

Geoadditive Regression

Jochen Wieland

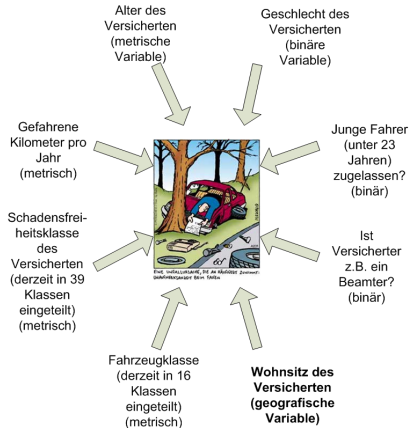
Seminar: Stochastische Geometrie und ihre Anwendungen - Zufallsfelder
Universität Ulm

27.01.2009

Inhalt

- 1 Einleitung
- 2 Modellbildung
- 3 P-Splines
 - B-Splines
 - Penalisierung
- 4 Kriging
 - Idee
 - Variogramm und Kovarianz
 - Gewöhnliches Kriging
 - Ansatz für das geoadditive Modell
- 5 Markov-Zufallsfelder
- 6 Zusammenfassung

Krafftfahrtversicherung: Mögliche Einflussgrößen auf die Schäden



...und viele weitere Einflussgrößen, die aktuell untersucht werden.

Ziel

Ein Regressionsmodell zu finden, das erlaubt alle relevanten Einflüsse angemessen zu berücksichtigen.

Herausforderungen

- Empirische Verteilung einzelner Einflüsse lassen erkennen, dass oft eine nichtlineare Abhängigkeit besteht
- Es wird eine Methode benötigt, die geografische Daten behandelt

Inhalt

- 1 Einleitung
- 2 Modellbildung
- 3 P-Splines
 - B-Splines
 - Penalisierung
- 4 Kriging
 - Idee
 - Variogramm und Kovarianz
 - Gewöhnliches Kriging
 - Ansatz für das geoadditive Modell
- 5 Markov-Zufallsfelder
- 6 Zusammenfassung

Das geoadditive Modell

Das geoadditive Modell vereint lineare, nichtlineare und räumliche Effekte. Wir betrachten also Regressionsdaten

$$(y_i, \mathbf{x}_i, \mathbf{z}_i, s_i), i = 1, \dots, n$$

wobei

- i der Beobachtungsindex
- y_i die Zielvariable
- \mathbf{x}_i ein Vektor mit linearen Einflussgrößen (darunter auch die binären Kovariablen)
- \mathbf{z}_i ein Vektor mit nichtlinearen Einflussgrößen
- s_i der Ort der Beobachtung

Das geoadditive Modell

und es soll gelten

$$y_i = \underbrace{\sum_{j=1}^m \beta_j x_{ij}}_{\text{linearer Einfluss}} + \underbrace{\sum_{j=1}^p f_j(z_{ij})}_{\text{nichtlinearer Einfluss}} + \underbrace{f_{geo}(s_i)}_{\text{räumlicher Einfluss}} + \epsilon_i,$$

wobei

- β_j die (unbekannten) Koeffizienten
- f_j (unbekannte) Funktionen
- f_{geo} der (unbekannte) räumliche Einfluss
- ϵ_i eine Störgröße.

Das geoadditive Modell

- Während die linearen Effekte durch ein multivariates lineares Modell geschätzt werden sollen, sollen die nichtlinearen Effekte durch möglichst glatte Funktionen mit Hilfe von P-Splines modelliert werden.
- Die räumlichen Effekte werden - abhängig davon, ob sie durch stetige oder diskrete Variablen beschrieben werden - durch Kriging bzw. durch Markov-Zufallsfelder modelliert.
- Ziel ist es aber letztlich, alle Einflüsse auf lineare Modelle zurückzuführen.

Inhalt

- 1 Einleitung
- 2 Modellbildung
- 3 **P-Splines**
 - B-Splines
 - Penalisierung
- 4 Kriging
 - Idee
 - Variogramm und Kovarianz
 - Gewöhnliches Kriging
 - Ansatz für das geoadditive Modell
- 5 Markov-Zufallsfelder
- 6 Zusammenfassung

B-Splines

P-Splines sind Polynom-Splines auf der Grundlage von B(asic)-Splines.

B-Splines werden aus Basisfunktionen gebildet, die folgendermaßen definiert sind:

Definition: B-Splines

Es sei ein reelles Intervall $[a, b]$ mit $d + 1$ Knoten $a = \kappa_1 < \dots < \kappa_{d+1} = b$ gegeben.

- B-Spline-Basisfunktionen vom Grad $l = 0$ sind definiert durch

$$B_j^0(z) = \mathbb{1}_{[\kappa_j, \kappa_{j+1})}(z) = \begin{cases} 1, & \text{wenn } z \in [\kappa_j, \kappa_{j+1}) \\ 0, & \text{sonst} \end{cases}, j = 1, \dots, d$$

- B-Spline-Basisfunktionen vom Grad $l = 1$ sind definiert durch

$$B_j^1(z) = \frac{z - \kappa_j}{\kappa_{j+1} - \kappa_j} \mathbb{1}_{[\kappa_j, \kappa_{j+1})}(z) + \frac{\kappa_{j+2} - z}{\kappa_{j+2} - \kappa_{j+1}} \mathbb{1}_{[\kappa_{j+1}, \kappa_{j+2})}(z)$$

Definition (Forts.)

- B-Spline-Basisfunktionen vom Grad $l > 1$ sind definiert durch

$$B_j^l(z) = \frac{z - \kappa_j}{\kappa_{j+l} - \kappa_j} B_j^{l-1}(z) + \frac{\kappa_{j+l+1} - z}{\kappa_{j+l+1} - \kappa_{j+1}} B_{j+1}^{l-1}(z)$$

B-Spline-Basisfunktionen

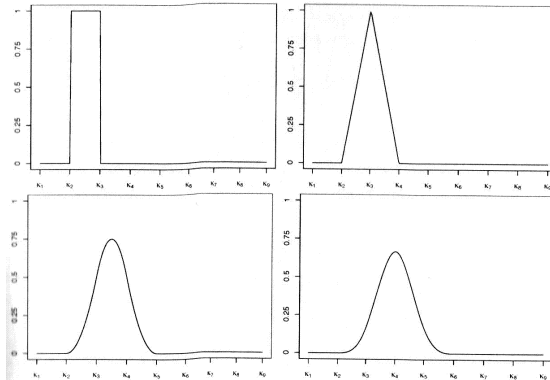


Abbildung: B-Spline-Basisfunktionen vom Grad $l = 0, 1, 2, 3$ (äquidistante Knoten)

Komplette B-Spline-Basis

Eine B-Spline-Basis vom Grad l ist entsprechend die Familie der einzelnen Basisfunktionen auf dem Intervall $[a, b]$.

