



Räumliche Statistik

Prof. Dr. Volker Schmid

Clara Happ

Wintersemester 2015/16

Institut für Statistik

Ludwig-Maximilians-Universität München

Nach einer Vorlage von Prof. Dr. Ludwig Fahrmeir. Mit Beiträgen von Michael Höhle, Thomas Kneib, Martina Feilke. \LaTeX -Satz von Eva Kraus und Wiktor Olszowy.

Version vom 21. Dezember 2015.

Inhaltsverzeichnis

1	Einführung	1
1.1	Räumliche Daten: Beispiele	1
1.1.1	Geostatistische Daten	1
1.1.2	Gitterdaten (areal data, lattice data)	2
1.1.3	Räumliche Punktmuster	2
1.2	Räumliche stochastische Prozesse	3
2	Markovzufallsfelder	5
2.1	Daten	5
2.2	Markov-Zufallsfelder (MRF, Markov Random Fields)	6
2.2.1	Nachbarschaftssysteme	7
2.2.2	Hammersley-Clifford-Theorem	10
2.2.3	Besag's Lemma	12
2.3	Gauß-Markov Zufallsfelder (GMRF)	13
2.3.1	Definition	13
2.3.2	Eigenschaften	14
2.3.3	Arten des GMRF	15
2.3.4	Ziehen aus GMRF	16
2.3.5	GMRF auf regulären Gittern	17
2.4	Latente Markov-Zufallsfelder	19
2.4.1	Besag-York-Mollie Modell	20
2.4.2	Hilfsvariablenansatz	22
2.4.3	Latente Ising/Potts-Felder	24
3	Geostatistik	25
3.1	Gauß-Prozesse	25
3.1.1	Stationarität	25
3.1.2	Theoretisches (Ko-)Variogramm und Korrelationsfunktion	26
3.2	Empirisches Variogramm	27
3.2.1	Schätzer	27

3.2.2	Bezeichnungen	29
3.3	Parametrische Modelle für Korrelationsfkt. und (Ko-)Variogramme	30
3.3.1	Gültigkeit von Korrelationsfunktionen	30
3.3.2	Übliche Korrelationsfunktionen	31
3.4	Anisotrope und nicht-stationäre Prozesse	32
3.5	Kriging	34
3.5.1	Einfaches Kriging	34
3.5.2	Kriging im Gauß-Modell	36
3.5.3	Inferenz	37
3.5.4	Generalisiertes Kriging	39
3.6	Diskrete Approximation von Gaussfeldern	42
4	Räumliche Regularisierung	44
4.1	Nichtparametrische Schätzung von Oberflächen	44
4.1.1	Nichtparametrische Regression mit penalisierten Splines	45
4.1.2	Oberflächenschätzung mit radialen Basisfunktionen . .	49
4.2	Geoadditive Regression	51
4.2.1	Gauß-Fall	51
4.2.2	Generalisierte Geoadditive Modelle	53
4.2.3	Deviance Information Criterion	56
5	Räumliche Punkt- und Zählprozesse	57
5.1	Einführung in räumliche Punktprozesse	57
5.2	Binomiale und Poisson-Prozesse	60
5.2.1	Bernoulli- und binomiale Prozesse	60
5.2.2	Homogener Poissonprozess	61
5.2.3	Inhomogener Poissonprozess	61
5.2.4	Simulation aus Poissonprozessen	63
5.3	Maßzahlen, explorative Analyse und Inferenz	64
5.3.1	Test auf rein zufälliges räumliches Muster	64
5.3.2	Schätzung der Intensitätsfunktion	66
5.3.3	Maßzahlen 2. Ordnung für räumliche Interaktion	67
5.3.4	Edge correction	69
5.4	Markierte Punktprozesse	72
5.4.1	Multivariate Markierungen	72
5.4.2	Quantitative Markierungen	73
5.5	Cluster-Modelle	75
5.5.1	Poisson-Cluster-Prozesse	75
5.5.2	Cox-Prozesse	76
5.5.3	Nicht-stationäre Prozesse mit Interaktionen	78
5.6	Log-Gauß-Cox-Prozesse	82

5.6.1	Data-Augmentation-Ansatz für latentes GRF	83
5.6.2	Modellierung der Intensität über latentes GMRF	84

Kapitel 1

Einführung

1.1 Räumliche Daten: Beispiele

1.1.1 Geostatistische Daten

Gegeben seien x/y - bzw. $x/y/z$ -Koordinaten in stetigem Bereich $D \subseteq \mathbb{R}^2, \mathbb{R}^3$; s variiert stetig in D :

$$s = (s_x, s_y) \in \mathbb{R}^2 \quad \text{bzw.} \quad s = (s_x, s_y, s_z) \in \mathbb{R}^3$$

Daten: $y(s_1), \dots, y(s_n)$, an n vorgegebenen Lokationen s_1, \dots, s_n .

Beispiel 1. (Schweizer Regenfalldaten)

Messung der Regenfallmenge;

s_1, \dots, s_n : (Mess-) Stationen mit stetigen Koordinaten;

$Y(s_1), \dots, Y(s_n)$: stetige geostatistische Daten.

Beispiel 2. (Rongelap-Daten)

Messung der Radioaktivität nach Atombombentests;

s_1, \dots, s_n : (Mess-) Stationen mit stetigen Koordinaten;

$Y(s_1), \dots, Y(s_n)$: Zähldaten.

Beispiel 3. (Waldschäden im Spessart)

kategorialer Schädigungsgrad von Bäumen;

s_1, \dots, s_n : Standort der Bäume (in x/y -Koordinaten)

$X_1 = X(s_1), \dots, X_n = X(s_n)$: Vektor von standort-/baumspezifischen Kovariablen

$Y_1 = Y(s_1), \dots, Y_n = Y(s_n)$: Schädigungsgrad von $n = 83$ Bäumen.

Beispiel 4. (Überlebenszeit nach Leukämie-Diagnose)

s_1, \dots, s_n : Wohnort der n Individuen (in x/y -Koordinaten);

$Y_i = Y(s_i)$: Überlebenszeiten

X_i : Kovariablen, $i = 1, \dots, n$

Beispiel 5. (Signale aus dem Universum)

Messungen eines Signals aus verschiedenen Winkelrichtungen; Ziel: Rekonstruktion der wahren 3d-Signale;

Y : Signalintensität; Y stetig

s_1, \dots, s_n : Messungen

1.1.2 Gitterdaten (areal data, lattice data)

$D \subset \mathbb{R}^d$, $d \geq 2$, partitioniert in eine vorgegebene, endliche, aber relativ große Anzahl räumlicher Einheiten mit wohldefinierten Grenzen. $s \in \{1, \dots, S\}$ räumlicher Index für die Regionen. „Gitter“ kann unregelmäßig (z.B. administrative Einheiten) oder regelmäßig (z.B. Pixel in einem Bild) sein.

Daten: y_1, \dots, y_S , in den Gitterzellen $1, \dots, S$.

Beispiel 6. (Wildschweine in Brandenburg)

Test auf PRV an frei lebenden Wildschweinen;

s : Index der Gemeinde;

$Y(s) = Y_s$: Zahl der infizierten Wildschweine in Gemeinde s ;

$N(s) = N_s$: Zahl der untersuchten Wildschweine in Gemeinde s .

Beispiel 7. (Disease Mapping)

s : Nummer der Region

$Y(s)$: (Standardisierte) Mortalitätsrate in Region s ; Zählvariable.

Beispiel 8. (Bildanalyse)

Regelmäßiges 2d- oder (3d-)Gitter;

$s \hat{=} (i, j)$: Nummer des Pixels (oder Voxels) mit Koordinate $i, j, (k)$;

$Y_s = Y_{ij}$

Fortsetzung Beispiel 4 (Überlebenszeit nach Leukämie-Diagnose)

s_1, \dots, s_n : Region des Wohnorts von n Individuen;

$Y_i = Y(s_i)$: Überlebenszeiten

X_i : Kovariablen, $i = 1, \dots, n$

1.1.3 Räumliche Punktmuster

s_1, \dots, s_n : Lokationen von zufälligen Ereignissen

$D = \{s_1, \dots, s_n\}$: „zufällige“ Menge

$y(s) = 1$: „Ereignis aufgetreten“ oder mit zusätzlicher Kovariablen-Information

(„Markierung“),

z.B. $y(s) = k \in \{1, \dots, K\}$, „Ereignis vom Typ k aufgetreten“.

Beispiel 9. (Phlebocarya filifolia)

Daten: Positionen von 207 Pflanzen.

Fragen: ist die Verteilung zufällig? Wurde auf einem regulären Gitter angepflanzt? Bilden die Pflanzen Cluster oder machen sie sich gegenseitig Konkurrenz?

Beispiel 10. (Norwegische Fichten)

Positionen von norwegischen Fichten in einem $56 \times 38\text{m}$ Ausschnitt. Radien der Kreise = $5 \times$ Stammdurchmesser \approx Einflusszone eines Baumes.

Fragen: Ist die räumliche Verteilung der Bäume zufällig? Wie hängt sie mit dem Umfang zusammen?

Fortsetzung Beispiel 5 (Sternverteilung)

Ziel: Analyse der Verteilung der Sterne im Universum.

1.2 Räumliche stochastische Prozesse

Räumliche Daten $y(s)$ werden als Realisierungen eines räumlichen stochastischen Prozesses $\{Y(s), s \in D\}$, $D \subset \mathbb{R}^d$ bzw. eines Zufallsfeldes angesehen.

Definition 1.1. (*Stochastischer Prozess*)

Sei (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum und T eine Indexmenge (i.d.R. $T \in \{\mathbb{N}_0, \mathbb{R}_+\}$). Sei $\{Y_t, t \in T\}$ eine Familie von Zufallsvariablen:

$$Y_t : (\Omega, \mathcal{F}, P) \rightarrow (Z, \mathcal{F}')$$

$\forall t$ mit Zustandsraum Z mit σ -Algebra \mathcal{F}' . Dann heißt $(\Omega, \mathcal{F}, P, (Y_t, t \in T))$ ein stochastischer Prozess.

- Oft auch einfach $\{Y_t, t \in T\}$ ist stochastischer Prozess.
- Z i.d.R. abzählbar (diskret), reellwertig oder vektoriell.
- Spezifizierung des SP oft über gemeinsame Verteilung von $(Y_{t_1}, \dots, Y_{t_n})$ für alle mögliche $t_1, \dots, t_n \in T$.

Definition 1.2. (*Zufallsfeld*)

Sei nun S ein topologischer Raum. Sonst wie in Definition 1.1. Dann heißt $(\Omega, \mathcal{F}, P; (Y(s), s \in S))$ Zufallsfeld.

Drei Typen von Daten: D stetig / diskret / zufällig. Dazu passen drei Klassen räumlicher Prozesse:

- (Stationäre) Zufallsfelder, z.B. Gauß-Zufallsfelder (Gaussian random fields, GRF): $\{Y(s_1), \dots, Y(s_n)\}$ für beliebige n , s_1, \dots, s_n , multivariat (Gauß-)verteilt (und stationär).
Grundmodell der **Geostatistik** (Kapitel 3).
- **Markov-Zufallsfelder** (Markov random fields, MRF, Kapitel 2)
 $\{Y(s), s \in \{1, \dots, S\}\}$ besitzt eine räumliche Markoveigenschaft (in Verallgemeinerung der Markoveigenschaft in diskreter Zeit). Wichtigster Fall: Gauß-Markov-Zufallsfelder (Gaussian Markov random fields, GMRF): Gemeinsame Verteilung der Y ist eine multivariate Normalverteilung mit einer Kovarianzmatrix, die sich aus der Nachbarschaftsstruktur der Daten ergibt.
- **Räumliche Punkt-Prozesse** (Kapitel 5)
Verallgemeinerung zeitlicher Punkt- bzw. Zählprozesse, z.B. des Poisson-Prozesses.

Ergänzungen:

- Alle genannten räumliche Prozesse lassen sich prinzipiell auf räumlich-zeitliche Prozesse erweitern.
- Die Prozess-Typen werden oft nicht für die Daten $Y(s)$ selbst, sondern als latente räumliche Komponente in einem hierarchischen Modell verwendet.
- Aus den Ansätzen für latente Modelle lassen sich Methoden zur räumlichen Regularisierung in Regressionsmodellen entwickeln (siehe Kapitel 4).

Kapitel 2

Markovzufallsfelder

2.1 Daten

Fortsetzung Beispiel 8 (Ratbert)

Farbintensitäten auf regulärem Gitter.

Fortsetzung Beispiel 7 (Disease Mapping)

- Anzahl von Erkrankungen oder Todesfällen in bestimmten Zeitperioden, geschichtet nach Geschlecht, Altersgruppen und Regionen.
- Keine individuellen Kovariablen.
- y_{sj} sei die beobachtete Anzahl von Fällen in Region $s = 1, \dots, S$ und Altersgruppe $j = 1, \dots, J$.
- n_{sj} sei die Anzahl von Personen unter Risiko (Bevölkerungszahl in Region s und Altersgruppe j).

- Binomialmodell

$$Y_{sj} \sim \text{Bin}(n_{sj}, \pi_{sj})$$

approximiert durch Poisson-Modell

$$Y_s \sim \text{Po}(e_s \lambda_s)$$

- e_s erwartete Anzahl von Fällen in Region s
- p_j altersspezifische Referenzwahrscheinlichkeiten

$$e_s = \sum_j n_{sj} p_j$$

- ML-Schätzer für λ_s bei Unabhängigkeitsann. für $\{Y_s, s = 1, \dots, S\}$:
Standardisierte Mortalitäts-Rate SMR

$$SMR_s = \frac{y_s}{e_s}, \quad s = 1, \dots, S$$

$$\mathbb{E}[SMR_s] = \frac{\mathbb{E}[Y_s]}{e_s} = \frac{e_s \lambda_s}{e_s} = \lambda_s$$

- Der Schätzer ist erwartungstreu, aber

$$\text{Var}(SMR_s) = \text{Var}\left(\frac{Y_s}{e_s}\right) = \frac{1}{e_s^2} \text{Var}(Y_s) = \frac{\lambda_s}{e_s}$$

\Rightarrow ML-Schätzer sehr unzuverlässig bei seltenen Krankheiten (e_s klein)

Ziel: Stabilisierung der Schätzung durch Annahme räumlicher Korrelation und damit Berücksichtigung von Nachbarschaftsinformation.

Beispiel 11. (Mietspiegel)

Y_i Nettomiete,
 x_i Kovariablen (Baujahr, Wohnfläche, binäre Variablen);
 s_i Subbezirksviertel der Wohnung i in München.

Allgemein

y_i Wert der Zielvariablen für Individuum i
 x_i Kovariablen
 $s_i \in \{1, \dots, S\}$ räumliche Einheit in der i „lebt“
 $Y_i | x_i, s_i \sim$ Gauß, Binomial, Poisson, ...

2.2 Markov-Zufallsfelder (MRF, Markov Random Fields)

Wir betrachten diskrete Zufallsfelder mit Markoveigenschaft:

Definition 2.1. Sei $Y(s)$ ein Zufallsfeld mit $s = \{1, \dots, S\}$. Dann nennt man mit $Y_i = Y(s_i)$ (siehe Abb. 2.2).

- Paarweise Markoveigenschaft: $Y_i \perp Y_j | Y_{-ij}$ falls $i \neq j$ und i kein Nachbar von j ,

- Lokale Markoveigenschaft $Y_i \perp Y_{-\{i, n \in (i)\}} \mid Y_{n \in (i)}, n \in (i) = \{j : i \sim j\}$
Nachbarn von i ,
- Globale Markoveigenschaft: $Y_A \perp Y_B \mid Y_C$ für alle disjunkten Mengen A, B und C , wobei C die (nichtleeren) Mengen A und B separiert.

Erfüllt $Y(s)$ eine Markoveigenschaft, so nennt man $Y(s)$ ein **Markovzufallsfeld**.

Anmerkung:

- Die Markoveigenschaften sind (im Allgemeinen) nicht äquivalent. Die globale Markoveigenschaft ist die stärkste, die paarweise die schwächste.
- Die Markoveigenschaft für zeitliche Prozesse (Gedächtnislosigkeit) ergibt sich als Spezialfall.
- Für räumliche Prozesse müssen wir erst definieren, was überhaupt ein Nachbar ist.

2.2.1 Nachbarschaftssysteme

$n(s)$ Menge der Nachbarn (nächste Pixel, angrenzende Regionen)

D Parameterraum von s , $D = \{1, \dots, S\}$ oder $D \subseteq \mathbb{Z}^d$, $d = 1, 2, 3, \dots$

Definition 2.2. (Nachbarschaftssystem)

Eine Menge von Orten $\partial = \{\partial(s) : s \in D\}$ ist ein Nachbarschaftssystem, wenn gilt:

$$s \notin \partial(s)$$

$$v \in \partial(s) \Leftrightarrow s \in \partial(v)$$

Alle $v \in \partial(s)$ heißen Nachbarn von s ($v \sim s$).

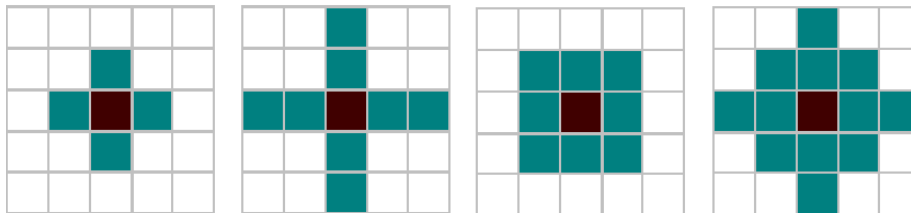


Abbildung 2.1: Nachbarschaftssysteme auf regulären Gittern. Von links nach rechts: Erste Nachbarn, zweite Nachbarn, zweite diagonale Nachbarn, zweite und diagonale Nachbarn.

Fortsetzung Beispiel 8 (Bildanalyse)

Auf regulären Gittern siehe Abb. 2.1.

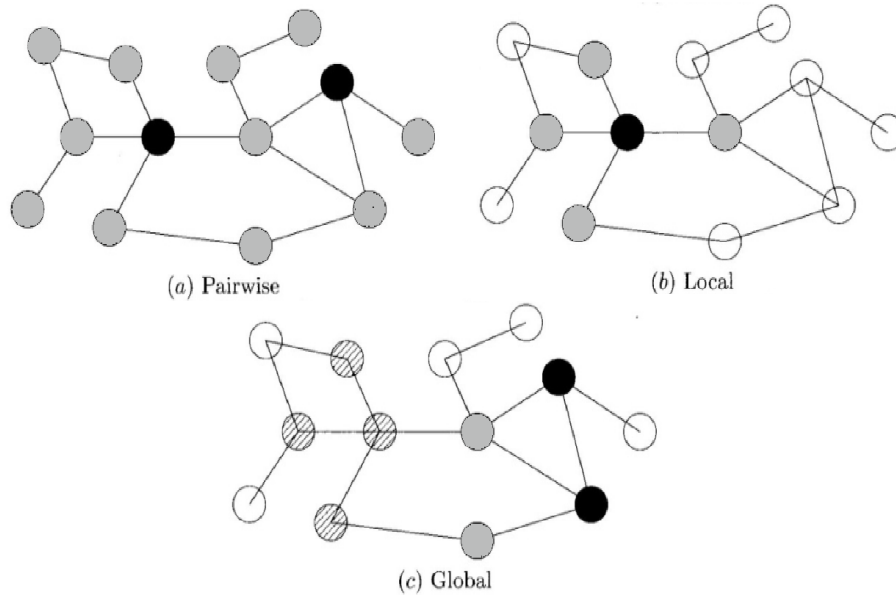
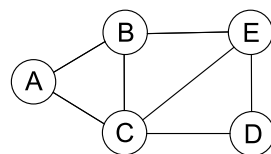


Abbildung 2.2: Markov-Eigenschaften: (a) Paarweise Markov-Eigenschaft, die schwarzen Knoten sind gegeben die grauen Knoten bedingt unabhängig. (b) Lokale Markov-Eigenschaft, schwarze und weiße Knoten sind bedingt unabhängig gegeben die grauen Knoten. (c) Globale Markov-Eigenschaft, schwarze und gestrichelte Knoten sind bedingt unabhängig gegeben den grauen Knoten.

Fortsetzung Beispiel 7 (Irreguläre Gitter)

Auf irregulären Gittern wie Landkarten in der Regel: Nachbarn haben gemeinsame Grenze.

Beispiel 12. (Ungerichtete Graphen/Markov-Netzwerke)



Definition 2.3. (Clique)

Eine Menge C von Knoten (Indizes, Regionen, Pixeln, ...) heißt Clique:

\Leftrightarrow Jedes Paar (i, j) von Knoten sind Nachbarn $i \sim j$.

Oft versteht man unter Clique auch die *maximale Clique*.

Fortsetzung Beispiel 8 (Reguläre Gitter)

Erste Nachbarn

$$C_1 = \left\{ \emptyset, \square, \square-\square, \begin{array}{c} \square \\ | \\ \square \end{array} \right\}$$

Zweite Nachbarn

$$C_2 = C_1 \cup \left\{ \square-\square-\square, \begin{array}{c} \square \\ | \\ \square \\ | \\ \square \end{array} \right\}; \quad \text{nicht: } \begin{array}{c} \square-\square \\ | \\ \square \end{array}$$

Definition 2.4 (Markov-Zufallsfeld). $Y = \{Y_s, s \in D\}$ heißt MRF: \Leftrightarrow

$$p(y_s | y_{s'}, s' \neq s) = p(y_s | y_{s'}, s' \in \partial(s))$$

für alle $s, s' \in D$

Beispiel 13. (Autoregressive Prozesse)

- Zeitlich:

$$Y_t = \phi Y_{t-1} + \varepsilon_t \Rightarrow Y_t | Y_{t-1}, \dots, Y_0 \sim Y_t | Y_{t-1} \sim N(\phi Y_{t-1}, \sigma^2)$$

was dem Random Walk im Fall $\phi = 1$ entspricht.

- Räumlich:

$$Y_s = \sum_{s' \in n(s)} w_{ss'} Y_{s'} + \varepsilon_s \Rightarrow$$

$$n(s) \ni s' \sim N \left(\sum_{s' \in n(s)} w_{ss'} Y_{s'}, \sigma^2 \right)$$

Wird durch Vorgabe eines räumlich autoregressiven Modells überhaupt in konsistenter Weise eine gemeinsame Verteilung definiert? Allgemeiner: Lässt sich durch Vorgabe bedingter Verteilungen in der MRF-Definition eine gemeinsame Verteilung konsistent definieren?

Beispiel 14. (Gegenbeispiel)

$$Y_1 | Y_2 \sim N(\alpha_0 + \alpha_1 Y_2, \sigma_1^2), \quad Y_2 | Y_1 \sim N(\beta_0 + \beta_1 Y_1^3, \sigma_2^2)$$

$$\Rightarrow \begin{aligned} \mathbb{E}[Y_1] &= \mathbb{E}[\mathbb{E}[Y_1 | Y_2]] = \mathbb{E}[\alpha_0 + \alpha_1 Y_2] = \alpha_0 + \alpha_1 \mathbb{E}[Y_2] \\ \mathbb{E}[Y_2] &= \mathbb{E}[\mathbb{E}[Y_2 | Y_1]] = \mathbb{E}[\beta_0 + \beta_1 Y_1^3] = \beta_0 + \beta_1 \mathbb{E}[Y_1^3] \end{aligned}$$

Bis auf Spezialfälle: inkompatibel! D.h. $f(y_1|y_2)$ und $f(y_2|y_1)$ definieren keine gemeinsame Dichte $f(y_1, y_2)$.

2.2.2 Hammersley-Clifford-Theorem

Nachbarschaftsstruktur durch Graphen gegeben: i, j Nachbarn: \Leftrightarrow direkte Kante zwischen Knoten $i \sim j$

Definition 2.5. (Positivität)

Seien T_1, \dots, T_n die Träger von Y_1, \dots, Y_n , d.h. $T_i = \{y : p(y_i) > 0\}$. Falls $T = T_1 \times \dots \times T_n$ der Träger von $Y = (Y_1, \dots, Y_n)$, spricht man von Positivität.

