$, keine gute Alternative zu den Größen zweiter Ordnung zu sein, wenn es um die explorative Datenanalyse von Clusterprozessen geht.

3 Schätzer der Paarkorrelationsfunktion

Die Paarkorrelationsfunktion ist eines der wichtigsten Werkzeuge bei der explorativen Analyse von Punktprozessen. Deshalb kommt ihrer Schätzung eine große Bedeutung zu. Es wurden zahlreiche Artikel zu diesem Thema veröffentlicht, was die Wichtigkeit dieser Funktion unterstreicht. Andererseits zeigt dies aber auch, wie schwierig ihre Schätzung ist und dass noch keine Einigkeit darüber herrscht, welches der optimale Schätzer ist.

Die meisten Schätzer haben im Prinzip die folgende Struktur

$$\widehat{g}(r) = \frac{\widehat{\rho^{(2)}}(r)}{\widehat{\lambda^2}}, \quad (4)$$

die sich von der Beziehung

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

ableitet. Hierbei bedeutet das Symbol „ $\widehat{}$ “ jeweils, dass es sich bei der Größe um einen Schätzer handelt.

In diesem Kapitel wird zunächst ein möglichst guter Schätzer für $\rho^{(2)}$ entwickelt, d. h. ein Schätzer mit möglichst geringem Bias und geringer Varianz. Dann werden verschiedene Ideen diskutiert, welcher λ^2 -Schätzer optimal dazu passt.

Ratio-Schätzer wie in (4) führen selbstverständlich nicht zwangsläufig zu guten (und schon gar nicht zu erwartungstreuen) Schätzern, wenn die einzelnen Schätzer $\widehat{\rho^{(2)}}(r)$ und $\widehat{\lambda^2}$ gute Eigenschaften (wie z. B. Erwartungstreue) besitzen. Ihre Verteilungsparameter sind i. a. nicht direkt berechenbar. Bei der Berechnung der Varianz wird z. B. $\widehat{\lambda^2}$ häufig als konstant postuliert, siehe z. B. [2] und [34].

Diese Vorgehensweise führt manchmal zu großen Fehlern, siehe Kapitel 3.2.2. Ein anderer Ansatz, der aber auch nicht ohne Vereinfachungen auskommt, wird später in Kapitel 4.2 diskutiert.

3.1 $\rho^{(2)}$ -Schätzer

In diesem Kapitel werden alle Teile eines $\rho^{(2)}(r)$ -Schätzers untersucht und für sie eine optimale Repräsentation gesucht. Die üblichen $\rho^{(2)}(r)$ -Schätzer haben die Struktur

$$\widehat{\rho}(r) = \frac{1}{db_d \nu_d(W)} \sum_{x,y \in \Phi}^{\neq} \mathbf{1}_W(x) \mathbf{1}_W(y) k_h(|x-y|-r) \frac{R(x,y)}{r^{d-1}}. \quad (5)$$

Hierbei ist W das Beobachtungsfenster, $k_h(v)$ eine Kernfunktion mit Bandweite h (üblicherweise ist h wesentlich kleiner als r) und $R(x,y)$ eine Funktion, die eine Randkorrektur bewirkt. Das \neq -Zeichen bedeutet, dass über alle Punktepaare mit

unterschiedlichen Punkten summiert wird. Der Term r^{d-1} kann auch durch $|x-y|^{d-1}$ ersetzt werden, was zu einem leicht verringerten Bias führt (siehe Kapitel 3.1.3).

Die Randkorrektur gleicht die Benachteiligung großer Zwischenpunktabstände gegenüber kleineren aus. Große Zwischenpunktabstände sind wegen der Beschränktheit des Beobachtungsfensters in den Beobachtungen unterrepräsentiert.

Die Kernfunktion ist eine Dichtefunktion, also eine nicht-negative reelle Funktion mit der Gesamtmasse 1, die nur in dem kleinen Intervall $[-h, h]$ positiv ist. Ein Beispiel ist die unter Physikern populäre Rechteckkernfunktion

$$k_h(v) = \frac{1}{2h} \mathbf{1}_{[-h, h]}(v).$$

Eine weitere Kernfunktion, die zu glatteren Schätzern führt, ist die Epanečnikov-Kernfunktion

$$k_h(v) = \frac{3}{4h} \left(1 - \left(\frac{v}{h}\right)^2\right) \mathbf{1}_{[-h, h]}(v).$$

Sie wurde für Schätzer der Paarkorrelationsfunktion durch Fiksel (1988) eingeführt und in vielen darauffolgenden Veröffentlichungen benutzt. In dieser Arbeit wird bei Schätzungen allerdings immer die Rechteckkernfunktion benutzt, da sie in einem gewissen Sinn die optimale Kernfunktion darstellt, siehe Kapitel 3.1.2.

Neben Schätzern der oben genannten Art (5) werden in der Astronomie noch weitere benutzt. Diese sind teilweise mit den hier behandelten Schätzern äquivalent. Sie werden in Kapitel 4.2 genauer behandelt.

3.1.1 Randkorrektur

Die beiden wichtigsten Randkorrekturen werden durch die Funktionen

$$R_1(x, y) = \frac{\nu_d(W)}{\bar{\gamma}_W(r)}$$

und

$$R_2(x, y) = \frac{\nu_d(W)}{\nu_d(\bar{W}_x \cap \bar{W}_y)}$$

repräsentiert. Während $R_1(x, y)$ nur für isotrope Punktprozesse geeignet ist, kann $R_2(x, y)$ immer benutzt werden.

Für Quader bzw. Rechtecke ist $\bar{\gamma}_W(r)$ bekannt. Ist r kleiner als die kürzeste Seitenlänge, so gilt im zweidimensionalen Fall

$$\bar{\gamma}_W(r) = w_1 w_2 - \frac{r}{\pi} (2w_1 + 2w_2 - r)$$

und im dreidimensionalen Fall

$$\begin{aligned} \bar{\gamma}_W(r) = & -\frac{1}{4\pi} r^3 + \frac{2}{3\pi} r^2 (w_1 + w_2 + w_3) - \frac{1}{2} r (w_1 w_2 + w_1 w_3 + w_2 w_3) \\ & + w_1 w_2 w_3, \end{aligned}$$

wobei die w_i die Seitenlängen des Quaders bzw. Rechtecks sind. Ohser (1991) hat gezeigt, daß beide Randkorrekturen im isotropen Fall ungefähr zu den gleichen Erwartungswerten führen. Simulationsstudien zeigen, daß sich auch die Schätzvarianz nicht wesentlich durch den Einsatz von $\tilde{\gamma}_W(r)$ ändert. Im praktischen Gebrauch ist sicher die allgemeine Randkorrektur $R_2(x, y)$ vorteilhafter, da sie auch im anisotropen Fall noch zu richtigen Ergebnissen führt.

3.1.2 Optimale Kernfunktion

Die Kernfunktion $k_h(v)$ in (5) bewirkt, dass nur solche Punktepaare berücksichtigt werden, die einen Zwischenpunkt Abstand von ungefähr r haben. Je nach dem, wie stark dieser Abstand von r abweicht, können die Punktepaare noch entsprechend gewichtet werden. Eine optimale Gewichtung ist dann erreicht, wenn der Fehler der $\varrho^{(2)}(r)$ -Schätzung minimal ist.

Sinnvollerweise wird die Bandweite h viel kleiner als r gewählt. In diesem Fall ist $\hat{\varrho}(r)$ ungefähr erwartungstreu, siehe z. B. Kapitel 3.1.3. Deshalb ist der mittlere quadratische Fehler von $\hat{\varrho}(r)$ im wesentlichen seine Schätzvarianz. Wie man in Kapitel 4.3 sehen wird (siehe die dortige Bemerkung S. 60), hängt die Schätzvarianz bezüglich der Kernfunktion in erster Näherung nur vom Wert des Integrals

$$\int_{-h}^h k^2(v) dv \quad (6)$$

ab. Zur Minimierung des Fehlers wird also die Kernfunktion gesucht, die das Integral (6) minimiert.

Angelehnt an die Vorgehensweise in der Variationsrechnung sei $k_m(v)$ diese optimale Funktion. Dann kann jede Kernfunktion $k(v)$ geschrieben werden als

$$k(v) = k_m(v) + \varepsilon f(v)$$

mit

$$\int_{-h}^h f(v) dv = 0 \quad (7)$$

und geeignetem ε . Das zu minimierende Integral lautet nun

$$\int_{-h}^h (k_m(v) + \varepsilon f(v))^2 dv. \quad (8)$$

Da (8) bei $\varepsilon = 0$ minimal ist, verschwindet dort seine Ableitung nach ε

$$\left. \frac{\partial}{\partial \varepsilon} \int_{-h}^h k^2(v) dv \right|_{\varepsilon=0} = 0.$$

Somit ergibt sich

$$\begin{aligned}
 0 &= \left. \frac{\partial}{\partial \varepsilon} \int_{-h}^h k^2(v) dv \right|_{\varepsilon=0} \\
 &= \left. \frac{\partial}{\partial \varepsilon} \int_{-h}^h (k_m(v) + \varepsilon f(v))^2 dv \right|_{\varepsilon=0} \\
 &= \left. \frac{\partial}{\partial \varepsilon} \int_{-h}^h k_m^2(v) dv + 2\varepsilon \int_{-h}^h k_m(v) f(v) dv + \varepsilon^2 \int_{-h}^h f^2(v) dv \right|_{\varepsilon=0} \\
 &= \left. 2 \int_{-h}^h k_m(v) f(v) dv + 2\varepsilon \int_{-h}^h f^2(v) dv \right|_{\varepsilon=0} \\
 &= 2 \int_{-h}^h k_m(v) f(v) dv
 \end{aligned}$$

für jedes $f(v)$ mit (7). Es ist leicht zu sehen, dass dieses innerhalb des Intervalls $[-h, h]$ zu

$$k_m(v) = \text{const} = \frac{1}{2h} 1_{[-h, h]}(v)$$

führt, der Rechteckkernfunktion. Für die zweite Ableitung gilt

$$\begin{aligned}
 \left. \frac{\partial^2}{\partial \varepsilon^2} \int_{-h}^h k^2(v) dv \right|_{\varepsilon=0} &= \left. \frac{\partial}{\partial \varepsilon} 2 \int_{-h}^h k_m(v) f(v) dv + 2\varepsilon \int_{-h}^h f^2(v) dv \right|_{\varepsilon=0} \\
 &= 2 \int_{-h}^h f^2(v) dv \\
 &> 0.
 \end{aligned}$$

Es handelt sich bei $k_m(v)$ also um das gesuchte Minimum.

Simulationsstudien bestätigen dieses Ergebnis. Sie haben gezeigt, dass der Einsatz der Rechteckkernfunktion gegenüber der Epanečnikovkernfunktion die Schätzvarianz wesentlich verringert.

3.1.3 Angepasste Bandweite

Abbildung 2 zeigt, dass der $\varrho^{(2)}(r)$ -Schätzer einen beträchtlichen Bias für $r < h$ besitzt. Da h im Normalfall sehr klein ist, ist dieser Bias bisher nicht stärker aufgefallen. Wenn allerdings das asymptotische Verhalten von $\varrho^{(2)}(r)$ für $r \rightarrow 0$ untersucht werden soll, kann diese Abweichung zu großen Fehlern führen, siehe Kapitel 5.

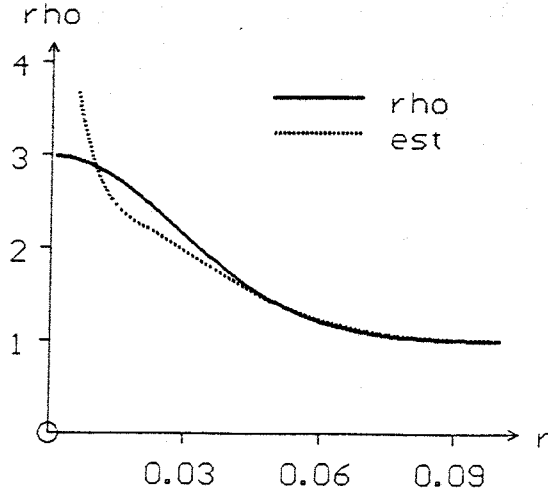


Abbildung 2: Die Produktdichtefunktion $\varrho^{(2)}(r)$ ('rho') verglichen mit ihrem Schätzer $\widehat{\varrho}(r)$ ('est') mit $h = 0.02$ im Fall eines zweidimensionalen Thomasprozesses mit $\lambda_p = 100$, $\lambda_d = 10$ und $\sigma = 0.02$. (Die y -Werte sind in 10^6 -Einheiten angegeben.) Der Schätzer besitzt für kleine r einen beträchtlichen Bias.

Für den Erwartungswert von $\widehat{\varrho}(r)$ (mit anisotroper Randkorrektur) gilt

$$\begin{aligned}
 \mathbf{E}\widehat{\varrho}(r) &= \frac{1}{db_d} \int_{W^2} \frac{\varrho^{(2)}(|x-y|)k_h(|x-y|-r)}{r^{d-1}\nu_d(W_x \cap W_y)} d(x,y) \\
 &= \frac{1}{db_d r^{d-1}} \int \varrho^{(2)}(|x|)k_h(|x|-r) \int \frac{\mathbf{1}_W(x+y)\mathbf{1}_W(y)}{\nu_d(W_{x+y} \cap W_y)} dy dx \\
 &= \frac{1}{db_d r^{d-1}} \int \varrho^{(2)}(|x|)k_h(|x|-r) \int_{W_x \cap W} \frac{1}{\nu_d(W_x \cap W)} dy dx \\
 &= \frac{1}{r^{d-1}} \int_{(r-h)_+}^{r+h} v^{d-1} \varrho^{(2)}(v)k_h(v-r) dv \tag{9}
 \end{aligned}$$

mit $x_+ = \max(0, x)$. Für positives r ist der Grenzwert für $h \rightarrow 0$

$$\begin{aligned}
 \lim_{h \rightarrow 0} \mathbf{E}\widehat{\varrho}(r) &= \lim_{h \rightarrow 0} \int_{(r-h)_+}^{r+h} \left(\frac{v}{r}\right)^{d-1} \varrho^{(2)}(v)k_h(v-r) dv \\
 &= \varrho^{(2)}(r) \lim_{h \rightarrow 0} \int_{(r-h)_+}^{r+h} k_h(v-r) dv \\
 &= \varrho^{(2)}(r).
 \end{aligned}$$

Ist allerdings h fest und $r \rightarrow 0$, so divergiert $\widehat{\varrho}(r)$

$$\begin{aligned}\lim_{r \rightarrow 0} \mathbf{E}\widehat{\varrho}(r) &= \lim_{r \rightarrow 0} \int_{(r-h)_+}^{r+h} \left(\frac{v}{r}\right)^{d-1} \varrho^{(2)}(v) k_h(v-r) dv \\ &= \int_0^h v^{d-1} \varrho^{(2)}(v) k_h(v) dv \cdot \lim_{r \rightarrow 0} \frac{1}{r^{d-1}}.\end{aligned}$$

Ersetzt man den Faktor $1/r^{d-1}$ durch $1/|x-y|^{d-1}$, so existiert zwar der Erwartungswert von $\widehat{\varrho}(r)$ an der Stelle $r=0$ (bei geeigneten Prozessen), der Bias ist aber nach wie vor beträchtlich

$$\begin{aligned}\lim_{r \rightarrow 0} \mathbf{E}\widehat{\varrho}(r) &= \lim_{r \rightarrow 0} \int_{(r-h)_+}^{r+h} \varrho^{(2)}(v) k_h(v-r) dv \\ &\approx \int_0^h \varrho^{(2)}(v) k_h(v) dv \\ &\approx \frac{1}{2} \varrho^{(2)}(h/2)\end{aligned}$$

(falls die Kernfunktion symmetrisch ist).

Um den Bias des Schätzers zu reduzieren, kann man statt einer konstanten Bandweite eine angepasste verwenden. Eine Möglichkeit ist $h^*(r) = \min\{h, r\}$ mit einem geeigneten festen h für größere Werte von r . Damit ergibt sich für $r \leq h$ (unter Benutzung von $1/r^{d-1}$)

$$\begin{aligned}\mathbf{E}\widehat{\varrho}(r) &= \int_0^{2r} \left(\frac{v}{r}\right)^{d-1} \varrho^{(2)}(v) k_{h^*}(v-r) dv \\ &\approx \varrho^{(2)}(r)\end{aligned}$$

bzw. (unter Benutzung von $1/|x-y|^{d-1}$)

$$\begin{aligned}\mathbf{E}\widehat{\varrho}(r) &= \int_0^{2r} \varrho^{(2)}(v) k_{h^*}(v-r) dv \\ &\approx \varrho^{(2)}(r).\end{aligned}$$

(Hierbei ist es im Grunde egal, welche spezielle Kernfunktion benutzt wird.)

Abbildung 3 zeigt, dass die angepasste Bandweite zu einem nahezu unverfälschten Schätzer führt (das gilt auch für viele andere Clusterprozesse mit Ausnahme von

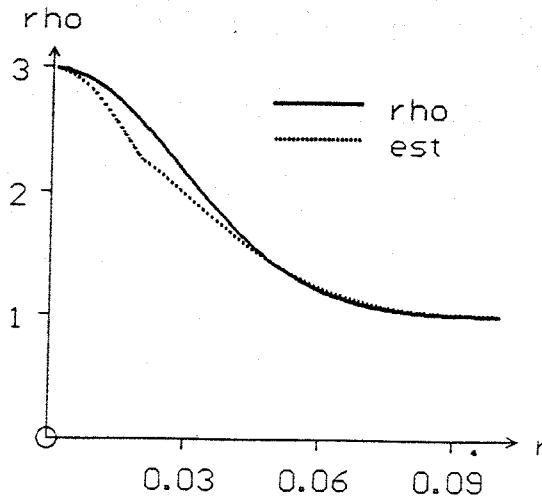


Abbildung 3: Die Paarkorrelationsfunktion ('rho') verglichen mit ihrem Schätzer $\hat{\rho}(r)$ ('est') mit angepasster Bandweite in derselben Situation wie in Abbildung 2 S. 24. Der Schätzer hat nur noch einen geringen Bias.

solchen, deren $\rho^{(2)}(r)$ -Funktion einen Pol an der Stelle $r = 0$ besitzen). Die Erfahrung zeigt, dass es hierbei ziemlich unerheblich ist, welcher der beiden Faktoren $1/|x - y|^{d-1}$ oder $1/r^{d-1}$ benutzt wird. Im ersten Fall ist der Bias gegenüber dem zweiten leicht verringert, was aber durch eine leichte Erhöhung der Schätzvarianz wieder ausgeglichen wird. Aus diesem Grund wird im Folgenden meistens nicht extra darauf hingewiesen, welcher Faktor benutzt wird.

Die Schätzvarianz ist nach wie vor in beiden Fällen für kleine r erheblich hoch. Eventuell kann sie durch eine andere Kernfunktion verringert werden. So sind z. B. asymmetrische Funktionen denkbar, deren Bandweite konstant bleibt. Im nächsten Kapitel wird eine spezielle Kernfunktion vorgestellt, die zu absolut erwartungstreuen Schätzern von $\rho^{(2)}(r)$ führt.

3.1.4 Erwartungstreuer Schätzer

Es ist möglich, einen erwartungstreuen Schätzer für $\rho^{(2)}(r)$ zu erhalten. Dazu muss eine beliebige Kernfunktion mit dem Faktor

$$\rho^{(2)}(r)/\rho^{(2)}(|x - y|)$$

multipliziert werden. Es gilt dann nämlich

$$\mathbf{E}\hat{\rho}(r) = \int_{r-h}^{r+h} k_h(v-r)\rho^{(2)}(v)\frac{\rho^{(2)}(r)}{\rho^{(2)}(v)}dv$$

$$\begin{aligned}
&= \varrho^{(2)}(r) \int_{-h}^h k_h(v) dv \\
&= \varrho^{(2)}(r).
\end{aligned}$$

Offensichtlich muss also die wahre $\varrho^{(2)}(r)$ -Funktion schon bekannt sein. Obwohl es auf den ersten Blick unsinnig erscheint, eine schon bekannte Funktion zu schätzen, kann dieser Schätzer doch Vorteile bringen. Man könnte zum Beispiel statt der wahren $\varrho^{(2)}(r)$ -Funktion eine Schätzung einsetzen, die man auf die übliche Art erhalten hat. Die neue Schätzung könnte bessere Eigenschaften haben als die erste. Auf eine systematische Untersuchung dieses Schätzers wird hier verzichtet. Zum einen wäre hierfür der Aufwand zu groß, zum anderen erbrachte dieser Schätzer bei bisherigen Simulationsstudien keinen wesentlichen Vorteil.