Damit eine Basis des Grades l auf $[a, b]$ komplett ist, muss man jedoch zusätzlich $2l$ Knoten außerhalb des Intervalls $[a, b]$ und entsprechende Basisfunktionen definieren. Man hat also eine erweiterte Knotenfolge:

$$\kappa_{1-l} < \dots < a = \kappa_1 < \dots < \kappa_{d+1} = b < \dots < \kappa_{d+l+1}$$

Und die komplette B-Spline-Basis vom Grad l ist gegeben durch

$$(B'_{1-l}, \dots, B'_0, B'_1, \dots, B'_d)$$

B-Spline-Basen

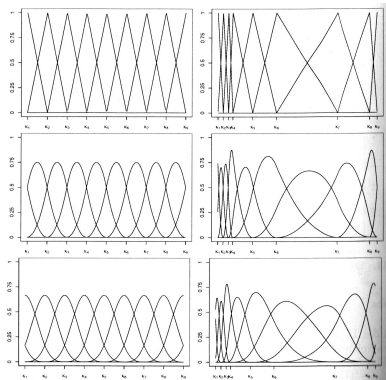


Abbildung: Vollständige B-Spline-Basen vom Grad $l = 1, 2, 3$ (links: äquidistante Knoten, rechts: ungleichmäßig verteilte Knoten)

Darstellung des Polynom-Spline

Gegeben sei eine B-Spline-Basis vom Grad l . Ein Polynom-Spline $f(z)$ ergibt sich dann als Linearkombination der Basisfunktionen:

$$f(z) = \sum_{j=1}^d \gamma_j B_j^l(z),$$

wobei γ_j (unbekannte) Koeffizienten zur Gewichtung der einzelnen Basisfunktionen B_j^l sind ($j = 1, \dots, d$).

Schätzung eines nichtlinearen Effekts

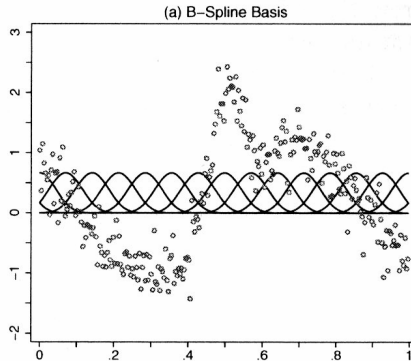


Abbildung: Schema: Schätzung eines nichtlinearen Effekts mit B-Splines

Schätzung eines nichtlinearen Effekts

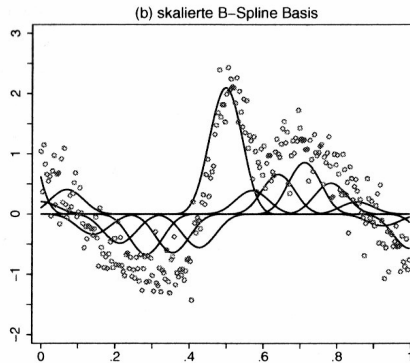


Abbildung: Schema: Schätzung eines nichtlinearen Effekts mit B-Splines

Schätzung eines nichtlinearen Effekts

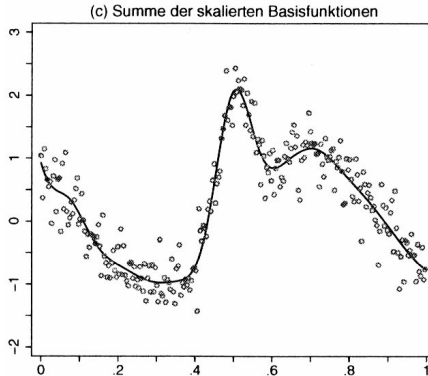


Abbildung: Schema: Schätzung eines nichtlinearen Effekts mit B-Splines

Schätzung der Koeffizienten

Somit lässt sich die Schätzung des Polynom-Spline auf die Schätzung eines linearen Modells (mit i.d.R. großer Parameterzahl) zurückführen.

Designmatrix:

$$Z = \begin{pmatrix} B_1^l(z_1) & \cdots & B_d^l(z_1) \\ \vdots & \ddots & \vdots \\ B_1^l(z_n) & \cdots & B_d^l(z_n) \end{pmatrix}$$

Minimierungsproblem:

$$(y - Z\gamma)^T (y - Z\gamma) \rightarrow \min$$

Normalengleichung:

$$Z^T Z \gamma = Z^T y$$

KQ-Schätzer (sofern $Z^T Z$ invertierbar):

$$\hat{\gamma} = (Z^T Z)^{-1} Z^T y$$

Eigenschaften von B-Splines

Gegeben sei eine B-Spline-Basis vom Grad l mit der erweiterten Knotenfolge

$$\kappa_{1-l} < \dots < a = \kappa_1 < \dots < \kappa_{d+1} = b < \dots < \kappa_{d+l+1}.$$

Eigenschaften von B-Splines

Gegeben sei eine B-Spline-Basis vom Grad l mit der erweiterten Knotenfolge

$$\kappa_{1-l} < \dots < a = \kappa_1 < \dots < \kappa_{d+1} = b < \dots < \kappa_{d+l+1}.$$

- Beschränktheit: Die Basisfunktionen sind beschränkt.
⇒ Berechnung numerisch stabiler.

Eigenschaften von B-Splines

Gegeben sei eine B-Spline-Basis vom Grad l mit der erweiterten Knotenfolge

$$\kappa_{1-l} < \dots < a = \kappa_1 < \dots < \kappa_{d+1} = b < \dots < \kappa_{d+l+1}.$$

- Beschränktheit: Die Basisfunktionen sind beschränkt.
⇒ Berechnung numerisch stabiler.
- Differenzierbarkeit: Die Basisfunktionen (und somit auch der Polynom-Spline) sind $(l - 1)$ -mal stetig differenzierbar, denn es gilt:

$$\frac{\partial}{\partial z} B_j^l(z) = l \cdot \left(\frac{1}{\kappa_{j+l} - \kappa_j} B_j^{l-1}(z) - \frac{1}{\kappa_{j+l+1} - \kappa_{j+1}} B_{j+1}^{l-1}(z) \right)$$

Eigenschaften von B-Splines (Forts.)

- B-Splines bilden lokale Basis: Jede Basisfunktion ist nur in einem von $l + 2$ Knoten gebildeten Bereich positiv. An einer beliebigen Stelle $z \in [a, b]$ sind genau $l + 1$ Basisfunktionen positiv.
 \implies Designmatrix Z dünn besetzt \implies Normalengleichung numerisch effizient lösbar.