Satz 2.1. (Hammersley-Clifford-Theorem)

$p(y)$ besitze die Positivitätseigenschaft, d.h.

$p(y) > 0$ für beliebige $y_1 \in T_1, \dots, y_i \in T_i, \dots, y_n \in T_n$. Zu einer gegebenen Nachbarschaftsstruktur ist Y ein MRF \Leftrightarrow Die gemeinsame Dichte $p(y)$ hat bis auf die Normalisierungskonstante die Form

$$p(y) \propto \exp \left(\sum_C \phi_C(y_C) \right)$$

wobei C alle einzelnen Elemente $i \in \{1, \dots, n\}$ und alle Cliques $C \subseteq \{1, \dots, n\}$ durchläuft, y_C der entsprechende Teilvektor von C ist und die Funktionen ϕ_C invariant gegenüber Vertauschungen / Permutationen der Komponenten von y_C sind.

Bemerkungen:

- Damit ist zulässige Form von gemeinsamen Dichten bei MRF charakterisiert. Die explizite Berechnung der Normalisierungskonstanten ist i.A. nicht in geschlossener analytischer Form möglich.

- Bei uns bestehen Cliques meist nur aus 2 Elementen. Dann

$$p(y) \propto \exp \left(\sum_{i=1}^n \phi_i(y_i) + \sum_{i \sim j} \phi_{i,j}(y_i, y_j) \right)$$

- Dichten dieser Art werden auch Gibbs-Feld oder Gibbs-Maß genannt. ϕ heißt Gibbs-Energie.

Beispiel 15. (Auto-logistisches Modell)

Y_i binär, $T_i = \{0, 1\}$, $i = 1, \dots, n$

Nachbarschaftsstruktur so, dass nur Cliques mit zwei Elementen (i, j) auftreten.

Hammersley-Clifford-Theorem \Rightarrow

$$p(y) \propto \exp \left(\sum_i \alpha_i y_i + \sum_{i \sim j} \beta_{ij} y_i y_j \right)$$

zulässiges MRF:

$$p(y_i | y_{-i}) \propto p(y) \propto \exp \left(\alpha_i y_i + \sum_{j \in \partial(i)} \beta_{ij} y_i y_j \right)$$

$$\Leftrightarrow \log \frac{\mathbb{P}(Y_i = 1 | Y_{-i})}{\mathbb{P}(Y_i = 0 | Y_{-i})} = \alpha_i + \sum_{j \in \partial(i)} \beta_{ij} Y_j$$

- Logit Modell; mit Kovariablen: $\alpha_i = x_i' \alpha$
- Probit-Modell nicht zulässig! (\Rightarrow latentes GMRF mit Probitlink)

Beispiel 16. (Ising-Modell für Schwarz-Weiß-Bilder)

4 nächste Nachbarn $y_i \in \{0, 1\}$

$i = 1, \dots, n$ kodiert weiß/schwarz für Pixel i

$$p(y) \propto \exp \left(\beta \sum_{i \sim j} \mathbf{1}(y_i = y_j) \right) \mathbf{1} \text{ Indikatorfunktion}$$

entspricht Bsp. 15.

Beispiel 17. (Potts-Modell für Farbstufen)

$$y_i \in \{1, \dots, k\}$$

$$p(y) \propto \exp \left(\beta \sum_{i \sim j} \mathbb{1}(y_i = y_j) \right)$$

$\Rightarrow p(y_i | y_{-i})$ multinomiales Logit-Modell zulässig. Andere multinomiale Modelle, z.B. multinomiales Probit-Modell, ordinale Modelle, unzulässig!

Beispiel 18. (Gauß-Markov-Zufallsfeld)

Für erste Nachbarn:

$$p(y) \propto \exp \left(-\frac{1}{2} \sum_{i \sim j} w_{ij} (y_i - y_j)^2 \right)$$

Beispiel 19. (Räumlich autoregressiver Prozess bei Exponential-Familie mit kanonischem Link)

$$p(y_i | y_{-i}) \propto \exp \left(\frac{\theta_i y_i - b(\theta_i)}{\phi} \right)$$

$$\theta_i = \alpha_i y_i + \sum_{j \sim i} \beta_{ij} y_i y_j$$

2.2.3 Besag's Lemma

$Y = (Y_1, \dots, Y_n)$ n -dimensionale ZV

$p(y_1, \dots, y_n)$ gemeinsame Dichte (bzw. W.fkt.)

$p(y_i | y_{-i}) = p(y_i | y_j, j \neq i)$ vollständig bedingte Dichten

Bei gegebener gemeinsamer Dichte sind die bedingten Dichten eindeutig bestimmt. Umgekehrt:

- Bestimmen die vollständig bedingten Dichten die gemeinsame Dichte?
- Sind Restriktionen zu beachten in der Wahl eines Modells für $p(y_i | y_{-i})$?
- Wie berechnet man $p(y_1, \dots, y_n)$?

Satz 2.2. (*Besag's Lemma/Brook's Lemma*)

$$\frac{p(y)}{p(\tilde{y})} = \prod_{i=1}^n \frac{p(y_i | y_1, \dots, y_{i-1}, \tilde{y}_{i+1}, \dots, \tilde{y}_n)}{p(\tilde{y}_i | y_1, \dots, y_{i-1}, \tilde{y}_{i+1}, \dots, \tilde{y}_n)}$$

Bemerkungen:

- Setze \tilde{y} fest, dann kann $p(y)$ aus bedingten Verteilungen (bis auf die Normalisierungskonstante) berechnet werden.

- Satz ist unabhängig von Anordnung ($n!$ Möglichkeiten)
 \Rightarrow Kompatibilitätsbedingung.

Beweis:

$$\frac{p(y_n|y_1, \dots, y_{n-1}) p(y_1, \dots, y_{n-1})}{p(\tilde{y}_n|y_1, \dots, y_n) p(y_1, \dots, y_{n-1})} = \frac{p(y_1, \dots, y_{n-1}, y_n)}{p(y_1, \dots, y_{n-1}, \tilde{y}_n)}$$

Daraus folgt

$$p(y_1, \dots, y_n) = \frac{p(y_n|y_1, \dots, y_{n-1})}{p(\tilde{y}_n|y_1, \dots, y_{n-1})} p(y_1, \dots, y_{n-1}, \tilde{y}_n)$$

Entwickle analog

$$p(y_1, \dots, y_{n-1}, \tilde{y}_n) = \frac{p(y_{n-1}|y_1, \dots, y_{n-2}, y_n)}{p(\tilde{y}_{n-1}|y_1, \dots, y_{n-2}, \tilde{y}_n)} p(y_1, \dots, y_{n-2}, \tilde{y}_{n-1}, \tilde{y}_n)$$

□

2.3 Gauß-Markov Zufallsfelder (GMRF)

2.3.1 Definition

Definition 2.6. Sei $Y \sim N(\mu, \Sigma)$. Dann heißt $P = \Sigma^{-1}$ Präzisionsmatrix. Dabei darf Σ auch singulär sein, also $\text{Rang}(\Sigma) < n$. Dann ist Σ^{-1} als verallgemeinerte Inverse aufzufassen.

Satz 2.3. Y sei normalverteilt mit $\mathbb{E}[Y] = \mu$ und $P = \Sigma^{-1} > 0$

$$\Rightarrow Y_i \perp Y_j | Y_{-ij} \Leftrightarrow p_{ij} = 0 \quad \text{für alle } i \neq j$$

Die Elemente $p_{ij} \neq 0$ in P bestimmen also den Nachbarschaftsgraphen.

Definition 2.7 (Gauß-Markov-Zufallsfeld). $Y \sim N(\mu, P^{-1})$, heißt GMRF bezüglich der Nachbarschaftsstruktur eines Graphen: \Leftrightarrow

$$p(y) = (2\pi)^{-n/2} |P|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(y - \mu)' P (y - \mu)\right)$$

und $p_{ij} \neq 0 \Leftrightarrow i \sim j$ (i und j sind Nachbarn).

$P > 0$ heißt Präzisionsmatrix (Inverse der Kovarianzmatrix).

Beispiel 20. (GMRF auf deutschen Ländern)

Definition 2.8 (Kanonische Schreibweise der Normalverteilung).

$$\begin{aligned} Y \sim N^k(m, P) &= N(P^{-1}m, P^{-1}) \\ \Leftrightarrow p(y) &\propto |P|^{\frac{1}{2}} \exp\left(-\frac{1}{2}y'Py + m'y\right) \end{aligned}$$

2.3.2 Eigenschaften

Satz 2.4 (Bedingte Verteilungen eines GMRF). $Y \sim N(\mu, P^{-1})$ sei GMRF bezüglich eines Nachbarschaftsgraphen. Dann gilt

$$\begin{aligned} \mathbb{E}[Y_i | Y_{-i}] &= \mu_i - \frac{1}{p_{ii}} \sum_{j:j \sim i} p_{ij}(Y_j - \mu_j) \\ \text{Var}(Y_i | Y_{-i}) &= \frac{1}{p_{ii}} \\ \text{Cov}(Y_i, Y_j | Y_{-ij}) &= -\frac{p_{ij}}{\sqrt{p_{ii}p_{jj}}} \end{aligned}$$

Satz 2.5 (Markov-Eigenschaften von GMRF). Sei $Y(s)$ ein GMRF. Dann besitzt Y die in Def. 2.1 genannten Markov-Eigenschaften. Siehe Abb. 2.2)

Satz 2.6 (Spezifikation durch vollständig bedingte Dichten).

$$\begin{aligned} \mathbb{E}[Y_i | Y_{-i}] &= \mu_i + \sum_{j:i \sim j} c_{ij}(Y_j - \mu_j) \\ \text{Var}(Y_i | Y_{-i}) &= \sigma_i^2 > 0 \\ \Rightarrow Y &\sim N(\mu, P^{-1}), \text{ also GMRF mit} \\ p_{ij} &= \begin{cases} -c_{ij}/(\sigma_i\sigma_j), & i \neq j \\ 1/\sigma_i^2, & i = j \end{cases} \end{aligned}$$

wobei c_{ij} den Gewichtungsfaktor bezeichnet.

Beweis: Besag's Lemma.

Sei $Y_i | Y_{-i}$ normalverteilt mit

$$\mathbb{E}[Y_i | Y_{-i}] = \mu_i + \sum_{j=1}^n c_{ij}(Y_j - \mu_j) \quad \text{und} \quad \text{Var}(Y_i | Y_{-i}) = \sigma_i^2$$

Wir zeigen mit Besag's Lemma, dass $Y = (Y_1, \dots, Y_n)'$ multivariat normalverteilt ist. Sei $C = (c_{ij})$, $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$, $Y_0 = \mu = (\mu_1, \dots, \mu_n)'$. Dann gilt für die (logarithmierte) Besag's Zerlegung

$$\begin{aligned}
\log \left(\frac{p(y)}{p(y_0)} \right) &= \log \left(\prod_{i=1}^n \frac{p(y_i | \mu_1, \dots, \mu_{i-1}, y_{i+1}, \dots, y_n)}{p(\mu_i | \mu_1, \dots, \mu_{i-1}, y_{i+1}, \dots, y_n)} \right) = \\
&= \log \left(\prod_{i=1}^n \frac{p(y_i | y_1, \dots, y_{i-1}, \mu_{i+1}, \dots, \mu_n)}{p(\mu_i | y_1, \dots, y_{i-1}, \mu_{i+1}, \dots, \mu_n)} \right) = \\
&= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \left(y_i - \mu_i - \sum_{j=1}^{i-1} c_{ij} (y_j - \mu_j) \right)^2 + \\
&\quad + \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \left(\sum_{j=1}^{i-1} c_{ij} (y_j - \mu_j) \right)^2 = \\
&= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \left[(y_i - \mu_i)^2 - 2(y_i - \mu_i) \sum_{j=1}^{i-1} c_{ij} (y_j - \mu_j) + \right. \\
&\quad \left. + \left(\sum_{j=1}^{i-1} c_{ij} (y_j - \mu_j) \right)^2 \right] + \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \left(\sum_{j=1}^{i-1} c_{ij} (y_j - \mu_j) \right)^2 = \\
&= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} (y_i - \mu_i)^2 + \sum_{i=1}^n \sum_{j=1}^{i-1} \frac{1}{\sigma_i^2} c_{ij} (y_i - \mu_i) (y_j - \mu_j) = \\
&= -\frac{1}{2} (y - \mu)' \Sigma^{-1} (y - \mu) + (y - \mu)' \Sigma^{-1} C (y - \mu) = \\
&= -\frac{1}{2} (y - \mu)' \Sigma^{-1} (\mathbb{I} - C) (y - \mu)
\end{aligned}$$

$\Rightarrow Y$ multivariat normalverteilt mit $\mathbb{E}[Y] = \mu$ und Präzisionsmatrix

$$P = \Sigma^{-1}(\mathbb{I} - C), \quad p_{ii} = \frac{1}{\sigma_i^2}, \quad p_{ij} = -\frac{c_{ij}}{\sigma_i^2}, \quad i \neq j$$

Vertauscht man die Reihenfolge der Komponenten y_i und y_j , muss in der Zerlegung i und j vertauscht werden $\Rightarrow c_{ij} = c_{ji}$. Man kann zeigen, dass keine weiteren Restriktionen auftreten. \square

2.3.3 Arten des GMRF

(a) **Conditional autoregressive model (CAR)**

$$\mu = 0 \quad \Rightarrow \quad Y_i | Y_{-i} \sim N \left(\sum_{j \sim i} \frac{\lambda w_{ij} Y_j}{w_{i \bullet}}, \frac{\sigma^2}{w_{i \bullet}} \right) \quad \text{mit} \quad w_{i \bullet} = \sum_{j \sim i} w_{ij}$$

$$Y \sim N(0, \sigma^2 P^{-1}) \quad \text{mit} \quad p_{ij} = \begin{cases} -\lambda w_{ij} & i \sim j \\ w_{i\bullet} & i = j \end{cases}$$

Üblicherweise $w_{ij} \equiv 1 \Rightarrow \text{CAR}$

(b) **Intrinsic autoregressive model (IAR/IGMRF)** CAR mit

$$\lambda = 1 \quad \Rightarrow \quad p_{ij} = \begin{cases} -1 & i \sim j \\ |\partial_i| & i = j \end{cases}$$

$\text{Rg}(P) = n - 1!$ Es liegt also eine partiell impropere Verteilung vor.

$$p(y) \propto \sigma^{-\text{Rg}(P)} \exp\left(-\frac{1}{2\sigma^2} y' P y\right)$$

Vgl. Bsp.20.

(c) **Simultaneous autoregressive model (SAR)**

Sei $\nu \sim N(0, \Sigma)$ mit $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$

$$Y(s_i) = \mu(s_i) + \sum_{j=1}^n b_{ij}(Y(s_j) - \mu(s_j)) + \nu(s_i)$$

mit $b_{ii} = 0$; μ ist der Erwartungswertvektor

B muss nicht symmetrisch sein, aber $(\mathbb{I} - B)$ muss invertierbar sein.

$$\Rightarrow Y \sim N(\mu, (\mathbb{I} - B)^{-1} \Sigma (\mathbb{I} - B)^{-1}),$$

also: $P = (\mathbb{I} - B) \Sigma^{-1} (\mathbb{I} - B)$

2.3.4 Ziehen aus GMRF

Sei $\theta \sim N^K(b, Q)$

Rue-Block-Algorithmus

1. Berechne die Cholesky-Zerlegung $Q = LL'$
2. Löse $Lw = b$
3. Löse $L'u = w$
4. Ziehe einen Zufallsvektor $z \sim N(0, \mathbb{I}_n)$
5. Löse $L'v = z$
6. Berechne $\theta = u + v$

Effiziente Algorithmen zur Cholesky-Zerlegung existieren insbesondere für Bandmatrizen. Für räumliche Nachbarschaftsmatrizen läßt sich durch Umordnen der Regionen eine Reduzierung der Bandweite erreichen.

Restringierte Ziehung Eine restringierte Ziehung unter $A\theta = e$ erhält man durch

$$\theta^* = \theta - Q^{-1}A'(AQ^{-1}A')^{-1}(A\theta - e),$$

wobei θ die unrestringierte Zufallsvariable ist.

2.3.5 GMRF auf regulären Gittern

Auf regulären Gittern lassen sich GMRF als Interaktionen von Random Walks darstellen (im Folgenden o.B.d.A. intrinsische GMRF).

(a) **Random Walk erster Ordnung (1D-GMRF)**

$$Y_i | Y_{i-1} \sim N(Y_{i-1}, \tau^{-1}) \quad \forall i = 2, \dots, n$$

$$p(y_1) \propto \text{const} \quad \Leftrightarrow \quad p(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} \tau^{(n-1)/2} \exp\left(-\frac{\tau}{2} \sum_{i=2}^n (y_i - y_{i-1})^2\right)$$

$$\Leftrightarrow Y \sim N\left(0, (\tau R^{(1)})^{-1}\right)$$

$$\Leftrightarrow p(y) \propto \tau^{(n-1)/2} \exp\left(-\frac{\tau}{2} y' R^{(1)} y\right)$$

wobei

$$R^{(1)} = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & & \\ & & & \ddots & \\ & & & & \ddots \end{pmatrix}$$

Eigenwerte von $R^{(1)}$:

$$\lambda_i = 2 \left[1 - \cos \left(\frac{\pi(i-1)}{n} \right) \right]$$

vgl. ?. $\text{Rg}(R^{(1)}) = n - 1$.

(b) **Random Walk zweiter Ordnung**

$$Y_i | Y_{i-1}, Y_{i-2} \sim N(2Y_{i-1} - Y_{i-2}, \tau^{-1}) \quad \forall i \geq 3$$

$$\begin{aligned} p(y_1) \propto p(y_2) \propto \text{const} &\Leftrightarrow Y \sim N\left(0, (\tau R^{(2)})^{-1}\right) \\ &\Leftrightarrow p(y) \propto \tau^{(n-2)/2} \exp\left(-\frac{\tau}{2} y' R^{(2)} y\right) \end{aligned}$$

$$R^{(2)} = \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & 1 & -4 & 6 & -4 & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & \ddots \end{pmatrix}$$

wobei $\text{Rg}(R^{(2)}) = n - 2$.

(c) **2D-reguläres Gitter, erste Nachbarn**

Nach Hammersley-Clifford:

$$\begin{aligned} p(y) &= \left[\left(\frac{1}{\sqrt{2\pi}} \right) \tau \right]^{(IJ-1)/2} \cdot \\ &\exp \left(-\frac{\tau}{2} \left(\sum_{i=1}^I \sum_{j=2}^J (y_{ij} - y_{ij-1})^2 + \sum_{i=2}^I \sum_{j=1}^J (y_{ij} - y_{i-1j})^2 \right) \right) \\ &\propto \tau^{(IJ-1)/2} \exp \left(-\frac{\tau}{2} \left(\sum_i y'_{i\bullet} R_J^{(1)} y_{i\bullet} + \sum_j y'_{\bullet j} R_I^{(1)} y_{\bullet j} \right) \right) \\ &\propto \tau^{(IJ-1)/2} \exp \left(-\frac{\tau}{2} \left(y' \left(R_J^{(1)} \oplus R_I^{(1)} \right) y \right) \right) \end{aligned}$$

Mit $Y_{\bullet j}, Y_{i\bullet}$ und Y passend definiert. Es gilt:

$$\begin{aligned} \text{Rg}(R_J^{(1)} \oplus R_I^{(1)}) &= IJ - 1 \\ R_J^{(1)} \oplus R_I^{(1)} &:= \mathbb{I}_I \otimes R_J^{(1)} + R_I^{(1)} \otimes \mathbb{I}_J \end{aligned}$$

? schlagen Richtungswichtung vor

$$Q = \gamma R_J^{(1)} \oplus (2 - \gamma) R_I^{(1)}$$

Es gilt: Erwartungswerte von Q sind Kroneckersummen der Erwartungswerte von $\gamma R_J^{(1)}$ und $(2 - \gamma) R_I^{(1)}$ als Vektoren.

Vorsicht: Modelle mit unterschiedlichen Gewichten zwischen allen Paaren lassen sich so nicht darstellen.

(d) **2D-reguläres Gitter mit zweiten Nachbarn**

Präzisionsmatrix: $P^{(2)} = R_J^{(2)} \oplus R_I^{(2)}$, $\text{Rg}(P^{(2)}) = IJ - 4$

(e) **2D-reguläres Gitter mit diagonalen Nachbarn**

$P^{(D)} = R_J^{(1)} \otimes R_I^{(1)}$, $\text{Rg}(P^{(D)}) = (I - 1)(J - 1)$ (starker Rangabfall)

(f) **Kombination von (d) und (e)**

$P^{(12)} = R_J^{(2)} \otimes R_I^{(2)} - R_J^{(1)} \otimes R_I^{(1)}$, $\text{Rg}(P^{(12)}) = IJ - 1$

(g) **3D-Gitter mit ersten Nachbarn**

$P^{(3D)} = R_K^{(1)} \otimes \mathbb{I}_J \otimes \mathbb{I}_I + \mathbb{I}_K \otimes R_J^{(1)} \otimes \mathbb{I}_I + \mathbb{I}_K \otimes \mathbb{I}_J \otimes R_I^{(1)}$

Analog für andere Nachbarschaften oder höhere Dimensionen.

(h) **Räumlich-zeitliche Felder**

Nach ? lassen sich Interaktion von (diskretem) Raum und (diskreter) Zeit als Kroneckerprodukt schreiben. ? unterscheidet vier Typen:

I	unabh. Interaktion	$\text{Rg}(I_n \otimes I_T) = n \cdot T$
II	unabh. Random Walks pro Region	$\text{Rg}(I_n \otimes R) = n \cdot \text{Rg}(R)$
III	unabh. GMRF pro Zeitpunkt	$\text{Rg}(Q \otimes I_T) = \text{Rg}(Q) \cdot T$
IV	abhängige Interaktion	$\text{Rg}(Q \otimes R) = \text{Rg}(Q) \cdot \text{Rg}(R)$

Dies ist für beliebige Präzisionsmatrizen Q und R anwendbar.

2.4 Latente Markov-Zufallsfelder

Wie in 2.2 gesehen, sind Nicht-Gauß-MRF oft schwierig zu konstruieren. Statt dessen werden oft latente (nicht direkt beobachtbare) GMRFs benutzt.

Fortsetzung Beispiel 7 (Disease Mapping)

y_i beobachtete Fälle in Region i (Zählvariable)

e_i erwartete Fälle in Region i (gegeben)

offensichtlich sinnvoll: $Y_i \sim \text{Po}(\lambda_i e_i)$

λ_i Rate mit räumlicher Abhängigkeit \Rightarrow transformiere $\lambda_i \rightarrow \log(\lambda_i) = \eta_i$
 und benutze GMRF für (latentes) η_i (Linkfunktion)

Als Hierarchisches Modell:

$$\begin{aligned} Y_i | \lambda_i &\stackrel{\text{i.n.d.}}{\sim} \text{Po}(\lambda_i e_i) & \forall i = 1, \dots, n \\ \log(\lambda_i) &\sim \text{GMRF}(\tau) & \forall i = 1, \dots, n \quad \text{bzw.} \quad \log(\lambda) \sim N_n(0, \tau^{-1} Q^{-1}) \end{aligned}$$

Alternativ als GAM darstellbar, dazu mehr in Kapitel 4.