3.1.5 Mehrere unabhängige Beobachtungsfenster

Gegeben seien n Beobachtungsfenster W_1, \dots, W_n und ein stationärer Punktprozess Φ . Der Abstand zwischen je zwei Fenstern sei so groß, dass die Realisierungen in ihnen praktisch unabhängig voneinander ablaufen. Genauso gut kann man sich statt der vielen Fenster auch ein einziges vorstellen, in dem mehrere unabhängige Realisierungen desselben Punktprozesses beobachtet werden. Es stellt sich die Frage, ob bei einer Schätzprozedur alle Fenster zu einem einzigen zusammengefasst werden sollten, $W = \bigcup_{i=1}^n W_i$, oder ob es von Vorteil ist, die Schätzung für jedes Fenster einzeln durchzuführen und anschließend geeignet zu mitteln.

Gegeben sei der globale $\varrho^{(2)}$ -Schätzer

$$\begin{aligned}
\hat{\varrho}(r) &= \frac{1}{db_d r^{d-1} \bar{\gamma}_W(r)} \sum_{x,y \in \Phi \cap W}^{\neq} k_h(r - |x - y|) \\
&= \frac{1}{\bar{\gamma}_W(r)} \sum_{x,y \in \Phi \cap W}^{\neq} h(x, y),
\end{aligned} \tag{10}$$

mit einer geeigneten Funktion $h(x, y)$, und sein lokales Äquivalent

$$\hat{\varrho}_*(r) = \sum_{i=1}^n a_i(r) \hat{\varrho}_i(r)$$

mit positiven Gewichten $a_i(r)$ mit $\sum_{i=1}^n a_i(r) = 1$. Hierbei ist $\hat{\varrho}_i(r)$ äquivalent zu $\hat{\varrho}(r)$ definiert mit W_i statt W . (In (10) kann der Term r^{d-1} auch durch $|x - y|^{d-1}$ ersetzt werden.) Es scheint natürlich, $a_i(r) = \frac{\bar{\gamma}_{W_i}(r)}{\bar{\gamma}_W(r)}$ zu setzen, was auf

$$\hat{\varrho}_*(r) = \sum_{i=1}^n \frac{\bar{\gamma}_{W_i}(r)}{\bar{\gamma}_W(r)} \hat{\varrho}_i(r)$$

führt. Damit folgt

$$\begin{aligned}
 \hat{\rho}_*(r) &= \sum_{i=1}^n \frac{\bar{\gamma}_{W_i}(r)}{\bar{\gamma}_W(r)} \hat{\rho}_i(r) \\
 &= \frac{1}{\bar{\gamma}_W(r)} \sum_{i=1}^n \sum_{x,y \in \Phi \cap W_i}^{\neq} h(x,y) \\
 &= \frac{1}{\bar{\gamma}_W(r)} \sum_{x,y \in \Phi \cap W}^{\neq} h(x,y) \\
 &= \hat{\rho}(r).
 \end{aligned}$$

Mit der hier gewählten Wichtung stimmen also lokaler und globaler Schätzer exakt überein.

Neben der hier gewählten Wichtung sind natürlich auch noch andere denkbar. Es ist allerdings nicht leicht festzustellen, welche davon die optimale Wichtung darstellt. Das liegt am relativ komplizierten Zusammenhang zwischen der Varianz des $\rho^{(2)}(r)$ -Schätzers und dem Beobachtungsfenster W , siehe Kapitel 4.2.

3.2 λ^2 -Schätzer

3.2.1 Bias von $\widehat{\lambda^2}$

Das Quadrat der Intensität wird gewöhnlich durch das Quadrat eines einfachen Intensitätsschätzers geschätzt

$$\widehat{\lambda^2} = \left(\frac{N}{\nu_d(W)} \right)^2 = \left(\frac{1}{\nu_d(W)} \sum_{x \in \Phi} \mathbf{1}_W(x) \right)^2,$$

wobei N die Anzahl der beobachteten Punkte in W ist. Ohser (1991) empfiehlt

$$\widehat{\lambda^2} = \frac{N(N-1)}{\nu_d^2(W)} = \frac{1}{\nu_d^2(W)} \sum_{x,y \in \Phi}^{\neq} \mathbf{1}_W(x) \mathbf{1}_W(y),$$

da dieses ein erwartungstreuer Schätzer im Falle eines homogenen Poissonprozesses ist.

Im Folgenden soll eine Abschätzung für den Fehler von $\widehat{\lambda^2}$ durchgeführt werden. Für den Erwartungswert des ersten Schätzers gilt

$$\begin{aligned}
 \mathbf{E} \widehat{\lambda^2} &= \frac{1}{\nu_d^2(W)} \int \rho^{(2)}(|x-y|) \mathbf{1}_W(x) \mathbf{1}_W(y) dx dy \\
 &\quad + \frac{1}{\nu_d^2(W)} \int \lambda \mathbf{1}_W(x) dx
 \end{aligned}$$

$$\begin{aligned}
&= \frac{\lambda^2}{\nu_d^2(W)} \int (1 + w(|x-y|)) \mathbf{1}_W(x) \mathbf{1}_W(y) dx dy \\
&\quad + \frac{\lambda}{\nu_d^2(W)} \nu_d(W) \\
&= \lambda^2 + \frac{\lambda^2}{\nu_d^2(W)} \int w(|x|) \int \mathbf{1}_W(x+y) \mathbf{1}_W(y) dy dx \\
&\quad + \frac{\lambda}{\nu_d(W)} \\
&= \lambda^2 + \frac{\lambda}{\nu_d(W)} + \frac{\lambda^2}{\nu_d^2(W)} \int w(|x|) \nu_d(W \cap W_x) dx \\
&< \lambda^2 + \frac{\lambda}{\nu_d(W)} + \frac{\lambda^2}{\nu_d(W)} \int w(|x|) dx \\
&= \lambda^2 + \frac{\lambda}{\nu_d(W)} + \frac{\lambda^2}{\nu_d(W)} \int db_d r^{d-1} w(r) dr.
\end{aligned}$$

Im Fall eines Neyman-Scott-Prozesses ist

$$w(r) = \frac{f(r)}{db_d r^{d-1} \lambda_p},$$

also

$$\begin{aligned}
\mathbf{E}\widehat{\lambda}^2 &< \lambda^2 + \frac{\lambda}{\nu_d(W)} + \frac{\lambda^2}{\nu_d(W)} \frac{1}{\lambda_p} \\
&= \lambda^2 (1 + 1/\mathbf{E}N + 1/\mathbf{E}N_p),
\end{aligned}$$

wobei $\mathbf{E}N$ bzw. $\mathbf{E}N_p$ die erwarteten Anzahlen von Punkten bzw. Clustern in W sind. Die entsprechenden Rechnungen für den Ohser-Schätzer führen zu dem Fehler

$$\mathbf{E}\widehat{\lambda}^2 < \lambda^2 (1 + 1/\mathbf{E}N_p).$$

Es handelt sich bei dem Ohser-Schätzer also um eine Verbesserung. Wie erwartet ist diese Verbesserung um so deutlicher, je weniger Punkte in den Clustern liegen, da der Charakter des Punktprozesses dabei immer stärker dem eines Poissonprozesses ähnelt.

3.2.2 Kenntnis des wahren λ

Angenommen, die wahre Intensität eines Punktprozesses ist bekannt. Benutzt man sie statt einer Schätzung bei der Schätzung der Paarkorrelationsfunktion, so sollte man erwarten, dass sich die Schätzung verbessert. Überraschenderweise ist das Gegenteil der Fall. Während sich der Erwartungswert des $g(r)$ -Schätzers nicht wesentlich ändert, nimmt die Varianz erheblich zu, siehe Abbildung 4.

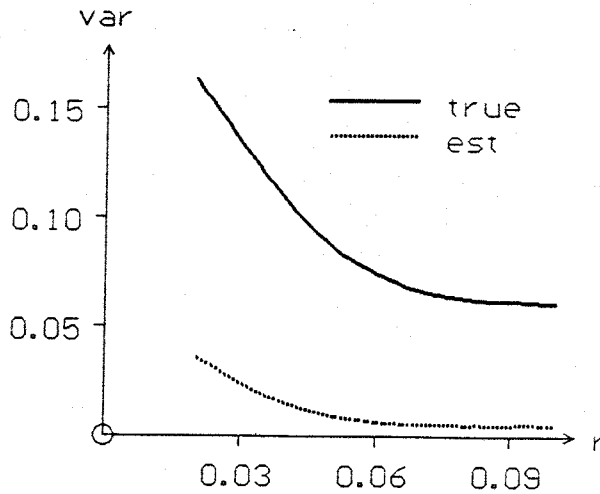


Abbildung 4: Der negative Einfluss der wahren Intensität λ^* auf die Varianz des Schätzers der Paarkorrelationsfunktion in derselben Situation wie in Abbildung 2 S. 24. Das eine Mal wurde die wahre Intensität benutzt ('true'), das andere Mal wurde sie geschätzt ('est').

Dieser Effekt rührt von der positiven Korrelation zwischen $\hat{\rho}$ und $\widehat{\lambda^2}$ her. Zufällige Schwankungen bei der Anzahl der beobachteten Punkte führen für beide Schätzer zur gleichen Tendenz. Werden z. B. mehr Punkte in W beobachtet als im Mittel erwartet, so wird in der Regel sowohl $\rho^{(2)}$ als auch λ^2 überschätzt. Die zufälligen Schwankungen der Schätzer werden teilweise kompensiert.

Hieraus ergeben sich u. a. Konsequenzen für die analytische Bestimmung der Varianz des Schätzers von $g(r)$. Üblicherweise wird hierbei nämlich vereinfachend angenommen, dass der Intensitätsschätzer konstant ist. Anhand der obigen Ausführungen wird klar, dass dieser Weg zu großen Fehlern führen kann.

3.2.3 Gewichteter Schätzer

Der gewöhnliche Intensitätsschätzer wichtet jeden beobachteten Punkt gleich stark. Das ist nicht ideal, wenn diese Schätzung für den Paarkorrelationsfunktionsschätzer benutzt wird. Dieser Schätzer ist ein Quotient, dessen Zähler all solche Punktpaare zählt, deren Zwischenpunkt Abstand ungefähr r beträgt.

Sei ein Punkt $x \in W \cap \Phi$ gegeben. Er bildet nur mit solchen Punkten $y \in W$ zählbare Paare (x, y) , die sich in unmittelbarer Nähe der Sphäre mit Radius r und Mittelpunkt x aufhalten. Befindet sich x in der Nähe einer Kante von W , so liegt nur etwa die Hälfte der Sphäre in W . Im Vergleich zu Punkten, die im Inneren von W liegen,

gibt es also im Mittel nur etwa halb so viele zählbare Punktepaare (x, y) in W . Die Schwankung des Zählers wird nur halb so stark von der Existenz oder Nichtexistenz solcher am Rand von W liegenden Punkte beeinflusst. Um die Korrelation zwischen Zähler und Nenner zu maximieren, (was die Schwankung des Quotienten minimiert,) sollten solche Punkte demnach auch im Nenner eine entsprechende Gewichtung bekommen, also auch nur zur Hälfte gezählt werden. Dieses ist die zentrale Idee des von Stoyan und Stoyan (1998) entwickelten Intensitätsschätzers $\widehat{\lambda}_S(r)$.

Sei p eine nichtnegative Gewichtsfunktion mit

$$0 < C_p(r) = \int_W p(x, r) dx < \infty, \quad r \geq 0. \quad (11)$$

Dann ist

$$\widehat{\lambda}(r) = \sum_{x \in \Phi} \mathbf{1}_W(x) p(x, r) / C_p(r)$$

ein erwartungstreuer Intensitätsschätzer. Entsprechend der vorhergehenden Darstellung sollte für die Schätzung der Paarkorrelationsfunktion eine Oberflächenwichtung benutzt werden, nämlich

$$p(x, r) = \frac{S(W \cap \partial b(x, r))}{S(\partial b(o, r))}.$$

In dieser Arbeit wird die einfachere nichtnormierte (äquivalente) Form

$$p_S(x, r) = S(W \cap \partial b(x, r))$$

gewählt. Der dazugehörige Intensitätsschätzer wird $\widehat{\lambda}_S(r)$ bezeichnet, der Schätzer der Paarkorrelationsfunktion $\widehat{g}_S(r)$. In Übereinstimmung mit Stoyan und Stoyan (1998) zeigen weitere Simulationsstudien, dass der Einsatz von $\widehat{\lambda}_S^2(r)$ gegenüber dem des gewöhnlichen $\widehat{\lambda}^2$ die Schätzvarianz für große und mittlere r erheblich reduziert. Abbildung 5 zeigt dies für ein Beispiel. Die Simulationen haben auch gezeigt, dass sich der Erwartungswert des Schätzers der Paarkorrelationsfunktion durch den Einsatz von $\widehat{\lambda}_S^2(r)$ nicht wesentlich ändert.

Bemerkung:

Die Berechnung der Konstanten $C_p(r)$ kann zu Problemen führen. Da das Integral (11) analytisch häufig nur sehr schwer zu lösen ist, wird es in diesen Fällen numerisch approximiert. Wird hierzu $p(x, r)$ auf einem regelmäßigen Gitter ausgewertet, so kommt es zu einem systematischen Fehler. Eine zufällige Störung der Gitterpunkte oder eine Realisierung eines homogenen Poissonprozesses statt des Gitters liefern bessere Ergebnisse. Ist die isotropierte Mengenkovarianzfunktion $\bar{\gamma}_W(r)$ bekannt, kann die Beziehung

$$C_S(r) = db_d r^{d-1} \bar{\gamma}_W(r)$$

benutzt werden (siehe Kapitel 2.4).

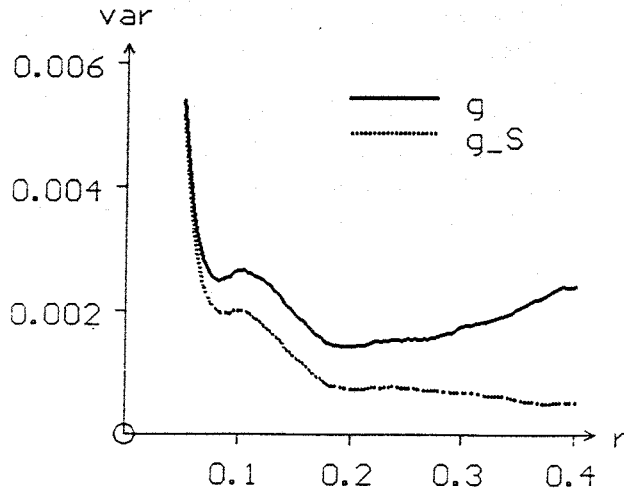


Abbildung 5: Varianzen der Schätzer der Paarkorrelationsfunktion mit gewöhnlichem ('g') und gewichtetem Intensitätsschätzer $\widehat{\lambda}_S(r)$ ('g-S') in der selben Situation wie in Abbildung 2 S. 24.

3.2.4 Erwartungstreuer λ^2 -Schätzer

Der Nenner des Schätzers der Paarkorrelationsfunktion (4) ist ein Schätzer von λ^2 . In den meisten Fällen wird hierfür einfach das Quadrat eines Intensitätsschätzers genommen. Das Quadrat eines erwartungstreuen λ -Schätzers ist i. a. aber kein erwartungstreuer Schätzer von λ^2 . Für einen Poissonprozess wurde in Kapitel 3.2.1 ein erwartungstreuer Schätzer für λ^2 vorgestellt. Für den allgemeinen Fall entwickelten Stoyan und Stoyan (1998) eine Prozedur, die zu besseren λ^2 -Schätzungen führen kann.

Ein λ^2 -Schätzer ist gegeben durch

$$\widehat{\lambda^2} = \frac{1}{\nu_d^2(W)} \sum_{x,y \in \Phi}^{\neq} \frac{\mathbf{1}_W(x)\mathbf{1}_W(y)}{h(|x-y|)}, \quad (12)$$

wobei $h(\cdot)$ eine geeignete Gewichtsfunktion ist. Wählt man für $h(\cdot)$ die Paarkorrelationsfunktion, so ergibt sich ein erwartungstreuer Schätzer

$$\begin{aligned} \mathbf{E} \frac{1}{\nu_d^2(W)} \sum_{x,y \in \Phi}^{\neq} \frac{\mathbf{1}_W(x)\mathbf{1}_W(y)}{g(|x-y|)} &= \frac{1}{\nu_d^2(W)} \int_{W^2} \varrho^{(2)}(|x-y|) \frac{\mathbf{1}_W(x)\mathbf{1}_W(y)}{g(|x-y|)} dx dy \\ &= \frac{\lambda^2}{\nu_d^2(W)} \int_{W^2} \mathbf{1}_W(x)\mathbf{1}_W(y) dx dy \\ &= \lambda^2. \end{aligned}$$

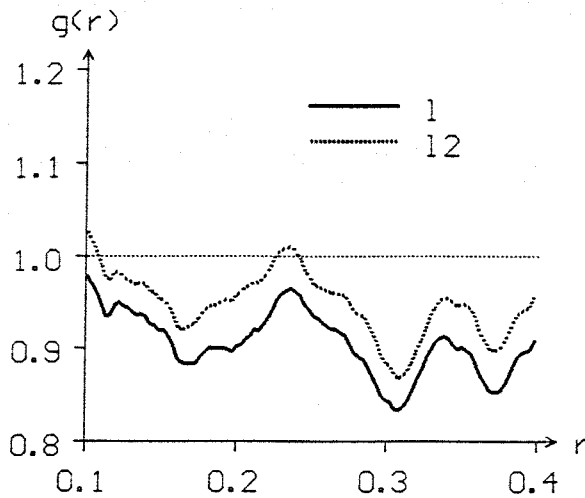


Abbildung 6: Der Einfluss des λ^2 -Schätzers $\widehat{\lambda}^2$ ('12') unter Einsatz einer ersten gewöhnlichen Schätzung von $g(r)$ auf die Schätzung der Paarkorrelationsfunktion verglichen mit dem des gewöhnlichen λ^2 -Schätzers ('1'). Der wahre Wert von $g(r)$ ist in diesem Bereich 1. Die benutzte Punktprozess ist der gleiche wie in Abbildung 2 S. 24.

(Ist $g(r) \equiv 1$, so liefert (12) den Ohser-Schätzer, siehe Kapitel 3.2.1.)

Da $g(r)$ im Allgemeinen unbekannt ist, kann man ähnlich wie in Kapitel 3.1.4 eine erste gewöhnliche Schätzung anstelle des wahren $g(r)$ einsetzen. Abbildung 6 demonstriert den Einfluss, den $\widehat{\lambda}^2$ (siehe (12)) unter Einsatz einer ersten gewöhnlichen Schätzung von $g(r)$ auf die Schätzung der Paarkorrelationsfunktion hat. Offensichtlich wird der Bias durch den obigen λ^2 -Schätzer für mittlere und große r verringert.

Es ist auch möglich, die Idee des gewichteten Intensitätsschätzers (siehe vorhergehendes Kapitel) mit dem erwartungstreuen λ^2 -Schätzer zu kombinieren

$$\widehat{\lambda}^2(r) = \sum_{x,y \in \Phi}^{\neq} \mathbf{1}_W(x) \mathbf{1}_W(y) \frac{p_S(r,x) p_S(r,y)}{g(|x-y|)} \Big/ C_S^2(r).$$

Beim dazugehörigen Schätzer der Paarkorrelationsfunktion können sich beide Effekte addieren, also sowohl die Schätzvarianz als auch der Bias können für mittlere und große r wesentlich geringer sein. Simulationsstudien haben allerdings gezeigt, dass dies nicht bei allen Punktprozessen der Fall ist. Der Einsatz einer $g(r)$ -Schätzung bei $\widehat{\lambda}^2$ erhöht etwas die Unsicherheit, was im Resultat zu einem schlechteren Schätzer führen kann.