Eigenschaften von B-Splines (Forts.)

- B-Splines bilden lokale Basis: Jede Basisfunktion ist nur in einem von $l + 2$ Knoten gebildeten Bereich positiv. An einer beliebigen Stelle $z \in [a, b]$ sind genau $l + 1$ Basisfunktionen positiv.
 \implies Designmatrix Z dünn besetzt \implies Normalengleichung numerisch effizient lösbar.
- Überlappung: Jede Basisfunktion überlappt sich innerhalb $[a, b]$ mit genau $2l$ benachbarten Basisfunktionen.

Eigenschaften von B-Splines (Forts.)

- B-Splines bilden lokale Basis: Jede Basisfunktion ist nur in einem von $l + 2$ Knoten gebildeten Bereich positiv. An einer beliebigen Stelle $z \in [a, b]$ sind genau $l + 1$ Basisfunktionen positiv.
 \implies Designmatrix Z dünn besetzt \implies Normalengleichung numerisch effizient lösbar.
- Überlappung: Jede Basisfunktion überlappt sich innerhalb $[a, b]$ mit genau $2l$ benachbarten Basisfunktionen.
- Zerlegung der Einheit: An jeder Stelle $z \in [a, b]$ und für jedes $l \geq 0$ gilt $\sum_{j=1-l}^d B_j^l(z) = 1$.

Beweis der Zerlegung der Einheit

Vollständige Induktion:

Für $l = 0$ gilt:

$$\sum_{j=1}^d B_j^0(z) = \mathbb{1}_{[\kappa_1, \kappa_2)}(z) + \mathbb{1}_{[\kappa_2, \kappa_3)}(z) + \dots + \mathbb{1}_{[\kappa_d, \kappa_{d+1})}(z) = 1$$

Gelte $\sum_{j=1-l}^d B_j^l(z) = 1$ für ein $l \geq 0$. Für $l+1$ gilt dann:

$$\sum_{j=1-(l+1)}^d B_j^{l+1}(z) = \sum_{j=-l}^d B_j^{l+1}(z)$$

Beweis (Forts.)

=

$$\begin{aligned}
 & \frac{Z - \kappa_{-l}}{\kappa_{-l+(l+1)} - \kappa_{-l}} B_{-l}^l(z) + \frac{\kappa_{-l+(l+1)+1} - Z}{\kappa_{-l+(l+1)+1} - \kappa_{-l+1}} B_{-l+1}^l(z) + \\
 & \frac{Z - \kappa_{-l+1}}{\kappa_{-l+(l+2)} - \kappa_{-l+1}} B_{-l+1}^l(z) + \frac{\kappa_{-l+(l+1)+2} - Z}{\kappa_{-l+(l+1)+2} - \kappa_{-l+2}} B_{-l+2}^l(z) + \\
 & \quad + \dots + \\
 & \frac{Z - \kappa_d}{\kappa_{d+(l+1)} - \kappa_d} B_d^l(z) + \frac{\kappa_{d+(l+1)+1} - Z}{\kappa_{d+(l+1)+1} - \kappa_{d+1}} B_{d+1}^l(z)
 \end{aligned}$$

Beweis (Forts.)

$$\begin{aligned}
 &= \\
 &\quad \frac{Z - \kappa_{-l}}{\kappa_1 - \kappa_{-l}} B_{-l}^l(z) + \\
 &\quad \underbrace{\frac{\kappa_2 - Z}{\kappa_2 - \kappa_{-l+1}} B_{-l+1}^l(z) + \frac{Z - \kappa_{-l+1}}{\kappa_2 - \kappa_{-l+1}} B_{-l+1}^l(z)}_{=B_{-l+1}^l(z)} + \underbrace{\frac{\kappa_3 - Z}{\kappa_3 - \kappa_{-l+2}} B_{-l+2}^l(z) + \frac{Z - \kappa_{-l+2}}{\kappa_3 - \kappa_{-l+2}} B_{-l+2}^l(z)}_{=B_{-l+2}^l(z)} \\
 &\quad + \frac{\kappa_4 - Z}{\kappa_4 - \kappa_{-l+3}} B_{-l+3}^l(z) \\
 &\quad + \dots + \\
 &\quad \frac{Z - \kappa_d}{\kappa_{d+l+1} - \kappa_d} B_d^l(z) + \frac{\kappa_{d+l+2} - Z}{\kappa_{d+l+2} - \kappa_{d+1}} B_{d+1}^l(z)
 \end{aligned}$$

Beweis (Forts.)

=

$$\begin{aligned}
 & \frac{z - \kappa_{-l}}{\kappa_1 - \kappa_{-l}} \underbrace{B_{-l}^l(z)}_{=0 \ \forall z \notin [\kappa_{-l}, \kappa_1]} + \sum_{j=-l+1}^d B_j^l(z) + \frac{\kappa_{d+l+2} - z}{\kappa_{d+l+2} - \kappa_{d+1}} \underbrace{B_{d+1}^l(z)}_{=0 \ \forall z \notin [\kappa_{d+1}, \kappa_{d+l+2}]} \\
 & = \sum_{j=-l+1}^d B_j^l(z) = 1
 \end{aligned}$$

Unzulänglichkeit der B-Splines

Die Genauigkeit der Schätzung nichtlinearer Effekte durch B-Splines hängt stark von der Anzahl der Knoten ab. Ein Problem ist dabei, dass

- bei zu wenigen Knoten der Spline die Daten unzureichend annähert
- bei zu vielen Knoten der Spline zu variabel wird und keine sinnvolle Interpretation ermöglicht.