2.4.1 Besag-York-Mollie Modell

Nach ? als BYM-Modell benannt:

$$\begin{aligned} Y_i | \lambda_i &\sim \text{Po}(\lambda_i e_i) \\ \lambda_i &= \exp(\gamma_i) = \exp(u_i + v_i) \\ u_i &\sim N\left(\frac{1}{|\partial_i|} \sum_{j \sim i} u_j, \frac{1}{\tau |\partial_i|}\right) && \text{Strukturierter Effekt} \\ v_i &\stackrel{\text{i.i.d.}}{\sim} N(0, \nu^{-1}) && \text{Unstrukturierter Effekt} \\ \tau &\sim \text{Ga}(a, b) \\ \nu &\sim \text{Ga}(c, d) \end{aligned}$$

Im Folgenden $\theta = (u, v, \tau, \nu)$. Die Posteriori ergibt sich zu

$$\begin{aligned} p(\theta | y) &\propto \prod_{i=1}^n (e_i \lambda_i)^{y_i} \exp(-e_i \lambda_i) \\ &\quad \cdot \tau^{\text{Rg}(Q)/2} \exp\left(-\frac{\tau}{2} u' Q u\right) \\ &\quad \cdot \nu^{n/2} \exp\left(-\frac{\nu}{2} v' v\right) \\ &\quad \cdot \tau^{a-1} \exp(-b\tau) \\ &\quad \cdot \nu^{c-1} \exp(-d\nu) \end{aligned}$$

(a) Inferenz über MCMC

$$\tau | \cdot \sim \text{Ga}, \quad \nu | \cdot \sim \text{Ga} \quad \rightarrow \quad \text{Gibbs-Schritte}$$

$$\begin{aligned}
 p(u|\cdot) &\propto \prod_{i=1}^n \exp\left(y_i u_i - e_i \exp(u_i + v_i) - \frac{\tau}{2} u' Q u\right) \\
 p(u_i|\cdot) &\propto \exp\left(\underbrace{y_i u_i - \frac{\tau}{2} \sum_{j \sim i} (u_i - u_j)^2}_{\text{passt zu Gaußverteilung}} - e_i \exp(u_i + v_i)\right)
 \end{aligned}$$

Metropolis-Hastings-Schritt: verwende als Vorschlagsdichte eine Normalverteilung, die „nahe“ an Full conditional liegt, z.B. über Taylor-Entwicklung (analog für andere Verteilungen möglich):

$$f(u_i) := e_i \exp(u_i + v_i) \quad \text{für ein } i$$

$$\tilde{f}(u_i) = f(u_0) + (u_i - u_0) f'(u_0) + \frac{1}{2} (u_i - u_0)^2 f''(u_0)$$

Alternativ: IWLS-Proposal.

Damit Akzeptanzwahrscheinlichkeiten nahe 1. Analog für v . Aus $q(u_i|\cdot)$ lässt sich $q(u|\cdot)$ konstruieren mit identischer Struktur der Präzisionsmatrix:

$$q(u|u_0) \propto \exp\left(-\frac{1}{2} u' (\tau Q + H(u_0)) u + u' (y - f'(u))\right)$$

(b) **Reparametrisierungstrick**

$\tilde{\theta} = (u, \gamma, \tau, \nu)$. Full conditionals:

$$\begin{array}{lll}
 u|\gamma, \tau &\propto N_n(\cdot, \cdot) &\rightarrow \text{Gibbs-Schritt} \\
 \gamma_i|u_i, \nu_i &\stackrel{\text{i.i.d.}}{\sim} (\cdot, \cdot) &\rightarrow \text{Metropolis-Hastings-Schritt} \\
 \tau|u &\propto \text{Ga}(\cdot, \cdot) & \\
 \nu|\gamma &\propto \text{Ga}(\cdot, \cdot) &
 \end{array}$$

$$\begin{aligned}
 p(u|\gamma, \tau) &\propto \exp\left(-\frac{\tau}{2} u' Q u - \frac{\nu}{2} (\gamma - u)' \mathbb{I} (\gamma - u)\right) \\
 &\propto \exp\left(-\frac{\tau}{2} u' (\tau Q + \nu \mathbb{I}) + \nu \gamma' u\right)
 \end{aligned}$$

$$u|\cdot \sim N^K(\nu \gamma, \underbrace{\tau Q + \nu \mathbb{I}}_{\text{voller Rang für } \nu > 0}) \quad \text{Rue-Block-Update (vgl. 2.3.4)}$$

Bemerkungen

- Ziehen aus $p(u_i|\cdot) \propto p(u_i|\partial u_i, \gamma_i, \tau)$ (single-site). Erlaubt parallelisieren, z.B. Schachbrettmuster bei regulärem Gitter.

- Gemeinsamer Update von Parametern verbessert Mixing, siehe ?:
z.B. ziehe $\tilde{\tau}|u$ und $\tilde{u}|\tilde{\tau}, \gamma$ und akzeptiere/verwerfe gemeinsam.
- u und v sind nicht identifizierbar! Nur $\gamma = u + v$ kann interpretiert werden. I.d.R. Priorerwartungswert für ν hoch, damit für v niedrig.

(c) **Inferenz über penalisierte Likelihood**

GLM-Ansatz: $\log(\mathbb{E}[Y_i]) = u_i + v_i$

Penalisierung auf Differenzen von u_i : $\text{Pen}(u) = \lambda_u \sum_{i \sim j} (u_i - u_j)^2$

Penalisierung auf Differenzen von v_i : $\text{Pen}(v) = \lambda_v \sum_{i=1}^n v_i^2$ (Ridge!)

\Rightarrow Pen. log-Likelihood entspricht marginaler log-Posteriori $\log(p(u, v|y))$.
Schätze λ_u und λ_v über Kreuzvalidierung.

(d) **Inferenz über empirischen Bayes/REML**

Schätze ν und τ aus den Daten, d.h. aus marginaler Posteriori, dann ML für u und v „einfach“.

(e) **Inferenz über INLA**

Idee: – Approximiere marginale Posteriori für ν und τ numerisch.
– Bedingte Posteriori $u, v | \nu, \tau, Y$ „fast“ normal.

Details zu INLA später.

2.4.2 Hilfsvariablenansatz

Formulierung eines Binomialmodells auf Basis der Individuen (nach ?):

$$\begin{aligned} Y_j &\sim \text{Bernoulli}(\exp(\eta_j)) \\ \eta_j &= u_{i(j)} + v_{i(j)} \\ u &\sim \text{N}(0, (\tau M)^{-1}) \\ v_i &\sim \text{N}(0, \delta) \quad \text{für alle } i = 1, \dots, n \end{aligned}$$

mit $Y_j \in \{0, 1\}$ für $j = 1, \dots, n$ binäre Responsevariable des Individuums j ,
 $i(j)$ Index der Region des Individuums j .

Das Modell läßt sich wie folgt mit Hilfsvariablen formulieren:

$$\begin{aligned} Y_j &= \begin{cases} 1 & \text{für } z_j > 0 \\ 0 & \text{sonst} \end{cases} \\ z_j &= u_{i(j)} + v_{i(j)} + \varepsilon_j \end{aligned}$$

$$\begin{aligned}
\varepsilon_j &\sim N(0, \lambda_j) \\
\lambda_j &= (2\psi_j)^2 \\
\psi_j &\sim \text{Kolmogorov-Smirnov} \\
u &\sim N(0, (\tau M)^{-1}) \\
v_i &\sim N(0, \sigma_v^2) \quad \text{für alle } i = 1, \dots, n
\end{aligned}$$

mit $\lambda_j = 1$ für $j = 1, \dots, n$ erhält man ein Probitmodell. Durch geeignete Wahl der Verteilung von λ sind auch andere Links möglich. Wichtig: Die marginale Likelihood $L(u, v|y)$ ist in beiden Modellformulierungen BYM und Holmes/Held-Modell gleich.

Full conditionals:

$$\begin{aligned}
u|v, z, \lambda, Y &\sim N(V^{-1}b, V^{-1}) \\
b &= (b_1, \dots, b_I) \\
b_i &= \sum_{j \in R_i} \frac{z_j - v_i}{\lambda_j} \\
V &= \tau M + W \\
W &= \text{diag} \left(\sum_{j \in R_1} \lambda_j^{-1}, \dots, \sum_{j \in R_I} \lambda_j^{-1} \right) \\
v|u, z, \lambda, Y &\sim N(\tilde{V}^{-1}\tilde{b}, \tilde{V}^{-1}) \\
\tilde{b} &= (\tilde{b}_1, \dots, \tilde{b}_I) \\
\tilde{b}_i &= \sum_{j \in R_i} \frac{z_j - u_i}{\lambda_j} \\
\tilde{V} &= \delta I + W \\
\tau &\sim \text{Ga} \left(a + \frac{n-1}{2}, b + \frac{1}{2} u' M u \right) \\
\delta &\sim \text{Ga} \left(a + \frac{n}{2}, b + \frac{1}{2} \sum_{i=1}^n v_i^2 \right)
\end{aligned}$$

Im Unterschied zum BYM-Modell kann v_i hier in jeder Region $i \in \{1, \dots, I\}$ geschätzt werden. Liegen in einer Region keine Daten vor, so ist die Posteriori von v_i gleich der Priori.

2.4.3 Latente Ising/Potts-Felder

Beispiel 21. (Segmentierung)

Auf regulärem (Pixel-)Gitter: $Y_i = \delta_i \alpha_i + (1 - \delta_i) \beta_i$; $\delta_i \in \{0, 1\}$

$$\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$$

$$\beta_i \sim N(\mu_\beta, \sigma_\beta^2)$$

$$p(\delta) \propto \exp \left(\beta_1 \sum_{i \sim j} \delta_i \delta_j + \beta_2 \sum_{i \sim j} (1 - \delta_i)(1 - \delta_j) \right)$$

Inferenz über:

- MCMC
- Simulated Annealing ($\beta_\bullet \rightarrow \infty$)
- MAP (Maximum a Posteriori)
- EM-Algorithmus (Gemischte Verteilung)

Vgl. Vorlesung BioImaging.

Kapitel 3

Geostatistik

3.1 Gauß-Prozesse

Definition 3.1. (Gauß-Zufallsfelder)

$Y = \{Y(s), s \in D \subseteq \mathbb{R}^d\}$ heißt Gauß-Zufallsfeld der Dimension d (Gauß-ZF oder Gaussian random field (GRF) oder Gauß-Prozess): \Leftrightarrow Für beliebige Lokationen s_1, \dots, s_n ist $(Y(s_1), \dots, Y(s_n))'$ multivariat normalverteilt.

Bezeichnungen:

- $\mu(s) = \mathbb{E}[Y(s)]$: Erwartungswertfunktion, räumlicher Trend
- $\sigma^2(s) = \text{Var}(Y(s))$: Varianzfunktion
- $c(s, s') = \text{Cov}(Y(s), Y(s'))$: Kovarianzfunktion

Satz 3.1.

Ein Gauß-Prozess wird durch $\mu(s)$ und $c(s, s')$ vollständig spezifiziert.

Beweis:

Für alle $s_1, \dots, s_n \in D$ gilt:

$$(Y(s_1), \dots, Y(s_n))' \sim N_n \left(\begin{pmatrix} \mu(s_1) \\ \vdots \\ \mu(s_n) \end{pmatrix}, \begin{pmatrix} c(s_i, s_j)_{i=1, \dots, n; j=1, \dots, n} \end{pmatrix} \right)$$

□

3.1.1 Stationarität

Definition 3.2. (Schwach stationär)

Ein räumlicher stochastischer Prozess $Y = \{Y(s), s \in D\}$ heißt schwach

stationär: $\Leftrightarrow \mathbb{E}[Y(s)] = \mu = \text{const}$, $\text{Cov}(Y(s), Y(s')) = c(s - s')$ für alle $s, s' \in D$

Definition 3.3. (Strenge Stationarität)

Ein räumlicher stochastischer Prozess Y heißt streng stationär: \Leftrightarrow alle endlich-dimensionale Verteilungen sind verschiebungsinvariant.

Satz 3.2. (Strenge und schwache Stationarität)

1. Y streng stationär $\Rightarrow Y$ schwach stationär
2. Y sei Gauß-Prozess. Dann gilt: Y streng stationär $\Leftrightarrow Y$ schwach stationär.

3.1.2 Theoretisches (Ko-)Variogramm und Korrelationsfunktion

Definition 3.4. Sei $Y = \{Y(s), s \in D\}$ ein räumlicher Prozess. Dann heißt

$$c(s, s') := \text{Cov}(Y(s), Y(s'))$$

Kovariogramm oder Kovarianzfunktion und

$$\gamma(s, s') := \frac{1}{2} \text{Var}(Y(s) - Y(s'))$$

(Semi-)Variogramm.

Es gilt:

$$\gamma(s, s') = \frac{1}{2} \text{Var}(Y(s)) + \frac{1}{2} \text{Var}(Y(s')) - c(s, s')$$

Definition 3.5. Sei $Y = \{Y(s), s \in D\}$ ein stationärer Prozess und $h := s - s'$. Dann heißt $c(h) = c(s - s') = c(s, s')$ stationäre Kovarianzfunktion und $\rho(h) := \frac{c(h)}{c(0)} = \frac{c(h)}{\sigma^2}$ Korrelogramm oder Korrelationsfunktion.

Es gilt:

$$\begin{aligned} c(0) &= \sigma^2 = \text{Var}(Y(s)) = \text{const.} \\ c(h) &= c(-h) \quad \rho(h) = \rho(-h) \end{aligned}$$

Definition 3.6. (Isotropie und Anisotropie)

Ist $c(s - s') = c(\|s - s'\|)$ nur eine Funktion des euklidischen Abstands von s und s' , so heißt Y bzw. c isotrop; ansonsten anisotrop.

Definition 3.7. (Intrinsisch stationär)

Y intrinsisch stationär: $\Leftrightarrow \mathbb{E}[Y(s)] = \mu, \quad \gamma(s, s') = \gamma(s - s')$.

$\gamma(h)$ heißt stationäres Variogramm.

Es gilt: $\gamma(h) = \gamma(-h), \quad \gamma(0) = 0$.

Hängt $\gamma(h)$ nur von $\|h\|$ ab, so heißt γ isotrop.

Satz 3.3.

Y schwach stationär $\Rightarrow Y$ intrinsisch stationär mit $\gamma(h) = c(0) - c(h) = \sigma^2(1 - \rho(h))$.

Definition 3.8. (L^2 -Stetigkeit, L^2 -Differenzierbarkeit)

Y heißt L^2 -stetig: $\Leftrightarrow \mathbb{E}[Y(s+h) - Y(s)]^2 \rightarrow 0$ für $h \rightarrow 0$

Y heißt L^2 -differenzierbar: \Leftrightarrow Es existiert ein SP Y' mit

$$\mathbb{E} \left[\frac{Y(s+h) - Y(s)}{h} - Y'(s) \right]^2 \rightarrow 0 \quad \text{für } h \rightarrow 0.$$

Später: wenn Y nicht L^2 -stetig, liegt ein Nugget Effekt vor.

Wir interessieren uns (erstmal) nur für stationäre Prozesse.

Eigenschaften: Y stationär \Rightarrow

- Y ist L^2 -stetig $\Leftrightarrow \rho(h)$ stetig in $h = 0$
- Y k -mal L^2 -differenzierbar $\Leftrightarrow \rho(h)$ $2k$ -mal in $h = 0$ differenzierbar.
- Y ist L^2 -stetig $\Leftrightarrow \gamma(h)$ stetig in $h = 0$
- $\gamma(h) = \text{const}$ für $h > 0 \Rightarrow \text{Cov}(Y(s), Y(s')) = 0$ für beliebige $s \neq s'$.

Die räumliche Korrelation wird bei stationären GRF also durch $\rho(h)$ bzw. $\gamma(h)$ bestimmt.

3.2 Empirisches Variogramm

In diesem Abschnitt nichtparametrische Schätzung des (Ko-)Variogramms als exploratives Werkzeug. Später definieren wir uns parametrische Modelle, die auch parametrische Schätzung erlauben.

3.2.1 Schätzer

(a) Empirisches Variogramm

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} (y(s_i) - y(s_j))^2$$

mit $N(h) = \{(s_i, s_j) : s_i - s_j = h; i, j = 1, \dots, n\}$

Herleitung:

$$\begin{aligned}
 2\gamma(h) &= \text{Var}(Y(s_i) - Y(s_j)) \Leftrightarrow \\
 \gamma(h) &= \frac{1}{2} \text{Var}(Y(s_i) - Y(s_j)) \\
 &= \frac{1}{2} (\mathbb{E}[(Y(s_i) - Y(s_j))^2] - (\mathbb{E}[Y(s_i) - Y(s_j)])^2) \\
 &= \frac{1}{2} (\mathbb{E}[(Y(s_i) - Y(s_j))^2] - (\mathbb{E}[Y(s_i)] - \mathbb{E}[Y(s_j)])^2) \\
 &= \frac{1}{2} (\mathbb{E}[(Y(s_i) - Y(s_j))^2] - (\mu - \mu)^2) \\
 &= \frac{1}{2} (\mathbb{E}[(Y(s_i) - Y(s_j))^2]) \Rightarrow \\
 \hat{\gamma}(h) &= \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} (y(s_i) - y(s_j))^2
 \end{aligned}$$

- Analoger Schätzer für Kovariogramm verzerrt.
- Nachteil: $|N(h)|$ meist klein, oft $|N(h)| = 1 \Rightarrow$ Schätzer instabil.

(b) **Empirisches Variogramm nach Matheron**

Für isotrope, intrinsisch stationäre Prozesse. Einteilung der euklidischen Distanz $d_{ij} = d(s_i, s_j)$ zwischen zwei Lokationen in K Klassen (Bins):

$$H_k = \{d_{ij} : h_{k-1} \leq d_{ij} < h_k\}, \quad k = 1, \dots, K; \quad h_0 = 0$$

$$\hat{\gamma}_k = \frac{1}{2|N_k|} \sum_{i,j \in N_k} (Y(s_i) - Y(s_j))^2$$

$$N_k = \{(s_i, s_j) : s_i - s_j \in H_k\}$$

Fragen / Probleme: $K = ?$, Wahl der Klassengrenzen.

(c) **Constant nearest-neighbour estimator**

$$\hat{\gamma}_{CNN}(h) = \frac{\sum_{i < j} \mathbb{1}\{-\delta < h - \|s_i - s_j\| < \delta\} (Y(s_i) - Y(s_j))^2}{\sum_{i < j} \mathbb{1}\{-\delta < h - \|s_i - s_j\| < \delta\}}$$

(d) **Variable nearest-neighbour estimator**

$$\hat{\gamma}_{VNN}(h) = \frac{\sum_{i < j} \frac{1}{\delta_0(\|s_i - s_j\|)} K \left(\frac{h - \|s_i - s_j\|}{\delta \delta_0(\|s_i - s_j\|)} \right) (Y(s_i) - Y(s_j))^2}{\sum_{i < j} \frac{1}{\delta_0(\|s_i - s_j\|)} K \left(\frac{h - \|s_i - s_j\|}{\delta \delta_0(\|s_i - s_j\|)} \right)},$$

wobei K den Kern und δ die Bandbreite bezeichnen. δ_0 erlaubt bessere Anpassung, oft wird $\delta_0 \equiv 1$ angenommen. δ kann AMISE-optimal gewählt werden.

3.2.2 Bezeichnungen

(a) **Nugget-Effekt**

Empirisch ist oft zu beobachten, dass das Variogramm bei $\|h\| = 0$ springt. Sei $Y(s)$ der wahre, stationäre Prozess und $Z(s)$ ein beobachteter Prozess mit

$$Z(s) = Y(s) + \varepsilon(s), \quad \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

Dann gilt:

$$c_Z(h) = \begin{cases} c_Y(h) + \sigma_\varepsilon^2, & \text{falls } \|h\| > 0 \\ c_Y(0), & \text{falls } \|h\| = 0 \end{cases}$$

(b) **Schwellenwert/Sill**

Mit *Schwellenwert* oder *Sill* bezeichnet man den Wert, den das Variogramm erreicht, wenn der Abstand h gegen unendlich geht. Zieht man von diesem Wert den Nugget-Effekt ab, so erhält man den *Partial Sill*.

(c) **Range**

Mit Range wird der Abstand bezeichnet, ab dem das Variogramm den Schwellenwert fast erreicht. Punkte, deren Abstand größer ist als der Range werden als unabhängig voneinander betrachtet. Das hat zur Folge, dass dann, wenn das Variogramm positiv konstant ist (außer am Ursprung, wo es null ist), kein Zusammenhang zwischen $Z(s_i)$ und $Z(s_j)$ für alle s_i und s_j besteht.

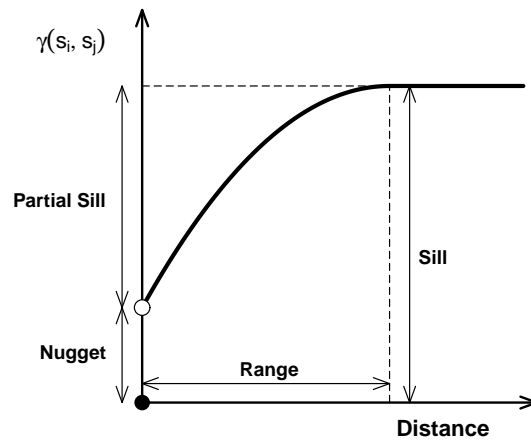


Abbildung 3.1: Variogramm

3.3 Parametrische Modelle für Korrelationsfunktionen und (Ko-)Variogramme

3.3.1 Gültigkeit von Korrelationsfunktionen

Gültige Modelle für Kovariogramme müssen positiv semidefinit sein, d.h.

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j c(s_i - s_j) \geq 0 \quad \text{für beliebige } a_i, a_j, s_i, s_j$$

erfüllen. Völlig analog für Korrelationsfunktionen. Dies sichert:

$$\text{Var}(a_1 Y(s_1) + \dots + a_n Y(s_n)) \geq 0.$$

Es gilt:

$c_1(h)$ und $c_2(h)$ gültige Kovarianzfunktionen \Rightarrow

$$b_1 c_1(h) + b_2 c_2(h) \quad \text{für } b_1 > 0 \text{ und } b_2 > 0$$

$$c_1(h) \cdot c_2(h)$$

sind gültige Korrelationsfunktionen.

Wünschenswerte Eigenschaften:

- $\rho(h)$ fallend in h , $\rho(h) \rightarrow 0$ für $h \rightarrow \infty$.
- Wenigstens 1 Parameter steuert, wie schnell $\rho(h) \rightarrow 0$ für $h \rightarrow \infty$

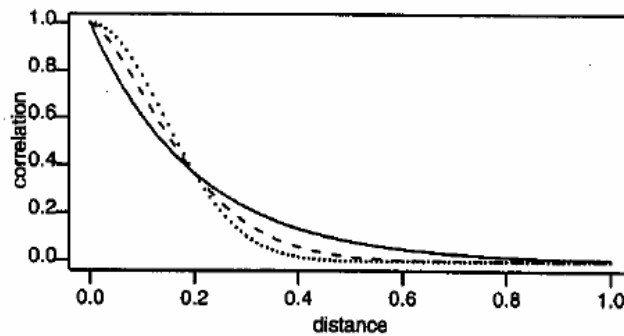


Abbildung 3.2: Powered-Exponential Korrelationsfunktion mit $\phi = 0.2$ und $\kappa = 1$ (durchgezogene Linie), $\kappa = 1.5$ (gestrichelte Linie) und $\kappa = 2$ (gepunktete Linie).

geht.

Satz 3.4. (Bochner-Kriterium für isotrope Korrelationsfunktionen)

$\rho(u)$ ist gültige isotrope Korrelationsfunktion \Leftrightarrow

$\rho(u)$ ist charakteristische Funktion einer symmetrischen ZV $X \Leftrightarrow$

$$\rho(u) = \int_{-\infty}^{+\infty} \exp(iux) f(x) dx$$

Weiterführende Literatur: ?

3.3.2 Übliche Korrelationsfunktionen

Erinnerung: Es gilt unter Isotropie: $\gamma(u) = c(0) - c(u) = \sigma^2(1 - \rho(u))$.

(a) **Potenz-Exponentialfamilie** (Abb. 3.2)

$$\rho(u) = \exp\{-(u/\phi)^\kappa\}, \quad \phi > 0, \quad 0 < \kappa \leq 2$$

$\kappa = 1$: Exponential-Korrelationsfunktion,

$\kappa = 2$: Gauß-Korrelationsfunktion.