3.2.5 Mehrere unabhängige Beobachtungsfenster

Gegeben sei die Situation aus Kapitel 3.1.5, also $W = \bigcup_{i=1}^n W_i$. Es ist

$$\hat{g}(r) = \frac{\hat{\rho}(r)}{\lambda^2} = \frac{\hat{\rho}_*(r)}{\lambda^2} = \sum_{i=1}^n \frac{\bar{\gamma}_{W_i}(r) \hat{\rho}_i(r)}{\bar{\gamma}_W(r) \lambda^2}. \quad (13)$$

Die lokale Version von $\hat{g}(r)$ sieht sehr ähnlich aus

$$\hat{g}_*(r) = \sum_{i=1}^n \frac{\bar{\gamma}_{W_i}(r) \hat{\rho}_i(r)}{\bar{\gamma}_W(r) \lambda_i^2}.$$

Der Unterschied zur globalen Version ist nur, dass der lokale $\rho^{(2)}(r)$ -Schätzer durch den lokalen λ^2 -Schätzer geteilt wird und nicht durch den globalen.

Ist die wahre Intensität λ bekannt, ist eine dritte Version denkbar

$$\hat{g}_\lambda(r) = \sum_{i=1}^n \frac{\bar{\gamma}_{W_i}(r) \hat{\rho}_i}{\bar{\gamma}_W(r) \lambda^2}.$$

(Wieder liegt der Unterschied nur in der unterschiedlichen λ^2 -Schätzung.)

Die Varianz des globalen Schätzers scheint offensichtlich zwischen den Varianzen der beiden anderen Schätzer zu liegen: steigt die Anzahl der Fenster, so strebt der globale λ^2 -Schätzer gegen das wahre λ^2 , während er für $n \rightarrow 1$ gegen den lokalen λ^2 -Schätzer strebt. Da die wahre Intensität laut Kapitel 3.2.2 zu einer erhöhten Schätzvarianz führt, ist die Varianz des lokalen Summanden $\hat{\rho}_i / \lambda_i^2$ wesentlich geringer als die desjenigen mit der wahren Intensität $\hat{\rho}_i / \lambda^2$.

Wären alle Summanden in (13) unabhängig (was nicht der Fall ist, da jeder Summand die gleiche Schätzung für λ^2 beinhaltet), so könnte man deshalb die folgende Ungleichungskette annehmen

$$\text{var} \hat{g}_* < \text{var} \hat{g} < \text{var} \hat{g}_\lambda.$$

Simulationsstudien haben gezeigt, dass zumindest

$$\text{var} \hat{g}_* \leq \text{var} \hat{g},$$

gilt. Simulationsstudien haben gezeigt, dass der lokale Schätzer umso mehr im Vorteil ist, je unterschiedlicher die Volumina der Teile des Beobachtungsfensters sind, siehe Abbildung (7). Sie zeigt die Varianzen des lokalen und des globalen Schätzers für die gleiche Situation wie in Abbildung (2) mit jeweils unterschiedlichen Beobachtungsfenstern. Diese bestehen jeweils aus dem Einheitsquader und einem Quader des Volumens 1, 2 oder 5. Die Simulationsstudien haben weiterhin gezeigt, dass sich die Erwartungswerte beider Versionen nicht wesentlich unterscheiden.

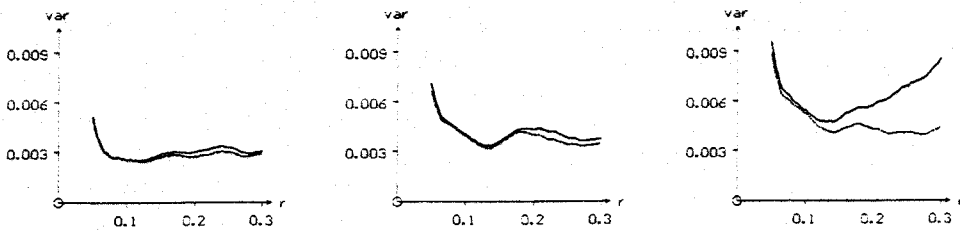


Abbildung 7: Die Varianzen des globalen (durchgezogene Linie) und des lokalen (gestrichelte Linie) Schätzers der Paarkorrelationsfunktion in der gleichen Situation wie in Abbildung 2 S. 24 mit unterschiedlichen Beobachtungsfenstern. Die Fenster setzen sich aus zwei (hinreichend weit voneinander entfernten) Teilen zusammen. Ein Teil ist jeweils der Einheitsquader, der zweite Teil ein Quader des Volumens 1 (linke Graphik), 0.5 (mittlere Graphik) oder 0.2 (rechte Graphik). Je mehr sich die Volumina der beiden Teile des Beobachtungsfensters unterscheiden, um so stärker unterscheiden sich die Schätzvarianzen des lokalen und des globalen Schätzers.

4 Variabilität der Schätzer der Paarkorrelationsfunktion

Die Variabilität eines Schätzers kann z. B. mit Hilfe von Konfidenzintervallen, der Schätzvarianz oder dem Bias charakterisiert werden. Diese Größen bestimmen unter anderem die Qualität des Schätzers. Außerdem benötigt man sie, um die Zuverlässigkeit einer gegebenen Schätzung richtig bewerten zu können.

Sind, abgesehen von einer Realisierung, nur wenige Informationen über den gegebenen Punktprozess bekannt, kann es schwierig sein, diese Größen exakt zu bestimmen. In diesem Kapitel werden einige Näherungen dazu vorgestellt und analysiert.

Das nächste Kapitel entstammt einer Veröffentlichung des Autors, siehe Sneath (1999), und wurde deshalb unverändert in diese Arbeit übernommen.

4.1 Bootstrap

Recently, bootstrap is a popular tool in many branches of statistics, also for stochastic processes. Thus it is natural to ask whether bootstrap techniques could be helpful also in point process statistics. Indeed, some authors have developed statistical procedures using bootstrap techniques, see e.g. [1], [2], [4] and [22]. All these papers deal with the estimation of the accuracy of estimators of point process characteristics. In the first three papers the estimation of variance of pair correlation function-estimators is treated. The last one presents a procedure to determine confidence regions for the intensity function of an inhomogeneous Poisson process.

The fundamental idea of bootstrap to resample given data to obtain ‘new’ pseudo data appears also in statistics of stochastic processes, in particular in the analysis of time series, see e.g. [19]. In some variants of the method, called the blockwise bootstrap, the time series is partitioned into several parts, which are then resampled. A similar idea is also applied in [10] in the statistical analysis of a planar random set. Clearly, the partitioning procedure can also be adapted to point process statistics. However, partition can destroy point structures or add new artificial structures to the point pattern. In the one-dimensional case the error resulting from this loss of information may be still acceptable, but in higher dimensions it will be serious. Thus in spatial point process statistics another method is used which is quite similar to the application of bootstrap in case of classical statistics: the points of the process (including their places, which are assumed to be pairwise different) are resampled. The pseudo pattern then consists of n points x_1^*, \dots, x_n^* which are obtained by sampling randomly with replacement n times from the original data $\{x_1, \dots, x_n\}$.

Naturally, the pseudo patterns generated by this method have always multiple points. Thus they have a character different to that of the original, which does not have multiple points.

Consequently, it would be surprising if quantities of such point processes would produce good estimators for quantities of the original point process.

This paper analyses the pointwise resampling technique for some examples of point process statistics. Section 4.1.1 discusses the main ideas of the paper [2] which presents a procedure for estimating the standard error of an estimator of the pair correlation function. In Section 4.1.2 a method (drawn from [4]) to determine confidence intervals for the intensity function of an inhomogeneous Poisson process is considered. Finally, an easier method is presented which yields confidence regions for the intensity function of an inhomogeneous Poisson process without bootstrap.

4.1.1 Variance of estimators of the product density

This part discusses the main ideas of [2], where bootstrap techniques are used to approximate the standard error of a pair correlation function estimator. The calculations are presented in an abridged form; the complete calculations are given in the Appendix.

Fundamentals

Given a stationary and isotropic point process Φ and an estimator of $\rho^{(2)}(r)$

$$\hat{\rho}(r) = \frac{1}{C_S(r)} \sum_{x,y \in \Phi \cap W}^{\neq} k_h(r - \|x - y\|).$$

The situation can be generalized to the case of any 'two-point estimator'

$$\hat{\theta} = \sum_{x,y \in \Phi}^{\neq} f(x,y)$$

with f being symmetrical in its arguments and of the form

$$f(x,y) = \mathbf{1}_W(x)\mathbf{1}_W(y)h(x,y)$$

with some function h . As the special form of f leading to $\hat{\rho}(r)$ is unimportant, the following calculations are carried out for a general $\hat{\theta}$.

The quantity of interest is the variance of $\hat{\theta}$ which is given by

$$\text{var}\hat{\theta} = \mathbf{E}\hat{\theta}^2 - (\mathbf{E}\hat{\theta})^2 = s_4 + 4s_3 + 2s_2 - (\mathbf{E}\hat{\theta})^2 \quad (14)$$

with

$$s_i = \int \varrho^{(i)}(x_1, \dots, x_i) f(x_1, x_2) f(x_{i-1}, x_i) dx_1 \dots dx_i,$$

where $\varrho^{(i)}$ is the i th order product density function of Φ , see the Appendix.

Bootstrap version of $\hat{\theta}$

Assume that a sample of Φ is given which consists of n points x_1, \dots, x_n in the observation window W . It is resampled N times to obtain N 'new' point patterns. Each pseudo pattern consists of n points x_1^*, \dots, x_n^* which are obtained by sampling randomly with replacement n times from $\{x_1, \dots, x_n\}$. Thus it happens that in the pseudo samples some points of the original point pattern do not occur while others occur twice or even more. Let the number of occurrences of x_i in the k th sample be $w_k(i)$. Then the k th sample can be represented by the vector $w_k = (w_k(1), \dots, w_k(n))$ which has a multinomial distribution. This distribution depends only on n . In the limiting case $n \rightarrow \infty$ the components $w_k(i)$ of w_k are independent and Poisson distributed with mean $\mu = 1$.

The bootstrap estimate for the k th pseudo sample is

$$\hat{\theta}_k^* = \sum_{i,j=1}^n f(x_i, x_j) w_k(i) w_k(j), \quad k = 1, \dots, N$$

where the summation goes over all pairs (i, j) with $i \neq j$. The variance of $\hat{\theta}$ is estimated by the usual variance estimator corresponding to the $\hat{\theta}_k^*$'s,

$$\widehat{v}_N^* = \frac{1}{N-1} \sum_{k=1}^N \left(\hat{\theta}_k^* - \frac{1}{N} \sum_{i=1}^N \hat{\theta}_i^* \right)^2.$$

Since the $\widehat{\theta}_k^*$ are (conditionally on x_1, \dots, x_n) independent and identically distributed, it is

$$\begin{aligned}
\lim_{N \rightarrow \infty} \widehat{v}_N^* &= \mathbf{V}\widehat{\theta}_1^* \\
&= \mathbf{E}\widehat{\theta}_1^{*2} - (\mathbf{E}\widehat{\theta}_1^*)^2 \\
&= \alpha_4 \sum_{i,j,k,l=1}^n \neq f(x_i, x_j)f(x_k, x_l) \\
&\quad + 4\alpha_3 \sum_{i,j,k=1}^n \neq f(x_i, x_j)f(x_i, x_k) \\
&\quad + 2\alpha_2 \sum_{i,j=1}^n \neq (f(x_i, x_j))^2
\end{aligned}$$

with

$$\begin{aligned}
\alpha_4 &= \left[\mathbf{E}w_1(1)w_1(2)w_1(3)w_1(4) - (\mathbf{E}w_1(1)w_1(2))^2 \right] \\
\alpha_3 &= \left[\mathbf{E}(w_1(1))^2w_1(2)w_1(3) - (\mathbf{E}w_1(1)w_1(2))^2 \right] \\
\alpha_2 &= \left[\mathbf{E}(w_1(1)w_1(2))^2 - (\mathbf{E}w_1(1)w_1(2))^2 \right],
\end{aligned}$$

where the expectations are conditionally on fixed x_1, \dots, x_n . All the α_i can be calculated numerically and depend only on n (see the Appendix). Thus the result of the whole bootstrap procedure for $N \rightarrow \infty$ can be simply obtained by direct computation.

*Expectation of \widehat{v}_N^**

The futility of \widehat{v}_N^* is demonstrated by the fact that it does neither estimate what is hoped (the variance of $\widehat{\theta}$) nor a multiple with a fixed factor. To show this, the unconditional expectation of \widehat{v}_N^* is determined, see the Appendix. Since the result is not very transparent, here an approximation is given which makes it possible to characterize the quality of \widehat{v}_N^* .

Assume that the $w_k(i)$ are independent and Poisson distributed with parameter $\mu = 1$; this simplifying assumption is exact in the limiting case $n \rightarrow \infty$, see above. This leads to a result which is close to the exact value for large n and is easy to interpret. By the way, the simplification is equivalent to replacement of n by n^* in each pseudo sample where n^* is a Poisson distributed number with mean $\mu = n$. In this scheme each pseudo sample consists of a random number of points. The result is

$$\lim_{N \rightarrow \infty} \mathbf{E}\widehat{v}_N^* = \mathbf{E} \lim_{N \rightarrow \infty} \widehat{v}_N^* = 4s_3 + 6s_2, \quad (15)$$

see the Appendix, while the desired result, given by (14), is

$$\text{var}\widehat{\theta} = s_4 + 4s_3 + 2s_2 - (\mathbf{E}\widehat{\theta})^2.$$

4.1.2 Confidence regions for the intensity function of an inhomogeneous Poisson process

The paper [4] presents a procedure which uses bootstrap techniques to determine confidence regions for the intensity function of an inhomogeneous Poisson process. The confidence regions are estimated using a kernel estimator. The following discusses the main idea of that paper and shows that, as above, it is not necessary to carry out the bootstrap procedure.

Fundamentals

For simplicity, the following calculations are carried out for an one-dimensional point process, but they could be easily generalized to higher-dimensional processes.

Consider an inhomogeneous Poisson point process Φ with unknown intensity function $\lambda(x)$ in the interval $(0, 1)$, with points $0 < x_1 \leq x_2 \leq \dots \leq x_n < 1$. A kernel estimator for $\lambda(x)$ is used as

$$\widehat{\lambda}(x) = \sum_{i=1}^n k_h(|x - x_i|), \quad x \in (0, 1),$$

where k_h is a kernel function with bandwidth h . Define

$$T(x) = \frac{\widehat{\lambda}(x) - \mathbf{E}\widehat{\lambda}(x)}{\sqrt{\widehat{\lambda}(x)}}, \quad 0 < x < 1,$$

and, for a given α with $0 \leq \alpha \leq 1$,

$$t_\alpha(x) = \min_{t \in \mathbb{R}^+} \{t : \mathbf{P}\{|T(x)| \leq t\} \geq 1 - \alpha\}, \quad 0 < x < 1.$$

Then an estimate of a confidence region for $\lambda(x)$ of level $1 - \alpha$ is the interval

$$C(x) = \left[\widehat{\lambda}(x) - t_\alpha(x) \sqrt{\widehat{\lambda}(x)}, \widehat{\lambda}(x) + t_\alpha(x) \sqrt{\widehat{\lambda}(x)} \right], \quad 0 < x < 1,$$

where the left border is set on 0 if it is negative.

Bootstrap versions

Since the distribution of T is not available (because the intensity function is unknown) it is approximated by simulation of pseudo data, see [4]. A set of pseudo data is obtained by drawing x_1^*, \dots, x_n^* by sampling randomly with replacement n^* times from $\{x_1, \dots, x_n\}$, where n^* has a Poisson distribution with mean n (this is method 2 in [4] and similar to the simplified case in Section 4.1.1). The number of occurrences of x_i in the k th sample is a random variable, denoted as above by $w_k(i)$. All the $w_k(i)$ are independent and Poisson distributed with mean $\lambda = 1$ for $i = 1, \dots, n$.

For given α with $0 \leq \alpha \leq 1$ the bootstrap versions of the quantities defined above are

$$\begin{aligned}\widehat{\lambda}_k^*(x) &= \sum_{i=1}^n k_h(|x - x_i|) w_k(i), \\ T_k^*(x) &= \frac{\widehat{\lambda}_k^*(x) - \widehat{\lambda}(x)}{\sqrt{\widehat{\lambda}_k^*(x)}}, \quad x \in (0, 1), \\ t_\alpha^*(x) &= \min_{t \in \mathbb{R}^+} \{t : \mathbf{P}^* \{|T^*(x)| \leq t\} \geq 1 - \alpha\}, \\ C^*(x) &= \left[\widehat{\lambda}(x) - t_\alpha^*(x) \sqrt{\widehat{\lambda}(x)}, \widehat{\lambda}(x) + t_\alpha^*(x) \sqrt{\widehat{\lambda}(x)} \right],\end{aligned}$$

where $\mathbf{P}^*(\cdot) = \mathbf{P}(\cdot | \{x_1, \dots, x_n\})$ is the distribution conditionally on $\{x_1, \dots, x_n\}$.

The determination of $t_\alpha^*(x)$ can be carried out by simulation. However, a faster and simpler possibility uses the well-known fact that the sum of independent Poisson distributed random variables is also Poisson distributed. It is demonstrated here for the simple rectangular kernel function

$$k_h(x) = \frac{1}{2h} \cdot \mathbf{1}_{[-h, h]}(x).$$

For other kernels, similar calculations are possible. Let $p(x)$ be the number of observed points in the interval $[x - h, x + h]$. Then its bootstrap version $p^*(x)$ is a random variable which is Poisson distributed with mean $p(x)$. Its cumulative distribution function is denoted by \mathbf{F}^* . Thus, for given α ,

$$\begin{aligned}t_\alpha^*(x) &= \min_{t \in \mathbb{R}^+} \{t : \mathbf{P}^* \{|T_1^*(x)| \leq t\} \geq 1 - \alpha\} \\ &= \min_{t \in \mathbb{R}^+} \left\{ t : \mathbf{P}^* \left\{ \left| \sum_{i=1}^{p(x)} w_1(i) - a_x(t) \right| \leq b_x(t) \right\} \geq 1 - \alpha \right\} \\ &= \min_{t \in \mathbb{R}^+} \{t : \mathbf{F}^*(a_x(t) + b_x(t)) - \mathbf{F}^*(a_x(t) - b_x(t)) \geq 1 - \alpha\},\end{aligned}$$

with

$$\begin{aligned} a_x(t) &= p(x) + ht^2, \\ b_x(t) &= t\sqrt{2hp(x) + h^2t^2} \end{aligned}$$

The calculations are of an elementary nature using that

$$\widehat{\lambda}_k^* = \frac{1}{2h} \sum_{i=1}^{p(x)} w_k(i).$$

Confidence regions without bootstrap

Of course, in case of an inhomogeneous Poisson process it is easy to build confidence regions without the bootstrap methodology. Assume that the intensity function $\lambda(x)$ is approximately linear in the interval $[x - h, x + h]$. Then $2h\widehat{\lambda}(x)$ using the rectangular kernel is Poisson distributed with mean $2h\lambda(x)$. (The rectangular kernel could be replaced by another kernel; then the corresponding calculations become a bit more difficult.) Therefore, known confidence regions for the Poisson parameter can be used, see for example [30]. Thus it is easy to build the desired confidence region for $\lambda(x)$.

This result corresponds to a general observation. If a parametric statistic problem is given, then parametric estimators lead usually to better results than bootstrap techniques.

Appendix

Here the derivation of some equations of Section 4.1.1 is given.