Unzulänglichkeit der B-Splines

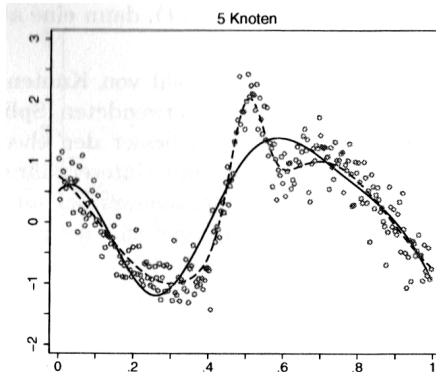


Abbildung: Einfluss der Knotenzahl auf die Schätzung - 5 Knoten

Unzulänglichkeit der B-Splines

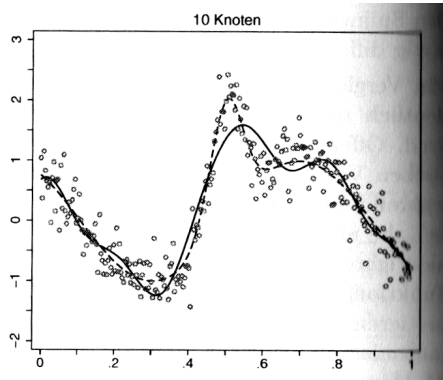


Abbildung: Einfluss der Knotenzahl auf die Schätzung - 10 Knoten

Unzulänglichkeit der B-Splines

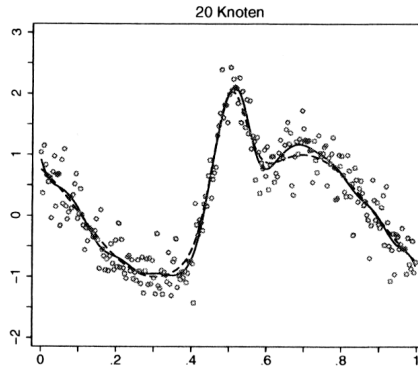


Abbildung: Einfluss der Knotenzahl auf die Schätzung - 20 Knoten

Unzulänglichkeit der B-Splines

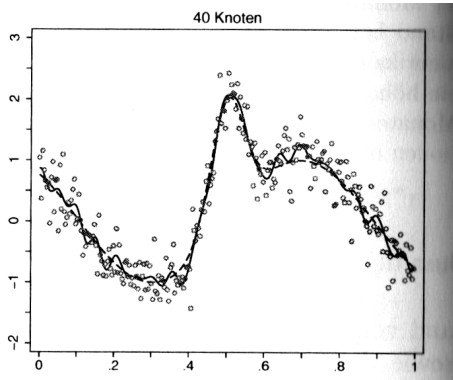


Abbildung: Einfluss der Knotenzahl auf die Schätzung - 40 Knoten

Unzulänglichkeit der B-Splines

Die Genauigkeit der Schätzung nichtlinearer Effekte durch B-Splines hängt stark von der Anzahl der Knoten ab. Ein Problem ist dabei, dass

- bei zu wenigen Knoten der Spline die Daten unzureichend annähert
- bei zu vielen Knoten der Spline zu variabel wird und keine sinnvolle Interpretation ermöglicht.

⇒ Lösung: Man fügt eine Penalisierung hinzu.

Idee der Penalisierung

Die grundlegende Idee von penalisierten Splines (P-Splines) ist dazu:

- Einen Polynom-Spline mit einer großen Knotenzahl zur Schätzung zu verwenden, womit der Spline prinzipiell flexibel genug ist, auch stark variierende Funktionen darzustellen.
- Einen zusätzlichen Strafterm einzuführen, der eine zu große Variabilität der Schätzung bestraft.

Strafterm

Betrachtet man nochmals die Darstellung des Polynom-Splines $f(z) = \sum_{j=1}^d \gamma_j B_j^l(z)$, ist ersichtlich, dass es sinnvoll ist, zu große Differenzen benachbarter Parameter γ_j zu bestrafen. Dies lässt sich auch aus den Ableitungen von f herleiten, da aufgrund der Eigenschaften der B-Splines gilt:

$$\frac{\partial}{\partial z} f(z) = \frac{\partial}{\partial z} \sum_{j=1}^d \gamma_j B_j^l(z) = l \cdot \sum_{j=1}^d \frac{\gamma_j - \gamma_{j-1}}{\kappa_{j+l} - \kappa_j} B_j^{l-1}(z)$$

Strafterm

Daraus und aus den höheren Ableitungen erhält man:

- Differenzen 1. Ordnung: $\Delta^1 \gamma_j = \gamma_j - \gamma_{j-1}$
- Differenzen 2. Ordnung:
$$\Delta^2 \gamma_j = \Delta^1 \gamma_j - \Delta^1 \gamma_{j-1} = \gamma_j - 2\gamma_{j-1} + \gamma_{j-2}$$
- Allgemein Differenzen k-ter Ordnung:
$$\Delta^k \gamma_j = \Delta^{k-1} \gamma_j - \Delta^{k-1} \gamma_{j-1}$$

Strafterm

Der Strafterm hat dann die Darstellung:

$$\lambda \sum_{j=k+1}^d (\Delta^k \gamma_j)^2 = \lambda \gamma^T D_k^T D_k \gamma = \lambda \gamma^T K_k \gamma,$$

wobei

- $\lambda \geq 0$ der Glattheitsparameter, der das Verhältnis von Datentreue zu Glattheit steuert,
- $K_k = D_k^T D_k$ die Strafmatrix und D_k die Differenzen k -ter Ordnung zusammenfasst.

Strafterm

In den Fällen $k = 1$ und $k = 2$ ist D_k eine $(d - 1) \times d$ -Matrix bzw. eine $(d - 2) \times d$ -Matrix mit folgender Gestalt:

$$D_1 = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix}, D_2 = \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{pmatrix}$$

Penalisierter KQ-Schätzer

Aus dem bekannten Minimierungsproblem des KQ-Schätzers

$$(y - Z\gamma)^T(y - Z\gamma) \rightarrow \min$$

wird

$$(y - Z\gamma)^T(y - Z\gamma) + \lambda\gamma^T K\gamma \rightarrow \min$$

Normalengleichung:

$$(Z^T Z + \lambda K)\gamma = Z^T y$$

Penalisierter KQ-Schätzer (falls $(Z^T Z + \lambda K)$ invertierbar):

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

Herleitung der Normalengleichung

Es gilt:

$$\begin{aligned} & \frac{d}{d\gamma}((y - Z\gamma)^T(y - Z\gamma) + \lambda\gamma^T K\gamma) \\ &= \frac{d}{d\gamma}(y^T y - y^T Z\gamma - (Z\gamma)^T y + \gamma^T(Z^T Z + \lambda K)\gamma) \\ &= -2Z^T y + 2(Z^T Z + \lambda K)\gamma = \mathbf{0} \\ &\Leftrightarrow (Z^T Z + \lambda K)\gamma = Z^T y \end{aligned}$$

Außerdem $\frac{d^2}{d^2\gamma}((y - Z\gamma)^T(y - Z\gamma) + \lambda\gamma^T K\gamma) = 2(Z^T Z + \lambda K)$
positiv definit bei vollem Rang.

Inhalt

- 1 Einleitung
- 2 Modellbildung
- 3 P-Splines
 - B-Splines
 - Penalisierung
- 4 Kriging**
 - Idee
 - Variogramm und Kovarianz
 - Gewöhnliches Kriging
 - Ansatz für das geoadditive Modell
- 5 Markov-Zufallsfelder
- 6 Zusammenfassung

Idee

Das Verfahren wurde entwickelt von dem südafrikanischem Geostatistiker Daniel Krige. Gegeben sei

- ein zufälliges Feld $\{Z(s), s \in \mathbb{R}^2\}$, also eine Familie von (reellen) Zufallsvariablen $Z(s)$ über einem gemeinsamen Wahrscheinlichkeitsraum
- eine endliche Menge von Beobachtungspunkten $\{s_1, \dots, s_d\}$.