(b) **Sphärische Familie** (Abb. 3.3)

$$\rho(u) = \begin{cases} 1 - \frac{3}{2}(u/\phi) + \frac{1}{2}(u/\phi)^3, & 0 \leq u \leq \phi \\ 0, & u > \phi \end{cases}$$

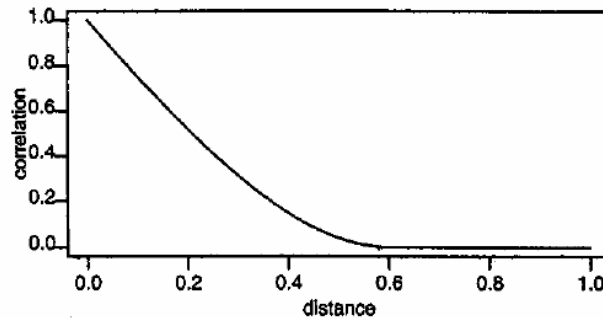


Abbildung 3.3: Sphärische Korrelationsfunktion mit $\phi = 0.64$.

Problem: Nicht differentierbar in $u = \phi \Rightarrow$ Schwierigkeiten bei ML-Schätzung.

(c) **Matérn-Familie** (Abb. 3.4)

$$\rho(u) = \frac{1}{2^{\kappa-1} \Gamma(\kappa)} (u/\phi)^\kappa K_\kappa(u/\phi), \quad \phi > 0, \quad \kappa > 0$$

K_κ : Besselfunktion, i.A. nicht in geschlossener Form anzugeben. Numerische Approximationen in ?. Explizite Formeln für $\kappa = 0.5, 1.5, \dots$ möglich:

$$\begin{aligned} \rho(u; \phi = 1, \kappa = 0.5) &= \exp(-u) \\ \rho(u; \phi = 1, \kappa = 1.5) &= \exp(-u)(1 + u) \\ \rho(u; \phi = 1, \kappa = 2.5) &= \exp(-u) \left(1 + u + \frac{1}{3}u^2 \right) \\ \rho(u; \phi = 1, \kappa = 3.5) &= \exp(-u) \left(1 + u + \frac{2}{5}u^2 + \frac{1}{15}u^3 \right) \end{aligned}$$

$\kappa \rightarrow \infty$: Gauß-Korrelationsfunktion

Obwohl nur numerisch greifbar gebräuchlichste Korrelationsfunktion, vgl. Abschnitt 3.6.

3.4 Anisotrope und nicht-stationäre Prozesse

Arten der Anisotropie:

- Geometrische Anisotropie: Sill identisch, Range unterschiedlich
- Zonale Anisotropie: Sill unterschiedlich

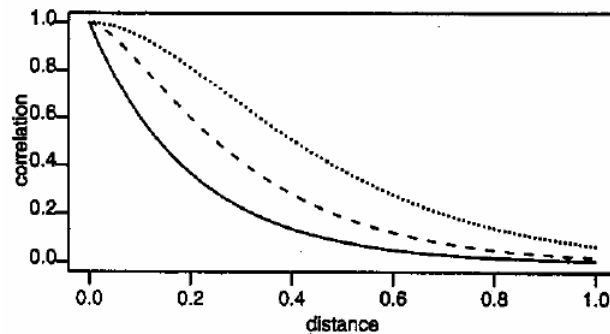


Abbildung 3.4: Matérn-Korrelationsfunktion mit $\phi = 0.2$ und $\kappa = 1$ (durchgezogene Linie), $\kappa = 1.5$ (gestrichelte Linie) und $\kappa = 2$ (gepunktete Linie).

Erkennung: Berechne Variogramme in verschiedenen Raumrichtungen.

Geometrische (Range-)Anisotropie

Wähle in isotroper Korrelationsfunktion $\rho(u)$ für u statt $u^2 = \|h\|^2$ eine nicht-euklidische Distanz, z.B.:

$$u^2 = h' R(\psi) S(\delta) R(\psi) h$$

mit

$$R(\psi) = \begin{pmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{pmatrix} \quad \text{Rotationsmatrix, } \psi \in (0, 2\pi)$$

$$S(\delta) = \begin{pmatrix} \delta^{-1} & 0 \\ 0 & 1 \end{pmatrix} \quad \text{Dehnungsmatrix, } \delta \geq 1$$

Zonale Anisotropie (siehe ?)

$\rho_l(u)$, $l = 1, \dots, p$ seien gültige 2D isotrope Korrelationsfunktionen $\Rightarrow \rho(h) = \rho_1(\sqrt{h' B_1 h}), \dots, \rho_p(\sqrt{h' B_p h})$ mit B_l pos. definit ist gültige Korrelationsfunktion in \mathbb{R}^2 .

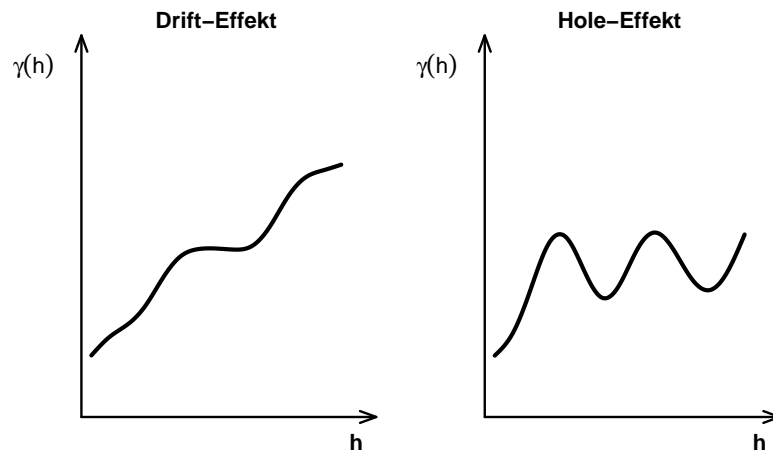


Abbildung 3.5: Der Drift Effekt weist auf Nicht-Stationarität hin. Der Hole Effekt weist auf regelmäßige Strukturen hin.

3.5 Kriging

Ziel: Schätzung der (wahren) Oberfläche. Vorerst seien (deterministische) Erwartungswertfunktion und Korrelationsfunktion bekannt.

3.5.1 Einfaches Kriging

Gegeben: Daten $y(s_1), \dots, y(s_n)$

Daten-Modell: $Y(s_i) = Z(s_i) + \varepsilon(s_i)$

Z sei ein (stationäres) Zufallsfeld mit $\mu(s)$ und $c(s, s') = \sigma^2 \rho(s, s')$

$\varepsilon(s_i)$ sei i.i.d. mit $\mathbb{E}[\varepsilon(s_i)] = 0$ und Varianz σ_ε^2

Gesucht: linearer Prädiktor für $Z(s_0)$ für alle $s_0 \in D$ mit minimalem mittleren quadratischen Prädiktionsfehler.

$$\hat{Z}(s_0) = l'Y + K$$

Wegen Verschiebungssatz gilt:

$$\text{MSPE}(l, K) = \mathbb{E}[Z(s_0) - l'Y - K]^2 = \text{Var}(Z(s_0) - l'Y - K) + \{\mathbb{E}[Z(s_0) - l'Y - K]\}^2$$

$$\mathbb{E}[Z(s_0) - l'Y - K] = \mu(s_0) - l'\mu_Y - K, \quad \text{wobei } \mu_Y = (\mu(s_1), \dots, \mu(s_n))$$

$$\text{Var}(Z(s_0) - l'Y - K) = \text{Var}(Z(s_0)) - 2l'\text{Cov}(Z(s_0), Y) + l'C_Y l$$

wobei K fest ist.

$$\text{Cov}(Z(s_0), Y) = (c(s_0, s_1), c(s_0, s_2), \dots, c(s_0, s_n))' =: c(s_0)$$

$$C_Y = (C_Y(s_i, s_j))_{\substack{i=1, \dots, n \\ j=1, \dots, n}} \quad \text{mit} \quad C_Y(s_i, s_j) = \begin{cases} c(s_i, s_i) + \sigma_\varepsilon^2, & \text{für } i = j \\ c(s_i, s_j), & \text{sonst} \end{cases}$$

$$\text{Ableiten nach } l \quad \Rightarrow \quad l^* = C_Y^{-1}c(s_0)$$

$$\text{Ableiten nach } K \quad \Rightarrow \quad K^* = \mu(s_0) - c(s_0)'C_Y^{-1}\mu_Y$$

Daraus folgt, der lineare Prädiktor ist

$$\widehat{Z}(s_0) = \mu(s_0) + c(s_0)'C_Y^{-1}(Y - \mu_Y)$$

Die Kriging-Varianz ist dann

$$\sigma_Y^2(s_0) = \text{MSPE}(l^*, K^*) = c(s_0, s_0) - c(s_0)'C_Y^{-1}c(s_0)$$

Bemerkungen:

(1) Fall $\mu(s) \equiv \mu$ heißt gewöhnliches (ordinary) Kriging.

(2) Fall $\mu(s) = x(s)'\beta$ mit bekanntem β heißt universelles Kriging.

Beispiel 22. Sei $\mu(s) = \mu = 0$, $\text{Var}(Z(s)) = 1$ und $\sigma_\varepsilon^2 = 0.5$.

Gegeben: Zwei Punkte $s_1 = (1, 0)$ und $s_2 = (2, 0)$, s_1 näher an $s_0 = (0.5, 0)$ als s_2 .

Annahme: es gelte die Exponentialkorrelationsfunktion mit $\sigma^2 = 1$.

$$c(s_1, s_0) = 0.61, \quad c(s_2, s_0) = 0.22 \quad \text{und} \quad c(s_1, s_2) = 0.37$$

Also

$$c(s_0) = (0.61, 0.22)'$$

und

$$C_Y = \begin{pmatrix} 1 & 0.37 \\ 0.37 & 1 \end{pmatrix} + 0.5\mathbb{I} = \begin{pmatrix} 1.5 & 0.37 \\ 0.37 & 1.5 \end{pmatrix}$$

Kriging ergibt

$$\widehat{Z} = 0.395 Y(s_1) + 0.049 Y(s_2)$$

wobei 0.395 und 0.049 die Kriging-„Gewichte“ sind. Die zugehörige Varianz ist

$$\sigma_Y^2 = 1 - (0.61, 0.22) \begin{pmatrix} 1.5 & 0.37 \\ 0.37 & 1.5 \end{pmatrix}^{-1} \begin{pmatrix} 0.61 \\ 0.22 \end{pmatrix} = 0.7508$$

also unabhängig von Y !

Bemerkungen:

- Größerer Nugget-Effekt \rightarrow Größere Kriging-Varianz.
- Größerer Sill \rightarrow Größere Kriging-Varianz.
- Größerer Range \rightarrow Kleinere Kriging-Varianz.
- Kriging-Varianz klein, wo mehr Daten (Trick: erhebe weitere Daten dort, wo Kriging-Varianz groß).

3.5.2 Kriging im Gauß-Modell

Nun Normalverteilungsannahme.

Erster Ansatz: additives Modell

$$Y(s) = \mu(s) + Z(s) + \varepsilon(s)$$

$\mu(s)$: deterministisch

$Z(s)$: stationäres Gaußzufallsfeld mit $\mathbb{E}[Z(s)] = 0$ und $\text{Var}(Z(s)) = \tau^2$

$\varepsilon(s)$: i.i.d. mit $\mathbb{E}[\varepsilon(s)] = 0$ und $\text{Var}(\varepsilon(s)) = \sigma^2$

Gelegentlich auch

$$Y(s) = \mu(s) + W(s) + V(s) + \varepsilon(s)$$

$W(s)$ und $V(s)$ unabhängige und stationäre GZF mit ρ_W und ρ_V gegeben und ρ_V geht schnell gegen 0.

\rightarrow „Mikroskalenvariation“, vgl. Bsp. 2 (Rongelap Daten).

Oft aber $V(s)$ und $W(s)$ nicht unterscheidbar.

Alternativer Ansatz: hierarchisches Modell

$$Y(s)|Z(s), \sigma_\varepsilon^2 \stackrel{\text{i.n.d.}}{\sim} N(Z(s), \sigma_\varepsilon^2)$$

$Z(s)$ Gauss-Zufallsfeld mit $\mathbb{E}[Z(s)] = \mu(s)$, $\text{Var}(Z(s)) = \tau^2$, $\rho_Z(h)$ gegeben.

Die beiden Ansätze sind identisch. Es gilt jeweils:

$$\mathbb{E}[Y(s)] = \mu(s) \text{ und } \text{Cov}(Y(s_i), Y(s_j)) \text{ wie oben}$$

$$\begin{aligned} Y &= (Y(s_1), \dots, Y(s_n))' \\ R &= (v_{ij}) = (\rho(s_i - s_j)) \end{aligned}$$

$$\begin{aligned}\mu &= (\mu(s_1), \dots, \mu(s_n))' \\ Y &\sim N(\mu, \tau R + \sigma^2 \mathbb{I}) \\ r &:= (r_1, \dots, r_n)'; \quad r_i = \rho(s_0 - s_i)\end{aligned}$$

Dann gilt auch für $Z_0 = Z(s_0)$

$$\begin{pmatrix} Z_0 \\ Y \end{pmatrix} \sim N \left(\begin{pmatrix} \mu(s_0) \\ \mu \end{pmatrix}, \begin{pmatrix} \tau^2 & \tau^2 r' \\ \tau^2 r' & \tau^2 R + \sigma^2 \mathbb{I} \end{pmatrix} \right)$$

$Z_0|Y$ ist normalverteilt mit:

$$\begin{aligned}\mathbb{E}[Z_0|Y] &= \mu(s_0) + \tau^2 r' (\tau^2 R + \sigma^2 \mathbb{I})^{-1} (Y - \mu) =: \hat{Z}_0 \\ \text{Var}(Z_0|Y) &= \tau^2 - \tau^2 r' (\tau^2 R + \sigma^2 \mathbb{I})^{-1} \tau^2 r\end{aligned}$$

Entspricht also dem einfachen Kriging-Schätzer und dessen Varianz.

3.5.3 Inferenz

Seien nun $\mu(s) = X'(s)\beta$ und $\rho(h)$ unbekannt.

Bemerkung: X kann parametrisches Modell für Oberfläche, etwa $\mu(s) = \beta_0 + \beta_1 s_X + \beta_2 s_Y + \beta_3 s_X^2 + \beta_4 s_Y^2 + \beta_5 s_X s_Y$, sein und/oder räumlich vorliegende Kovariable (für die Prädiktion müssen diese aber auf ganz D vorliegen).

(a) **Maximum-Likelihood-Schätzer:**

$$Y(s_i) = X'(s_i)\beta + Z(s_i) + \varepsilon(s_i)$$

θ Korrelationsparameter, $\alpha := (\sigma^2, \tau^2, \theta)$

$$\text{Cov}(Y) = \sigma^2 \mathbb{I} + \tau^2 R(\theta) =: \Sigma(\alpha)$$

Das allgemeine lineare Modell ergibt sich zu

$$Y \sim N(X\beta, \Sigma(\alpha))$$

Log-Likelihood (bei Normalverteilungsannahme)

$$l(\beta, \alpha) = -\frac{1}{2} (\log |\Sigma(\alpha)| + (y - X\beta)' \Sigma^{-1}(\alpha) (y - X\beta))$$

Für gegebenes α ist

$$\hat{\beta}(\alpha) = (X' \Sigma^{-1}(\alpha) X)^{-1} X' \Sigma^{-1}(\alpha) y$$

ML-Schätzer für β . Einsetzen in $l(\beta, \alpha)$ ergibt die *Profil-log-Likelihood* für α

$$l(\alpha) := l(\widehat{\beta}(\alpha), \alpha).$$

Maximieren liefert ML-Schätzer $\widehat{\alpha}$. Einsetzen in $\widehat{\beta}(\alpha)$ liefert $\widehat{\beta}(\widehat{\alpha})$

Problem: Verzerrung in kleineren Stichproben \Rightarrow Übergang zu REML.

(b) Restringierte ML (REML) - Schätzung

Ziel: Bias-Reduktion

Herleitung 1

- Transformiere $Y = X\beta + Z + \varepsilon$ linear zu $Y^* = AY$ so, dass Verteilung von Y^* nicht von β abhängt
 $\Rightarrow A$ so, dass $AX = 0$. Geht z.B. mit $A = \mathbb{I} - X(X'X)^{-1}X'$.
- Schätze α nach ML-Prinzip für Y^* .
- Ergebnis: $\widehat{\alpha}_{REML}$ maximiert die restringierte log-Likelihood $l^*(\alpha) =$

$$-\frac{1}{2} \left(\log |\Sigma(\alpha)| + \log |X'\Sigma^{-1}(\alpha)X| + (y - X\widehat{\beta}(\alpha))' \Sigma^{-1}(\alpha) (y - X\widehat{\beta}(\alpha)) \right)$$

(Alternative: ersetze $\widehat{\beta}(\alpha)$ durch β und schätze α und β simultan.)

Herleitung 2

- Für das hierarchische Modell gilt

$$\begin{aligned} Y|\beta, Z, \alpha &\sim N(X\beta + Z, \sigma^2\mathbb{I}) \\ Z &\sim N(0, \tau^2 R(\theta)) \end{aligned}$$

- Sei $p(y|\beta, Z, \alpha)$ zugehörige Dichte. Bilde marginale Dichte (mit $p(\beta) \propto \text{const}$)

$$p(\alpha|y) = \int p(y|\beta, Z, \alpha) d\beta dp(Z)$$

und maximiere bezüglich α .

- Es gilt (?) bis auf multiplikative Konstanten

$$\log p(\alpha|y) = l^*(\alpha)$$

Nach ? ist REML in Ordnung für gewöhnliches Kriging, ansonsten ist die MSE-Güte aber unklar; REML sensitiv bei Missspezifikation von $\mu(s)$ (allgemeines Problem).

Manchmal wird

$$l_{re}(\alpha, \beta) = -\frac{1}{2} (\log |\Sigma(\alpha)| + \log |X' \Sigma^{-1}(\alpha) X| + (y - X\beta)' \Sigma^{-1}(\alpha) (y - X\beta))$$

als restringierte log-Likelihood bezeichnet und zum simultanen Schätzen von α, β maximiert.

(c) Bayes-Schätzer

Als Hierarchisches Modell:

$$Y(s) | Z(s), \sigma_\varepsilon^2 \sim N(Z(s), \sigma_\varepsilon^2)$$

Prioris:

$$\left. \begin{array}{l} \sigma_\varepsilon^2 \sim \text{IG}(\cdot, \cdot) \\ Z(s) \sim \text{GZF}(\tau^2, \theta) \\ \tau^2 \sim \text{IG}(\cdot, \cdot) \\ \theta \sim ? \end{array} \right\} \text{semi-konjugierte}$$

Posteriori \propto Priori \cdot Likelihood

Vollständig bedingte Posteriori:

$$\begin{array}{l} Z(s) | \cdot \sim N(\cdot, \cdot) \\ \tau^2 | \cdot \sim \text{IG}(\cdot, \cdot) \\ \sigma^2 | \cdot \sim \text{IG}(\cdot, \cdot) \\ \theta | \cdot \sim ? \quad (\text{oft } \kappa \text{ fix}) \end{array}$$

Prädiktive Posteriori

$$Y(s_0) | Z, \sigma^2 \sim N(Z(s_0), \sigma_\varepsilon^2)$$

3.5.4 Generalisiertes Kriging

Geostatistische Modelle für nicht-normalverteilte Zielvariablen, z.B. Y binär, binomial, kategorial oder Zählvariable, bzw. Erweiterung des generalisierten linearen Modells um geostatistische Komponente. Vgl. Beispiele 2 (Rongelap) und 3 (Waldschäden).

Hierarchisches Modell

- (1) $Y(s_i) | W(s_i)$, $i = 1, \dots, n$ bedingt unabhängig, Dichten aus Exponentialfamilien
- (2) W stationärer Gauß-Prozess, $\mathbb{E}[W(s)] = 0$
- (3) $\mu_i = \mathbb{E}[Y(s_i) | W(s_i)] = h(\eta_i)$, $\eta_i = x_i' \beta + W(s_i)$, $x_i = x(s_i)$

Speziell: $x_i' \beta = \mu$ **Gewöhnliches Kriging**

$$\widehat{W}(s) = \mathbb{E}[W(s) | Y], \quad Y = (Y(s_1), \dots, Y(s_n))'$$

bzw.: $\widehat{\eta}(s) = x_i' \widehat{\beta} + \widehat{W}(s)$ ist generalisierter linearer Kriging-Prädiktor für W bzw. η an der Stelle s .

Statistische Inferenz

Zu schätzen sind: β Regressionsparameter, θ Kovarianzparameter, $W(s)$ an Beobachtungsstellen s_1, \dots, s_n und an Vorhersagestelle s .

Sei die (bedingte) Likelihood, mit $W = (W(s_1), \dots, W(s_n))'$

$$p(y | W, \beta) = \prod_{i=1}^n f_i(y(s_i) | W(s_i), \beta)$$

z.B.

$$f_i(y(s_i) | W(s_i), \beta) = \pi_i^{y(s_i)} (1 - \pi_i)^{1-y(s_i)} \quad \text{für binäre } Y(s_i)$$

und

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad \text{bei Logit-Modell}$$

(a) **Bayes-Ansatz**

$$\begin{aligned} W | \theta &\sim N_n(\cdot, \cdot) \\ p(\beta) &\propto \text{const} \quad \text{oder} \quad \beta \sim N(\cdot, \cdot) \\ \sigma^2, \tau^2 &\sim \text{IG}(\cdot, \cdot) \\ \theta &\sim ? \end{aligned}$$

MCMC mit Sampling aus vollständig bedingten Dichten:

$(\beta \cdot)$	Metropolis-Hastings
$(W \cdot)$	Metropolis-Hastings
$(\tau^2 \cdot)$	IG mit aufdatierten Parametern
$(\sigma^2 \cdot)$	IG mit aufdatierten Parametern
$(\theta \cdot)$?

(b) **Penalisierter Likelihood-Ansatz**

$$(1) \quad l_{pen}(\beta, W | \theta) = l(\beta, W) + \underbrace{\frac{1}{\tau^2} W' R^{-1}(\theta) W}_{\text{vgl. Kapitel 4.2}} \rightarrow \max_{\beta, W}$$

$$\begin{aligned} l(\beta, W) &= \log p(y | W, \beta) \\ r_{ij} &= \rho(\|s_i - s_j\| | \theta) \end{aligned}$$

(2) τ^2, θ, σ^2 mit REML

(1) und (2) iterieren

(c) **MCML: Monte Carlo Maximum Likelihood (?)**Mit φ wird im Folgenden der Vektor der Parameter bezeichnet.

$$\begin{aligned} L(\varphi) &= f(y|\varphi) = \int f(y|W)f(W|\varphi)dW \\ &\propto \int \frac{f(y|W)f(W|\varphi)}{\tilde{f}(y)}dW = \int \frac{f(y|W)f(W|\varphi)}{f(y|W)\tilde{f}(W)}\tilde{f}(W|y)dW \\ &= \tilde{\mathbb{E}} \left[\frac{f(W|\varphi)}{\tilde{f}(W)} \middle| y \right] \end{aligned}$$

 $\tilde{f}(W)$ ist eine Vorschlagsdichte für W auf \mathbb{R}^n .

$$\tilde{f}(y) = \int f(y|W)\tilde{f}(W)dW$$

(i) Monte Carlo Approximation

$$L_M(\varphi) = \frac{1}{m} \sum_{j=1}^m \frac{f(W(j)|\varphi)}{\tilde{f}(W(j))}$$

wobei $W(j)$ Ziehungen aus $\tilde{f}(\cdot|y)$ sind. Wahl von \tilde{f} kritisch, Vorschlag $\tilde{f} = f(\cdot|\varphi_0)$, einige Iterationen für neues φ .

(ii) Maximiere L_M über Newton-Raphson iterativ

$$\widehat{\beta}(W(j)) = (X'C^{-1}X)^{-1}C^{-1}W(j)$$

Varianz der Kovarianzstruktur

$$\widehat{\sigma}^2(W(j)) = \frac{1}{n}(W(j) - X\widehat{\beta}(W(j)))'C^{-1}(W(j) - X\widehat{\beta}(W(j)))$$

mit

$$C = R(\theta) + \tau^2\mathbb{I}_n$$

(iii) $\widehat{\beta}$ und $\widehat{\sigma}^2$ Plugin einsetzen in L_M . Maximiere $\tilde{L}_M(\theta) = L_M(\widehat{\beta}, \widehat{\sigma}^2, \theta)$. Maximiere numerisch (C^{-1} kritisch).