Equation (14)

For

$$\widehat{\theta} = \sum_{x,y \in \Phi}^{\neq} f(x,y) = \sum_{x,y \in \Phi}^{\neq} \mathbf{1}_W(x)\mathbf{1}_W(y)h(x,y)$$

it is

$$\begin{aligned} \mathbf{E}\widehat{\theta}^2 &= \mathbf{E}\left(\sum_{x,y \in \Phi}^{\neq} f(x,y)\right)^2 \\ &= \mathbf{E}\sum_{w,x,y,z \in \Phi}^{\neq} f(w,x)f(y,z) \end{aligned}$$

$$\begin{aligned}
& + 4\mathbf{E} \sum_{x,y,z \in \Phi}^{\neq} f(x,y)f(x,z) \\
& + 2\mathbf{E} \sum_{x,y \in \Phi}^{\neq} (f(x,y))^2 \\
= & \int \varrho^{(4)}(x_1, x_2, x_3, x_4) f(x_1, x_2) f(x_3, x_4) dx_1 dx_2 dx_3 dx_4 \\
& + 4 \int \varrho^{(3)}(x_1, x_2, x_3) f(x_1, x_2) f(x_1, x_3) dx_1 dx_2 dx_3 \\
& + 2 \int \varrho^{(2)}(x_1, x_2) (f(x_1, x_2))^2 dx_1 dx_2 \\
= & s_4 + 4s_3 + 2s_2
\end{aligned}$$

with

$$s_i = \int \varrho^{(i)}(x_1, \dots, x_i) f(x_1, x_2) f(x_{i-1}, x_i) dx_1 \dots dx_i.$$

Equation (15)

For

$$\widehat{\theta}_1^* = \sum_{i,j=1}^n \neq f(x_i, x_j) w_1(i) w_1(j)$$

it is

$$\begin{aligned}
\mathbf{E} \widehat{\theta}_1^{*2} & = \mathbf{E} \widehat{\theta}_1^{*2} - \mathbf{E} \widehat{\theta}_1^* \widehat{\theta}_2^* \\
& = \mathbf{E} \sum_{i,j,k,l=1}^n \neq f(x_i, x_j) f(x_k, x_l) w_1(i) w_1(j) w_1(k) w_1(l) \\
& \quad + 4\mathbf{E} \sum_{i,j,k=1}^n \neq f(x_i, x_j) f(x_i, x_k) (w_1(i))^2 w_1(j) w_1(k) \\
& \quad + 2\mathbf{E} \sum_{i,j=1}^n \neq (f(x_i, x_j))^2 (w_1(i) w_1(j))^2 \\
= & \sum_{i,j,k,l=1}^n \neq f(x_i, x_j) f(x_k, x_l) \mathbf{E} w_1(i) w_1(j) w_1(k) w_1(l) \\
& \quad + 4 \sum_{i,j,k=1}^n \neq f(x_i, x_j) f(x_i, x_k) \mathbf{E} (w_1(i))^2 w_1(j) w_1(k) \\
& \quad + 2 \sum_{i,j=1}^n \neq (f(x_i, x_j))^2 \mathbf{E} (w_1(i) w_1(j))^2
\end{aligned}$$

$$\begin{aligned}
&= \mathbf{E}w_1(1)w_1(2)w_1(3)w_1(4) \sum_{i,j,k,l=1}^n \neq f(x_i, x_j)f(x_k, x_l) \\
&+ 4\mathbf{E}(w_1(1))^2w_1(2)w_1(3) \sum_{i,j,k=1}^n \neq f(x_i, x_j)f(x_i, x_k) \\
&+ 2\mathbf{E}(w_1(1)w_1(2))^2 \sum_{i,j=1}^n \neq (f(x_i, x_j))^2
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{E}\widehat{\theta}_1^* \widehat{\theta}_2^* &= \mathbf{E} \sum_{i,j,k,l=1}^n \neq f(x_i, x_j)f(x_k, x_l)w_1(i)w_1(j)w_2(k)w_2(l) \\
&+ 4\mathbf{E} \sum_{i,j,k=1}^n \neq f(x_i, x_j)f(x_i, x_k)w_1(i)w_1(j)w_2(i)w_2(k) \\
&+ 2\mathbf{E} \sum_{i,j=1}^n \neq (f(x_i, x_j))^2 w_1(i)w_1(j)w_2(i)w_2(j) \\
&= (\mathbf{E}w_1(1)w_1(2))^2 \sum_{i,j,k,l=1}^n \neq f(x_i, x_j)f(x_k, x_l) \\
&+ 4(\mathbf{E}w_1(1)w_1(2))^2 \sum_{i,j,k=1}^n \neq f(x_i, x_j)f(x_i, x_k) \\
&+ 2(\mathbf{E}w_1(1)w_1(2))^2 \sum_{i,j=1}^n \neq (f(x_i, x_j))^2.
\end{aligned}$$

This yields

$$\begin{aligned}
\lim_{N \rightarrow \infty} \widehat{v}_N^* &= \left[\mathbf{E}w_1(1)w_1(2)w_1(3)w_1(4) - (\mathbf{E}w_1(1)w_1(2))^2 \right] \cdot \\
&\quad \cdot \sum_{w,x,y,z \in \Phi} \neq f(w, x)f(y, z) \\
&+ 4 \left[\mathbf{E}(w_1(1))^2w_1(2)w_1(3) - (\mathbf{E}w_1(1)w_1(2))^2 \right] \cdot \\
&\quad \cdot \sum_{x,y,z \in \Phi} \neq f(x, y)f(x, z) \\
&+ 2 \left[\mathbf{E}(w_1(1)w_1(2))^2 - (\mathbf{E}w_1(1)w_1(2))^2 \right] \cdot \\
&\quad \cdot \sum_{x,y \in \Phi} \neq (f(x, y))^2
\end{aligned}$$

$$\begin{aligned}
&= \alpha_4 \sum_{i,j,k,l=1}^n \neq f(x_i, x_j) f(x_k, x_l) \\
&\quad + 4\alpha_3 \sum_{i,j,k=1}^n \neq f(x_i, x_j) f(x_i, x_k) \\
&\quad + 2\alpha_2 \sum_{i,j=1}^n \neq (f(x_i, x_j))^2
\end{aligned}$$

with

$$\begin{aligned}
\alpha_4 &= (-4n^2 + 10n - 6)/n^3 \\
\alpha_3 &= (n^3 - 7n^2 + 12n - 6)/n^3 \\
\alpha_2 &= (3n^3 - 11n^2 + 14n - 6)/n^3
\end{aligned} \tag{16}$$

Equation (15)

The expectation value of \widehat{v}^* is

$$\begin{aligned}
\mathbf{E}\widehat{v}^* &= \alpha_4 \int \varrho^{(4)}(x_1, x_2, x_3, x_4) f(x_1, x_2) f(x_3, x_4) dx_1 dx_2 dx_3 dx_4 \\
&\quad + 4\alpha_3 \int \varrho^{(3)}(x_1, x_2, x_3) f(x_1, x_2) f(x_1, x_3) dx_1 dx_2 dx_3 \\
&\quad + 2\alpha_2 \int \varrho^{(2)}(x_1, x_2) (f(x_1, x_2))^2 dx_1 dx_2 \\
&= s_4 \alpha_4 + 4s_3 \alpha_3 + 2s_2 \alpha_2
\end{aligned}$$

In the limiting case ($n \rightarrow \infty$) it is

$$\mathbf{E}\widehat{v}^* = 4s_3 + 6s_2$$

(see Equation (16)).

4.2 Quotientenschätzer

Wie schon oft vorher beschrieben, ist der Schätzer der Paarkorrelationsfunktion ein zufälliger Quotient. Er ist häufig von der Form

$$q = \frac{A}{B^2},$$

wobei A und B zwei Zufallsvariablen sind. Die Verteilungsparameter solcher zufälligen Quotienten lassen sich im Allgemeinen nur schwer berechnen. Im Fall der Paarkorrelationsfunktion wird der Erwartungswert häufig durch den Quotienten der einzelnen Erwartungswerte approximiert

$$\mathbf{E} \frac{A}{B^2} \approx \frac{\mathbf{E}A}{\mathbf{E}B^2}.$$

Die Erfahrung zeigt, dass dieses Vorgehen nicht zu großen Fehlern führt.

Anders ist das bei der Varianz. Für den Schätzer der Paarkorrelationsfunktion wird sie manchmal durch die Varianz des Zählers approximiert, da vermutet wird, dass der Nenner nur in geringem Maße zur Variabilität des Quotienten beiträgt. Dieses ist ein gravierender Fehler. Gerade in diesem Fall treten nicht unwesentliche Korrelationen zwischen dem Zähler und Nenner auf, die bei einer solchen Approximation unberücksichtigt bleiben, siehe Kapitel 3.2.2.

Zwei Astrophysiker, Landy und Szalay (1993), haben ein einfaches Verfahren entwickelt, das es ermöglicht, diese Korrelation mit in die Rechnungen einzubeziehen. Ursprünglich war es für die Berechnung des Erwartungswertes und der Varianz des Schätzers der Paarkorrelationsfunktion gedacht. Es kann aber im Grunde auch zur Berechnung von Momenten beliebiger zufälliger Quotienten angewandt werden.

Im Folgenden werden mit Hilfe dieses Verfahrens der Erwartungswert und die Varianz einer Reihe verschiedener Schätzer der Paarkorrelationsfunktion bestimmt. Ein Vergleich der erhaltenden Abschätzungen der mittleren quadratischen Fehler liefert dann den in diesem Sinn optimalen Schätzer.

4.2.1 Reihenentwicklung

Gegeben sei also der zufällige Quotient

$$q = \frac{A}{B^2}, \quad (17)$$

wobei A und B Zufallsvariablen sind. Die Variabilität von A und B sei durch die Zufallsvariablen α und β ausgedrückt

$$\begin{aligned} A &= (1 + \alpha)EA \\ B &= (1 + \beta)EB \end{aligned}$$

mit $\mathbf{E}\alpha = \mathbf{E}\beta = 0$. Dann ist

$$\begin{aligned} \mathbf{E}q &= \frac{\mathbf{E}A}{(\mathbf{E}B)^2} \mathbf{E} \left(\frac{1 + \alpha}{(1 + \beta)^2} \right), \\ \mathbf{var}q &= \left(\frac{\mathbf{E}A}{(\mathbf{E}B)^2} \right)^2 \mathbf{var} \left(\frac{1 + \alpha}{(1 + \beta)^2} \right). \end{aligned}$$

Die Nenner der zufälligen Quotienten sollen nun in eine Potenzreihe bzgl. $1 - (1 + \beta)^2 = -2\beta - \beta^2$ entwickelt werden. Allgemein gilt für $|x| < 1$ bekanntlich

$$\frac{1}{1 - x} = \sum_{i=0}^{\infty} x^i.$$

Ist also die Variabilität von $-2\beta - \beta^2$ nicht zu groß, so ergibt sich ungefähr

$$\begin{aligned} \mathbf{E} \left(\frac{1 + \alpha}{(1 + \beta)^2} \right) &= \mathbf{E} \left((1 + \alpha) \sum_{n=0}^{\infty} (-2\beta - \beta^2)^n \right) \\ &= \sum_{n=0}^{\infty} \mathbf{E}(-2\beta - \beta^2)^n + \sum_{n=0}^{\infty} \mathbf{E}\alpha(-2\beta - \beta^2)^n. \end{aligned}$$

(Es ist $|-2\beta - \beta^2| < 1$ genau dann, wenn $0 \leq B < \sqrt{2\mathbf{E}B}$.) Dieser Ausdruck wird nun insofern vereinfacht, dass nur Terme von maximal zweiter Ordnung berücksichtigt werden, d. h. $\mathbf{E}1$, $\mathbf{E}\alpha$ ($= 0$), $\mathbf{E}\beta$ ($= 0$), $\mathbf{E}\alpha\beta$, $\mathbf{E}\alpha^2$ und $\mathbf{E}\beta^2$. Es handelt sich hierbei um eine starke Vereinfachung, weshalb das Ergebnis auch nur mehr oder weniger, abhängig vom gegebenem Punktprozess, als Näherung angesehen werden kann. Es folgt

$$\begin{aligned} \mathbf{E} \left(\frac{1 + \alpha}{(1 + \beta)^2} \right) &= \sum_{n=0}^{\infty} \mathbf{E}(-2\beta - \beta^2)^n + \sum_{n=0}^{\infty} \mathbf{E}\alpha(-2\beta - \beta^2)^n \\ &= \mathbf{E}1 + \mathbf{E}(-2\beta - \beta^2) + \mathbf{E}(-2\beta - \beta^2)^2 + \mathbf{E}\alpha + \mathbf{E}(\alpha(-2\beta - \beta^2)) \\ &= \mathbf{E}1 - 2\mathbf{E}\beta - \mathbf{E}\beta^2 + 4\mathbf{E}\beta^2 + \mathbf{E}\alpha - 2\mathbf{E}\alpha\beta \\ &= 1 - 2\mathbf{E}\alpha\beta + 3\mathbf{E}\beta^2. \end{aligned}$$

Auf die gleiche Art und Weise folgt für die Varianz

$$\text{var} \left(\frac{1 + \alpha}{(1 + \beta)^2} \right) = \mathbf{E}\alpha^2 - 4\mathbf{E}\alpha\beta + 4\mathbf{E}\beta^2.$$

(Für den mittleren quadratischen Fehler ergibt sich das Gleiche wie für die Varianz.) Es gilt

$$\begin{aligned} \mathbf{E}\alpha^2 &= \frac{\text{var} A}{(\mathbf{E}A)^2}, \\ \mathbf{E}\beta^2 &= \frac{\text{var} B}{(\mathbf{E}B)^2}, \\ \mathbf{E}\alpha\beta &= \frac{\mathbf{E}((A - \mathbf{E}A)(B - \mathbf{E}B))}{(\mathbf{E}A)(\mathbf{E}B)}. \end{aligned}$$

Man benötigt also die ersten beiden Momente und die Kovarianz der Zufallsvariablen A und B . Diese werden im Folgenden für Schätzer der Paarkorrelationsfunktion der Art (17) ausführlich berechnet. Es wird dabei immer ein Neyman-Scott-Prozess vorausgesetzt.

Im Anhang finden sich die entsprechenden Berechnungen für weitere, ähnlich wie q gebaute Schätzer.

4.2.2 Intensitätsschätzer

Der Nenner des Schätzers der Paarkorrelationsfunktion (4) ist üblicherweise von der Art

$$(B(r))^2 = \left(\sum_{x \in \Phi \cap W} \frac{p(x, r)}{C_p(r)} \right)^2 \quad (18)$$

mit $C_p(r) = \int_W p(x, r) dx$. Ist $p(x, r) \equiv 1$, so führt das zum primitiven Intensitätsschätzer $\hat{\lambda}$. Im Fall $p(x, r) = p_S(x, r)$ wird die Intensität durch $\widehat{\lambda}_S(r)$ geschätzt, siehe Kapitel 3.2.3. Neben diesen beiden speziellen p -Funktionen sollen hier zunächst auch andere zugelassen sein.

Um eine bessere Vergleichbarkeit zu erreichen, sei $p(x, r)$ so, daß $C_p(r) = \nu_d(W)$. Das kann immer erreicht werden, indem man an Stelle von $p(x, r)$ die Funktion

$$p^*(x, r) = p(x, r) \frac{\nu_d(W)}{C_p(r)}$$

benutzt. (Diese Konvention bedeutet keine Einschränkung in der Wahl der p -Funktion.)

Es ist

$$\begin{aligned} \mathbf{E}B(r) &= \frac{\lambda}{\nu_d(W)} \int_W p(x, r) dx \\ &= \lambda. \end{aligned}$$

Weiterhin gilt unter Verwendung der Vereinfachungen aus Kapitel 2.4

$$\begin{aligned} \mathbf{E}(B(r))^2 &= \mathbf{E} \sum_{x, y \in \Phi \cap W}^{\neq} \frac{p(x, r)}{C_p(r)} \frac{p(y, r)}{C_p(r)} + \mathbf{E} \sum_{x \in \Phi \cap W} \frac{p^2(x, r)}{C_p^2(r)} \\ &= \frac{1}{\nu_d^2(W)} \int_{W^2} \rho^{(2)}(x, y) p(x, r) p(y, r) dx dy \\ &\quad + \frac{\lambda}{\nu_d^2(W)} \int_W p^2(x, r) dx \\ &\approx \frac{\lambda^2}{\nu_d^2(W)} \left(\int_W p(x, r) dx \right)^2 + \frac{\lambda}{\nu_d^2(W)} \mathbf{I}p^2 \\ &= \lambda^2 + \frac{\lambda}{\nu_d^2(W)} \mathbf{I}p^2 \end{aligned}$$

mit

$$\mathbf{I}f = \int_W f(x, r) dx.$$

Für die Varianz gilt dann

$$\text{var}B(r) = \frac{\lambda p^2}{\nu_d^2(W)}$$

und

$$\mathbf{E}\beta^2 = \frac{Ip^2}{\lambda\nu_d^2(W)}.$$

4.2.3 Schätzer der Produktdichtefunktion

Der Zähler des Schätzers der Paarkorrelationsfunktion ist ein Schätzer der Produktdichtefunktion wie

$$\begin{aligned} A(r) &= \frac{1}{db_d r^{d-1} \bar{\gamma}_W(r)} \sum_{x,y \in \Phi \cap W}^{\neq} k_h(r - |x - y|) \\ &= \frac{1}{C_S(r)} \sum_{x,y \in \Phi \cap W}^{\neq} k_h(r - |x - y|). \end{aligned} \quad (19)$$

In Kapitel 3.1 wurde gezeigt, dass sein Verhalten nicht wesentlich von der Wahl der Randkorrektur oder des Faktors $1/r^{d-1}$ bzw. $1/|x - y|^{d-1}$ abhängt. Die folgenden Berechnungen gelten also näherungsweise auch für die entsprechenden anderen Versionen des Schätzers.