Ziel des Kriging ist, mittels Kovarianzen und geeigneter Gewichtung der Beobachtungspunkte, einen Schätzer für den (unbekannten) Wert an einer Stelle s_0 zu konstruieren.

Definition: Variogramm

Die räumliche Variation einer Zufallsvariablen $Z(s)$ kann durch den „Zuwachs“ $Z(s+h) - Z(s)$ zwischen den Werten zweier Punkte $s, s+h \in \mathbb{R}^2$ beschrieben werden.

$\gamma(h) := \frac{1}{2} \mathbb{E}[(Z(h) - Z(0))^2]$, $\forall h \in \mathbb{R}^2$ heißt (theoretisches) Variogramm, falls für jedes Paar $s, s+h \in \mathbb{R}^2$ gilt:

- $\mathbb{E}[Z(s+h) - Z(s)] =: m(h) \equiv 0$, d.h. Erwartungswert der Zuwächse konstant,
- $\text{Var}[Z(s+h) - Z(s)] = 2\gamma(h) < \infty$, d.h. Varianz der Zuwächse hängt nur von Länge und Richtung des Vektors zwischen den 2 Punkten ab.

(Intrinsische Stationarität 2. Ordnung)

Eigenschaften

- $\gamma(\mathbf{o}) = 0$
- $\gamma(h) \geq 0$
- $\gamma(-h) = \gamma(h)$
- $\lim_{\|h\| \rightarrow \infty} \frac{\gamma(h)}{\|h\|^2} = 0$ (sonst Widerspruch zu $m(h) \equiv 0$)
- Falls $\gamma(h)$ nur von $\|h\|$ abhängt, d.h. Varianz der Zuwächse hängt nur von der Länge von h ab, so heißt $\{Z(s), s \in \mathbb{R}^2\}$ isotrop.

Experimentelles Variogramm

- $\{h : h = (s_i - s_j) \in \mathbb{R}^2, 1 \leq i < j \leq d\}$ seien die Differenzenvektoren zwischen den Beobachtungspunkten $\{s_1, \dots, s_d\}$
- Bilden von k Klassen H_k der Vektoren h nach ihrer Länge und Richtung
- experimentelles Variogramm:
$$\gamma^*(H_k) = \frac{1}{2n_k} \sum_{h=(s_i-s_j) \in H_k} (z(s_i) - z(s_j))^2$$
 mit n_k die Anzahl der Vektoren in der Klasse H_k
- Das experimentelle Variogramm dient zur Strukturierung der Daten, muss aber für das weitere Vorgehen durch ein geeignetes theoretisches ersetzt werden.

Experimentelles Variogramm

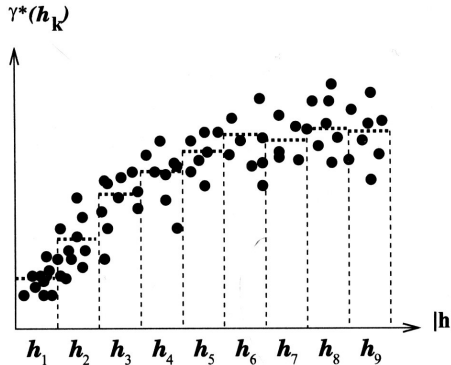


Abbildung: Experimentelles Variogramm - h nur nach Länge klassifiziert, unter Annahme von Isotropie

Passendes theoretisches Variogramm

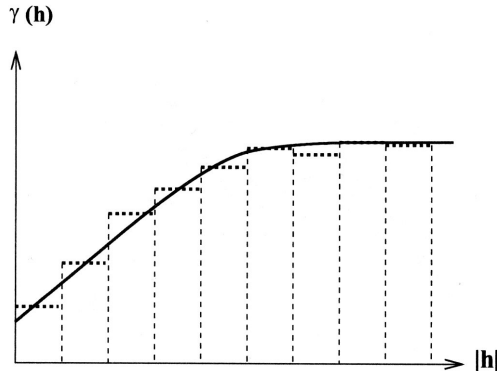


Abbildung: Passendes theoretisches Variogramm wird über das experimentelle gelegt

Definition: Kovarianzfunktion

Falls in $\{Z(s), s \in \mathbb{R}^2\}$ $\mathbb{E}[Z(s)], \mathbb{E}[Z^2(s)] < \infty$ und die Stationarität der ersten beiden Momente gilt, d.h. falls

- $\mathbb{E}[Z(s)] \equiv m$ konstant $\forall s \in \mathbb{R}^2$
- $\mathbb{E}[Z(s)Z(s+h)] - m^2 =: C(h)$ nur abhängig von h
 $\forall s, s+h \in \mathbb{R}^2$,

dann heißt $C(h)$ eine Kovarianzfunktion.

Zusammenhang $\gamma(h)$ und $C(h)$

Ein Variogramm kann von einer Kovarianzfunktion abgeleitet werden, da gilt

$$\gamma(h) = C(0) - C(h) \quad \forall h.$$

Definitheit

Eine Kovarianzfunktion ist positiv definit, da

$$\text{Var} \left[\sum_{i=0}^n \omega_i Z(s_i) \right] = \sum_{i=0}^n \sum_{j=0}^n \omega_i \omega_j C(s_i - s_j) \geq 0$$

für jede Familie von Punkten (s_0, \dots, s_n) und Gewichten $(\omega_0, \dots, \omega_n)$.
Daraus folgt, dass ein Variogramm eine bedingt negativ definite Funktion ist, unter der Bedingung $\sum_{i=0}^n \omega_i = 0$:

$$\begin{aligned} \text{Var} \left[\sum_{i=0}^n \omega_i Z(s_i) \right] &= \sum_{i=0}^n \sum_{j=0}^n \omega_i \omega_j (C(0) - \gamma(s_i - s_j)) \\ &= C(0) \sum_{i=0}^n \omega_i \underbrace{\sum_{j=0}^n \omega_j}_{=0} - \sum_{i=0}^n \sum_{j=0}^n \omega_i \omega_j \gamma(s_i - s_j) \geq 0 \end{aligned}$$