3.6 Diskrete Approximation von Gaussfeldern

Was passiert, wenn die Grösse von diskreten Gitterfeldern gegen Null geht. Wird dann aus einem Gauss-Markov-Zufallsfeld ein Gaussfeld? Und wenn ja, mit welcher Korrelationsfunktion?

Umgekehrt: Lässt sich ein GRF durch ein (numerisch vorteilhafteres) GMRF approximieren?

Beispiel 23. Geburten in Porto Alegre

Erste Überlegungen:

(a) Gegeben sei ein CAR(1)-Modell (nur erste Nachbarn)

$$\mathbb{E}[\phi_{ij}^* | \phi_{-(ij)}^*] = \frac{1}{w}(\phi_{i-1j}^* + \phi_{i+1j}^* + \phi_{ij-1}^* + \phi_{ij+1}^*)$$

$$\text{Var}(\phi_{ij}^* | \phi_{-(ij)}^*) = \frac{1}{w}$$

$$w > 4 \text{ (!)}$$

Mit $l = \|(ij) - (i'j')\|$ gilt (nach ?)

$$\text{Cov}(\phi_{ij}^*, \phi_{i'j'}^*) \approx \frac{w}{2\pi} \underbrace{K_0}_{\text{Bessel}}(l\sqrt{w-4})$$

Das entspricht einer Matern-Kovarianzfunktion mit $\kappa \rightarrow 0$, $\phi = \sqrt{w-4}$ und $\sigma^2 = \frac{w}{4\pi}$.

(b) CAR(2)-Modell (mit zweiten Nachbarn) $\hat{=}$ Matern-GRF mit $\kappa = 1$ und $\sigma^2 = \frac{1}{4\pi(w-4)}$.

(c) CAR($\nu+1$)-Modell $\hat{=}$ Matern mit $\kappa = \nu$ und $\sigma^2 = \frac{1}{4\pi\nu(w-4)^\nu}$.

Der Zusammenhang lässt sich über Stochastische Partielle Differentialgleichungen zeigen:

$$(\phi^2 - \Delta)^{\alpha/2} x(u) = w(u), \quad u \in \mathbb{R}^d, \quad \alpha = \kappa + \frac{d}{2}$$

$$\Delta := \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}, \quad w \text{ weißes Rauschen}$$

Stationäre Lösung dieser Gleichung ist ein Matern-Gauß-Feld. Spektral Darstellung (über Fourier-Transformation) zeigt, dass CAR's MGF's approximieren. Approximation gilt für beliebige Gitter, besser für „finite Elemente“-Darstellung (klassisches Verfahren zur Lösung partieller-DGL mit Triangulierung). Details in ?.

Kapitel 4

Räumliche Regularisierung

4.1 Nichtparametrische Schätzung von Oberflächen

Ziel dieses Abschnitts: Beziehung zwischen Kriging und der Schätzung von Oberflächen mittels Basisfunktionen (2D-Splines, radialen Basisfunktionen) aufzeigen.

Daten: $y_i = y(s_i)$ $i = 1, \dots, n$
Modelle: $Y_i = \mu(s_i) + \varepsilon_i$ $\varepsilon_i = \varepsilon(s_i) \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$
 $\mu(s) = f(s)$ 2D- (oder 3D-)Oberfläche,
 $s = (s_x, s_y)$ räumliche Lokation (bzw. (s_x, s_y, s_z))

- Erweiterung auf nicht-normalverteilte Zielvariablen im GLM-Rahmen
- s skalar: Nichtparametrische Regression bzw. Kurvenschätzung
- Ansätze mit Basisfunktionen:

$$f(s) = \sum_{k=1}^K \gamma_k B_k(s)$$

$B_k(s)$ Basisfunktionen $\mathbb{R}^d \rightarrow \mathbb{R}$, z.B. 2D-Splines, radiale Basisfunktionen.

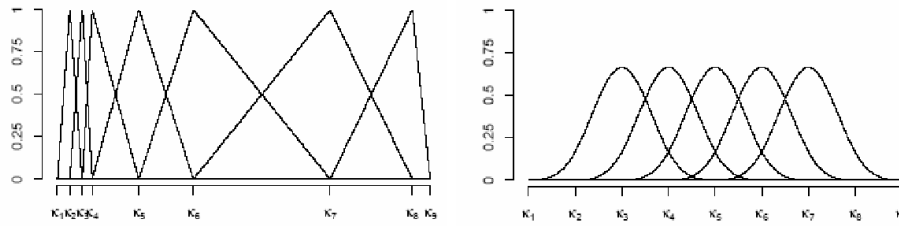


Abbildung 4.1: B-Spline Basisfunktionen. Links: erster Ordnung mit nicht äquidistanten Knoten, rechts: dritter Ordnung mit äquidistanten Knoten.

4.1.1 Nichtparametrische Regression mit penalisierten Splines

1D-Polynom-Splines

- $a = \xi_0 < \xi_1 < \dots < \xi_m < \xi_{m+1} = b$ Gitter auf $[a, b]$
- Die Menge der Polynom-Splines vom Grade l mit gegebenen m Knoten bildet $(m + l + 1)$ -dimensionalen Vektorraum
 \Rightarrow Jeder Polynom-Spline ist durch $m+l+1$ Basisfunktionen darstellbar

$$f(x) = \gamma_0 B_0(x) + \gamma_1 B_1(x) + \dots + \gamma_{m+l} B_{m+l}(x)$$

Nichtparametrische Regression mit Polynom-Splines

- Modell: $Y_i = f(x_i) + \varepsilon_i$ mit $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$
- Annahme: $f(x)$ kann durch einen Polynom-Spline approximiert werden:

$$f(x) = \sum_{k=0}^{m+l} \gamma_k B_k(x)$$

$$\Rightarrow Y_i = \gamma_0 B_0(x_i) + \dots + \gamma_{m+l} B_{m+l}(x_i) + \varepsilon_i$$

Lineares Modell

$$Y = Z\gamma + \varepsilon \quad \text{mit} \quad \gamma = (\gamma_0, \dots, \gamma_{m+l})'$$

und Design-Matrix

$$Z \quad \text{mit} \quad Z[i, j] = B_j(x_i)$$

- KQ-Schätzer: $\hat{\gamma} = (Z'Z)^{-1}Z'y$

Probleme:

- Grad $l = ?$ Meist $l = 3$, kubische Splines
- Wieviele Knoten und an welchen Positionen?
zu wenig Knoten: zu glatte Schätzung
zu viele Knoten: zu raue Schätzung („Overfitting“).

Penalisierte Regressions-Splines

- Definiere eine relativ große Zahl (äquidistanter) Knoten, um ausreichende Flexibilität zu erreichen
- Bestrafe Abweichungen benachbarter B-Spline Koeffizienten durch die Summe quadrierter erster bzw. zweiter Differenzen

$$\text{Pen}(\gamma) = \sum_k (\gamma_k - \gamma_{k-1})^2$$

bzw.
$$\text{Pen}(\gamma) = \sum_k (\gamma_k - 2\gamma_{k-1} + \gamma_{k-2})^2$$

In Matrixnotation:

$$\text{Pen}(\gamma) = \gamma'R^{(1)}\gamma \quad \text{bzw.} \quad \text{Pen}(\gamma) = \gamma'R^{(2)}\gamma$$

mit den Strafmatrizen $R^{(1)}$ bzw. $R^{(2)}$ (wie in 2.3.5, siehe unten bei Bayes). Statt der Residuenquadratsumme wird nun die penalisierte Residuenquadratsumme bezüglich β und γ minimiert (penalisierte KQ-Schätzung):

$$\begin{aligned} \text{KQ}_{pen}(\gamma) &= (y - Z\gamma)'(y - Z\gamma) + \lambda\gamma'R\gamma \\ &= \text{KQ}(\gamma) + \lambda\text{Pen}(\gamma) \rightarrow \min \end{aligned}$$

- λ steuert Kompromiss zwischen Datentreue und Glattheit.
- Lösung:

$$\hat{\gamma} = (Z'Z + \lambda R)^{-1} Z'y$$

kann numerisch effizient berechnet werden, weil $Z'Z$ und K Bandmatrizen sind. Schätzung vom Ridge-Typ!

Bayes-Version: Differenzen werden durch Random Walk-Modelle ersetzt:
Random Walk erster Ordnung:

$$\gamma_k = \gamma_{k-1} + \nu_k, \nu_k \stackrel{\text{i.i.d.}}{\sim} \text{N}(0, \tau^2), p(\gamma_0) \propto \text{const.}$$

Random Walk zweiter Ordnung:

$$\gamma_k = 2\gamma_{k-1} - \gamma_{k-2} + \nu_k, \nu_k \stackrel{\text{i.i.d.}}{\sim} \text{N}(0, \tau^2), p(\gamma_0) \propto p(\gamma_1) \propto \text{const.}$$

In Matrizenform:

$$p(\gamma) \propto \exp\left(-\frac{1}{2\tau^2}\gamma'R\gamma\right) \quad \text{Priori für } \gamma$$

mit $R = R^{(1)}$ für RW1 und $R = R^{(2)}$ für RW2 wie in Abschnitt 2.3.5.
 Posteriori:

$$p(\gamma|y) \propto p(y|\gamma)p(\gamma) \quad \text{Normalverteilung}$$

Logarithmieren und Maximieren liefert Posteriori-Erwartungswert. Äquivalent zu $-l_{pen}(\gamma) \rightarrow \min$ bzw.

$$\text{KQ}_{pen}(\gamma) = (y - Z\gamma)'(y - Z\gamma) + \frac{\sigma^2}{\tau^2} \gamma'R\gamma \rightarrow \min$$

\Rightarrow Mit $\lambda = \frac{\sigma^2}{\tau^2}$ sind der penalisierte KQ-Schätzer $\hat{\gamma}$ und der Posteriori-Erwartungswert als Punktschätzer identisch. Aber: Kovarianzmatrizen bei frequentistischem und Bayesianischem Ansatz verschieden!

- Frequentistisch: $\hat{\gamma} = (Z'Z + \lambda R)^{-1}Z'y \Rightarrow$

$$\begin{aligned} \text{Cov}(\hat{\gamma}) &= (Z'Z + \lambda R)^{-1}Z'\sigma^2\mathbb{I}Z(Z'Z + \lambda R)^{-1} \\ &= \sigma^2(Z'Z + \lambda R)^{-1}Z'Z(Z'Z + \lambda R)^{-1} \quad \text{Sandwich-Matrix} \end{aligned}$$

- Bayesianisch: Posteriori-Kovarianzmatrix aus Posteriori-Verteilung

$$\text{Cov}(\gamma|y) = \sigma^2V^{-1} = \sigma^2(Z'Z + \lambda R)^{-1}$$

Erweiterung auf 2D: 2D-Tensorprodukt-B-Splines

$$\begin{aligned} Y_i &= f(s_i, t_i) + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{N}(0, \sigma^2) \\ f(s_i, t_i) &= \sum_{j,k} \gamma_{jk} B_j(s) B_k(t) \end{aligned}$$

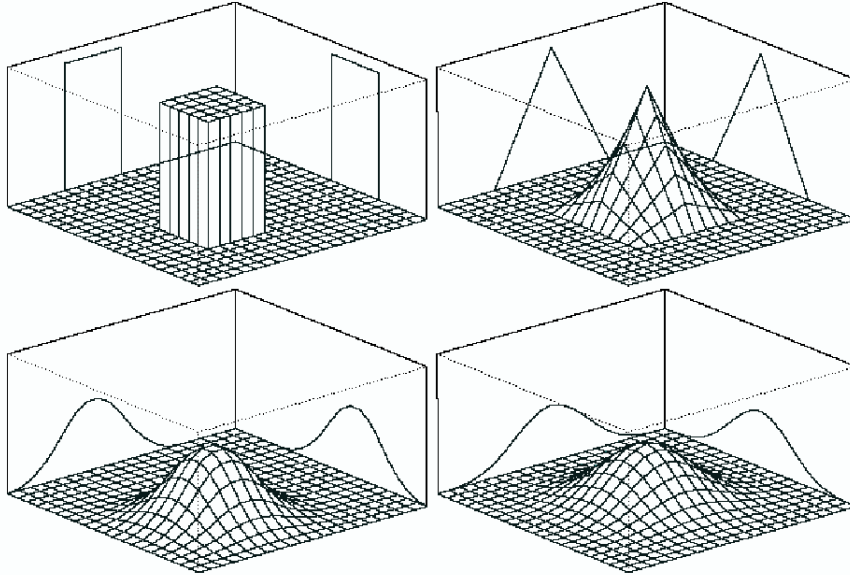


Abbildung 4.2: Tensorprodukt der B-Spline-Basisfunktionen vom Grad 0 bis 3.

$B_j(s), \dots, B_k(t)$ eindimensionale B-Spline-Basisfunktionen, paarweise Produkte der Basen (Tensor-Produkt).

- Bei geeigneter Anordnung und Definition von Designmatrizen und Parametern wieder in der Form

$$Y = Z\gamma + \varepsilon$$

darstellbar

- Schätzung mit penalisiertem KQ-Ansatz oder mit Bayes-Ansatz
- Penalisierung über Differenzen von benachbarten Knotenpunkten, im Bayes-Ansatz über Gauß-Markov-Zufallsfeld als Priori (siehe 2.3.5 (c) bis (f)).

4.1.2 Oberflächenschätzung mit radialen Basisfunktionen

Definition 4.1. (Radiale Basisfunktion)

Eine radiale Basisfunktion ist eine reellwertige Funktion, deren Wert nur vom Abstand von einem Zentrum (Knoten) abhängt.

$$C(s, \xi^s) = C(\|s - \xi\|)$$

Basisfunktion ist also radialsymmetrisch um Knoten ξ .

- Traditionelle radiale Basisfunktionen

$$C(r) = \begin{cases} |r|^{2m-1} & \text{(stückweise Polynome)} \\ |r|^{2m} \log |r| & \end{cases}$$

z.B.: $C(r) = |r|^2 \log |r|$ (Thin Plate Splines).

- Je größer m , desto glatter die (geschätzte) Oberfläche.
- $m = 1$:

$$f(t) = \beta_0 + \beta_1 t + \sum_{j=1}^n \gamma_j |t - t_j|, \quad Y_i = f(t_i) + \varepsilon_i$$

Knoten $\xi_t = t_j$ an jeder Beobachtungsstelle $t_j!$ $\Rightarrow n \leq p$

Datenmodell:

$$Y = X\beta + Z\gamma + \varepsilon, \quad Z = (z_{ij}) = (|t_i - t_j|) \quad n \times n, \quad \text{symmetrisch}$$

Regularisierung! Vorschlag

$$KQ_{pen}(\beta, \gamma) = \|y - X\beta - Z\gamma\|^2 + \lambda\gamma'Z\gamma \rightarrow \min$$

Sieht ad hoc aus; jedoch enge Beziehung zu Glättungs-Splines (Daten=Knoten).

Parametersparsamer ist LR (low rank)-Glättung:

Wähle Knoten $\xi_1 < \dots < \xi_j < \dots < \xi_m$, $m < n$. Approximiere

$$f(t) = \beta_0 + \beta_1 t + \sum_{j=1}^m \gamma_j C(|t - \xi_j|)$$

$$\Rightarrow Z = (z_{ij}) = (C(|t_i - \xi_j|)) \quad n \times m, \quad \text{Rg}(Z) = m$$

Schätzen:

$$\text{KQ}_{\text{pen}}(\beta, \gamma) = \|y - X\beta - Z\gamma\|^2 + \lambda\gamma'K\gamma \rightarrow \min$$

$$k_{ij} = C(\|\xi_i - \xi_j\|)$$

oder mit Bayes-Ansatz

$$Y|\gamma \sim N(X\beta + Z\gamma, \sigma^2\mathbf{I}), \quad p(\beta) \propto \text{const}$$

$$p(\gamma) \propto \exp\left(-\frac{1}{2\tau^2}\gamma'K\gamma\right)$$

Logarithmieren ergibt

$$l_{\text{pen}}(\beta, \gamma) = -\frac{1}{2\sigma^2}\|y - X\beta - Z\gamma\|^2 - \frac{1}{\tau^2}\gamma'K\gamma \rightarrow \max$$

$$\Leftrightarrow \text{KQ}_{\text{pen}}(\beta, \gamma) \rightarrow \min \quad \text{mit} \quad \lambda = \frac{\sigma^2}{\tau^2}$$

$\Rightarrow \hat{\beta}, \hat{\gamma}$ Posteriori-Modus und -Erwartungswert

Erweiterung auf $d \geq 2$ konzeptionell einfach

$$f(s, t) = \beta_0 + \beta_1s + \beta_2t + \sum_{j=1}^m \gamma_j C(\|(s, t) - \xi_j\|)$$

$$\Rightarrow Y = X\beta + Z\gamma + \varepsilon$$

(LR-) Glättung analog wie für $d = 1$, X , Z und K geeignet definiert.

Kriging als Oberflächenschätzung

Wähle als radiale Basisfunktionen isotrope Korrelationsfunktionen

$$C(\|r\|) = \rho(\|r\| |\theta)$$

z.B. ρ aus der Matérn-Familie.

Mit $Y = (Y(s_1), \dots, Y(s_n))'$, $W = (W(s_1), \dots, W(s_n))'$ gilt beim Kriging

$$Y = X\beta + W + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2\mathbf{I})$$

$$W \sim N(0, \tau^2R), \quad r_{ij} = \rho(\|s_i - s_j\| |\theta)$$

$$\Leftrightarrow Y = X\beta + Z\gamma + \varepsilon, \quad Z = R, \quad \gamma \sim N(0, \tau^2R^{-1})$$

Fazit: Kriging ist Bayes- (bzw. stochastischer Prozesse-)Ansatz zum Oberflächenschätzen mit Korrelationsfunktionen als radialen Basisfunktionen. Kriging Formel (als Punktschätzung) auch ableitbar über die penalisierte KQ-Schätzung.

4.2 Geoadditive Regression

Datensituation: Zu jeder individuellen Beobachtung $i = 1, \dots, n$ werden eine Zielvariable Y_i , Kovariablen x_i zu räumliche Lokationen s_i erhoben. s_i kann dabei geostatistisch sein (also stetige Koordinaten) oder ein Gitterdatum (diskretes Gitter).

Vergleiche Beispiele 11 (Mietspiegel, Gauß-Fall) und 3 (Waldschäden, generalisierte Regression).

4.2.1 Gauß-Fall

Y stetig, (approximativ) normalverteilt

x Kovariablen mit linearem Effekt

z stet. Kovariablen mit nicht-linearem Effekt (oder andere Regularisierung)

s räumliche Lokation

Additives Modell:

$$Y_i = x_i\beta + \underbrace{\sum_{j=1}^p f_j(z_{ij})}_{\text{glatte (1D)-Effekte}} + f_{geo}(s_i) + \varepsilon_i \quad \text{mit} \quad \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

Bemerkung: Für die glatten Effekte Zentrierung um Null nötig zur Sicherung der Identifizierbarkeit.

Inferenz über Bayes oder über penalisierten KQ-Ansatz. Die Likelihood lautet

$$f(\theta|y) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - x_i\beta - \sum_{j=1}^p f_j(z_{ij}) - f_{geo}(s_i)\right)^2\right)$$

Für das Kleinst-Quadrate-Kriterium gilt:

$$\text{KQ} = -\sigma^2 \log(f(\theta|y)) + \text{const}$$

Für die glatte Effekten $f_j(z) = z_j\gamma_j$ benutzen wir Penalisierungsterme $\lambda_j\gamma_j K_j\gamma_j$ oder Priors $\gamma_j \sim N(0, (\tau_j K_j)^{-1})$ (siehe 4.1). Für den räumlichen Effekt bieten sich folgende Modellierungsmöglichkeiten an:

(a) **Gitterdaten:** $f_{geo}(s_i) = \phi_K$ für $s_i = K$. GMRF als Priori:

$$p(\phi) \propto \exp\left(-\frac{\tau_{geo}}{2}\phi'Q\phi\right)$$

bzw. Penalisierungsterm: $\lambda_{geo}\phi'Q\phi$

(b) **Geostatistisch/Kriging (über radiale Basisfunktionen)**

$$f_{geo}(s) = R\phi \quad \text{mit} \quad \phi \sim N(0, (\tau_{geo}R)^{-1})$$

bzw. Penalisierungsterm: $\lambda_{geo}\phi'R\phi$

(c) **Als 2D-Oberfläche/2D-Spline**

$$f_{geo} = B\phi \quad \text{mit} \quad p(\phi) \propto \exp\left(-\frac{\tau_{geo}}{2}\phi'Q\phi\right)$$

bzw. Penalisierungsterm: $\lambda_{geo}\phi'Q\phi$

In allen Fällen: $(\phi \equiv \gamma_{geo})$

$$f_{geo}(s) = z_{geo}\gamma_{geo} \quad \text{mit} \quad \gamma_{geo} \sim N(0, (\tau_{geo}K_{geo})^{-1})$$

Q ist spärlich besetzt (sparse), damit numerisch effizienter.

Insgesamt lässt sich das additive Modell schreiben als:

$$Y = x\beta + \sum_{j=1}^p z_j\gamma_j + \varepsilon \quad \text{mit} \quad \gamma_j \sim N(0, (\tau_j K_j)^{-1})$$

Bemerkungen

- Geoadditive Modelle lassen sich wie Additive Modelle ohne räumliche Komponente dahingehend erweitern, dass Interaktionen einbezogen werden. Eine Interaktion zwischen einer (binären) Kovariablen x und einer metrischen Kovariablen z von der Form

$$\dots + f(z)x + \dots$$

kann als mit z variierender Effekt von x interpretiert werden (VCM, „varying coefficient model“). Interaktionen zwischen zwei metrischen Variablen z_1 und z_2 (Interaktion zwischen Raum und Zeit wie in 2.3.5 (h)) können durch

$$\dots + f_{12}(z_1, z_2) + \dots$$

z.B. mit 2D-Tensorprodukt-Splines modelliert werden.

- Bei Verwendung von penalisierten B-Splines und Tensorprodukt-Splines hat die Präzisionsmatrix K_j in der Regel keinen vollen Rang ($\text{Rg}(K_j) < \dim(\gamma_j)$).

Details zur Inferenz

$$Y | \eta \sim N(\eta, \sigma^2 \mathbb{I}), \quad \eta = X\beta + Z_1\gamma_1 + \dots + Z_p\gamma_p + Z_{geo}\gamma_{geo} = X\beta + Z\gamma$$

$$p(\gamma_j) \propto \exp\left(-\frac{1}{2\tau_j^2}\gamma_j'K_j\gamma_j\right), \quad j = 1, \dots, p, \text{ geo}$$

- **Penalisierte Likelihood-/empirische Bayes-Inferenz**

- (1) Minimiere bei gegebenen $\sigma^2, \tau_1^2, \dots, \tau_p^2, \tau_{geo}^2$

$$\text{KQ}_{pen}(\beta, \gamma_1, \dots, \gamma_p, \gamma_{geo}) = \frac{1}{\sigma^2} \|y - X\beta - Z\gamma\|^2 + \sum_{j=1}^p \frac{1}{\tau_j^2} \gamma_j' K_j \gamma_j$$

- (2) Schätze $\sigma^2, \tau_1^2, \dots, \tau_p^2, \tau_{geo}^2$ mit REML (oder anderem Kriterium wie Kreuzvalidierung).

Iteriere (1) und (2).