Ist h wesentlich kleiner als r (was in der Regel der Fall sein sollte), so gilt

$$\begin{aligned} \mathbf{E}A(r) &= \frac{1}{C_S(r)} \int_{W^2} \varrho^{(2)}(x, y) k_h(r - |x - y|) dx dy \\ &\approx \frac{1}{C_S(r)} \varrho^{(2)}(r) \int_{W^2} k_h(r - |x - y|) dx dy \\ &= \frac{1}{C_S(r)} \varrho^{(2)}(r) \int k_h(r - |x|) \int \mathbf{1}_W(x + y) \mathbf{1}_W(y) dy dx \\ &= \frac{1}{C_S(r)} \varrho^{(2)}(r) \int k_h(r - |x|) V(W \cap W_{-x}) dx \\ &= \frac{1}{C_S(r)} \varrho^{(2)}(r) \int_{r-h}^{r+h} k_h(v) db_d v^{d-1} \bar{\gamma}_W(v) dv \\ &\approx \frac{db_d r^{d-1} \bar{\gamma}_W(r)}{C_S(r)} \varrho^{(2)}(r) \int_{r-h}^{r+h} k_h(v) dv \\ &= \varrho^{(2)}(r). \end{aligned} \quad (20)$$

Weiterhin gilt

$$(A(r))^2 = \frac{1}{C_S^2(r)} \sum_{x_1, x_2, x_3, x_4 \in \Phi \cap W}^{\neq} k_h(r - |x_1 - x_2|) k_h(r - |x_3 - x_4|)$$

$$\begin{aligned}
& + \frac{4}{C_S^2(r)} \sum_{x_1, x_2, x_3 \in \Phi \cap W}^{\neq} k_h(r - |x_1 - x_2|) k_h(r - |x_1 - x_3|) \\
& + \frac{2}{C_S^2(r)} \sum_{x_1, x_2 \in \Phi \cap W}^{\neq} k_h^2(r - |x_1 - x_2|)
\end{aligned}$$

und somit

$$\begin{aligned}
\mathbf{E}(A(r))^2 &= \frac{1}{C_S^2(r)} \int_{W^4} \varrho^{(4)}(x_1, x_2, x_3, x_4) k_h(r - |x_1 - x_2|) \\
&\quad k_h(r - |x_3 - x_4|) dx_1 dx_2 dx_3 dx_4 \\
&+ \frac{4}{C_S^2(r)} \int_{W^3} \varrho^{(3)}(x_1, x_2, x_3) k_h(r - |x_1 - x_2|) \\
&\quad k_h(r - |x_1 - x_3|) dx_1 dx_2 dx_3 \\
&+ \frac{2}{C_S^2(r)} \int_{W^2} \varrho^{(2)}(x_1, x_2) k_h^2(r - |x_1 - x_2|) dx_1 dx_2.
\end{aligned} \tag{21}$$

Man kann diese Summe vereinfachen. Es ist

$$\begin{aligned}
\varrho^{(4)}(x_1, x_2, x_3, x_4) &= \lambda^4 \left(1 + w_{12}^{(2)} + w_{13}^{(2)} + w_{14}^{(2)} \right. \\
&\quad + w_{23}^{(2)} + w_{24}^{(2)} + w_{34}^{(2)} \\
&\quad + w_{123}^{(3)} + w_{124}^{(3)} + w_{134}^{(3)} + w_{234}^{(3)} \\
&\quad + w_{12}^{(2)} w_{34}^{(2)} + w_{13}^{(2)} w_{24}^{(2)} + w_{14}^{(2)} w_{23}^{(2)} \\
&\quad \left. + w_{1234}^{(4)} \right),
\end{aligned}$$

siehe Kapitel 2.4. Die Punktepaare (x_1, x_2) und (x_3, x_4) haben jeweils einen Zwischenpunkt Abstand von ungefähr r . Die Punkte aller anderen Punktepaare sind im gewissen Sinne „gleichverteilt“ in W , ihre Zwischenpunkt Abstände sind unbestimmt. Ähnliches gilt auch für alle Punktetripel und das Quadrupel, bei denen sich auch jeweils mindestens ein Punkt „gleichverteilt“ in W aufhält. Gemäß Kapitel 2.4 können also alle $w^{(\cdot)}$ -Funktionen mit Ausnahme von $w_{12}^{(2)}$ und $w_{34}^{(2)}$ vernachlässigt werden. Es bleibt

$$\begin{aligned}
\varrho^{(4)}(x_1, x_2, x_3, x_4) &\approx \lambda^4 \left(1 + w_{12}^{(2)} + w_{34}^{(2)} + w_{12}^{(2)} w_{34}^{(2)} \right) \\
&= \lambda^2 \left(1 + w_{12}^{(2)} \right) \lambda^2 \left(1 + w_{34}^{(2)} \right) \\
&= \varrho^{(2)}(x_1, x_2) \varrho^{(2)}(x_3, x_4).
\end{aligned}$$

Damit ist

$$\frac{1}{C_S^2(r)} \int_{W^4} \varrho^{(4)}(x_1, \dots, x_4) k_h(r - |x_1 - x_2|) k_h(r - |x_3 - x_4|) dx_1, \dots, dx_4$$

$$\begin{aligned}
&= \frac{1}{C_S^2(r)} \left(\int_{W^2} \varrho^{(2)}(x_1, x_2) k_h(r - |x_1 - x_2|) dx_1 dx_2 \right)^2 \\
&= (\mathbf{EA}(r))^2.
\end{aligned}$$

Beim zweiten Summanden von (21) liegen die Punkte x_2 und x_3 in einer Kugelschale mit Radius $r \pm h$ und Mittelpunkt x_1 . Die Punktepaare (x_1, x_2) und (x_1, x_3) haben also einen Zwischenpunktabstand von ungefähr r . Der Abstand zwischen x_2 und x_3 liegt zwischen 0 und $2r$, er beträgt also in schwacher Näherung im Mittel etwa r . Für $\varrho^{(3)}(x_1, x_2, x_3)$ gilt somit

$$\begin{aligned}
\varrho^{(3)}(x_1, x_2, x_3) &= \lambda^3 \left(1 + w_{12}^{(2)} + w_{13}^{(2)} + w_{23}^{(2)} + w_{123}^{(3)} \right) \\
&\approx \lambda^3 \left(1 + 3w(r) + w_{123}^{(3)} \right).
\end{aligned}$$

Für den Teil $w_{123}^{(3)}$ soll in schwacher Näherung ein Durchschnittswert, bezeichnet mit $w^{(3)}(r)$, angenommen werden. Insgesamt ergibt das

$$\begin{aligned}
&\frac{4}{C_S^2(r)} \int_{W^3} \varrho^{(3)}(x_1, x_2, x_3) k_h(r - |x_1 - x_2|) k_h(r - |x_1 - x_3|) dx_1 dx_2 dx_3 \\
&= \frac{4}{C_S^2(r)} \lambda^3 \left(1 + 3w(r) + w^{(3)}(r) \right) \int_W \left(\int_W k_h(r - |x_1 - x_2|) dx_2 \right)^2 dx_1
\end{aligned}$$

und im Fall einer Rechteckkernfunktion

$$\approx \frac{4}{C_S^2(r)} \lambda^3 \left(1 + 3w(r) + w^{(3)}(r) \right) I p_S^2.$$

Der dritte Teil aus (21) ist im Fall einer Rechteckkernfunktion bis auf einen Faktor identisch mit (20), dem Erwartungswert von $A(r)$

$$\frac{2}{C_S^2(r)} \int_{W^2} \varrho^{(2)}(x_1, x_2) k_h^2(r - |x_1 - x_2|) dx_1 dx_2 = \frac{\mathbf{EA}(r)}{h C_S(r)}.$$

Insgesamt erhält man also

$$\text{var} A(r) \approx \frac{4\lambda^3 I p_S^2}{C_S^2(r)} \left(1 + 3w(r) + w^{(3)}(r) \right) + \frac{\mathbf{EA}(r)}{h C_S(r)}$$

und

$$\begin{aligned}
\mathbf{E}\alpha^2 &\approx \left(\frac{1}{h C_S(r)} + \frac{4\lambda I p_S^2}{C_S^2(r)} \right) / \varrho^{(2)}(r) \\
&\quad + \frac{4\lambda^3 I p_S^2}{C_S^2(r)} \left(2w(r) + w^{(3)}(r) \right) / \left(\varrho^{(2)}(r) \right)^2.
\end{aligned} \tag{22}$$

4.2.4 Gemischte Größen

Es ist

$$A(r)B(r) = \frac{1}{\nu_d(W)C_S(r)} \sum_{x,y,z \in \Phi \cap W}^{\neq} k_h(r - |x - y|)p(z, r) \\ + \frac{2}{\nu_d(W)C_S(r)} \sum_{x,y \in \Phi \cap W}^{\neq} k_h(r - |x - y|)p(x, r)$$

und so

$$\mathbf{E}A(r)B(r) = \frac{1}{\nu_d(W)C_S(r)} \int_{W^3} \varrho^{(3)}(x_1, x_2, x_3) k_h(r - |x_1 - x_2|) p(x_3, r) dx_1 dx_2 dx_3 \\ + \frac{2}{\nu_d(W)C_S(r)} \int_{W^2} \varrho^{(2)}(x_1, x_2) k_h(r - |x_1 - x_2|) p(x_1, r) dx_1 dx_2.$$

Hier hat jetzt jeweils das Punktepaare (x_1, x_2) einen Zwischenpunktabstand von ungefähr r (während x_3 im ersten Integral „gleichverteilt“ ist), deshalb gilt

$$\mathbf{E}A(r)B(r) \approx \frac{1}{\nu_d(W)C_S(r)} \int_{W^3} \lambda^3 \left(1 + w_{12}^{(2)}\right) k_h(r - |x_1 - x_2|) p(x_3, r) dx_1 dx_2 dx_3 \\ + \frac{2\varrho^{(2)}(r)}{\nu_d(W)C_S(r)} \int_{W^2} k_h(r - |x_1 - x_2|) p(x_1, r) dx_1 dx_2 \\ = \frac{\lambda}{C_S(r)} \int_{W^2} \varrho^{(2)}(x_1, x_2) k_h(r - |x_1 - x_2|) dx_1 dx_2 \\ + \frac{2\varrho^{(2)}(r)}{\nu_d(W)C_S(r)} \int_{W^2} k_h(r - |x_1 - x_2|) p(x_1, r) dx_1 dx_2 \\ = \lambda \mathbf{E}A(r) + \frac{2\varrho^{(2)}(r)}{\nu_d(W)C_S(r)} \int_{W^2} k_h(r - |x_1 - x_2|) p(x_1, r) dx_1 dx_2.$$

Im Fall einer Rechteckkernfunktion ergibt das (vgl. (3) in Kapitel 2.4)

$$\mathbf{E}A(r)B(r) \approx \lambda \mathbf{E}A(r) + \frac{2\varrho^{(2)}(r)}{\nu_d(W)C_S(r)} \int_W p_S(x, r) p(x, r) dx \\ = \lambda \mathbf{E}A(r) + \frac{2\varrho^{(2)}(r)}{\nu_d(W)C_S(r)} \mathbf{I}(pp_S)$$

und

$$\mathbf{E}\alpha\beta = \frac{2\mathbf{I}(pp_S)}{\lambda \nu_d(W)C_S(r)}.$$

4.2.5 Optimale $p(x, r)$ -Funktion

Faßt man die Ergebnisse zusammen, so erhält man

$$\begin{aligned} \mathbf{E}q &\approx g(r) \left(1 - \frac{4\mathbf{I}(pps)}{\lambda\nu_d(W)C_S(r)} + 3\frac{Ip^2}{\lambda\nu_d^2(W)} \right) \\ \mathbf{var}q &\approx g^2(r) \left(\mathbf{E}\alpha^2 + \frac{4}{\lambda\nu_d(W)} \left(\frac{Ip^2}{\nu_d(W)} - \frac{2\mathbf{I}(pps)}{C_S(r)} \right) \right). \end{aligned} \quad (23)$$

Es soll jetzt zunächst untersucht werden, welche $p(x, r)$ -Funktion den mittleren quadratischen Fehler minimiert.

Berechnet man den mittleren quadratischen Fehler (MSE) gemäß Kapitel 4.2.1, so erhält man $\text{MSE} = \mathbf{var}q$. Die Varianz wird bezüglich $p(x, r)$ minimiert, wenn der Ausdruck

$$\left(\frac{Ip^2}{\nu_d(W)} - \frac{2\mathbf{I}(pps)}{C_S(r)} \right)$$

aus (23) minimiert wird.

Analog zur Vorgehensweise in der Variationsrechnung (siehe auch Kapitel 3.1.2) sei nun $p_m(x, r)$ die minimierende Funktion. Dann kann jede $p(x, r)$ -Funktion geschrieben werden als

$$p(x, r) = p_m(x, r) + \varepsilon f(x, r)$$

mit

$$\int_W f(x, r) dx = 0 \quad (24)$$

und geeignetem ε . Dann ist

$$\begin{aligned} 0 &= \frac{\partial}{\partial \varepsilon} \left(\frac{Ip^2}{\nu_d(W)} - \frac{2\mathbf{I}(pps)}{C_S(r)} \right) \Bigg|_{\varepsilon=0} \\ &= \frac{\partial}{\partial \varepsilon} \left(\frac{1}{\nu_d(W)} \int_W (p_m(x, r) + \varepsilon f(x, r))^2 dx \right. \\ &\quad \left. - \frac{2}{C_S(r)} \int_W (p_m(x, r) + \varepsilon f(x, r)) p_S(x, r) dx \right) \Bigg|_{\varepsilon=0} \\ &= \int_W \frac{2f(x, r)p_m(x, r)}{\nu_d(W)} - \frac{2f(x, r)p_S(x, r)}{C_S(r)} dx \\ &= \int_W 2f(x, r) \left(\frac{p_m(x, r)}{\nu_d(W)} - \frac{p_S(x, r)}{C_S(r)} \right) dx \end{aligned}$$

für alle $f(x, r)$ mit (24). (Für weitere Erklärungen vgl. mit der Argumentation aus Kapitel 3.1.2.) Daraus folgt für $x \in W$

$$\frac{p_m(x, r)}{\nu_d(W)} - \frac{p_S(x, r)}{C_S(r)} = \text{const}(r),$$

und so

$$p_m(x, r) = \nu_d(W) \left(\text{const}(r) + \frac{p_S(x, r)}{C_S(r)} \right).$$

Für die zweite Ableitung gilt

$$\frac{\partial^2}{\partial \varepsilon^2} \left(\frac{Ip^2}{\nu_d(W)} - \frac{2I(pp_S)}{C_S(r)} \right) \Big|_{\varepsilon=0} = \frac{2}{\nu_d(W)} \int_W f^2(x, r) dx > 0,$$

es handelt sich bei $p_m(x, r)$ also um die gesuchte minimierende Funktion. Die Normierung ergibt

$$\begin{aligned} \nu_d(W) &= \int_W p_m(x, r) dx \\ &= \int_W \nu_d(W) \left(\text{const}(r) + \frac{p_S(x, r)}{C_S(r)} \right) dx \\ &= \nu_d^2(W) \text{const}(r) + \nu_d(W), \end{aligned}$$

also $\text{const}(r) \equiv 0$ und so

$$p_m(x, r) = \frac{\nu_d(W)}{C_S(r)} p_S(x, r).$$

Dieses Ergebnis bestätigt ein früheres Resultat. Setzt man nämlich $p_m(x, r)$ in $B(r)$ ein (siehe Gleichung (18) S. 47), so erhält man den in Kapitel 3.2.3 aus einer heuristischen Überlegung heraus entwickelten gewichteten Intensitätsschätzer $\hat{\lambda}_S(r)$.

4.2.6 Weitere Schätzer

Neben der im vorhergehenden Kapitel untersuchten Schätzerart tauchen in der physikalischen Fachliteratur noch andere Schätzer auf, die im Folgenden untersucht werden sollen. Sie setzen sich aus den fünf Bausteinen $DD(r)$, $DR(r)$, $RR(r)$, N und $N_{r,d}$ zusammen. Hierbei ist N die Anzahl der von Φ beobachteten Punkte in W

$$N = \sum_{x \in \Phi} \mathbf{1}_W(x) = \hat{\lambda} \nu_d(W)$$

und $DD(r)$ die Anzahl der beobachteten Punktepaare, die einen Zwischenpunktstand von $r \pm h$ haben

$$DD(r) = \sum_{x, y \in \Phi \cap W}^{\neq} \mathbf{1}_{[-h, h]}(r - |x - y|).$$

$RR(r)$ ist das Analogon von $DD(r)$ für einen homogenen Poissonprozess Ψ mit $N_{r,d}$ Punkten in W ($N_{r,d}$ soll hierbei möglichst groß sein). $DR(r)$ ist eine Mischung aus $DD(r)$ und $RR(r)$

$$DR(r) = \sum_{\substack{x \in \Phi \cap W \\ y \in \Psi \cap W}} \mathbf{1}_{[-h, h]}(r - |x - y|).$$

Man kann diese Bausteine mit den bisher bekannten Schätzern von $\varrho^{(2)}(r)$ und λ vergleichen. Für großes N_{rd} und kleines h gilt ungefähr

$$\begin{aligned} DD(r) &= 2hC_S(r)\widehat{\varrho}(r) \\ DR(r) &= 2hC_S(r)N_{rd}\widehat{\lambda}_S(r)/\nu_d(W) \\ RR(r) &= 2hC_S(r)N_{rd}^2/\nu_d^2(W), \end{aligned}$$

wobei $\widehat{\varrho}(r)$ der in den vorigen Kapiteln untersuchte Zähler $A(r)$ (siehe (19)) und $\widehat{\lambda}$ der Nenner $B(r)$ (siehe (18)) mit $p(x, r) \equiv 1$ ist. Mit diesen Näherungen gilt für einige in der Physik populäre Schätzer der Paarkorrelationsfunktion

$$\begin{aligned} \frac{DD(r)}{RR(r)} \cdot \frac{N_{rd}^2}{N^2} &= \frac{\widehat{\varrho}(r)}{\widehat{\lambda}^2} \quad (\text{P \& H}) \\ \frac{DD(r)}{DR(r)} \cdot \frac{N_{rd}}{N} &= \frac{\widehat{\varrho}(r)}{\widehat{\lambda}\widehat{\lambda}_S(r)} \quad (\text{D \& P}) \\ \frac{DD(r) \cdot RR(r)}{DR^2(r)} &= \frac{\widehat{\varrho}(r)}{\widehat{\lambda}_S^2(r)} \quad (\text{Ham}) \\ \frac{DD(r)}{RR(r)} \cdot \frac{N_{rd}^2}{N^2} - n \frac{DR(r)}{RR(r)} \cdot \frac{N_{rd}}{N} + n &= \frac{\widehat{\varrho}(r)}{\widehat{\lambda}^2} - n \frac{\widehat{\lambda}_S(r)}{\widehat{\lambda}} + n \quad (\text{HLS}), \end{aligned}$$

siehe Peebles und Hauser (1974), Davis und Peebles (1983), Hamilton (1993), Hewett (1982) (\widehat{g}_{HLS} mit $n = 1$) und Landy und Szalay (1993) (\widehat{g}_{HLS} mit $n = 2$). Offensichtlich stimmen in diesem Sinne die Schätzer von Stoyan und Hamilton, $\widehat{g}_S(r)$ und $\widehat{g}_{\text{Ham}}(r)$, überein.

4.2.7 Schätzervergleich

In diesem Kapitel sollen alle bisher aufgetretenen Schätzer miteinander verglichen werden. Die Rechnungen aus Kapitel 4.2.1 wurden entsprechend für alle Schätzer durchgeführt, siehe Anhang. Die Unterschiede im Erwartungswert und in der Varianz lassen sich durch die Größen R_1 und R_2 ausdrücken

$$\begin{aligned} R_1 &= \frac{1}{\lambda\nu_d(W)} \\ R_2 &= \frac{I p_S^2}{\lambda C_S^2(r)} - \frac{1}{\lambda\nu_d(W)}. \end{aligned}$$

Beide sind positiv, was für R_2 aus der Ungleichung

$$\int_W \left(\frac{\nu_d(W)}{C_S(r)} p_S(x, r) - 1 \right)^2 dx > 0$$

folgt.

Insgesamt ergibt sich

$$\begin{aligned}
 \mathbf{E}\widehat{g}_{\text{PH}}(r) &= g(r)(1 - R_1) \\
 \mathbf{E}\widehat{g}_{\text{DP}}(r) &= g(r)(1 - R_1 - R_2) \\
 \mathbf{E}\widehat{g}_{\text{Sto}}(r) &= g(r)(1 - R_1 - R_2) \\
 \mathbf{E}\widehat{g}_{\text{HLS}}(r) &= g(r)(1 - R_1)
 \end{aligned} \tag{25}$$

und

$$\begin{aligned}
 \text{var}\widehat{g}_{\text{PH}}(r) &= g^2(r)(\mathbf{E}\alpha^2 - 4R_1) \\
 \text{var}\widehat{g}_{\text{DP}}(r) &= g^2(r)(\mathbf{E}\alpha^2 - 4R_1 - 3R_2) \\
 \text{var}\widehat{g}_{\text{Sto}}(r) &= g^2(r)(\mathbf{E}\alpha^2 - 4R_1 - 4R_2) \\
 \text{var}\widehat{g}_{\text{HLS}}(r) &= g^2(r)(\mathbf{E}\alpha^2 - 4R_1 - \vartheta R_2)
 \end{aligned} \tag{26}$$

mit $\vartheta = 4ng(r) - n^2$. (Vgl. hierzu auch die Ergebnisse von Landy und Szalay (1993).)

Die Varianz des vierten Schätzers $\widehat{g}_{\text{HLS}}(r)$ hängt von der Wahl von n ab. Sie ist minimal an der Stelle $n = 2g(r)$. In diesem Fall ist

$$\text{var}\widehat{g}_{\text{HLS}}(r) = g^2(r)(\mathbf{E}\alpha^2 - 4R_1 - 4R_2).$$

Ist ein Poissonprozess gegeben, so ist $n = 2$ und $\widehat{g}_{\text{HLS}}(r) = \widehat{g}_{\text{LS}}(r)$ der Schätzer von Landy und Szalay.