Beispiele für Kovarianzfunktionen

- Isotrope Kovarianzfunktionen, d.h. $C(h)$ hängt nur von $\|h\|$, also nur der Länge von h , ab
- Exponentielle Kovarianzfunktion
 - $C_{exp}(h) = b \exp(-\frac{\|h\|^c}{a})$, $a, b > 0, 0 < c \leq 2$
 - a und c geben an, wie schnell die Kovarianz mit dem Abstand vom Ursprung fällt
 - b gibt den max. Wert der Kovarianz (am Ursprung) an
 - am Ursprung stetig, aber nicht differenzierbar
 - $C_{exp}(h) \rightarrow 0$ ($\|h\| \rightarrow \infty$)

Beispiele für Kovarianzfunktionen

- Sphärische Kovarianzfunktion

- $$C_{sph}(h) = \begin{cases} b(1 - \frac{3}{2} \frac{\|h\|}{a} + \frac{1}{2} \frac{\|h\|^3}{a^3}), & \text{für } 0 \leq \|h\| \leq a \\ 0, & \|h\| > a \end{cases} \quad a, b > 0$$
- a (range) gibt die Reichweite der Kovarianz an
- b (sill) gibt den max. Wert der Kovarianz (am Ursprung) an
- stetig, aber nicht differenzierbar
- monoton fallend für $0 \leq \|h\| \leq a$, dann verschwindend

Beispiele für Kovarianzfunktionen

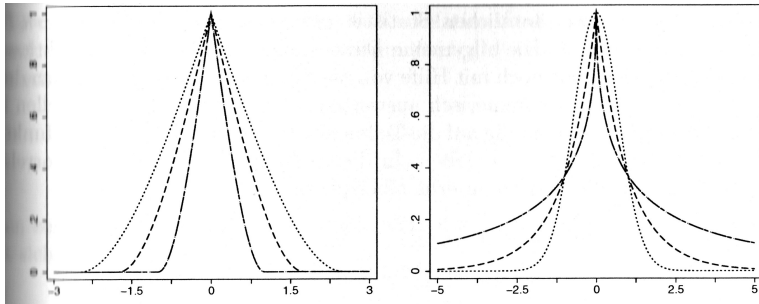


Abbildung: Beispiele für Kovarianzfunktionen (links: sphärische für $b = 1$ und $a = 1, 1.75, 2.5$, rechts: exponentielle für $a, b = 1$ und $c = 0.5, 1, 2$)

Beispiele für Kovarianzfunktionen

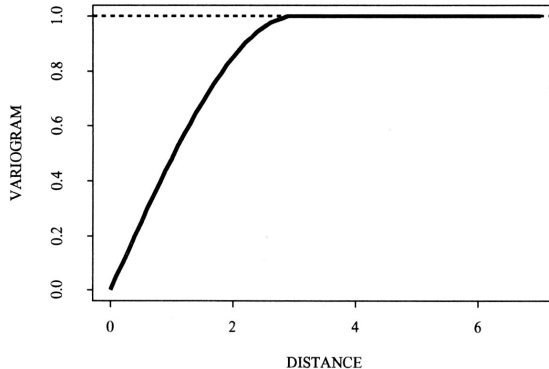


Abbildung: Sphärisches Variogramm für $b = 1$ und $a = 3$

Beispiele für Kovarianzfunktionen

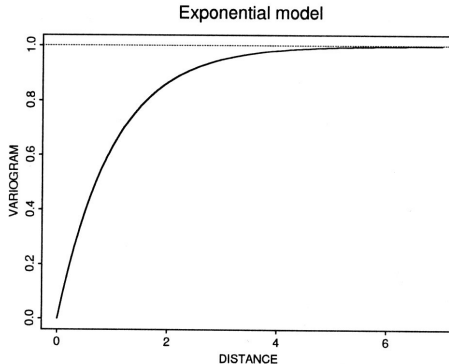


Abbildung: Exponentielles Variogramm für $a, b, c = 1$

Gewöhnliches Kriging

- Es soll nun der Wert am Punkt s_0 durch eine gewichtete Linearkombination der Werte an den Beobachtungspunkten geschätzt werden:

$$\hat{Z}(s_0) = \sum_{i=1}^d \omega_{i,s_0} Z(s_i)$$

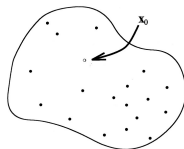


Abbildung: Gebiet mit unregelmäßig verteilten Beobachtungspunkten und s_0

Gewöhnliches Kriging

- Wir nehmen an, dass die Daten eine Realisation der Zufallsvariablen $Z(s)$ mit Variogramm $\gamma(h)$ sind.
- Unter der Voraussetzung $\sum_{i=1}^d \omega_i = 1$ erhalten wir einen erwartungstreuen Schätzer:

$$\begin{aligned}\mathbb{E}[\hat{Z}(s_0) - Z(s_0)] &= \mathbb{E}\left[\sum_{i=1}^d \omega_{i,s_0} Z(s_i) - Z(s_0) \underbrace{\sum_{i=1}^d \omega_{i,s_0}}_{=1}\right] \\ &= \sum_{i=1}^d \omega_{i,s_0} \mathbb{E}[Z(s_i) - Z(s_0)] = 0\end{aligned}$$

Gewöhnliches Kriging

- Die Schätzvarianz ist:

$$\begin{aligned}\hat{\sigma}^2 &= \mathbb{E}[(\hat{Z}(s_0) - Z(s_0))^2] \\ &= -\gamma(s_0 - s_0) - \sum_{i=1}^d \sum_{j=1}^d \omega_{i,s_0} \omega_{j,s_0} \gamma(s_i - s_j) - 2 \sum_{i=1}^d \omega_{i,s_0} \gamma(s_i - s_0)\end{aligned}$$

- Die 1. Ableitung:

$$\frac{d}{d\omega_{i,s_0}} \hat{\sigma}^2 = -2 \sum_{j=1}^d \omega_{j,s_0} \gamma(s_i - s_j) - 2\gamma(s_i - s_0)$$

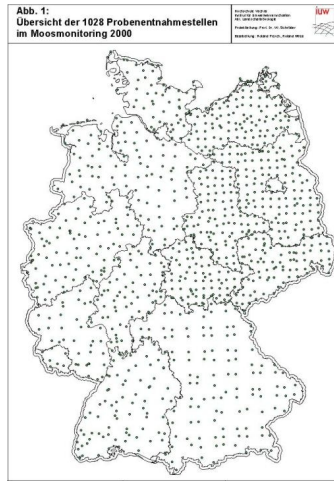
Schätzung der Gewichte

- Minimiert man die Schätzvarianz unter der Bedingung $\sum_{i=1}^d \omega_i = 1$, erhält man nach Lagrange (μ der Lagrange-Parameter) durch das Gleichungssystem

$$\begin{pmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_1 - \mathbf{s}_d) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{s}_d - \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_d - \mathbf{s}_d) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \omega_{1,s_0} \\ \vdots \\ \omega_{d,s_0} \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_0) \\ \vdots \\ \gamma(\mathbf{s}_d - \mathbf{s}_0) \\ 1 \end{pmatrix}$$