- **Volle Bayes-Inferenz mit Gibbs-Sampling**

Priori-Verteilungen $\sigma^2 \sim \text{IG}(a_0, b_0)$, $\tau_j^2 \sim \text{IG}(a_j, b_j)$. Iteratives Ziehen aus vollständig bedingten Dichten

- (1) $\beta | \cdot \sim N((X'X)^{-1}X'(y - \tilde{\eta}), \sigma^2(X'X)^{-1})$
mit $\tilde{\eta} = \eta - X\beta = Z\gamma$, d.h. Prädiktor ohne $X\beta$

- (2) Für $j = 1, \dots, p, \text{ geo}$: $\gamma | \cdot \sim N(m_j, \sigma^2 V_j)$

$$m_j = V_j Z_j' (y - \eta_{-j}), \quad V_j = \left(Z_j' Z_j + \frac{\sigma_j^2}{\tau_j^2} K_j \right)^{-1}, \quad \eta_{-j} = \eta - Z_j \gamma_j$$

- (3) $\sigma^2 | \cdot \sim \text{IG}(a'_0, b'_0)$, $a'_0 = a_0 + \frac{n}{2}$, $b'_0 = b_0 + \frac{1}{2}(y - \eta)'(y - \eta)$
 $\tau_j^2 | \cdot \sim \text{IG}(a'_j, b'_j)$, $a'_j = a_j + \frac{\text{Rg}(K_j)}{2}$, $b'_j = b_j + \frac{1}{2}\gamma_j' K_j \gamma_j$

4.2.2 Generalisierte Geoadditive Modelle

Daten: $y_i, x_i, z_i, s_i, i = 1, \dots, n$:

- Y_i nicht-normale, z.B. binäre Zielvariable
- x_i übliche Kovariablen mit linearen Effekten
- z_i stetige Kovariablen mit nichtlinearen Effekten
- s_i räumliche Koordinaten von Individuum bzw. Objekt i oder Index der Region i .

Modell:

$Y_i | x_i, z_i, s_i$ unabhängig,

$$\mathbb{E}[Y_i | \cdot] = \mu_i = h(\eta_i),$$

$$\eta_i = x_i' \beta + \sum_{j=1}^p f_j(z_{ij}) + f_{geo}(s_i),$$

f_1, \dots, f_p z.B. penalisierte B-Splines,

$f_{geo}(s_i) = W(s_i)$ stationäres GRF oder GMRF oder 2D-Tensorprodukt-Spline.

Außer der Likelihood hat sich also im Modell im Vergleich zum Gauß-Fall nichts geändert. Modellierung der glatten Effekte und der glatten räumlichen Effekte (und so weiter) wie oben.

Details zur Inferenz

(a) **Penalisierter Likelihoodansatz**

$$(1) \quad l_{pen}(\beta, \gamma) = l(\beta, \gamma) - \sum_{j=1}^{p,geo} \frac{1}{\tau_j} \gamma_j' K_j \gamma_j \quad \rightarrow \quad \max_{\beta, \gamma}$$

Lösung über: Fisher-Scoring (bei gegebenem τ_j^2)

(2) τ_j^2 durch REML

(b) **Voller Bayes-Ansatz**

Ziehen aus vollständig bedingten Dichten

- $(\beta | \cdot), (\gamma_j | \cdot)$ mit MH-Schritten oder wie bei latenten GMRF
- $(\tau_j^2 | \cdot) \sim \text{IG}(a'_j, b'_j)$

(c) **Integrated Nested Laplace Approximation (INLA)**

kompaktere Darstellung

$$\begin{aligned} Y_i | \phi_i, \theta &\stackrel{\text{i.i.d.}}{\sim} f(y_i | \phi_i, \theta) \\ \phi_i | \theta &\sim \text{N}(0, Q(\theta)^{-1}) \end{aligned}$$

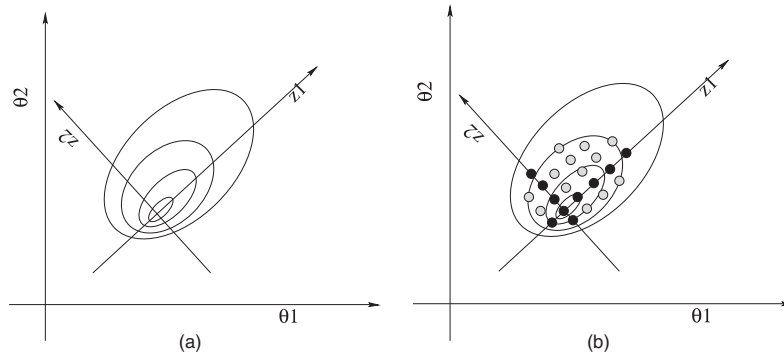


Abbildung 4.3: Berechnung der marginalen Posteriori $p(\theta)$: (a) Finde Modus und berechne Hesse-Matrix, rotiere Koordiantensystem; (b) ausgehend vom Modus wird $p(\theta)$ an Punkten (\bullet) entlang der Achsen berechnet, bis diese unter eine Schranke fallen; danach wird $p(\theta)$ an weiteren Punkten (\bullet) berechnet (übernommen aus ?).

$$\theta \sim p(\theta)$$

Posteriori:

$$\begin{aligned} p(\phi, \theta|y) &\propto p(\theta)p(\phi|\theta) \prod_{i=1}^n f(y_i|\phi_i, \theta) \\ &\propto p(\theta)|Q(\theta)|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\phi'Q(\theta)\phi + \sum_{i=1}^n \log f(y_i|\phi_i, \theta)\right) \end{aligned}$$

(1) Laplace-Approximation der marginalen Posteriori von ϕ

$$p(\theta|y) = \frac{p(\phi, \theta|y)}{p(\phi|\theta, y)} \approx \frac{p(\phi, \theta|y)}{\tilde{p}_G(\phi|\theta, y)} \Big|_{\phi=\phi^*(\theta)}$$

\tilde{p}_G Gauß-Approximation über iterative Taylor-Entwicklung am ϕ^* Modus von $\tilde{p}_G(\phi|\theta, y)$. Numerische Approximation über Interpolation an Integrationspunkten.

(2) Approximation von $p(\phi_i|\theta, y)$

- \tilde{p}_G (nicht sehr gut)
- $\tilde{p}_{LA}(\phi_i|\theta, y) \propto \frac{p(\phi, \theta|y)}{\tilde{p}_{GG}(\phi_{-i}|\phi_i, \theta, y)} \Big|_{\phi_{-i}=\phi_{-i}^*(\phi_i, \theta)}$

Hier wird $\phi_{-i}^*(\phi_i, \theta) = \mathbb{E}_{\tilde{p}_G}[\phi_{-i}|\phi_i]$ ersetzt (Berechnung einfa-

cher).

- \tilde{p}_{SLA} vereinfachte LA.

4.2.3 Deviance Information Criterion

In der Regel ist bei geadditiven Modellen $p > n$. Durch Regularisierung ist Modell schätzbar. Problem: Wie Modellwahl? Viele Parameter führen zu gutem Fit. AIC-Ansatz: Fit + Anzahl Parameter. Im Räumlichen: Abhängig von Glättungsparameter! Glättungspräzision:

- $\lambda \rightarrow \infty \Rightarrow$ Fläche komplett konstant, also nur ein Parameter ($p = 1$)
- $\lambda \rightarrow 0 \Rightarrow$ Räumliche Effekte i.i.d. $\hat{=} p = n$

Anzahl der Parameter hängt ab von den Daten!

? schlagen Deviance Information Criterion DIC vor.:

Saturierte Devianz

$$D = -2(l(\theta) - l(\hat{\theta}))$$

wobei l die log-Likelihood und $\hat{\theta}$ den geschätzten Parametervektor bezeichnen.

D kann nur geschätzt werden $\rightarrow \hat{D}$

$$\text{DIC} = \hat{D} + p_D$$

p_D ist die Anzahl effektiver Parameter, geschätzte Anzahl an Parametern

$$p_D = \hat{D} - D(\hat{\theta})$$

Beispiel 24. Räumliches Alters-Perioden-Kohorten-(APC-)Modell mit Interaktionen [?].

Kapitel 5

Räumliche Punkt- und Zählprozesse

5.1 Einführung in räumliche Punktprozesse

Ein räumlicher Punktprozess N ist eine zufällige, endliche oder abzählbare Menge

$$N = \{s_1, \dots, s_n, \dots\} \subset S \subseteq \mathbb{R}^d$$

S ist eine feste, vorgegebene Teilmenge von \mathbb{R}^d .

s_1, \dots, s_n sind zufällige Punkte (Lokationen von Ereignissen), $s_i \in S$.

- Vereinbarung: In einer beschränkten (Borel-)Menge B treten nur endlich viele Ereignisse auf; die Anzahl $n = N(B)$ ist jedoch zufällig, genauso wie die Lokationen. Für eine andere Realisierung, z.B. Aufzeichnung von Blitzschlägen an einem anderen Tag, ändert sich die Anzahl der Ereignisse (Punkte) und ihre Lokation.
- Beschränkung auf einfache Punktprozesse: An jeder Lokation tritt nur ein Ereignis auf. Multiple Ereignisse lassen sich mit markierten Punktprozessen behandeln.
- Markierte Punktprozesse: Mit jedem Ereignis s ist (die Realisierung) einer Zufallsvariable (Marke) $Y(s)$ verbunden, z.B. Typ der Zelle (0/1), Durchmesser der Fichten, Stärke des Blitzes, etc.
- Realisierung von Punktprozessen ist oft Momentaufnahme eines räumlich-zeitlichen Prozesses.
- Definiere zu räumlichem SP $\{Y(s), s \in S\}$ auf (dichtem Gitter) S den Indikator-Prozess $\mathbf{1}(s) \in \{0, 1\}$ mit $\mathbf{1}(s) = 1 \Leftrightarrow Y(s) > c$, c Schwel-

lenwert. Entferne Punkte s mit $\mathbb{1}(s) = 0 \Rightarrow$ verbleibende (Ereignis-) Punkte bilden Punktprozess.

- Typen räumlicher Punktmuster
 - rein zufälliges Muster
 - Muster mit Cluster-Bildung (Aggregation)
 - reguläre Muster (Abstoßung, Disaggregation)

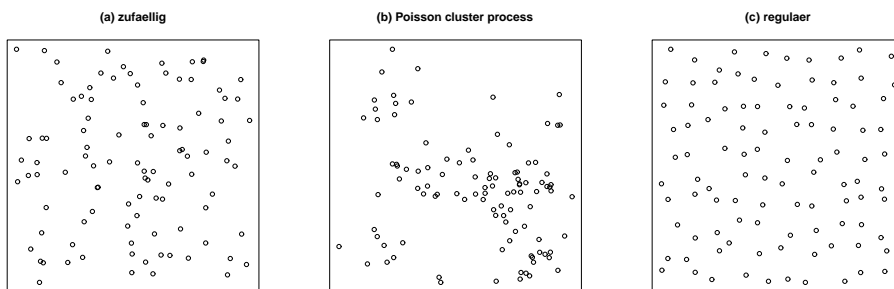


Abbildung 5.1: Realisationen von (a) einem komplett zufälligem Muster, (b) einem Poisson cluster Prozess, (c) einem regulärem Muster. Für alle Muster ist der Erwartungswert pro Einheitsquadrat gleich 10.

- Rein zufällige Muster: Gleiche erwartete Anzahl von Ereignissen pro Einheitsfläche in S (homogenes Muster); Anzahlen $N(B_1), \dots, N(B_k)$ in disjunkten Mengen $B_1, \dots, B_k \subset S$ sind unabhängig. Stochastisches Modell: Homogener Poisson-Prozess.
- Aggregierte Muster: Durchschnittliche Distanz von Ereignis s_i zum nächsten Ereignis kleiner als bei zufälligem Muster.
- Reguläre Muster: Durchschnittliche Distanz von Ereignis s_i zum nächsten Ereignis größer als bei zufälligem Muster.

Charakterisierung durch stochastische Modelle

- **Vollständige Charakterisierung als Punktprozess durch endlich-dimensionale (e.-d.) Verteilungen**

$\mathbb{P}(s_1 \in B_1, \dots, s_n \in B_n)$ für beliebige n , (Borel-)Mengen $B_1, \dots, B_n \subseteq S$; oder (äquivalent) als Zählprozess N , mit $N(B)$ zufällige Anzahl von Ereignissen in $B \subseteq S$ (wobei $N(B) < \infty$ für beschränkte B). Die endlich-dimensionalen Verteilungen $\mathbb{P}(N(B_1) = n_1, \dots, N(B_k) = n_k)$,

für beliebige $k = 1, 2, \dots, B_1, \dots, B_k$ (Borel-) Mengen, determinieren räumlichen Prozess vollständig.

Maßzahlen: $\mathbb{E}[N(B)] = \mu(B)$, $\text{Cov}(N(B_1), N(B_2)), \dots$

Es gilt: Endlich-dimensionale Verteilungen bereits determiniert durch die „Null-Wahrscheinlichkeiten“ (void probabilities) $\mathbb{P}(N(B) = 0)$, für beliebige beschränkte Borel-Mengen B .

- **Unvollständige Charakterisierung durch Intensitätsfunktionen:** ds kleine Menge (Kreis, Kugel) um s , $|ds|$ Fläche/Volumen (Lebesgue-Maß) von ds .

Definition 5.1. (Intensitätsfunktion 1. Ordnung)

$$\lambda(s) = \lim_{|ds| \rightarrow 0} \frac{\mathbb{E}[N(ds)]}{|ds|} \quad \text{Intensitätsfunktion 1. Ordnung}$$

Es gilt: $\mathbb{E}[N(B)] = \mu(B) = \int_B \lambda(s) ds$, d.h. $\lambda(s)$ ist Dichte des (lokal finiten, aber nicht normierten) Maßes $\mu(B)$ auf S bezüglich des Lebesgue-Maßes.

Interpretation: $\lambda(s) |ds| \approx \mathbb{E}[N(ds)] \approx \mathbb{P}(1 \text{ Ereignis in } ds)$.

Vergleiche Definition der Hazardrate $\lambda(t)$.

Definition 5.2. (Intensitätsfunktion 2. Ordnung)

$$\lambda_2(s, u) = \lim_{|ds|, |du| \rightarrow 0} \frac{\mathbb{E}[N(ds) N(du)]}{|ds| |du|} \quad \text{Intensitätsfunktion 2. Ordnung}$$

Interpretation: $\lambda_2(s, u) |ds| |du| \approx \mathbb{P}(\text{in } ds \text{ und } du \text{ tritt je ein Ereignis auf})$.

Es gilt:

$$\mathbb{E}[N(A) \cdot N(B)] = \int_A \int_B \lambda_2(s, u) dsdu + \mu(A \cap B),$$

$$\text{Cov}(N(A), N(B)) = \mathbb{E}[N(A) \cdot N(B)] - \mu(A) \mu(B)$$

$\Rightarrow \lambda$ und λ_2 determinieren die 1. und 2. Momente der Zufallsvariablen

$\{N(B), B \subseteq \mathbb{R}^d\}$.

$C(s, u) = \lambda_2(s, u) - \lambda(s) \lambda(u)$ Kovarianzdichtefunktion

Definition 5.3. (Stationarität und Isotropie)

Punkt-/Zählprozess N heißt (schwach) stationär:

$$\Leftrightarrow \begin{cases} \lambda(s) = \lambda = \text{const} \\ \lambda_2(s, u) = \lambda_2(s + \Delta, u + \Delta) \end{cases} \Leftrightarrow \lambda \text{ und } \lambda_2 \text{ sind translationsinvariant.}$$

- N heißt (schwach) isotrop: $\Leftrightarrow \lambda_2(s, u) = \lambda_2(\|s - u\|)$ ist rotationsinvariant.
 N heißt stark stationär: \Leftrightarrow E.-d. Verteilungen sind translationsinvariant.
 N heißt stark isotrop: \Leftrightarrow E.-d. Verteilungen sind rotationsinvariant.

Definition 5.4. (Sphärische Kontakt-Verteilungs-Funktion)

Die sphärische Kontakt-Verteilungs-Funktion („spherical contact distribution function“) ist

$$H_{s,o}(r) = 1 - \mathbb{P}(N(b(o, r)) = 0)$$

für $r > 0$. Mit $b(o, r)$ wird ein Ball um Ort o mit Radius r definiert.

5.2 Binomiale und Poisson-Prozesse

5.2.1 Bernoulli- und binomiale Prozesse

$|B| = \nu(B)$ = Fläche/Volumen (Lebesgue-Maß) einer (Borel-)Menge B .

- Bernoulli-Prozess: Die zufällige Menge D besteht nur aus einem einzigen Ereignis $s \in S$, d.h. $D = \{s\}$. Dabei ist s in S (beschränkt) verteilt gemäß $\mathbb{P}(s) = |B| / |S|$ für alle $B \subseteq S$.
- Binomial-Prozess: Überlagerung von n unabhängigen Bernoulli-Prozessen, mit $D = \{s_1, \dots, s_n\}$; s_1, \dots, s_n Ereignisse (Lokationen) der n Bernoulli-Prozesse; n fest.

$$\Rightarrow \mathbb{P}(s_1 \in B_1, \dots, s_n \in B_n) = \frac{|B_1| \cdot \dots \cdot |B_n|}{|S|^n}, \quad B_1, \dots, B_n \subseteq S$$

$N(B)$ Anzahl der Ereignisse (Lokationen) in $B \subseteq S$

$$\Rightarrow N(B) \sim \text{Bin}(n = N(S), \pi(B) = |B| / |S|)$$

Damit ergibt sich für die Intensitätsrate

$$\lambda(s) = \lim_{|ds| \rightarrow 0} \frac{\mathbb{E}[N(ds)]}{|ds|} = \lim_{|ds| \rightarrow 0} \frac{n |ds| / |S|}{|ds|} = \frac{n}{|S|} =: \lambda$$

$\Rightarrow \lambda(s) = \lambda$ unabhängig von s , d.h. der Binomial-Prozess ist ein homogener Prozess.

Ereignisse in disjunkten Teilmengen sind jedoch nicht unabhängig:

$n = N(S)$ ist fest. Sei $m = N(B)$ für $B \subset S \Rightarrow$

$N(S \setminus B) = n - m \Rightarrow N(B)$ und $N(S \setminus B)$ abhängig.

5.2.2 Homogener Poissonprozess

Definition 5.5. (*Homogener Poissonprozess*)

N homogener Poissonprozess: \Leftrightarrow

1. $N(B) \sim \text{Po}(\lambda|B|)$, mit konstanter Intensität $0 < \lambda < \infty$,
2. B_1, B_2 disjunkte (Borel-)Mengen $\Rightarrow N(B_1)$ und $N(B_2)$ sind unabhängig („independent scattering“).

Homogener Poisson-Prozess erzeugt (per Definition) rein zufälliges räumliches Muster.

Ziele der Statistik räumlicher Punktprozesse:

- Explorative Analysen, Maßzahlen, ... oft im Vergleich zu Poisson-Prozessen.
- Komplexere Punktprozess-Modelle, oft ausgehend von Poisson-Prozessen.

Grundlegende Eigenschaften

- (1) $\mathbb{P}(N(B) = n) = \frac{\lambda^n |B|^n}{n!} \exp(-\lambda|B|)$ folgt aus Eigenschaft 1. des Poisson-Prozesses.

- (2) Für die Null-Wahrscheinlichkeit („void probability“) gilt:

$$V_K = \mathbb{P}(N(K) = 0) = \exp(-\lambda|K|)$$

- (3) Für disjunkte Borelmengen B_1, \dots, B_K gilt: $N(B_1), \dots, N(B_K)$ sind unabhängige, Po-verteilte Zufallszahlen mit Mittelwerten $\lambda|B_1|, \dots, \lambda|B_K|$.
- (4) Homogene Poisson-Prozesse sind stationär und isotrop.
- (5) Unter der Restriktion $N(B) = n$ wird aus dem Poisson-Prozess ein Binomial-Prozess (siehe Møller / Waagepetersen, S.15).
- (6) Für Poisson-Prozesse gilt: $H_s(r) = 1 - \exp(-\lambda b_d r^d)$, also unabh. von o . d bezeichnet die Dimension.

5.2.3 Inhomogener Poissonprozess

Definition 5.6. (*Inhomogener Poissonprozess*)

N inhomogener Poisson-Prozess: \Leftrightarrow

- (1) Die Intensitätsfunktion $0 < \lambda(s) < \infty$ variiert räumlich, und

$$N(B) \sim \text{Po}(\mu(B)) \quad \text{mit} \quad \mu(B) = \int_B \lambda(s) ds$$

- (2) B_1, B_2 disjunkte (Borel-)Mengen $\Rightarrow N(B_1)$ und $N(B_2)$ sind unabhängig („independent scattering“).

N inhomogener Poisson-Prozess: \Leftrightarrow

- (1) $N(B) \sim \text{Po}(\mu(B))$ mit $\mu(B) = \int_B \lambda(s) ds$
 $\lambda(s) = \lim_{|ds| \rightarrow 0} \mathbb{E}[N(ds)]/|ds|$ Intensitätsrate, B beschränkte Borelmenge
 also:

$$\mathbb{P}(N(B) = n) = \frac{e^{-\mu(B)} (\mu(B))^n}{n!}$$

- (2) B_1, B_2, \dots, B_k disjunkte Borelmengen $\Rightarrow N(B_1), N(B_2), \dots, N(B_k)$ unabhängig

\Rightarrow Gegeben $N(B) = n$, ist das (geordnete) n -Tupel s_1, \dots, s_n eine i.i.d. Stichprobe aus der (bivariaten) Dichte

$$f_B(s) = \frac{\lambda(s)}{\int_B \lambda(u) du} = \frac{\lambda(s)}{\mu(B)}, \quad s \in B$$

\Rightarrow Bedingte Dichte von s_1, \dots, s_n gegeben $N(B) = n$

$$f((s_1, \dots, s_n) | N(B) = n) = \frac{\prod_{i=1}^n \lambda(s_i)}{(\mu(B))^n}$$

\Rightarrow Gemeinsame „Dichte“ von (s_1, \dots, s_n) und n ist

$$f((s_1, \dots, s_n), n) = \begin{cases} \exp(-\mu(B)), & n = 0 \\ \exp(-\mu(B)) \left[\prod_{i=1}^n \lambda(s_i) \right] / n!, & n \geq 1 \end{cases}$$

Summieren und Integrieren \Rightarrow

$$\begin{aligned} & \exp(-\mu(B)) + \sum_{n=1}^{\infty} \frac{\exp(-\mu(B))}{n!} \int_{B \times \dots \times B} \prod_{i=1}^n \lambda(s_i) ds_1 \dots ds_n = \\ & = \sum_{n=0}^{\infty} \frac{\exp(-\mu(B))}{n!} \left(\int_B \lambda(u) du \right)^n = \sum_{n=0}^{\infty} \frac{\exp(-\mu(B)) (\mu(B))^n}{n!} = 1 \end{aligned}$$

Aus 2. folgt für $s \neq u$

$$\begin{aligned} \lambda_2(s, u) &= \lim_{|ds|, |du| \rightarrow 0} \frac{\mathbb{E}[N(ds)N(du)]}{|ds| |du|} = \\ &= \lim_{|ds|, |du| \rightarrow 0} \frac{\mathbb{E}[N(ds)]\mathbb{E}[N(du)]}{|ds| |du|} = \lambda(s)\lambda(u) \end{aligned}$$

5.2.4 Simulation aus Poissonprozessen

- Charakteristika von räumlichen Prozessen sind oft schwer zu berechnen
- Simulation bietet Ausgangslage für Analyse komplexerer Punktprozessmodelle

Simulation eines homogenen Poissonprozesses

- (1) Ziehe $n \sim \text{Po}(|S|\lambda)$ (z.B. Inversionsverfahren)
- (2) Überlagere Realisationen von n unabhängigen Bernoulliprozessen

Simulation inhomogener Poissonprozesse

- Überlagerung:
 - $N_1(s), \dots, N_k(s)$ räumliche Punktprozesse
 - \Rightarrow Überlagerung $N(s) = \bigcup_{i=1}^k N_i(s)$ ist Punktprozess
 - Es gilt:
 - $N_1(s), \dots, N_k(s)$ unabhängige Poisson-Prozesse
 - mit Intensitäten $\lambda_1(s), \dots, \lambda_k(s) \Rightarrow$
 - $N(s) = \bigcup_{i=1}^k N_i(s)$ ist Poisson-Prozess mit Intensität $\lambda(s) = \sum_{i=1}^k \lambda_i(s)$
- Ausdünnung:
 - p -Ausdünnen: Jeder Punkt $s_i \in \{s_1, \dots, s_n\}$ wird mit W'keit p beibehalten und mit W'keit $1 - p$ eliminiert. Die $n = N(B)$ Bernoulliversuche sind dabei unabhängig.
 - $p(s)$ -Ausdünnen: Die W'keit für Beibehaltung („retention probability“) von Punkten s hängt von s ab, d.h. $0 \leq p(s) \leq 1$ ist deterministische Funktion.
 - $\pi(s)$ -Ausdünnen: $\pi(s)$ ist Zufallsfeld. Ausdünnen erfolgt durch Ziehen einer Realisierung $p(s)$ von $\pi(s)$ und anschließendem $p(s)$ - Ausdünnen.