Wie schon in Kapitel 4.2.1 bemerkt, ist der mittlere quadratische Fehler identisch mit der Varianz. Die in diesem Sinn optimalen Schätzer sind $\widehat{g}_{\text{HLS}}(r)$ mit $n = 2g(r)$ und $\widehat{g}_{\text{S}}(r)$.

In der Praxis ist $g(r)$ aber nicht bekannt, und das optimale n für $\widehat{g}_{\text{HLS}}(r)$ kann nicht genau bestimmt werden. Man müsste es mittels einer ersten einfachen Schätzung von $g(r)$ approximieren, wodurch die Gesamt-Schätzung wiederum etwas unsicherer wird. In der praktischen Anwendung hat also der Hamilton/Stoyan-Schätzer $\widehat{g}_{\text{S}}(r)$ bzw. $\widehat{g}_{\text{Ham}}(r)$ die geringste Schätzvarianz und damit auch den geringsten mittleren quadratischen Fehler. (Simulationsstudien haben allerdings gezeigt, dass $\widehat{g}_{\text{HLS}}(r)$ unter Benutzung einer ersten naiven Schätzung von $g(r)$ zur Bestimmung von $n = 2g(r)$ nur unwesentlich schlechter ist als $\widehat{g}_{\text{S}}(r)$. Im Gegensatz dazu kann $\widehat{g}_{\text{LS}}(r)$, also $\widehat{g}_{\text{HLS}}(r)$ mit $n = 2$, im allgemeinen nicht-Poissonschen Fall beträchtlich schlechter sein als $\widehat{g}_{\text{S}}(r)$.)

4.2.8 Simulationsergebnisse

Da die für die vorhergehende Bewertung wesentliche Abschätzung (23) auf einigen Vereinfachungen basiert, soll ihre Güte anhand von Simulationen demonstriert werden. Die dabei untersuchten Punktprozesse sind ein dreidimensionaler Poissonprozess mit $\lambda = 200$ und ein Thomas-Prozess (siehe Kapitel 2) mit $\sigma = 0.05$, $\lambda_p = 50$ und $\lambda_d = 5$, beobachtet im Einheitsquader.

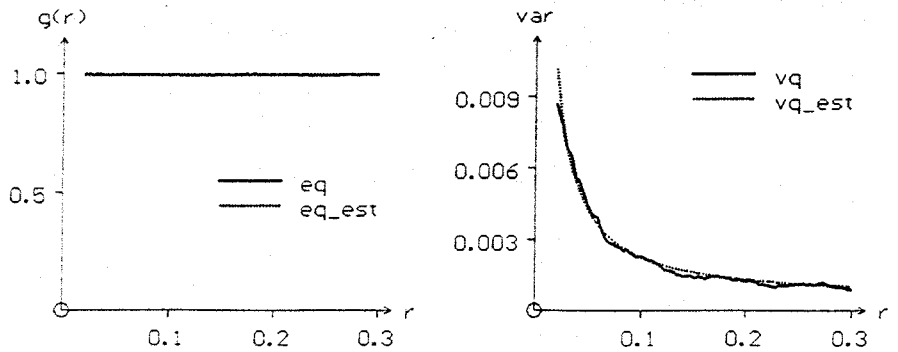


Abbildung 8: Der Erwartungswert und die Varianz von q ('eq' bzw. 'vq') verglichen mit den Abschätzungen aus (25) und (26) ('eq.est' bzw. 'vq.est') im Fall eines Poissonprozesses. Es ist kaum ein Unterschied festzustellen.

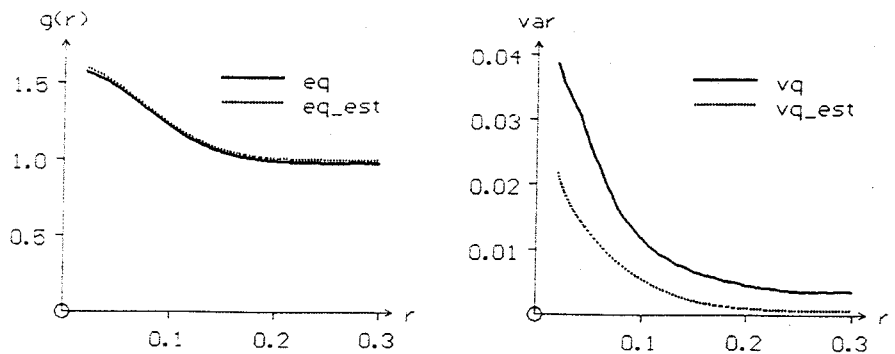


Abbildung 9: Die gleiche Situation wie in Abbildung (8) im Fall eines Thomasprozesses. Während die Abschätzung des Erwartungswertes gute Ergebnisse liefert, stimmt bei der Abschätzung der Varianz nur die Tendenz.

Die Abschätzungen liefern im Fall eines Poissonprozesses fast exakte Ergebnisse, siehe Abbildung 8. Für einen Thomas-Prozess liefert nur die Abschätzung des Erwartungswertes voll befriedigende Ergebnisse. Die Abschätzung der Varianz dagegen liefert zu kleine Werte. Allerdings besitzt sie immerhin die gleiche Tendenz wie die wahre Varianz. Weitere Simulationsstudien haben gezeigt, dass diese Aussagen nicht nur für die hier abgebildeten sondern auch für viele andere Clusterprozesse zutrifft.

Anhang

Sei hier

$$\begin{aligned} B_m &= (1 + \beta_m) \mathbf{E}B_m = \widehat{\lambda}_S(r), \\ B_0 &= (1 + \beta_0) \mathbf{E}B_0 = \widehat{\lambda} \end{aligned}$$

und $k_h(v)$ die Rechteckkernfunktion. Es ist

$$\begin{aligned} \mathbf{E}B_m &= \lambda \\ \mathbf{E}B_0 &= \lambda \\ \mathbf{E}\beta_m\beta_0 &= \frac{1}{\lambda\nu_d(W)} \\ \mathbf{E}\beta_m^2 &= \frac{Ip_S^2}{\lambda C_S^2(r)} \\ \mathbf{E}\beta_0^2 &= \frac{1}{\lambda\nu_d(W)} \\ \mathbf{E}\alpha\beta_m &= \frac{2Ip_S^2}{\lambda C_S^2(r)} \\ \mathbf{E}\alpha\beta_0 &= \frac{2}{\lambda\nu_d(W)}. \end{aligned}$$

Für $g_{DP}(r)$ gilt

$$\begin{aligned} \mathbf{E}g_{DP} &= \frac{\mathbf{E}A}{\mathbf{E}B_m\mathbf{E}B_0} (1 + \mathbf{E}\beta_m\beta_0 + \mathbf{E}\beta_m^2 + \mathbf{E}\beta_0^2 - \mathbf{E}\alpha\beta_m - \mathbf{E}\alpha\beta_0) \\ &= g(r) \left(1 - \frac{Ip_S^2}{\lambda C_S^2(r)} \right) \end{aligned}$$

und

$$\begin{aligned} \text{var}g_{DP} &= \left(\frac{\mathbf{E}A}{\mathbf{E}B_m\mathbf{E}B_0} \right)^2 (2\mathbf{E}\beta_m\beta_0 + \mathbf{E}\beta_m^2 + \mathbf{E}\beta_0^2 - 2\mathbf{E}\alpha\beta_m - 2\mathbf{E}\alpha\beta_0 + \mathbf{E}\alpha^2) \\ &= g^2(r) \left(\mathbf{E}\alpha^2 - \frac{1}{\lambda\nu_d(W)} - \frac{3Ip_S^2}{\lambda C_S^2(r)} \right). \end{aligned}$$

Für die Berechnung der Verteilungsparameter von $\widehat{g}_{HLS}(r)$ sei $X = Y + n - nZ$ mit $\mathbf{E}Z = 1$. Dann ist

$$\mathbf{E}X = \mathbf{E}Y$$

und

$$\text{var}X = \text{var}Y - n(n - n\mathbf{E}Z^2 + 2\mathbf{E}YZ - 2\mathbf{E}Y).$$

Es ist

$$\begin{aligned} \mathbf{E} \left(\frac{\widehat{\lambda}_S(r)}{\widehat{\lambda}} \right)^2 &= \left(\frac{\mathbf{E}\widehat{\lambda}_S(r)}{\mathbf{E}\widehat{\lambda}} \right)^2 (1 + 3\beta_0^2 - 4\beta_m\beta_0 + \beta_m^2) \\ &= 1 - \frac{1}{\lambda\nu_d(W)} + \frac{Ip_S^2}{\lambda C_S^2(r)} \end{aligned}$$

und

$$\begin{aligned} \mathbf{E} \frac{\widehat{\varrho}(r)\widehat{\lambda}_S(r)}{\widehat{\lambda}^3} &= \frac{\mathbf{E}\widehat{\varrho}(r)\mathbf{E}\widehat{\lambda}_S(r)}{(\mathbf{E}\widehat{\lambda})^3} (1 + 6\beta_0^2 + \alpha\beta_m - 3\alpha\beta_0 - 3\beta_m\beta_0) \\ &= g(r) \left(1 + \frac{2Ip_S^2}{\lambda C_S^2(r)} - \frac{3}{\lambda\nu_d(W)} \right), \end{aligned}$$

und so

$$\mathbf{E}\widehat{g}_{\text{HLS}}(r) = g(r) \left(1 - \frac{1}{\lambda\nu_d(W)} \right)$$

und

$$\text{var}\widehat{g}_{\text{HLS}}(r) = g^2(r) \left(\mathbf{E}\alpha^2 - \frac{4}{\lambda\nu_d(W)} + \left(\frac{Ip_S^2}{\lambda C_S^2} - \frac{1}{\lambda\nu_d(W)} \right) (n^2 - 4ng(r)) \right).$$

4.3 Eine heuristische Näherung

Die hier vorgestellte Näherung wurde von Stoyan, Bertram und Wendrock (1993) entwickelt. Obwohl sie von einem anderen Ansatz als im vorigen Kapitel ausgeht, führt sie im Poissonfall zum exakt gleichen Ergebnis wie (26).

Herleitung

Es sei der übliche Schätzer für $\varrho^{(2)}(r)$

$$\widehat{\varrho}(r) = \frac{1}{C_S(r)} \sum_{x,y \in \Phi \cap W}^{\neq} k_h(r - |x - y|),$$

gegeben. Berücksichtigt man jedes Punktepaar jeweils nur einmal, so ergibt das

$$\widehat{\varrho}(r) = \frac{2}{C_S(r)} \sum_{i=1}^{N_r} X_{r,i},$$

wobei N_r die Anzahl der Punktepaare ist, die einen Zwischenpunkt Abstand von $r \pm h$ haben. Die Variable $X_{r,i}$ entspricht dem Wert der Kernfunktion für das i te Punktepaar.

Die Verteilung der $(X_{r,i})_{i \leq N_r}$ ist unbekannt. Es wird vereinfachend angenommen, dass sie unabhängig und wie $k_h(Y)$ verteilt sind, wobei Y eine in $[-h, h]$ gleichverteilte Zufallsvariable ist. Darüber hinaus soll angenommen werden, dass auch N_r von den $X_{r,i}$ unabhängig ist. Es ergibt sich

$$\mathbf{E}X_{r,i} = \frac{1}{2h} \int_{-h}^h k_h(v) dv = \frac{1}{2h}$$

und so

$$\mathbf{E}\hat{\varrho}(r) = \frac{\mathbf{E}N_r}{hC_S(r)}.$$

Da $\hat{\varrho}(r)$ näherungsweise erwartungstreu ist (siehe Kapitel 4.2.3), folgt

$$\mathbf{E}N_r = \varrho^{(2)}(r)hC_S(r).$$

Für die Schätzvarianz folgt

$$\begin{aligned} \text{var}\hat{\varrho}(r) &= \frac{4}{C_S^2(r)} \left(\mathbf{E}N_r \text{var}X_{r,1} + \text{var}N_r (\mathbf{E}X_{r,1})^2 \right) \\ &= \frac{4}{C_S^2(r)} \left(\varrho^{(2)}(r)hC_S(r) \text{var}X_{r,1} + \frac{\text{var}N_r}{4h^2} \right). \end{aligned} \quad (27)$$

Über die Verteilung von N_r kann man nur spekulieren. Da das Ereignis sehr selten ist, dass der Zwischenpunkt Abstand eines zufällig gewählten Punktepaars im Intervall $[r-h, r+h]$ liegt, könnte man für N_r die Poissonverteilung („Verteilung der seltenen Ereignisse“) annehmen. Geht man darüber hinaus noch von der (falschen) Annahme aus, dass der λ -Schätzer keinen wesentlichen Einfluss auf die Schätzvarianz hat (und so als konstant mit $\hat{\lambda} = \lambda$ vorausgesetzt werden kann), so ergibt sich die folgende Abschätzung

$$\begin{aligned} \text{var}\hat{g}(r) &= \text{var} \frac{\hat{\varrho}(r)}{\lambda^2} \\ &\approx \text{var} \frac{\hat{\varrho}(r)}{\lambda^2} \\ &= \frac{\text{var}\hat{\varrho}(r)}{\lambda^4} \\ &= \frac{4}{\lambda^4 C_S^2(r)} \left(\varrho^{(2)}(r)hC_S(r) \text{var}X_{r,1} + \frac{\text{var}N_r}{4h^2} \right) \\ &= \frac{4}{\lambda^4 C_S^2(r)} \left(\varrho^{(2)}(r)hC_S(r) \text{var}X_{r,1} + \frac{\varrho^{(2)}(r)hC_S(r)}{4h^2} \right). \end{aligned}$$

Nimmt man für $k_h(v)$ die Rechteckkernfunktion an, so ist $X_{r,1}$ konstant (und damit $\text{var}X_{r,1} = 0$) und so

$$\begin{aligned} \text{var}\hat{g}(r) &\approx \frac{4}{\lambda^4 C_S^2(r)} \frac{\varrho^{(2)}(r)C_S(r)}{4h} \\ &= \frac{g(r)}{h\lambda^2 C_S(r)}. \end{aligned}$$

Dieses ist genau das Ergebnis (26) von Kapitel 4.2 im Fall eines Poissonprozesses für den optimalen Schätzer $\hat{g}_S(r)$. Danach ist nämlich

$$\text{var}\hat{g}_S(r) = g^2(r) (\mathbf{E}\alpha^2 - 4R_1 - 4R_2).$$

Mit (22) folgt dann für einen Poissonprozess

$$\text{var}\hat{g}_S(r) = \frac{1}{h\lambda^2 C_S(r)}.$$

Bemerkung:

Aus (27) folgt, dass die Schätzvarianz bezüglich der Kernfunktion nur vom zweiten Moment von $X_{r,1}$ abhängt, also von

$$\mathbf{E}X_{r,i}^2 = \frac{1}{2h} \int_{-h}^h k_h^2(v) dv.$$

Dieser Zusammenhang wurde zuvor in Kapitel 3.1.2 benötigt.

5 Bestimmung des Potenzgesetzes der Paarkorrelationsfunktion

Ein wichtiges Anwendungsgebiet der Punktprozessstatistik ist die Astronomie. Eine zentrale Frage ist hier die Entstehung des Weltalls, wozu es viele verschiedene Theorien gibt. Man kann eine Theorie bewerten, indem man geschätzte charakteristische Größen des realen Weltalls mit denjenigen vergleicht, die das nach der Theorie heute existierende Weltall haben sollte. Weichen die Schätzungen dieser Größen zu sehr voneinander ab, so ist das ein Hinweis darauf, dass die Theorie nicht zur Modellierung der Entstehung des Weltalls geeignet ist.

Das Weltall wird häufig durch das Punktmuster repräsentiert, das aus den Positionen seiner Galaxien besteht. Man geht davon aus, dass es eine Realisierung eines stationären und isotropen Punktprozesses ist. Ein wichtiges Hilfsmittel zur Charakterisierung dieses Punktprozesses ist seine Produktdichtefunktion zweiter Ordnung $\varrho^{(2)}(r)$. In der Astronomie ist man davon überzeugt, dass sie für kleine r proportional zu $r^{-\gamma}$ ist.

Der wahre Wert von γ ist unbekannt. Die in der Literatur dafür angegebenen Schätzwerte variieren zwischen 1 und 2. Die Schätzung von γ basiert meistens auf einer Schätzung von $\varrho^{(2)}(r)$. Deshalb wurde in bisher veröffentlichten Artikeln dazu hauptsächlich darauf geachtet, einen möglichst genauen Schätzer dieser Funktion zu entwickeln. Der Einfluss anderer Schritte bei der γ -Schätzung blieben weitgehend im Hintergrund. In diesem Kapitel wird deshalb die gesamte Schätzprozedur analysiert.

5.1 Grundlagen

Gegeben sei ein dreidimensionaler Punktprozess Φ , dessen Produktdichtefunktion $\varrho^{(2)}(r)$ für kleine r proportional zu $r^{-\gamma}$ mit $0 < \gamma$ ist

$$\varrho^{(2)}(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$$

mit einem geeigneten r_0 . (Die Dimension des Punktprozesses kann im Grunde beliebig gewählt werden. Da in den Anwendungen aber ausschließlich der dreidimensionale Fall von Interesse ist, wird hier und auch im folgenden Kapitel nur dieser betrachtet.)

Ausgehend von einer Schätzung von $\varrho^{(2)}(r)$ kann γ durch doppelt logarithmische lineare Regression bestimmt werden. Dazu sei die Funktion p definiert als

$$p: s \rightarrow \ln\left(\varrho^{(2)}(e^s)\right), \quad -\infty < s < \infty,$$

was äquivalent ist zu

$$p: \ln r \rightarrow \ln\left(\varrho^{(2)}(r)\right), \quad r > 0.$$

Im Folgenden wird die zweite, intuitivere Darstellung benutzt.

Im gegebenen Fall ist p eine Gerade der Steigung $-\gamma$

$$\ln \varrho^{(2)}(r) = -\gamma \ln r + \gamma \ln r_0.$$

Mittels linearer Regression angewandt auf eine Schätzung von p erhält man eine Schätzung für γ .

Bemerkung:

Aus $\varrho^{(2)}(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$ folgt $\gamma < d$. Es ist nämlich

$$\varrho^{(2)}(r) = \lambda^2 \frac{\partial K(r)}{\partial r} / (db_d r^{d-1})$$

(siehe Kapitel 2.1). Daraus folgt

$$K(r) = \frac{db_d}{\lambda^2} \int_0^r r_0^\gamma t^{-\gamma} t^{d-1} dt.$$

Da $K(r)$ endlich ist, gilt dies auch für das Integral. Es muss also $d - 1 - \gamma > -1$ und so $\gamma < d$ gelten.

5.2 Mittelwert von $\hat{\gamma}$

Lineare Regression

Die lineare Regression basiert üblicherweise auf Paaren $(t_i, x(t_i))$, wobei die t_i innerhalb eines geeigneten Intervalls $[a, b]$ liegen. Dies führt zu einer relativ komplexen Prozedur, weshalb hier jetzt vereinfachend angenommen wird, dass die t_i nur die Werte der zwei Intervallgrenzen annehmen, $t_1 = a$ und $t_2 = b$. Man kann sie schreiben als $t_1 = r - \varepsilon$ und $t_2 = r + \varepsilon$, wobei r der Mittelpunkt des Intervalls $[a, b]$ ist. Die lineare Regression ist dann äquivalent zur Bestimmung der Sekante um den Punkt r

$$\mathbf{E} \hat{\gamma}(r) \approx \mathbf{E} \frac{\Delta \hat{p}(\ln r)}{\Delta \ln r} \quad \text{mit} \quad \frac{\Delta f}{\Delta r} = \frac{f(r + \varepsilon) - f(r - \varepsilon)}{2\varepsilon}.$$

Hierbei ist $\hat{p}(\ln r)$ ein Schätzer von $p(\ln r)$.

Im Einzelfall kann die Sekante (durch die Endpunkte des Intervalls) stark von der auf vielen Punkten basierenden Regressionsgeraden abweichen. Es scheint aber offensichtlich, dass sich die Steigung der Sekante und die der Regressionsgeraden im Mittel kaum voneinander unterscheiden.