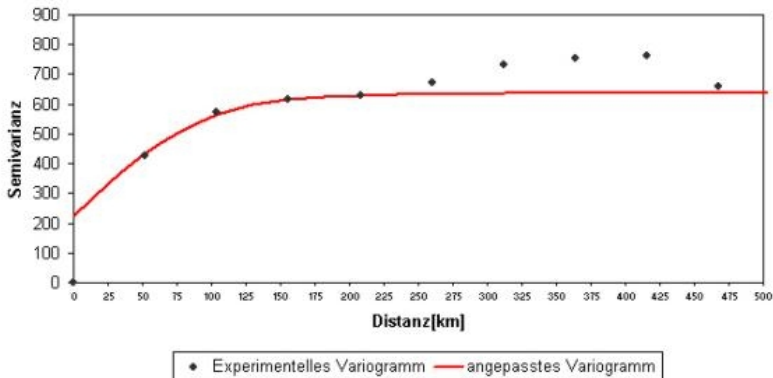
die Gewichte ω_{i,s_0} für einen erwartungstreuen und varianzminimalen Schätzer für den Wert an s_0 .

Beispiel: Bundesumweltamt Moosmonitoring

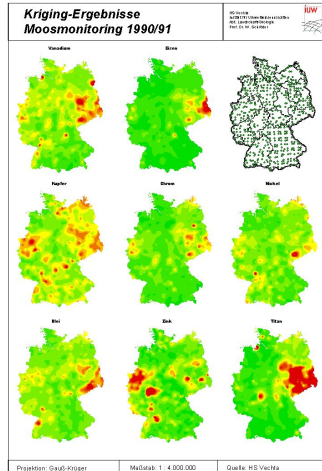


Beispiel: Bundesumweltamt Moosmonitoring

Variogrammanalyse Zink Moosmonitoring 1990/91 (16 Klassen)



Beispiel: Bundesumweltamt Moosmonitoring



Problematic

- Dieses „Gewöhnliche Kriging“ eignet sich nicht direkt zur Verwendung der Schätzung des räumlichen Einflusses $f_{geo}(s)$ in unserem geoadditiven Modell, weil
 - die Gewichte ω_{i,s_0} in Abhängigkeit von einem Punkt s_0 bestimmt werden,
 - der räumliche Einfluss an den Beobachtungsstellen $(Z(s_1), \dots, Z(s_d))$ zuvor bestimmt sein muss.
- Die Ansätze des Kriging und die der P-Splines lassen sich aber zu einem geeigneten Ansatz kombinieren.

Basisfunktionenansatz

Der räumliche Effekt lässt sich auch mit Hilfe von Basisfunktionen darstellen:

$$f_{geo}(s) = \sum_{j=1}^d \tilde{\gamma}_j B_j(s)$$

wobei

- $s_{(1)}, \dots, s_{(d)}$ gegebene Messpunkte (eine geeignet gewählte Teilmenge aller Beobachtungspunkte s_1, \dots, s_n)
- $B_j(s) = C(s - s_{(j)})$ eine (zuvor bestimmte) Kovarianzfunktion
- $\tilde{\gamma}_j$ (zu schätzende) Koeffizienten.

In Matrixschreibweise als lineares Modell:

$$\tilde{y} = \tilde{Z}\tilde{\gamma} + \epsilon$$

mit Designmatrix $\tilde{Z}[i, j] = C(s_i - s_{(j)})$ und Störgrößen ϵ .

Schätzung des räumlichen Effekts

Als Strafterm für $\tilde{\gamma}$ verwendet man analog zu den P-Splines $\lambda \tilde{\gamma}^T \tilde{K} \tilde{\gamma}$, wobei \tilde{K} hier eine Kovarianzmatrix mit $\tilde{K}[i, j] = C(s_{(i)} - s_{(j)})$ ist.

Damit ergibt sich der Schätzer analog zum vorigen Kapitel die Normalengleichung:

$$(\tilde{Z}^T \tilde{Z} + \lambda \tilde{K})^{-1} \tilde{\gamma} = \tilde{Z}^T \tilde{y}$$

Inhalt

- 1 Einleitung
- 2 Modellbildung
- 3 P-Splines
 - B-Splines
 - Penalisierung
- 4 Kriging
 - Idee
 - Variogramm und Kovarianz
 - Gewöhnliches Kriging
 - Ansatz für das geoadditive Modell
- 5 Markov-Zufallsfelder
- 6 Zusammenfassung

Definition: Markov-Zufallsfelder

Wir betrachten jetzt den Fall, dass die räumliche Information diskret, in Form von Regionen- oder Gitterpunktvariablen $s \in \{1, \dots, d\}$ vorliegt.

Bezeichne $D = \{1, \dots, s, \dots, d\}$ die Menge aller Regionen bzw. Gitterpunkte, so heißt $\gamma = \{\gamma_s, s \in D\}$ ein Markov-Zufallsfeld (mit γ_s Zufallsvariablen über gemeinsamem Wahrscheinlichkeitsraum), wenn für alle $s \in D$ die bedingte Verteilung von γ_s , gegeben alle übrigen Effekte $\gamma_t, t \neq s$, nur von den Nachbarn abhängt. Es lässt sich also schreiben:

$$P(\gamma_s | \gamma_t, t \neq s) = P(\gamma_s | \gamma_t, t \in N(s)) \quad \forall s \in D$$

wobei $N(s)$ eine (noch näher zu bestimmende) Nachbarschaft von s bezeichnet.

Nachbarschaften

- Solche Nachbarschaften werden üblicherweise und anschaulich über gemeinsame Grenzen definiert. (Nachbarschaften 1. Ordnung)
- Probleme treten aber wenn, wenn Regionen „Exklaven“ enthalten oder Teile des Beobachtungsbereichs räumlich separiert sind. Abhilfe können im Einzelfall Modifikationen dieser Definition leisten.
- Neben diesen Nachbarschaften 1. Ordnung können ergänzend „indirekte“ Nachbarn als Nachbarschaften höherer Ordnung betrachtet werden.
- Zwei benachbarte Regionen s und t werden im Folgenden mit der Notation $s \sim t$ bezeichnet.