Es gilt: $N(s)$ Poisson-Prozess mit Intensitätsrate $\lambda(s)$ wird $p(s)$ -ausgedünnt $\Rightarrow N_{dünn}(s)$ ist wieder Poisson-Prozess mit Intensitätsrate $\lambda_{dünn}(s) = \lambda(s)p(s)$, $N(s) \setminus N_{dünn}(s)$ ist P.P., mit Intensitätsrate $\lambda(s)(1 - p(s))$, und $N_{dünn}(s)$, $N(s) \setminus N_{dünn}(s)$ sind unabhängig.

Beweis: Møller / Waagepetersen, S. 23.

Folgerung: Ein inhomogener Poisson-Prozess auf S mit Intensität $\lambda(s)$ kann durch Ausdünnen eines homogenen Poisson-Prozesses mit Intensität $\lambda = \max_B \lambda(s)$ und $p(s) = \lambda(s)/\lambda$ generiert werden (vgl. Übung).

Beim inhomogenen Poisson-Prozess ergibt sich ein visueller Eindruck von Clustern in Regionen mit erhöhtem $\lambda(s)$.

5.3 Maßzahlen, explorative Analyse und Inferenz

Meist nonparametrische Schätzer und Tests für Maßzahlen und Hypothesen über Muster. In diesem Abschnitt exemplarische Auswahl.

5.3.1 Test auf rein zufälliges räumliches Muster

Nullhypothese: Complete Spatial Randomness, CSR \Leftrightarrow Muster ist Realisierung eines homogenen Poisson-Prozesses.

Zählmethoden Entspricht χ^2 - Anpassungstest.

Zerlege S (Rechteck) in $r \cdot c$ kleine Quadrate, $r =$ Anzahl Zeilen, $c =$ Anzahl Spalten. n_{ij} Anzahl von Ereignissen in Quadrat ij , $\bar{n} = \frac{n}{rc}$ erwartete Anzahl von Ereignissen unter der CSR-Hypothese in jedem Quadrat \Rightarrow

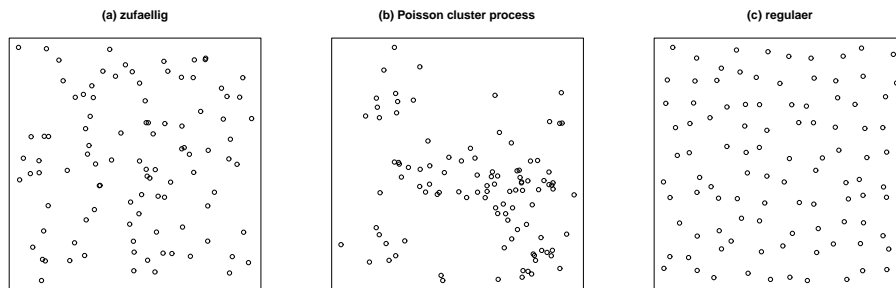
$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \bar{n})^2}{\bar{n}}$$

unter H_0 approximativ χ_{rc-1}^2 - verteilt.

Faustregel: $\bar{n} > 1$ und $rc > 6$

Beispiel 25. Simulierte Muster

$r = c = 5$, $n = 100 \Rightarrow \bar{n} = 4$



	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
5	4	3	4	7	2	7	1	0	1	2	2	4	3	2	3
4	3	5	5	3	4	0	1	3	2	2	5	4	5	4	2
3	5	7	5	6	3	5	9	4	14	10	3	4	3	3	4
2	2	2	6	3	2	1	0	1	8	2	5	5	2	5	4
1	4	3	5	4	3	5	1	0	3	9	4	5	2	4	4
$\chi^2 = 16.528$		$\chi^2 = 97.16854$					$\chi^2 = 7.843$								
$p = 0.868$		$p = 9.094e - 11$					$p = 0.999$								

Distanzmethoden

Manchmal sind Punktdaten (nur) durch den gegenseitigen Abstand gegeben (weil n groß), z.B. Baumlokationen in einem Wald.

Sei $D(r)$ der Nächste-Nachbar-Abstand. Für homogene Poisson-Prozesse gilt:

$$D(r) = 1 - \exp(-\lambda b_d r^d) \quad \text{für } r > 0.$$

Die rechte Seite ist theoretisch gleich $H_s(r)$.

Eine von Diggle (2003) vorgeschlagene Teststatistik ist daher

$$\int_0^t \left(\hat{H}_s(r) - \hat{D}(r) \right)^2 dr$$

Kritische Werte können nur über Simulation berechnet werden.

Alternative: J-Test

$$J(r) = \frac{1 - D(r)}{1 - H_s(r)}$$

Unter H_0 : Poisson-Prozess gilt: $J(r) = 1$

Test-Statistiken sind

$$\tau_{max} = \max_{r \leq s} \left| \hat{J}(r) - 1 \right|, \quad \tau_{int} = \int_0^s \left(\hat{J}(r) - 1 \right)^2 dr$$

Kritische Werte über Simulation.

5.3.2 Schätzung der Intensitätsfunktion

Parametrisch

$$\hat{\lambda} = \frac{N(B)}{|B|}$$

ist unverzerrter Maximum-Likelihood-Schätzer. Es gilt:

$$\text{Var}(\hat{\lambda}) = \frac{\lambda}{|B|},$$

daraus lässt sich (z.B. mit approximativer Normalverteilungsannahme) ein Konfidenzintervall angeben.

Nichtparametrisch

Sei $W \subseteq S$ das Fenster (window), in dem Ereignisse beobachtet wurden. Im inhomogenen Fall werden i.A. Kerndichteschätzer verwendet, z.B. mit $s_i = (x_i, y_i)$, $s = (x, y)$:

$$\hat{\lambda}(s) = \frac{1}{\nu(W)b_x b_y} \sum_{i=1}^n K\left(\frac{x_i - x}{b_x}\right) K\left(\frac{y_i - y}{b_y}\right)$$

mit univariatem Kern $K(\cdot)$ wie beim Dichteschätzen, Bandweiten b_x, b_y und bivariatem Produktkern $K(\cdot) K(\cdot)$; oder (Diggle, 1985) mit bivariatem Kern K und Randkorrektur sowie $b = b_x = b_y$:

$$\hat{\lambda}(s) = \frac{1}{p_b(s)} \sum_{i=1}^n \frac{1}{b^2} K\left(\frac{s_i - s}{b}\right), \quad p_b(s) = \int_W \frac{1}{b^2} K\left(\frac{s - u}{b}\right) du$$

Weitere bivariate Kerne z.B. in Fahrmeir, Hamerle, Tutz, „Multivariate Verfahren“, Kapitel 3.

Likelihood-Inferenz basierend auf inhomogenen Poissonprozessen

Sei $\lambda(s|\theta)$ durch θ parametrisiert, z.B. $\lambda(s|\theta) = \exp(z'\theta)$

„Dichte“ = Likelihood zu Daten $n = N(B)$, $\{s_1, \dots, s_n\}$ ist (bis auf Konstante)

$$L(\theta; B) = \prod_{i=1}^n \lambda(s_i | \theta) \exp \left\{ - \int_B \lambda(u | \theta) du \right\} = \prod_{i=1}^n \lambda(s_i | \theta) \exp(-\mu(B | \theta))$$

(vgl. Likelihood von unzensierten Lebensdauern).

Unter (etwas undurchsichtigen) Regularitätsbedingungen gilt:

$\hat{\theta}_{ML}$ asymptotisch normal, LQ- und Wald-Tests durchführbar.

Praktisches Problem: Algorithmische Umsetzung, $\int = ?$

z.B. bei bivariater Quadraturregel $\int_B \lambda(u | \theta) du \approx \sum_{j=1}^J w_j \lambda_j(u_j | \theta)$

mit Designpunkten $\{s_1, \dots, s_n\} \subseteq \{u_1, \dots, u_J\}$

$$l(\theta) = \sum_{i=1}^n \log \lambda(s_i | \theta) - \sum_{j=1}^J \lambda(u_j | \theta)$$

⇒ approx. log-Likelihood hat Form einer Poisson-log-Likelihood mit Gewichten.

⇒ GLM Software einsetzbar: Berman & Turner (1992), „Applied Statistics“; Baddeley & Turner (2000), Australian and N. Z. J. of Statistics; implementiert in spatstat R-Paket.

5.3.3 Maßzahlen 2. Ordnung für räumliche Interaktion

(a) Paar-Korrelationsfunktion (g -Funktion)

Als *pair correlation function* (pcf) bekannt, Alternative zu K-/L-Funktion. Wird z.B. von Illian et al. (2008) bevorzugt.

- Durch Normieren von λ_2 erhält man die Paar-Korrelationsfunktion

$$g(s, u) = \frac{\lambda_2(s, u)}{\lambda(s)\lambda(u)}$$

- Für einen Poisson-Prozess folgt aus Axiom 2: $g(s, u) = 1$.
- Aggregation (Clusterbildung) indiziert durch: $g(s, u) > 1$.
- Disaggregation (Regularität) durch: $g(s, u) < 1$.
- In der Regel geht man zu Abständen über: $g(s, u) = g(|s - u|)$.

(b) Ripley's K-Funktion für stationäre, isotrope Prozesse:

Für 2D:

$$K(h) = \frac{2\pi}{\lambda^2} \int_0^h x \lambda_2(x) dx, \quad h \geq 0, \quad x = \|s - u\|$$

Allgemein:

$$K(h) = \frac{\mathbb{E} \left[\sum_{i=1}^n \sum_{j=1, j \neq i}^n \frac{\mathbb{1}(\|s_i - s_j\| \leq h)}{\nu(W_{s_i} \cap W_{s_j})} \right]}{n(n-1)/\nu(W)^2}$$

Eigenschaften und Interpretation der K-Funktion:

- $\lambda K(h) = \mathbb{E}[h]$ = erwartete Anzahl von weiteren Ereignissen innerhalb eines Kreises mit Radius h um ein (beliebiges) Ereignis. Für den homogenen Poisson-Prozess gilt $K(h) = \pi h^2$.

$$\lambda_2(h) = \frac{\lambda^2}{2\pi h} \frac{dK(h)}{dh}, \quad \text{d.h. } K(h) \text{ determiniert } \lambda_2(h)$$

- Bei Clusterbildung ist zu erwarten, dass relativ viele weitere Punkte um ausgewählten Punkt liegen: $K(h)$ groß für h klein. Bei regulärem Muster: $K(h)$ klein für h klein.
- Ripley's Schätzer für $K(h)$, mit Punkten im Bereich B

$$\widehat{K}(h) = \frac{1}{\widehat{\lambda}} \widehat{\mathbb{E}}[h], \quad \widehat{\lambda} = \frac{N(B)}{|B|}$$

mit

$$\widehat{\mathbb{E}}[h] = \frac{1}{n} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \frac{\mathbb{1}(\|s_i - s_j\| \leq h)}{\nu(W_{s_i} \cap W_{s_j})}$$

Die pcf kann als Ableitung der K-Funktion interpretiert werden

$$g(r) = \frac{K'(r)}{d b_d r^{d-1}}, \quad d : \text{Dimensionen des Raums}$$

Schätzung über

- Glatte Schätzungen der Momente erster und zweiter Ordnung

$$\widehat{g}(s, u) = \frac{\widehat{\lambda}_2(s, u)}{\widehat{\lambda}(u)\widehat{\lambda}(s)}$$

- Ableitung der K-Funktion

$$\widehat{g}(r) = \frac{\widehat{K}'(r)}{db_d r^{d-1}}$$

(c) **L-Funktion**

$$L(h) = \sqrt{K(h)/\pi}$$

Für homogenen Poisson-Prozess: $L(h) = h$

Plots von $\widehat{L}(h) - h$ um Clusterbildung/Regularität zu entdecken:

$\widehat{L}(h) - h > 0$ indiziert für kleine h Clusterbildung

$\widehat{L}(h) - h < 0$ indiziert für kleine h Regularität

$\widehat{L}(h) = \sqrt{\widehat{K}(h)/\pi}$ Schätzer für L-Funktion

Test auf CSR möglich:

$$\tau = \max |\widehat{L}(h) - h|$$

Kritische Werte: $\tau_{0.05} = 1.45 \frac{\sqrt{a}}{n}$, $\tau_{0.01} = 1.75 \frac{\sqrt{a}}{n}$ mit $a = |W|$

5.3.4 Edge correction

Bei der Schätzung von Intensitäten, insbesondere aber auch bei der Schätzung der Abhängigkeiten zwischen Punkten (K-,L-, g -Funktion) sind Randeffekte zu beachten. Dies gilt besonders bei (relativ) kleinen oder stark irregulären Beobachtungsfenstern.

(a) **Grundideen:**

- Periodische Wiederholung
- Spiegelung
- Simulation
- Minus-Sampling
- Plus-Sampling (falls weitere Beobachtungen verfügbar)
- Nearest neighbour edge correction (translational oder isotrop)

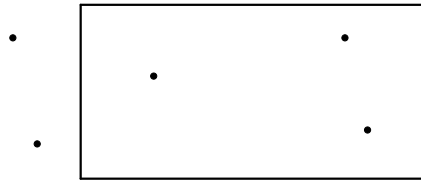
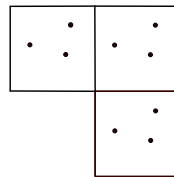


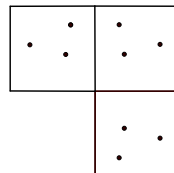
Abbildung 5.2: Edge-Korrektur.

Ideen:

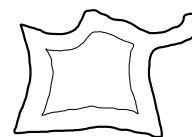
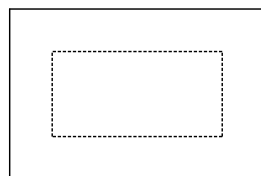
- **Plus-Sampling**
Zusätzliche Beobacht. außerhalb von W sind verfügbar (z.B. bei $D(r)$).
- **Keine Korrektur**
 W und n groß.
- **Periodisch**



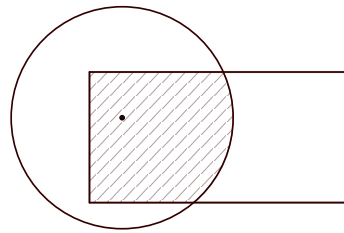
- **Reflection**



- **Simulation (bedingt)**
- **Minus-Sampling**



- **Nearest-Neighbour edge-correction** für $D(r)$
Gewichte Beobachtung mit Fläche des Kreises innerhalb des Fensters.

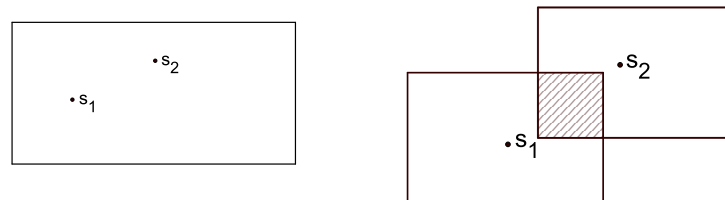


(b) **Second-order-edge-correction**

Bei Abhängigkeit von zwei Punkten K, L, *g*-Kurve:

- translational (Vorr.: Stationarität)

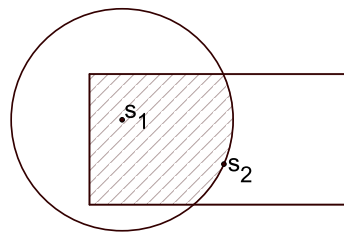
$$w = \frac{1}{\|w_{s_1} \cap w_{s_2}\|}$$



- isotrop (Vorr.: Stationarität + Isotropie)

$$w = \frac{1}{w(s_1, s_2)}$$

$$w(s_1, s_2) = \frac{\|b(s_1, \|s_1 - s_2\|) \cap w\|}{2\pi \|s_1 - s_2\|}$$



Bei *g*-Kurve: adaptiver Kern

$$\hat{g}(r) = \frac{1}{b_d r^{d-1} \|w\|} \sum_{i \neq j} \frac{K_b^*(r, \|s_i - s_j\|)}{\hat{\lambda}(s_i) \hat{\lambda}(s_j) \|w \cap w_{s_i - s_j}\|}$$

$$K_b^*(r, \|s_i - s_j\|) = K_{\min(b,r)}(r - \|s_i - s_j\|)$$

5.4 Markierte Punktprozesse

5.4.1 Multivariate Markierungen

		Abhängigkeit	
		<i>unabhängig</i>	<i>abhängig</i>
Dominanz	<i>symmetrisch</i>	unabhängig	abhängig
	<i>dominiert</i>	separiert	kontrolliert

Maßzahlen

- Clark-Evans aggregation index

$$CE_{12} = m_{D,12} \cdot 2\sqrt{\lambda_2}$$

mit $m_{D,12}$ mittlere Distanz von Punkten des Typs 1 zu nächstem Nachbar von Typ 2, und λ_2 Intensität der Punkte des Typs 2.

Beim Poisson-Prozess gilt:

$$\lambda = \lambda_1 + \lambda_2 \quad \Rightarrow \quad m_{D,12} = \frac{1}{2\sqrt{\lambda_2}} \quad \Rightarrow \quad CE_{12} = 1$$

Interpretation: $CE_{12} > 1 \Rightarrow$ Abstoßung (Repulsion)
 $CE_{12} < 1 \Rightarrow$ Anziehung (Attraction)

- Nearest-Neighbour-Tabelle

	Markierung des nächsten Nachbarn			
	1	2	Σ	
Markierung eines typischen Punktes	1	p_{11}	p_{12}	$p_{1\bullet}$
	2	p_{21}	p_{22}	$p_{2\bullet}$
	Σ	$p_{\bullet 1}$	$p_{\bullet 2}$	1

- Segregationskoeffizient

$$S = 1 - \frac{p_{12} + p_{21}}{p_{1\bullet}p_{\bullet 2} + p_{2\bullet}p_{\bullet 1}}$$

Bei Unabhängigkeit $S = 0$.

$S < 0 \Rightarrow$ Durchmischung; $S > 0 \Rightarrow$ Segregation

- **Mingling-Index** (Verfilzung)

$$M_k = \frac{1}{k} \mathbb{E} \left[\sum_{i=1}^k \mathbf{1}(m(o) \neq m(z_i(o))) \right]$$

Mittlerer Anteil von Punkten unter den k nächsten Nachbarn, deren Markierung unterschiedlich ist. Unter Unabhängigkeit gilt $M_k = 2p_1p_2$.

- **bivariate K, L- und g-Funktion**

$\lambda_j K_{ij}(r)$ erwartete Anzahl von Punkten des Typs j im Umkreis r von einem typischen Punkt des Typs i .

Daraus bivariate L-Funktionen und Cross-Pair-Correlation-Function.

- **Mark-Correlation-Function**

$$K_t(r) = \frac{C_t(r)}{C_t(\infty)} \text{ für } r > 0$$

mit $C_t(r) = \mathbb{E}_{or} [t(m(o), m(r))]$ und t eine Testfunktion, z.B.

- $t(m_1, m_2) = m_1 \cdot m_2$
- $t(m_1, m_2) = m_1$
- $t(m_1, m_2) = \frac{1}{2}(m_1 - m_2)^2$
- $t(m_1, m_2) = (m_1 - \mu_1)(m_2 - \mu_2)$

$K > 1 \Rightarrow$ „mutual stimulation“

$K < 1 \Rightarrow$ „inhibition“

5.4.2 Quantitative Markierungen

Statistiken für quantitative Markierungen basieren auf Gewichtung der Masszahlen für räumliche Interaktionen, z.B. der K-Funktion: Sei $m(s)$ die Markierung des Punktes s . Dann definiert man für einen typischen Punkt o

$$c_{mm}(r) = \mathbb{E}[m(o) \cdot m(R)],$$

wobei R alle Punkte mit Abstand r von o sind. Die *mark correlation function* ist definiert als

$$K_{mm}(r) = \frac{c_{mm}(r)}{\mu^2},$$

wobei $\mu^2 = \lim_{r \rightarrow \infty} c_{mm}(r)$ ist. Schätzer :

$$\widehat{K}_t(r) = \frac{\widehat{\rho}_t(r)}{\widehat{\rho}(r)\widehat{c}_t}$$

mit

$$\rho_t(r) = \sum_{u \neq s} \frac{m(u)m(s)K(\|u - s\| - r)}{db_d r^{d-1} |W_u \cap W_s|}$$

$$\rho(r) = \sum_{u \neq s} \frac{K(\|u - s\| - r)}{db_d r^{d-1} |W_u \cap W_s|}$$

$$\widehat{c}_t = \sum_{u,s} \frac{m(u)m(s)}{n^2}$$

5.5 Cluster-Modelle

Beispiele: Blitzeinschläge, Galaxien.

5.5.1 Poisson-Cluster-Prozesse

Idee: Ein Originalprozess („parent process“, Mutter-Prozess) erzeugt nach stochastischem Mechanismus einen Nachkommen-Prozess („offspring process“).

1. Der Mutterprozess ist ein inhomogener Poisson-Prozess mit Intensität $\lambda(s)$.
2. Jedes Mutterereignis s erzeugt eine zufäl. Anzahl $n(s)$ von Nachkommen.
3. Die Positionen der Nachkommen relativ zum Mutterereignis sind gemäß einer bivariaten Dichte $f(s)$ verteilt.
4. Die Nachkommen-Lokationen bilden den finalen Poisson-Cluster-Prozess.

Spezialfall: Neyman-Scott-Prozess

2. $\rightarrow 2^*$: Anzahl der Nachkommen jeder Mutter i.i.d. mit $\mathbb{P}(n(s) = k) = p_k$.
3. $\rightarrow 3^*$: Positionen der Nachkommen relativ zur Mutter i.i.d. verteilt.

Weitere Vereinfachung: Mutter-Prozess ist homogener Poisson-Prozess und $f(s)$ ist radial symmetrisch \Rightarrow Neyman-Scott-Prozess ist stationär mit (ohne Beweis)

$$\lambda_2(h) = \lambda^2 \mathbb{E}[n(s)]^2 + \lambda \mathbb{E}[n(s)(n(s) - 1)] f(h)$$

(h = euklidische Distanz zwischen Mutter und Nachkommen). Daraus kann die K-Funktion berechnet werden.

Beispiel: Matern-Cluster-Prozess

Stationärer Poisson-Prozess mit

$$f(s) \propto \text{const} \quad \text{für } \|s\| < R$$

Beispiel: Thomas-Prozess

$$f(s) \propto \exp\left(-\frac{1}{2\tau^2\|s\|}\right)$$

dann

$$K(h) = \pi^2 h + \frac{1}{\lambda} \left(1 - \exp\left(\frac{-h^2}{4\tau^2}\right)\right)$$

Allgemeiner: Shot-noise Cox-Prozesse.

5.5.2 Cox-Prozesse

Cox-Prozesse sind geeignete Modelle für aggregierte / geklumpete Punkt-Muster. Idee: Die Intensität $\lambda(s)$ eines inhomogenen Poisson-Prozesses ist nicht deterministisch, sondern ein Zufallsfeld (siehe Kapitel 2). Gegeben eine Realisierung $\lambda(s, \omega)$ ist ein Cox-Prozess ein inhomogener Poisson-Prozess. Anderer Name: Doppelt-stochastischer Poisson-Prozess.

Definition 5.7. (Cox-Prozess)

$\{\lambda(s), s \in S\}$ sei nichtnegatives Zufallsfeld mit lokal-integrierbaren Realisierungen $\lambda(s, \omega)$, d.h. $\int_B \lambda(s, \omega) < \infty$ für beschränkte Borelmengen B .