Da die Sekantenbildung ein lineares Funktional ist, gilt

$$\mathbf{E} \frac{\Delta \widehat{p}(\ln r)}{\Delta \ln r} = \frac{\Delta \mathbf{E} \widehat{p}(\ln r)}{\Delta \ln r},$$

was näherungsweise

$$\mathbf{E} \frac{\Delta \widehat{p}(\ln r)}{\Delta \ln r} \approx \frac{\partial \mathbf{E} \widehat{p}(\ln r)}{\partial \ln r} = r \frac{\partial \mathbf{E} \ln(\widehat{\varrho}(r))}{\partial r}$$

ergibt.

Es ist

$$\mathbf{E} \ln(\widehat{\varrho}(r)) = \ln(\mathbf{E} \widehat{\varrho}(r)) + \mathbf{E} \ln \frac{\widehat{\varrho}(r)}{\mathbf{E} \widehat{\varrho}(r)}.$$

Die Entwicklung des zweiten Summanden in eine Taylorreihe (hierbei werden nur die ersten beiden Terme berücksichtigt) ergibt

$$\begin{aligned} \mathbf{E} \ln \frac{\widehat{\varrho}(r)}{\mathbf{E} \widehat{\varrho}(r)} &\approx \mathbf{E} \left(\frac{\widehat{\varrho}(r)}{\mathbf{E} \widehat{\varrho}(r)} - 1 \right) - \frac{1}{2} \mathbf{E} \left(\frac{\widehat{\varrho}(r)}{\mathbf{E} \widehat{\varrho}(r)} - 1 \right)^2 \\ &= -\frac{1}{2} \mathbf{var} \frac{\widehat{\varrho}(r)}{\mathbf{E} \widehat{\varrho}(r)}. \end{aligned}$$

Landy und Szalay (1993) haben gezeigt, dass die Varianz von $\widehat{\varrho}(r) / \mathbf{E} \widehat{\varrho}(r)$ ein O -Term der Ordnung $1/\nu_d(W)$ ist. Das gleiche gilt auch für die restlichen Taylor-Glieder. Man kann sie deshalb vernachlässigen und erhält

$$\mathbf{E} \widehat{\gamma}(r) = \frac{r}{\mathbf{E} \widehat{\varrho}(r)} \frac{\partial}{\partial r} \mathbf{E} \widehat{\varrho}(r). \quad (28)$$

Der $\varrho^{(2)}(r)$ -Schätzer

Zugrunde gelegt sei der $\varrho^{(2)}(r)$ -Schätzer

$$\widehat{\varrho}(r) = \frac{1}{4\pi r^2 2h \bar{\gamma}_W(r)} \sum_{x,y \in \Phi \cap W}^{\neq} \mathbf{1}_{[r-h, r+h]}(|x-y|),$$

siehe Kapitel 3. Er lässt sich wegen seiner Einfachheit am leichtesten untersuchen. Die meisten anderen Schätzer unterscheiden sich hiervon im wesentlichen durch eine andere Randkorrektur. Diese wirkt sich aber für kleine r nur sehr schwach aus (siehe Kerscher 1999), weshalb die Ergebnisse dieses Kapitels angenähert auch für andere Schätzer von ähnlicher Bauart gelten.

Der Erwartungswert von $\widehat{\varrho}(r)$ ist

$$\mathbf{E} \widehat{\varrho}(r) = \frac{1}{2hr^2} \int_{(r-h)_+}^{r+h} \varrho^{(2)}(v) v^2 dv, \quad (29)$$

siehe (9) aus Kapitel 3.1.3. Im gegebenen Fall $\varrho^{(2)}(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$ führt (29) zu

$$\mathbf{E}\widehat{\varrho}(r) = \frac{r_0^\gamma}{2hr^2} \int_{(r-h)_+}^{r+h} v^{2-\gamma} dv = r_0^\gamma \frac{(r+h)^{3-\gamma} - ((r-h)_+)^{3-\gamma}}{2hr^2(3-\gamma)}.$$

Die Ableitung nach r ergibt

$$\frac{\partial}{\partial r} \mathbf{E}\widehat{\varrho}(r) = -\frac{2}{r} \mathbf{E}\widehat{\varrho}(r) + r_0^\gamma \frac{(r+h)^{2-\gamma} - ((r-h)_+)^{2-\gamma}}{2hr^2}.$$

Insgesamt ergibt sich mit (28) also

$$\mathbf{E}\widehat{\gamma}(r) = 2 - (3-\gamma)r \frac{(r+h)^{2-\gamma} - ((r-h)_+)^{2-\gamma}}{(r+h)^{3-\gamma} - ((r-h)_+)^{3-\gamma}}. \quad (30)$$

Dieser Ausdruck hängt nicht von den absoluten Werten von r und h , sondern nur von deren Verhältnis ab. Um dies zu sehen sei $q = h/r$. Dann folgt aus (30)

$$\mathbf{E}\widehat{\gamma}(r) = 2 - (3-\gamma) \frac{(1+q)^{2-\gamma} - ((1-q)_+)^{2-\gamma}}{(1+q)^{3-\gamma} - ((1-q)_+)^{3-\gamma}}.$$

Ist q wesentlich kleiner als 1, lässt sich der letzte Quotient annähern durch

$$\begin{aligned} \frac{(x^\alpha)'}{(x^\beta)'}(x_0) &= \lim_{\varepsilon \rightarrow 0} ((x_0 + \varepsilon)^\alpha - (x_0 - \varepsilon)^\alpha) / \varepsilon \\ &\quad \cdot \lim_{\varepsilon \rightarrow 0} \varepsilon / ((x_0 + \varepsilon)^\beta - (x_0 - \varepsilon)^\beta) \\ &= \lim_{\varepsilon \rightarrow 0} \frac{(x_0 + \varepsilon)^\alpha - (x_0 - \varepsilon)^\alpha}{(x_0 + \varepsilon)^\beta - (x_0 - \varepsilon)^\beta} \end{aligned}$$

mit $x_0 = 1$, $\alpha = 2 - \gamma$ und $\beta = 3 - \gamma$. Es ist

$$\frac{(x^{2-\gamma})'}{(x^{3-\gamma})'}(1) = \frac{2-\gamma}{3-\gamma}$$

und so

$$\mathbf{E}\widehat{\gamma}(r) = 2 - (3-\gamma) \frac{2-\gamma}{3-\gamma} = \gamma.$$

Ist also h klein im Vergleich zu r , so ist $\widehat{\gamma}$ angenähert erwartungstreu und unabhängig von r .

Das gilt nicht für den umgekehrten Fall, wenn r im Vergleich mit h klein ist. Dann liefert (28)

$$\mathbf{E}\widehat{\gamma}(r) \approx 2 - \frac{r}{h}(3-\gamma) \approx 2. \quad (31)$$

Eine naive Schätzung von $\varrho^{(2)}(r)$, bei der r beliebig klein wird, kann also ernsthafte Fehler verursachen.

In Kapitel 3.1.3 wurde die angepasste Bandweite $h^*(r) = \min\{h, r\}$ entwickelt, um gegenüber einer konstanten Bandweite den Bias des $\varrho^{(2)}(r)$ -Schätzers zu verringern. Auch wenn hier eine andere Situation vorliegt, stellt sich die Frage, ob die angepasste Bandweite auch in diesem Fall zu besseren Ergebnissen führt. Wird $h^*(r)$ als Bandweite benutzt, so liefert (29)

$$\mathbf{E}\hat{\varrho}(r) = c_\varrho \frac{(2r)^{3-\gamma}}{2r^3(3-\gamma)} = \frac{2^{2-\gamma}c_\varrho}{3-\gamma} r^{-\gamma}.$$

Die Ableitung nach r ergibt

$$\frac{\partial}{\partial r} \mathbf{E}\hat{\varrho}(r) = -\gamma \frac{2^{2-\gamma}c_\varrho}{3-\gamma} r^{-\gamma-1},$$

also nach (28)

$$\mathbf{E}\hat{\gamma}(r) = \gamma.$$

Obwohl der Einsatz von $h^*(r)$ gegenüber dem einer hinreichend kleinen konstanten Bandweite den Bias von $\hat{\varrho}(r)$ erhöhen könnte, führt er zu einer γ -Schätzung, die im Mittel nicht schlechter ist als bei einer konstanten Bandweite.

5.3 Varianz von $\hat{\gamma}$

Die vorhergehenden Berechnungen haben gezeigt, dass sowohl die angepasste als auch eine konstante Bandweite zu einem näherungsweise erwartungstreuen γ -Schätzer führen. Neben der Erwartungstreue interessiert aber auch noch die Varianz der γ -Schätzer. Es wird jetzt gezeigt, dass die Schätzvarianz im Fall der angepassten Bandweite gegenüber einer konstanten zum Teil erheblich geringer ist.

Der Schätzer von $\varrho^{(2)}(r)$

$$\hat{\varrho}(r) = \frac{1}{4\pi r^2 2h\hat{\gamma}_W(r)} \sum_{x,y \in \Phi \cap W}^{\neq} \mathbf{1}_{[r-h, r+h]}(|x-y|)$$

zählt im Wesentlichen all solche Punktepaare, deren Zwischenpunktabstand im Intervall $[r-h, r+h]$ liegt. Im Fall der angepassten Bandweite $h^*(r) = r$ ist dieses Intervall viel größer als bei einer konstanten, da die konstante Bandweite h wesentlich kleiner als r sein müsste. (Nur dadurch wird die angenäherte Erwartungstreue des γ -Schätzers gewährleistet, vgl. Kapitel 5.2. Ansonsten könnten gemäß (31) starke Abweichungen auftreten.) Deshalb umfasst die Schätzung unter Verwendung der angepassten Bandweite mehr Punktepaare, was sowohl bei $\hat{\varrho}(r)$ als auch bei $\hat{\gamma}$ zu stabileren Ergebnissen führen sollte.

Eine genauere analytische Untersuchung der Schätzvarianz ist sehr kompliziert. Es wird hier darauf verzichtet und statt dessen mit Simulationsstudien gearbeitet. Insgesamt wurde in über 3000 verschiedenen Ausgangssituationen überprüft, welche der beiden Bandweiten zu einer geringeren Schätzvarianz führt. Es wurden dabei

Punktprozesse zugrunde gelegt, deren Produktdichtefunktionen Pole unterschiedlicher Ordnung hatten, siehe 6.1 und 6.2. Die Länge und die Position des Regressionsintervalls wurden ebenso variiert wie auch die Länge der konstanten Bandweite. Um hierbei die (angenäherte) Erwartungstreue zu gewährleisten, war h immer kleiner als der kleinste Wert von r , für den $\varrho^{(2)}(r)$ geschätzt wurde.

Das Ergebnis dieser Studien ist eindeutig. In nahezu allen Fällen zeigte sich die angepasste Bandweite der konstanten überlegen. Die Schätzvarianzen unterscheiden sich in manchen Fällen sogar um ein vielfaches.

Die untenstehenden Tabellen dokumentieren dieses Ergebnis. Sie beinhalten die Varianzen der γ -Schätzung in Abhängigkeit der zugrundegelegten Regressionsintervalle für unterschiedliche Bandweiten. Die Bandweite beträgt 0,01 (linke Tabelle), 0,1 (rechte Tabelle) oder ist angepasst (mittlere Tabelle).

Der in allen Fällen zugrundegelegte Punktprozess ist ein Neyman-Scott-Prozess. Hierbei ist

$$d(r) = 0,2r^{-0,4} \mathbf{1}_{[0,1]}(r)$$

und $\lambda_p = 1$. Die Anzahl der Tochterpunkte je Cluster ist Poissonverteilt mit Parameter $\lambda_d = 5$. Das Beobachtungsfenster ist ein Würfel der Kantenlänge 10.

| $h = 0,01$ | r_{\max} | | | | $h = 0,1$ | r_{\max} | | | |
|----------------|------------|------|------|-------|----------------|------------|------|------|------|
| | 0,2 | 0,3 | 0,4 | 0,5 | | 0,2 | 0,3 | 0,4 | 0,5 |
| r_{\min} 0,1 | 0,61 | 0,21 | 0,12 | 0,08 | r_{\min} 0,1 | 0,06 | 0,04 | 0,04 | 0,03 |
| 0,2 | | 2,32 | 0,55 | 0,24 | 0,2 | | 0,20 | 0,13 | 0,10 |
| 0,3 | | | 5,02 | 1,15 | 0,3 | | | 0,31 | 0,20 |
| 0,4 | | | | 11,58 | 0,4 | | | | 0,51 |

| h angepasst | r_{\max} | | | |
|----------------|------------|-------|-------|-------|
| | 0,2 | 0,3 | 0,4 | 0,5 |
| r_{\min} 0,1 | 0,011 | 0,009 | 0,007 | 0,006 |
| 0,2 | | 0,023 | 0,015 | 0,012 |
| 0,3 | | | 0,036 | 0,021 |
| 0,4 | | | | 0,041 |

Tabelle 1: Die Varianzen der γ -Schätzung in Abhängigkeit der zugrundegelegten Regressionsintervalle für unterschiedliche Bandweiten.

In allen hier aufgeführten Fällen wird die Schätzvarianz durch eine angepasste Bandweite gegenüber der konstanten verringert.

6 Modelle für Clusterpunktprozesse

Bei analytischen Untersuchungen, die mit Punktprozessen zusammenhängen, stößt man häufig auf große Probleme. Alternativ zur analytischen Herangehensweise können die gegebenen Fragestellungen auch basierend auf Simulationsstudien durchgeführt werden. Damit es dabei zu glaubhaften Aussagen kommen kann, werden geeignete und hinreichend variable Punktprozesse benötigt.

Im letzten Kapitel wurde ein dreidimensionaler Punktprozess zugrunde gelegt, dessen Produktdichtefunktion für kleine r von der Form

$$\varrho^{(2)}(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \quad (32)$$

ist. Zum testen des dort untersuchten γ -Schätzers werden also Punktprozesse benötigt, deren Produktdichtefunktion von der Form (32) ist. Da (32) nur für kleine r gelten muß, reicht es auch schon aus, wenn die Produktdichtefunktion einen Pol an der Stelle $r = 0$ hat. Hat der Pol die Ordnung γ , so kann man davon ausgehen, dass sich $\varrho^{(2)}(r)$ für kleine r ungefähr proportional zu $r^{-\gamma}$ verhält. Einige solcher Punktprozesse, deren Produktdichtefunktionen einen Pol haben, werden im Folgenden vorgestellt.

6.1 Punktfeldmodelle

Buryak und Doroshkevich

Buryak und Doroshkevich (1996) führten einen Clusterprozess ein, dessen Cluster sich auf zufällig orientierten Geraden befinden. Die Cluster-Punkte bilden einen eindimensionalen Poissonprozess auf den Geraden. Die zugehörige Produktdichtefunktion hat einen Pol der Ordnung $\gamma = 2$.

Weygaert und Icke

Weygaert und Icke (1987) und (1989) schlagen den Punktprozess vor, der aus den Ecken eines Poisson-Voronoi-Mosaiks gebildet wird. Ihre Annahme, dass auch hier $\gamma = 2$ gilt, wurde von Heinrich et al. (1998) bewiesen.

Zwei-Punkt-Cluster

Ein weitaus einfacherer Punktprozess erscheint in Stoyan (1994). Die Punkte eines homogenen Poissonprozesses der Intensität λ_p bilden die Mittelpunkte von Clustern. Jedes Cluster besteht aus zwei Punkten. Die Produktdichtefunktion ist

$$\varrho^{(2)}(r) = \lambda^2 \left(1 + \frac{f(r)}{8\pi\lambda_p r^2}\right),$$

wobei $f(r)$ die Dichtefunktion des Abstandes der zwei Clusterpunkte ist. Die Ordnung des Pols von $\varrho^{(2)}(r)$ hängt vom asymptotischen Verhalten von $f(r)$ für $r \rightarrow 0$ ab. Verhält sich $f(r)$ für kleine r wie r^α , so ist $\gamma = 2 - \alpha$. Eine geeignete Wahl von $f(r)$ führt also zu jedem beliebigen γ zwischen 0 und 3.

Erneuerungsprozess auf Geraden

Eine Verallgemeinerung des Punktprozesses von Buryak und Doroshkevich führt zu einem weiteren Punktprozess mit beliebig wählbarem γ . Wieder befinden sich die Clusterpunkte auf zufällig orientierten Geraden. Diesmal bilden sie jedoch keinen Poisson- sondern einen Erneuerungsprozess (d. h. unabhängige und identisch verteilte Abstände zwischen den Punkten) mit Gamma-verteilten Zuwächsen, d. h. die Dichtefunktion der Zwischenpunktabstände ist

$$f(r) = \frac{a^p}{\Gamma(p)} r^{p-1} e^{-ar}, \quad p > 0.$$

Die Produktdichtefunktion dieses Prozesses hat einen Pol der Ordnung $3 - p$. Der Fall $p = 1$ ist äquivalent zu der Dichtefunktion der Exponentialverteilung und führt zu einem eindimensionalen Poissonprozess, also zum Punktprozess von Buryak und Doroshkevich.

Neyman-Scott-Prozesse

Eine variable Klasse von Punktprozessen, deren Produktdichtefunktionen Pole von beliebiger Ordnung haben können, bilden die Neyman-Scott-Prozesse, siehe Kapitel 2.1. Hier gilt

$$\varrho^{(2)}(r) = \lambda^2 \left(1 + \frac{f(r)}{4\pi r^2 \lambda_p} \right).$$

Wird die Clustergeometrie und somit $f(r)$ geeignet gewählt, so können Pole von beliebiger Ordnung erreicht werden.

Es kann allerdings sehr kompliziert sein, ausgehend von einer Clustergeometrie die zugehörige $f(r)$ -Funktion oder auch nur ihr asymptotisches Verhalten zu bestimmen, siehe Kapitel 2.3. Letzteres wird im nächsten Kapitel vollständig beantwortet für Neyman-Scott-Prozesse beliebiger Ordnung mit dreidimensionalen Clustern.

6.2 Störung

6.2.1 Das Grundprinzip

Der Ausgangspunkt sei hier zunächst ein dreidimensionaler Neyman-Scott-Prozess, bei dem das asymptotische Verhalten der Produktdichtefunktion als bekannt vorausgesetzt wird. Später wird dann auch gezeigt, wie dieses asymptotische Verhalten ausgehend von der $d(r)$ -Funktion berechnet werden kann.

Es wird im Folgenden untersucht, welchen Effekt eine zufällige Störung der Punkte auf die Ordnung eines vorhandenen Pols der Produktdichtefunktion hat. Unter der Störung ist hier eine zufällige Verschiebung zu verstehen. Die Länge der Verschiebung sei nach einer Dichtefunktion $d^*(r)$ verteilt, die Orientierung des Verschiebungsvektors sei gleichverteilt auf der Einheitssphäre.

Die Störung verändert die Verteilung des Abstandes zwischen den Punkten eines Clusters. Sei $f^*(r)$ die der $f(r)$ -Funktion entsprechende Dichtefunktion des Abstandes zweier Punkte desselben Clusters nach der Störung. Für die Produktdichtefunktion des Prozesses nach der Störung gilt dann

$$\rho^{(2)*}(r) = \lambda^2 \left(1 + \frac{f^*(r)}{db_d r^{d-1} \lambda_p} \right).$$

Ist das asymptotische Verhalten von $f^*(r)$ bekannt, so ergibt sich daraus auch das von $\rho^{(2)*}(r)$.

6.2.2 Asymptotisches Verhalten nach der Verwacklung

Gegeben sei also ein dreidimensionaler Neyman-Scott-Prozess, dessen Produktdichtefunktion einen Pol der Ordnung γ hat mit $0 < \gamma < 3$. Dann ist $f(r)$ asymptotisch äquivalent zu $r^{2-\gamma}$

$$f(r) \asymp r^{2-\gamma} \iff \lim_{r \rightarrow 0} f(r) / r^{2-\gamma} = c$$

mit einer geeigneten Konstante c . Weiterhin sei auch für die Störungsdichtefunktion $d^*(r)$ ein bestimmtes asymptotisches Verhalten vorausgesetzt

$$d^*(r) \asymp r^\alpha$$

mit $\alpha > -1$.