X heißt Cox-Prozess: \Leftrightarrow Der bedingte Prozess $X | \lambda(s)$, d.h. gegeben eine beliebige Realisierung $\lambda(s, \omega)$ von $\lambda(s)$, ist ein inhomogener Poisson-Prozess mit Intensität $\lambda(s) = \lambda(s, \omega)$.

Bemerkung: Damit wird $\mu(b) = \int_B \lambda(s) ds$ ebenfalls zu einer Zufallsvariablen.

Beispiele für Cox-Prozesse

- Gemischte Poisson-Prozesse („mixed P.p.“)

$\lambda(s) = \lambda_0$ ist eine (für alle s gleiche) positive Zufallsvariable, d.h. $X | \lambda_0$ ist ein homogener Poisson-Prozess mit Intensität λ_0 .

Vom Sonderfall $\lambda_0 = \text{const}$ abgesehen, sind $N(A)$ und $N(B)$ für disjunkte $A, B \subseteq S$ positiv korreliert, z.B.: $N(B) | \lambda \sim \text{Po}(\lambda(B))$, $\lambda \sim \text{Ga} \Rightarrow N(B)$ negativ-binomial-verteilt (Zusammenhang mit latenten Bayes-Modellen).
- Ausdünnen von Cox-Prozessen

X Cox-Prozess mit zufälliger Intensität $\lambda(s)$, $\{0 \leq \pi(s) \leq 1, s \in S\}$ Zufallsfeld. $\pi(s)$ -Ausdünnen von X ergibt einen Cox-Prozess $X_{\text{dünn}}$ mit zufälliger Intensität $\pi(s)\lambda(s)$, z.B.: $\pi(s)$ logistischer Prozess, d.h. $\pi(s) = \exp(Y(s)) / (1 + \exp(Y(s)))$, mit stat. Gauß-Prozess $Y(s)$, $\mathbb{E}[Y(s)] = 0$ und isotroper Kovarianzfunktion.
- Log-Gauß-Cox-Prozesse (LGCP)

$Y(s)$ (stationäres) Gauß-Feld. X heißt LGCP: $\Leftrightarrow X$ ist Cox-Prozess mit Intensität(-sfeld) $\lambda(s) = \exp(Y(s))$, d.h. $\log(\lambda(s))$ ist Gauß-Feld.

Bemerkung: In Anwendungen ist das Intensitätsfeld $\lambda(s)$ unbeobachtet (latent), d.h. falls nur ein realisiertes Muster $\{s_1, \dots, s_n\}$ beobachtet wird, kann der Cox-Prozess X nicht vom entsprechenden Poisson-Prozess $X | \lambda$ unterschieden werden. Welches der beiden Modelle geeigneter ist, d.h. ob λ als zufällig oder „deterministisch“ angesehen wird, hängt z.B. ab von

- spezifischer Fragestellung / etwa biologischem Modell
- Inferenzansatz: $\lambda(s)$ „zufällig“ kann als Bayes-Ansatz verstanden werden.

⇒ (Vorschlag:) Mixed-Model-Ansatz

$$\lambda(s) = \exp(\eta(s)) \quad , \quad \eta(s) = x(s)' \beta + W(s)$$

mit $\eta(s)$ wie bei generalisiertem Kriging, d.h. $x(s)' \beta$ systematische Komponente mit Kovariablen $x(s)$, W stationäres Gauß-Feld (später mehr).

Einige Eigenschaften von Cox-Prozessen

- Mit $\mu(B) = \int_B \lambda(s) ds$ ist

$$\mathbb{P}(N(B) = n) = \mathbb{E}_\lambda \left[\frac{1}{n!} (\mu(B))^n \exp(-\mu(B)) \right] \quad , \quad B \subseteq S$$

(mit $\mu(B)$ als Zufallsvariable, im Gegensatz zum inhomogenen Poisson-Prozess!)

$n = 0$ ergibt Null-Wahrscheinlichkeiten $\nu(B) = \mathbb{P}(N(B) = 0)$, $B \subseteq S$

- Intensitätsfunktionen: $\lambda_{\text{det}}(s) = \mathbb{E}[\lambda(s)]$, $\lambda_{\text{det},2}(s, u) = \mathbb{E}[\lambda_2(s, u)]$ sind die „üblichen, deterministischen“ Intensitätsfunktionen.
- Cox-Prozesse haben Überdispersion, d.h. $\text{Var}(N(B)) \geq \mathbb{E}[N(B)]$, und = nur für den entarteten Fall von Poisson-Prozessen.

$$\text{Cov}(N(A), N(B)) = \int_A \int_B \text{Cov}(\lambda(s), \lambda(u)) dsdu + \mu_{\text{det}}(A \cap B) \quad ,$$

mit $\mu_{\text{det}}(B) = \mathbb{E}[\int_B \lambda(s) ds] = \mathbb{E}[\mu(B)]$

$\lambda(s)$ stationär / isotrop ⇒ Cox-Prozess X stationär / isotrop.

Neyman-Scott-Prozesse als Cox-Prozesse

1. Sei C ein homogener Poisson-Prozess mit Intensität $\rho > 0$, der Clusterzentren $c = (c_1, c_2, \dots)$ in $S \subseteq \mathbb{R}^2$ erzeugt.
2. Für (eine Realisierung von) Clusterzentren seien N_{c_i} ($i = 1, 2, \dots$) unabhängige Poisson-Prozesse mit Intensitäten $\lambda_{c_i}(s) = \alpha K(s - c_i)$, $\alpha > 0$ Parameter, K Kernfunktion.

3. $X = \bigcup_{c_i} N_{c_i}$ sei Überlagerung der Poisson-Prozesse aus 2. \Rightarrow (ohne Beweis) X ist Cox-Prozess mit Intensitätsfunktion $\lambda(s) = \sum_{c_i} \alpha K(s - c_i)$.
 $\lambda(s)$ ist ein stationäres Feld, und isotrop falls $K(s) = K(\|s\|)$ isotrop.

Intensitäts- und Paarkorrelationsfunktion von LGCP $\{Y(s), s \in S\}$
 GRF mit $\mathbb{E}[Y(s)] = \mu(s)$, $\text{Cov}(Y(s), Y(u)) = c(s, u)$, $\lambda(s) = \exp(Y(s))$
 $\Rightarrow \lambda_{\text{det}}(s) = \exp(\mu(s) + c(s, s)/2)$, $g_{\text{det}}(s, u) = \exp(c(s, u))$

5.5.3 Nicht-stationäre Prozesse mit Interaktionen

Bezeichne $f(s_1, \dots, s_n) = f(\{s_1, \dots, s_n\}, n)$ die „Dichte“ eines Punktprozesses (in Bezug auf den homogenen Poisson-Prozess mit $\lambda = 1$).

Definition 5.8. (Prozess mit paarweiser Interaktion)

$b(s|\theta) \geq 0$, $s \in S$, Trendfunktion

$h(s, u|\theta) = h(u, s|\theta) \geq 0$, $u, s \in S$, Interaktionsfunktion

X heißt Punktprozess mit paarweiser Interaktion: \Leftrightarrow

$$f(s_1, \dots, s_n|\theta) = a(\theta) \prod_{i=1}^n b(s_i|\theta) \prod_{i<j} h(s_i, s_j|\theta)$$

$a(\theta) > 0$ Normierungskonstante, θ Parameter.

- Falls $b(s|\theta) \neq \text{const}$, impliziert die Funktion einen räumlichen Trend.
- $h(s, u|\theta)$ induziert Abhängigkeiten zwischen verschiedenen Punkten (Ereignissen) des Prozesses. Für $h(s, u|\theta) = 1$ ergibt sich ein inhomogener Poisson-Prozess mit Intensitätsfunktion $b(s|\theta)$.
- Die Normalisierungskonstante $a(\theta)$ ist i.A. eine analytisch unzugängliche Funktion von θ . Methoden zur numerischen Approximation: Taylorentwicklung, Monte-Carlo-Integration, EM-Typ-Approximationen.

Beispiel: Strauss-Prozess

Spezieller Interaktions-Prozess mit $b(s|\theta) = \beta$ und

$$h(s, u|\theta) = \begin{cases} \gamma & \text{für } \|s - u\| \leq r \\ 1 & \text{sonst} \end{cases}$$

$\beta > 0$, $0 \leq \gamma \leq 1$ Parameter, $r > 0$ „Interaktionsdistanz“

\Rightarrow „Dichte“ = Likelihood $L(\beta, \gamma) = \alpha(\beta, \gamma) \beta^n \gamma^m$

mit Normalisierungskonstante $\alpha(\beta, \gamma)$

und $m = \#\{(i, j) : i < j, \|s_i - s_j\| \leq r\}$ Anzahl von Punktepaaren mit einer Distanz kleiner / gleich r .

$\gamma = 1$: Homogener Poisson-Prozess

$\gamma = 0$: „Hard core“-Prozess: zwei Punkte haben mindestens die Distanz r bzw. es gibt keine zwei Punkte, die näher zueinander liegen

$0 < \gamma < 1$: erzeugt Regularität (Disaggregation, inhibition)

Beispiel: Gibbs-Prozess (n fest)

$$f(s_1, \dots, s_n) \propto \exp \left(-\alpha - \sum_{i \neq j} \Phi(\|s_i - s_j\|) \right)$$

α self potential, Φ pair potential (potentielle Energie durch Interaktion), z.B.

- $\Phi(r) = 0 \Rightarrow h(r) = 1 \Rightarrow$ Binomialprozess
- $\Phi(r) = 1/(\tau r^2) \Rightarrow$ Gauß-Verteilung
- $\Phi(r) = \gamma$ für $\|s_i - s_j\| \leq r \Rightarrow$ Strauss-Prozess

Simulation (n fest)

- Beginne auf regulärem Gitter
- Lösche zufälligen Punkt s_i (Gleichverteilung)
- Ziehe neuen Punkt \tilde{s}_i mit $f(\tilde{s}_i | s_{-i})$

Pseudolikelihood

Normalisierungskonstante $\alpha(\theta)$ mit Ausnahme des inhomogenen Poisson-Prozesses analytisch unzugänglich \Rightarrow ML-Schätzung von θ schwierig \Rightarrow „Pseudolikelihood“ (Besag, 1977)

- Bedingte Intensität (Papangelou-Intensität)

$$\lambda(u; s_1, \dots, s_n) = \frac{f(s_1, \dots, s_n, u)}{f(s_1, \dots, s_n)} \quad u \notin \{s_1, \dots, s_n\}$$

$$\lambda(s_i; s_1, \dots, s_n) = \frac{f(s_1, \dots, s_n)}{f(s_1, \dots, s_n \setminus s_i)} \quad s_i \in \{s_1, \dots, s_n\}$$

Informelle Interpretation: Bedingte Wahrscheinlichkeit, dass der Prozess einen Ereignispunkt u hat, gegeben die restlichen Ereignisse liegen an den Lokationen s_1, \dots, s_n .

- Ein inhomogener Poisson-Prozess mit Intensitätsfunktion $\lambda(s)$ hat die bedingte Intensitätsfunktion $\lambda(u; s_1, \dots, s_n) = \lambda(u)$ für alle $u \in S$.
- Ein Prozess mit paarweisen Interaktionen hat die bedingte Intensitätsfkt.

$$\lambda(u; s_1, \dots, s_n | \theta) = b(u | \theta) \prod_{\substack{i=1 \\ s_i \neq u}}^n h(u, s_i | \theta)$$

\Rightarrow Normalisierungskonstante $\alpha(\theta)$ eliminiert!

Definition 5.9. (Pseudolikelihood)

$\lambda(u; s_1, \dots, s_n | \theta)$ sei bedingte Intensität eines Punktprozesses $B \subseteq S$.

$$PL_B(\theta; s_1, \dots, s_n) = \prod_{s_i \in B} \lambda(s_i; s_1, \dots, s_n | \theta) \exp \left(- \int_B \lambda(u; s_1, \dots, s_n | \theta) du \right)$$

heißt Pseudolikelihood.

- Für einen inhomogenen Poisson-Prozess reduziert sich die Pseudolikelihood zur Likelihood.
- Für einen Prozess mit paarweiser Interaktion ist (im Fenster W)

$$PL_W(\theta; s_1, \dots, s_n) = \prod_{i=1}^n b(s_i | \theta) \prod_{i \neq j} h(s_i, s_j | \theta) \exp \left(- \int_W b(u | \theta) \prod_{i=1}^n h(u, s_i | \theta) du \right)$$

Normal.-Konstante $a(\theta)$ wird durch $\exp(-\dots)$ ersetzt, so als ob ein Poisson-Prozess vorläge. Approximation von \int nötig, z.B. durch Riemann-Summe.

Beispiel: Strauss-Prozess

Als bedingte Intensität ergibt sich $\lambda(u; s_1, \dots, s_n | \beta, \gamma) = \beta \gamma^{t(u; s_1, \dots, s_n)}$ mit $t(u; s_1, \dots, s_n) = \# \{s_i \in \{s_1, \dots, s_n\} : 0 < \|s_i - u\| \leq r\}$

Pseudolikelihood

$$PL_W(\beta, \gamma) = \beta^n \gamma^{2m} \exp \left(-\beta \int_W \gamma^{t(u; \bullet)} du \right)$$

- Modelle mit loglinearer bedingter Intensitätsfunktion
 $\lambda(u; s^*) = \exp(\theta' z(u; s^*))$, $s^* = (s_1, \dots, s_n)$
 $z(u; s^*)$ Vektor von räum. Kovariablen, die für jedes $u \in W$ definiert sind.

- Loglineare Modelle
 $b(u | \theta) = \exp(\theta' B(u))$, $h(u, v | \theta) = \exp(\theta' H(u, v))$ mit Vektoren $B(u)$, $H(u, v)$, die für alle $u, v \in W$ definiert sind.

$$\Rightarrow \lambda(u; s^* | \theta) = \exp\left(\theta' B(u) + \theta' \sum_{i=1}^n H(u, s_i)\right)$$

- MPLE $\hat{\theta}$ als Lösung der Pseudo-ML-Gleichungen

$$\frac{\partial}{\partial \theta} \log PL_W(\theta) = 0$$

Eigenschaften: Unter Regularitätsannahmen ist $\hat{\theta}$ asymptotisch normal, aber nicht mehr asymptotisch effizient.

Vermutung: ineffizient bei starker Interaktion (wegen Approximation als Poisson-Prozess).

- Berman-Turner-Approximation für MPL-Schätzung:
 Approximiere \int durch finite Summe mittels einer numerischen Quadraturregel, z.B. durch Riemann-Approximation auf einem Gitter in $W \Rightarrow$

$$\int_W \lambda(u; s^* | \theta) du \approx \sum_{j=1}^J \lambda(u_j; s^* | \theta) w_j, \quad J > n$$

u_j Designpunkte in W , w_j Quadraturgewichte, mit $\sum w_j = |W|$

$$\Rightarrow \log PL(\theta) \approx \sum_{i=1}^n \log \lambda(s_i; s^* | \theta) - \sum_{j=1}^J \lambda(u_j; s^* | \theta) w_j$$

Trick: Wähle Designpunkte $\{u_j, j = 1, \dots, J\}$ mit $J > n$ so, dass alle Datenpunkte $\{s_j, j = 1, \dots, n < J\}$ spezielle Designpunkte sind \Rightarrow

$$\log PL(\theta) \approx \sum_{j=1}^J (y_j \log \lambda_j - \lambda_j) w_j$$

mit $\lambda_j = \lambda(u_j, \bullet)$, $y_j = z_j/w_j$, und

$$z_j = \begin{cases} 1 & \text{falls } u_j \text{ Datenpunkt, d.h. } u_j \in \{s_1, \dots, s_n\} \\ 0 & \text{falls } u_j \text{ reiner Designpunkt, d.h. } u_j \notin \{s_1, \dots, s_n\} \end{cases}$$

Die rechte Seite hat die Form einer log-Likelihood eines loglinearen Poisson-Modells mit Gewichten w_j und unabhängigen Y_j .

⇒ Maximierung mit GLM-Software möglich, falls

1. gewichtete log-Likelihoods erlaubt,
2. nicht-ganzzahlige Werte für Y_j akzeptiert,
3. bedingte Intensitätsfunktion in loglinearer Form vorliegt.

Spezielle Modelle

- Soft Core Modell

$$b(u) \equiv \beta, \quad h(u, v) = \exp\left(-\left(\frac{\sigma}{\|u-v\|}\right)^{2/\kappa}\right), \quad u \neq v$$

$$\beta > 0, \quad 0 \leq \sigma < \infty, \quad 0 < \kappa < 1 \text{ (}\kappa \text{ vorgegeben)}$$

- Stückweise konstante Interaktion
 $h(u, v | \theta)$ stückweise konstant in $\|u - v\|$, etwa

$$\log h(u, v | \theta) = \begin{cases} \theta_l & \text{für } r_{l-1} < \|u - v\| \leq r_l \\ 0 & \text{für } \|u - v\| > r_k \end{cases}$$

auf Gitter $0 = r_0 < r_1 < r_2 < \dots < r_k$

5.6 Log-Gauß-Cox-Prozesse

- $X | \lambda(s)$ inhomogener Poissonprozess
- $\lambda(s) = \exp(Y(s))$
- $\{Y(s), s \in S\}$ stationäres Gauß-Feld
- Mit Vektor $z(s)$ von räumlichen Kovariablen: $\lambda(s) = \exp(z(s)' \beta + Y(s))$
- θ Vektor unbekannter Parameter ($\tau^2 = \text{Var}(Y(s))$, Parameter in Korrelationsfunktion) des GRF

5.6.1 Data-Augmentation-Ansatz für latentes GRF

Markov chain Monte Carlo-Algorithmus für latentes Gauß-Zufallsfeld Y . Für gegebenes $\{Y(s), s \in S\}$ ist die (bedingte) Likelihood $f(s_1, \dots, s_n | y(s), s \in S)$ die eines inhomogenen Poisson-Prozesses, d.h.

$$f(s_1, \dots, s_n | y, \beta, \theta) = \exp\left(|W| - \int_W \exp(z(s)' \beta + y(s)) ds\right) \cdot \prod_{i=1}^n \exp(z(s_i)' \beta + y(s_i))$$

W ist das Beobachtungsfenster.

Posteriori (mit Y als latentem Feld) wäre

$$p(\beta, \theta, y | s_1, \dots, s_n) \propto f(s_1, \dots, s_n | y, \beta) p(y | \theta) p(\beta) p(\theta)$$

Probleme: $\{Y(s), s \in S\}$ unendlich-dimensional,

$$\int_W \exp(z(s)' \beta + y(s)) ds = ?$$

Ein Lösungsansatz (vgl. Møller / Waagepetersen (2004), Ch. 8; Benes, Bodlak, Møller, Waagepetersen (2002) Techn. Report):

1. Approximiere $\{Y(s), s \in S\}$ durch endlich-dim. GRF $\{\tilde{Y}(s), s \in G\}$, G 2D-Gitter. Dabei gleiche Korrelationsfunktion.
2. Approximiere \int_W durch Riemann-Summe.

Damit: $p(\beta, \theta, \tilde{y} | s_1, \dots, s_n) \propto f(s_1, \dots, s_n | \tilde{y}, \beta) p(\tilde{y} | \theta) p(\beta) p(\theta)$

\Rightarrow MCMC mit MH-Schritten für Ziehen aus $p(\theta | \cdot)$, $p(\beta | \cdot)$, $p(\tilde{y} | \cdot)$

Details eines Lösungsvorschlags in Møller / Waagepetersen (2004), Ch. 8, und Benes et al. (2002).

Beispiel: Encephalitis in Böhmen (Benes et al.)

TBE tick-borne encephalitis (Hirnhautentzündung nach Zeckenstich)

446 Fälle, 1971-93 in Zentralböhmen. Nur 225 Lokationen sichtbar durch „ties“. Kovariablen: $z_1(s), \dots, z_5(s)$ Indikatorfunktionen für Regionen mit wenig / mittlerem Waldestand; Nadelwald / Mischwald / Laubwald; $z_6(s)$ = Höhe über dem Meeresspiegel. Um die unterschiedliche Bevölkerungsdichte zu berücksichtigen, wird folgendes Cox-Modell vorgeschlagen

$$\lambda(s) = \rho(s) \exp(z'(s)\beta + Y(s)) = \exp(\log \rho(s) + z'(s)\beta + Y(s)).$$

Dabei ist $\rho(s)$ die Bevölkerungsintensität vorausgeschätzt; $\log \rho(s)$ ist dann ein bekannter „Offset“ im Prädiktor. Der 2. Faktor ist die um die Bevölkerungsdichte bereinigte Intensität $\lambda_{\text{inf}}(s)$ für TBE-Infektionen. Für $Y(s)$ wird ein stationäres und isotropes GRF mit Exponentialkorrelationsfunktion

$$\rho(u) = \exp(-u/\alpha), \quad u = \|s - t\| \quad \text{und} \quad \text{Var}(Y(s)) = \sigma^2$$

angenommen.

5.6.2 Modellierung der Intensität über latentes GMRF

(Illian, Rue 2010)

Sei wieder $\lambda(s)$ ein Cox-Prozess mit $s \in S$.

Idee: Diskretisiere $S \Rightarrow$ GMRF. Gitter so klein, dass maximal ein Ereignis pro Pixel p_i . Es gilt

$$\lambda(s) \sim \text{GRF} \quad \Rightarrow \quad \lambda(s_i) \sim \text{GMRF}.$$

Zusätzlich: Konstruiere Kovariablen, die z.B. Abstoßung oder Anziehung von Punkten berücksichtigen (mit Dummyspunkten in „leeren“ Pixeln). Dann zerlege

$$\lambda(s) = \exp [Z(s) + U(s) + \beta_0 + \beta' z(s) + f(c(s))]$$

- Z und U strukturierte und unstrukturierte räumliche Komponenten
- $z(s)$ Kovariablen (z.B. Umwelteinflüsse)
- $f(c(s))$ glatte Funktion der konstruierten Kovariablen. z.B. Abstand zum nächsten Nachbarn, Abstand zum nächsten Clusterzentrum, vorher gesch. lokale Intensität.

Hierarchisches Bayes-Modell:

$$\begin{aligned} Y|\lambda &\sim \text{Po}(\lambda) \\ \log(\lambda) &= Z + U + \beta_0 I + \beta^T X + \gamma^T C \\ Z &\sim \text{N}(0, Q^{-1}) \\ U &\sim \text{N}(0, \sigma^2 \mathbb{I}) \\ p(\beta_0) &\propto \text{const} \\ p(\beta) &\propto \text{const} \\ \gamma &\sim \text{N}(0, K^{-1}) \end{aligned}$$

Inferenz z.B. über INLA oder Algorithmen für STAR-Modelle.

Beispiel: Cox Prozesse mit INLA

Beilschmiedia pendula Lauraceae, 3605 Lorbeer Bäume in Panama über 4 Jahre gezählt.

$$\lambda_i = \beta_0 + \beta_{ALT} \underbrace{\textit{Altitude}}_{\text{Höhe}} + \beta_{GRAD} \underbrace{\textit{Gradient}}_{\text{Gefälle}} + f_i^{(s)} + f_i^{(u)}$$

$$\begin{aligned} f_i^{(s)} &\sim \text{GMRF} && \text{second-order} \\ f_i^{(u)} &\stackrel{\text{i.i.d.}}{\sim} \text{N}(\cdot, \cdot) \end{aligned}$$

→ 4 Stunden Rechenzeit mit INLA.

Beispiel: Konstruiere Kovariablen

$$d_{np}(\underbrace{C(s_{ij})}_{\text{Zentraler Punkt}}) = \min_{\xi_i \in x} (\|C(s_{ij}) - \xi_i\|)$$

$$\eta_{ij} = \mu + f(d_{np}(C(s_{ij})))$$

entspricht sphärischer Kontakt-Verteilung.

d_{nc} : Cluster bestimmen, Abstand zum nächsten Cluster.

Beilschmiedia Daten

$$\eta_{ij} = f_s(s_{ij}) + u_{ij} + \beta_0 + \beta_1 \underbrace{z_{1ij}}_{ALT} + \beta_2 \underbrace{z_{2ij}}_{GRAD} + f(c_{ij})$$