Zunächst sei angenommen, dass jedes Cluster zwei Punkte enthalte und das nur einer von ihnen gestört wird. (In einem zweiten Schritt ist dann die Anzahl der Punkte je Cluster zufällig und alle Punkte werden gestört.) Die Verteilungsfunktion des Abstandes der zwei Punkte beträgt

$$F^*(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) d^*(z) \frac{A(x; z; r)}{4\pi z^2} dx dz, \quad (33)$$

wobei $A(x; z; r)$ der Oberflächeninhalt des Schnitts einer Kugel vom Radius r mit einer Sphäre vom Radius z ist, deren Mittelpunkte den Abstand x haben

$$A(x; z; r) = \begin{cases} 4\pi z^2, & x < r - z, r \geq z \\ \frac{\pi z}{x} (r^2 - (x - z)^2), & |r - z| \leq x \leq r + z \\ 0, & \text{sonst} \end{cases} .$$

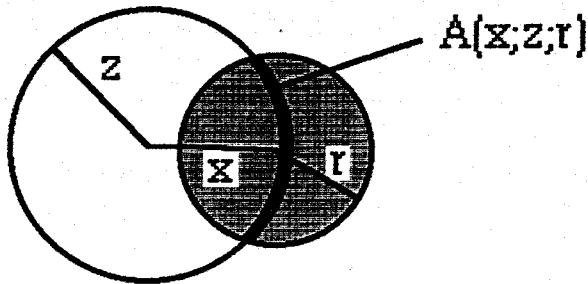


Abbildung 10: Demonstration der Größe $A(x; z; r)$ im zweidimensionalen Fall.

Es ist relativ schwierig, das Integral (33) zu lösen, siehe Anhang. Das Ergebnis ist, dass $f^*(r)$ asymptotisch äquivalent ist zu $r^{3-\gamma+\alpha}$, d. h., die Ordnung des Pols von $\varrho^{(2)*}(r)$ ist nach der Störung $\gamma - (\alpha + 1)$. Sie hat sich also verringert (da $\alpha + 1 > 0$).

Sei nun die Anzahl der Tochterpunkte je Cluster beliebig und seien alle Punkte von der Störung betroffen. Da die Tochterpunkte unabhängig sind, ist diese Situation im Grunde identisch mit der vorherigen. Es muss nur $f(r) \asymp r^{2-\gamma}$ durch $f(r) \asymp r^{3-\gamma+\alpha}$ ausgetauscht werden (letzteres ist die resultierende Dichtefunktion $f^*(r)$ nach der Störung des ersten Punktes). Es folgt unmittelbar

$$f^*(r) \asymp r^{4-\gamma+2\alpha}$$

was einem Pol der Produktdichtefunktion von der Ordnung $\gamma - 2(\alpha + 1)$ entspricht.

6.2.3 Neyman-Scott-Prozesse höherer Ordnung

Mit Hilfe der gewonnenen Resultate lassen sich auch die $f(r)$ -Funktionen von Neyman-Scott-Prozessen n -ter Ordnung berechnen. Das sind Punktprozesse, die wie Neyman-Scott-Prozesse gebildet werden, nur dass die Elternpunkte statt eines Poissonprozesses einen Neyman-Scott-Prozess $n - 1$ -ter Ordnung bilden. In diesem Sinn ist der Poissonprozess ein Neyman-Scott-Prozess 0-ter Ordnung.

Neyman-Scott-Prozess erster Ordnung

Gegeben sei zunächst ein Neyman-Scott-Prozess erster Ordnung Φ mit $d(r) \asymp r^\alpha$. Zur Bestimmung der $f(r)$ -Funktion wird dieser Prozess mit einem im Folgenden beschriebenen Prozess Ψ verglichen. Dieser geht aus einer Störung des Zwei-Punkt-Cluster-Prozesses hervor, siehe Kapitel 6.1.

Bei Ψ bestehe jedes Cluster aus zwei Punkten. Der Abstand zwischen den Punkten vor der Störung sei nach $d(r)$ verteilt. Die Störung werde bei jedem Cluster bei

jeweils einem der beiden Punkte gemäß $d(r)$ durchgeführt, was entsprechend dem vorherigen Kapitel zu der neuen Dichtefunktion des Abstandes zwischen den beiden Punkten

$$f^*(r) \asymp r^{1+2\alpha}$$

führt.

Der folgende Vergleich macht deutlich, dass die Verteilungen des Abstandes zweier Clusterpunkte (desselben Clusters) bei beiden Punktprozessen Φ und Ψ identisch sind. Seien dazu x_Φ und y_Φ zwei Punkte aus Φ und e_Φ ihr gemeinsamer Elternpunkt. Außerdem seien x_Ψ und y_Ψ die zwei Punkte eines Clusters aus Ψ , wobei y_Ψ durch Störung aus y'_Ψ hervorgegangen ist. Identifiziert man nun x_Φ , e_Φ und y_Φ der Reihe nach mit x_Ψ , y'_Ψ und y_Ψ , so wird klar, dass die entsprechenden Zwischenpunktabstände in beiden Gruppen identisch verteilt sind. Es gilt also auch für Φ

$$f(r) \asymp r^{1+2\alpha}.$$

Damit ist es also nun möglich, ausgehend von der $d(r)$ -Funktion, das asymptotische Verhalten von $f(r)$ zu bestimmen.

Neyman-Scott-Prozess n -ter Ordnung

Den gleichen Gedankengang kann man auch auf Neyman-Scott-Prozesse höherer Ordnung übertragen. Gegeben sei nun also ein Neyman-Scott-Prozess n -ter Ordnung. Die Tochterpunkte n -ter Ordnung seien gemäß der Dichtefunktion $d(r) \asymp r^\alpha$ um die jeweiligen Elternpunkte n -ter Ordnung (oder Tochterpunkte $(n-1)$ -ter Ordnung) verteilt. Der Abstand zweier Elternpunkte n -ter Ordnung, die demselben Poisson-Eltern-Punkt erster Ordnung entstammen, sei gemäß der Dichtefunktion $f_{n-1}(r) \asymp r^{2-\gamma}$ verteilt (die Produktdichtefunktion des Elternprozesses n -ter Ordnung hat also einen Pol der Ordnung γ).

Es werden nun zwei Tochterpunkte n -ter Ordnung desselben Primärclusters betrachtet, d. h. beide Punkte sind Tochterpunkte n -ter Ordnung desselben Elternpunktes erster Ordnung. Die Verteilungsfunktion des Abstandes zwischen den beiden Punkten setzt sich additiv aus zwei Teilen zusammen

$$F(r) = p_g F_g(r) + p_u F_u(r). \quad (34)$$

Hierbei ist p_g die Wahrscheinlichkeit, dass beide Punkte demselben Elternpunkt n -ter Ordnung entstammen und $F_g(r)$ die zugehörige Verteilungsfunktion des Abstandes zweier solcher Punkte. $p_u = 1 - p_g$ und $F_u(r)$ sind die entsprechenden Größen für den Fall, dass beide Tochterpunkte unterschiedlichen Elternpunkten n -ter Ordnung entstammen.

Aus (34) folgt für die entsprechenden Dichtefunktionen

$$f(r) = p_g f_g(r) + p_u f_u(r). \quad (35)$$

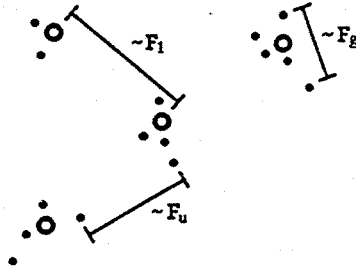


Abbildung 11: Geometrie eines Neyman-Scott-Prozesses zweiter Ordnung.

Entsprechend den obigen Überlegungen für Neyman-Scott-Prozesse erster Ordnung gilt

$$f_g(r) \asymp r^{1+2\alpha}.$$

Außerdem folgt aus dem vorherigen Kapitel

$$f_u(r) \asymp r^{4-\gamma+2\alpha}.$$

Der bezüglich der Asymptotik bestimmende Faktor in (35) ist demnach $f_g(r)$, da immer $1 + 2\alpha < 4 - \gamma + 2\alpha$. Es ist also

$$f(r) \asymp r^{1+2\alpha},$$

was zu einem Pol der Produktdichtefunktion der Ordnung $1 - 2\alpha$ führt. Offensichtlich hängt die Asymptotik von $\varrho^{(2)}(r)$ bei einem Neyman-Scott-Prozess n -ter Ordnung also nur von der Bildung der letzten Tochterpunkt-Generation ab.

Anhang

Hier wird das asymptotische Verhalten von $f^*(r)$ für $r \rightarrow 0$ bestimmt. Ausgangspunkt ist das Integral (33)

$$F^*(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) d^*(z) \frac{A(x; z; r)}{4\pi z^2} dx dz,$$

wobei $f(r) \asymp r^{2-\gamma}$ und $d^*(r) \asymp r^\alpha$ mit $\gamma < 3$ und $\alpha > -1$.

Man kann auf elementare Weise zeigen, dass das asymptotische Verhalten von $F^*(r)$ nicht von der speziellen Form von $f(r)$ und $d^*(r)$ abhängen. Die Rechnungen werden deshalb nur für den Fall $f(r) = r^{2-\gamma} \mathbf{1}_{[0,1]}(r)$ und $d^*(r) = r^\alpha \mathbf{1}_{[0,1]}(r)$ durchgeführt.

Es ist für hinreichend kleine r

$$\begin{aligned}
 F^*(r) &= \int_0^r \int_{r-z}^{r+z} \frac{r^2 - (x-z)^2}{4x^{\gamma-1}z^{1-\alpha}} dx dz \\
 &+ \int_r^1 \int_{z-r}^{z+r} \frac{r^2 - (x-z)^2}{4x^{\gamma-1}z^{1-\alpha}} dx dz \\
 &+ \int_0^r \int_0^{r-z} x^{2-\gamma} z^\alpha dx dz
 \end{aligned}$$

mit der Ableitung

$$\begin{aligned}
 f^*(r) &= \int_0^r \int_{r-z}^{r+z} \frac{r}{2x^{\gamma-1}z^{1-\alpha}} dx dz + \frac{r^{3-\gamma+\alpha}}{(3-\gamma)(4-\gamma)} - \frac{\Gamma(1+\alpha)\Gamma(3-\gamma)}{\Gamma(4+\alpha-\gamma)} r^{3-\gamma+\alpha} \\
 &+ \int_r^1 \int_{z-r}^{z+r} \frac{r}{2x^{\gamma-1}z^{1-\alpha}} dx dz - \frac{r^{3-\gamma+\alpha}}{2\gamma-2(3-\gamma)} \\
 &+ \frac{\Gamma(1+\alpha)\Gamma(3-\gamma)}{\Gamma(4+\alpha-\gamma)} r^{3-\gamma+\alpha} \\
 &= \frac{r}{2(2-\gamma)} \int_0^r \frac{(r+z)^{2-\gamma} - (r-z)^{2-\gamma}}{z^{1-\alpha}} dz \\
 &+ \frac{r}{2(2-\gamma)} \int_r^1 \frac{(z+r)^{2-\gamma} - (z-r)^{2-\gamma}}{z^{1-\alpha}} dz \\
 &+ \frac{r^{3-\gamma+\alpha}}{3-\gamma} \left(\frac{1}{4-\gamma} - 2^{2-\gamma} \right) \\
 &= \frac{r}{2(2-\gamma)} \int_0^1 \frac{(r+z)^{2-\gamma}}{z^{1-\alpha}} dz \\
 &- \frac{r}{2(2-\gamma)} \int_0^r \frac{(r-z)^{2-\gamma}}{z^{1-\alpha}} dz \\
 &- \frac{r}{2(2-\gamma)} \int_r^1 \frac{(z-r)^{2-\gamma}}{z^{1-\alpha}} dz \\
 &+ \frac{r^{3-\gamma+\alpha}}{3-\gamma} \left(\frac{1}{4-\gamma} - 2^{2-\gamma} \right) \\
 &= \frac{r}{2(2-\gamma)} {}_2F_1(\gamma-2, \alpha, 1+\alpha, -1/r) r^{2-\gamma}/\alpha
 \end{aligned}$$

$$\begin{aligned}
& - \frac{r}{2(2-\gamma)} \frac{\Gamma(3-\gamma)\Gamma(\alpha)}{\Gamma(3-\gamma+\alpha)} r^{2-\gamma+\alpha} \\
& - \frac{r}{2(2-\gamma)} \frac{(1-r)^{3-\gamma}}{3-\gamma} {}_2F_1(3-\gamma, 1-\alpha, 4-\gamma, 1-1/r) r^{-1+\alpha} \\
& + \frac{r^{3-\gamma+\alpha}}{3-\gamma} \left(\frac{1}{4-\gamma} - 2^{2-\gamma} \right).
\end{aligned}$$

Hier ist ${}_2F_1(\cdot)$ die Gauss'sche Hypergeometrische Funktion (siehe [18]). Es ist

$$\begin{aligned}
{}_2F_1(\gamma-2, \alpha, 1+\alpha, -1/r) & \asymp \begin{cases} \frac{\Gamma(1+\alpha)\Gamma(\gamma-2-\alpha)}{\Gamma(\gamma-2)} r^\alpha & 2-\gamma+\alpha < 0 \\ \frac{\alpha}{2-\gamma+\alpha} r^{\gamma-2} & 2-\gamma+\alpha \geq 0 \end{cases}, \\
{}_2F_1(3-\gamma, 1-\alpha, 4-\gamma, 1-1/r) & \asymp \begin{cases} \frac{\Gamma(4-\gamma)\Gamma(\gamma-2-\alpha)}{\Gamma(1-\alpha)(1-r)^{3-\gamma}} r^{3-\gamma} & 2-\gamma+\alpha < 0 \\ \frac{3-\gamma}{(2-\gamma+\alpha)(1-r)^{1-\alpha}} r^{1-\alpha} & 2-\gamma+\alpha \geq 0 \end{cases}.
\end{aligned}$$

Ist $2-\gamma+\alpha < 0$, so sind alle Terme asymptotisch äquivalent zu $r^{3-\gamma+\alpha}$. Im Fall $2-\gamma+\alpha \geq 0$ heben sich alle asymptotisch linearen Terme gegenseitig auf. Insgesamt gilt in jedem Fall

$$f^*(r) \asymp r^{3-\gamma+\alpha}.$$

Literatur

- [1] J. D. Barrow, S. P. Bhavsar und D. H. Sonoda: A bootstrap resampling analysis of galaxy clustering. *Mon. Not. R. astr. Soc.*, 210: 19P–23P, 1984.
- [2] G. Börner, H. J. Mo und Y. P. Ying: On the error estimates of correlation functions. *ApJ*, 392: 452–457, 1992.
- [3] O. Buryak und A. Doroshkevich: *A & A* 306: 1, 1996.
- [4] A. Cowling, P. Hall und M. J. Phillips: Bootstrap confidence regions for the intensity of a Poisson point process. *Am. Stat. Ass.*, 91(436): 1516–1524, 1998.
- [5] N. Cressie: *Spatial Statistics*. John Wiley & Sons, New York, Chichester, 1991.
- [6] D. J. Daley und P. Vere-Jones: An introduction to the theory of point processes. Springer-Verlag, New York, Berlin, Heidelberg, 1988.
- [7] M. Davis und P. J. E. Peebles: *ApJ*, 267: 465, 1983.
- [8] P. J. Diggle: *Statistical analysis of point processes*. Chapman and Hall, London, 1983.
- [9] T. Fiksel: Edge-corrected density estimators for point processes. *Statistics*, 19: 67–76, 1988.
- [10] P. Hall: Resampling a coverage process. *Stoch. Proc.*, 20: 231–246, 1985.
- [11] A. J. S. Hamilton *ApJ*, 417: 19, 1993.
- [12] L. Heinrich: Minimum contrast estimates for parameters of spatial ergodic point processes. In: *Transactions of the 11th Prague conference on random processes, information theory and statistical decision functions*, 479–492, 1992.
- [13] L. Heinrich: Asymptotic properties of minimum contrast estimators for parameters of Boolean models. *Metrika*, 40: 67–94, 1993.
- [14] Hewett: *MNRAS*, 201: 867, 1982.
- [15] L. Heinrich, R. Körner, N. Mehlhorn und L. Muche: Numerical and analytical computation of some second-order characteristics of spatial Poisson-Voronoi tessellations. *Statistics*, 31: 235–259, 1998.
- [16] M. Kerscher: The geometry of second-order statistics – biases in common estimators. *A & A*, 343: 333–347, 1999.
- [17] D. König und V. Schmidt: *Zufällige Punktprozesse*. B G Teubner, Stuttgart, 1992.

- [18] A. Kratzer und W. Franz: *Transzendente Funktionen*. Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1960.
- [19] H. R. Künsch: The jackknife and the bootstrap for general stationary observations. *Ann. Stat.*, 17(3): 1217–1241, 1989.
- [20] S. D. Landy und A. S. Szalay: *ApJ*, 412: 64–71, 1993.
- [21] G. Last und M. Holtmann: On the empty space function of some germ-grain models. *Pattern Recognition*, 32: 1587–1600, 1999.
- [22] E. N. Ling, C. S. Frenk und J. D. Barrow: Uncertainties in the cluster-cluster correlation function. *Mon. Not. R. astr. Soc.*, 223: 21p–27p, 1986.
- [23] B. Matérn: *Spatial Variation*. Lecture Notes in Statistics 36. Springer-Verlag, Berlin, Heidelberg, New York, 1986.
- [24] G. Matheron: *Random sets and integral geometry*. J. Wiley & Sons, New York, 1975.
- [25] J. Møller, A. R. Syversveen und R. P. Waagepetersen: Log Gaussian Cox processes. *Scand. J. Stat.*, 25(3): 451–482, 1998.
- [26] J. Ohser: On estimation of pair correlation function. *Research in Informatics*, 4: 147–152, 1991.
- [27] P. J. E. Peebles: *The large scale structure of the universe*. Princeton Univ. Press, Princeton, 1980.
- [28] P. J. E. Peebles und M. G. Hauser: *ApJS*, 28: 19, 1974.
- [29] M.-J. Pons-Bordería, V. J. Martínez, D. Stoyan, H. Stoyan und E. Saar: Comparing estimators of the galaxy correlation function. *ApJ*, 523: 480–491, 1999.
- [30] L. Sachs: *Angewandte Statistik*. Springer-Verlag, Berlin Heidelberg New York, 1974.
- [31] M. Sneath: Is bootstrap really helpful in Point process statistics? *Metrika*, 49: 245–255, 1999.
- [32] D. Stoyan: Statistical Estimation of Model Parameters of Planar Neyman-Scott Cluster Processes. *Metrika*, 39: 67–74, 1992.
- [33] D. Stoyan: Caution with “fractal” point patterns! *Statistics*, 25: 267–270, 1994.
- [34] D. Stoyan, U. Bertram und H. Wendrock: Estimation variances for estimators of product densities and pair correlation functions of planar point processes. *Ann. Inst. Statist. Math.*, 45 (2): 211–221, 1993.

- [35] D. Stoyan, W. S. Kendall und J. Mecke: *Stochastic geometry and its applications*. J. Wiley & Sons, Chichester, 1995.
- [36] D. Stoyan und H. Stoyan: *Fraktale-Formen-Punktfelder*. Akademie Verlag, Berlin, 1992.
- [37] D. Stoyan und H. Stoyan: Estimating Pair Correlation Functions of Planar Cluster Processes. *Biom. J.*, 38: 259-271, 1996.
- [38] D. Stoyan und H. Stoyan: Improving ratio estimators of second order point process characteristics. *Scand. J. Stat.*, 2000.
- [39] G. Upton und B. Fingleton: *Spatial data analysis by example* vol. 1. John Wiley & Sons, Chichester, 1985.
- [40] V. Icke und R. Van de Weygaert: *Astron. Astrophys.*, 184: 16, 1987.
- [41] R. Van de Weygaert und V. Icke: *Astron. Astrophys.*, 213: 1, 1989.
- [42] R. Van de Weygaert: *Monthly Notices Roy. Astron. Soc.*, 249: 159, 1991.