Nachbarschaften

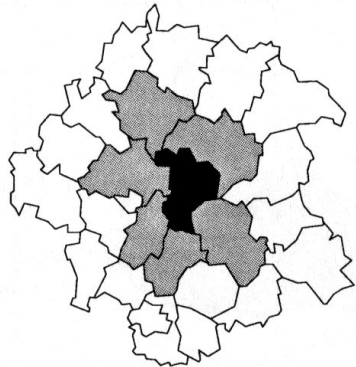
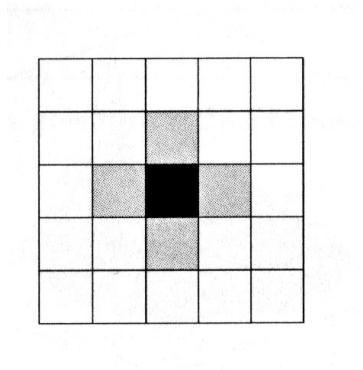


Abbildung: Nachbarschaften 1. Ordnung (links: auf einem regulären Gitter, rechts: für irregulär angeordnete Regionaldaten)

Ansatz für die Schätzfunktion

Wir weisen jeder Region einen Regressionskoeffizienten für den räumlichen Effekt zu:

$$f_{geo}(s) = \gamma_s, s = 1, \dots, d$$

Den Strafterm konstruieren wir auf der Basis der quadrierten Differenzen zwischen Parametern benachbarter Regionen mit dem Ziel einer glatten Regression.

Ansatz für die Schätzfunktion

Für den penalisierten KQ-Schätzer erhalten wir folglich das Minimierungsproblem:

$$\sum_{i=1}^n (y_i - \gamma_{s_i})^2 + \lambda \sum_{s=2}^d \sum_{t \in N(s), t < s} (\gamma_t - \gamma_s)^2 \rightarrow \min,$$

wobei

- y_i der Wert der i -ten Messung,
- s_i die Region, aus der die i -te Messung stammt,
- $N(s)$ die Menge aller Nachbarn der Region s und
- $\lambda \geq 0$ der Glattheitsparameter.

Schätzung des räumlichen Effektes

In Matrixnotation erhalten wir also wieder:

$$(y - Z\gamma)^T(y - Z\gamma) + \lambda\gamma^T K\gamma \rightarrow \min,$$

wobei die Designmatrix Z beschrieben wird mit

$$Z[i, s] = \begin{cases} 1 & \text{falls } y_i \text{ in Region } s \text{ beobachtet wurde} \\ 0 & \text{sonst} \end{cases}$$

und die Strafmatrix K mit

$$K[s, t] = \begin{cases} -1 & s \neq t, s \sim t \\ 0 & s \neq t, s \not\sim t \\ |N(s)| & s = t. \end{cases}$$

Die Normalengleichung ergibt sich wieder wie vorhin:

$$(Z^T Z + \lambda K)\gamma = Z^T y$$

Inhalt

- 1 Einleitung
- 2 Modellbildung
- 3 P-Splines
 - B-Splines
 - Penalisierung
- 4 Kriging
 - Idee
 - Variogramm und Kovarianz
 - Gewöhnliches Kriging
 - Ansatz für das geoadditive Modell
- 5 Markov-Zufallsfelder
- 6 Zusammenfassung

Gesamtmodell

Erinnern wir uns an das geoadditive Modell am Anfang:

$$y_i = \sum_{j=1}^m \beta_j x_{ij} + \sum_{j=1}^p f_j(z_{ij}) + f_{geo}(s_i) + \epsilon_i,$$

wobei

- β_j die (unbekannten) Koeffizienten
- f_j (unbekannte) Funktionen
- f_{geo} der (unbekannte) räumliche Einfluss
- ϵ_i eine Störgröße.

Gesamtmodell

Durch die Rückführung der nichtlinearen und räumlichen Einflüsse auf lineare Modelle erhalten wir:

$$y = \underbrace{X\beta}_{\text{linearer Einfluss}} + \underbrace{Z_1\gamma_1 + \dots + Z_p\gamma_p}_{\text{nichtlineare Einflüsse}} + \underbrace{Z_{geo}\gamma_{geo}}_{\text{räumlicher Einfluss}} + \epsilon = \tilde{X}\tilde{\beta} + \epsilon,$$

mit

$$\tilde{X} = (X \quad Z_1 \quad \dots \quad Z_p \quad Z_{geo}) \quad \text{und} \quad \tilde{\beta} = \begin{pmatrix} \beta \\ \gamma_1 \\ \vdots \\ \gamma_p \\ \gamma_{geo} \end{pmatrix}$$

Gesamtmodell

Das Minimierungsproblem:

$$(y - \tilde{X}\tilde{\beta})^T(y - \tilde{X}\tilde{\beta}) + \lambda\tilde{\beta}^T\tilde{K}\tilde{\beta} \rightarrow \min,$$

wobei die Strafmatrix \tilde{K} in folgender Gestalt aus den einzelnen Strafmatrizen K_1, \dots, K_p, K_{geo} der nichtlinearen Einflüsse bzw. des räumlichen Einflusses besteht:

$$\tilde{K} = \begin{pmatrix} \mathbf{0} & & & & \\ & K_1 & & & \\ & & \ddots & & \\ & & & K_p & \\ & & & & K_{geo} \end{pmatrix}$$

Normalengleichung:

$$(\tilde{X}^T\tilde{X} + \lambda\tilde{K})\tilde{\beta} = \tilde{X}^T\tilde{X}y$$

Kraftfahrtversicherung: Nichtlineare Einflüsse

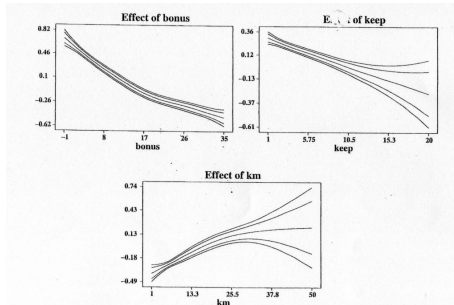


Abbildung: Geschätzte Funktionen für die nichtlinearen Einflüsse Schadensfreiheitsklasse, Kfz-Besitzdauer und gefahrene Kilometer/Jahr mit (punktweisen) 80%- und 95%-Konfidenzintervallen

Kraftfahrtversicherung: Räumlicher Einfluss

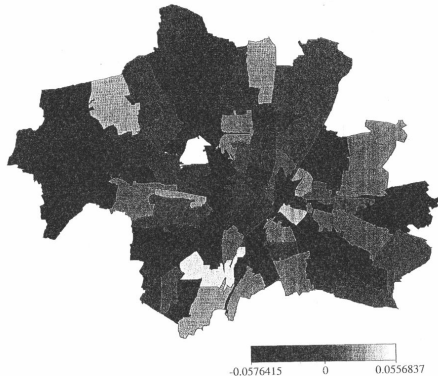


Abbildung: Geschätzter diskreter räumlicher Einfluss

Ausblick

Was hier nicht ausführlich dargestellt werden konnte:

- Bestmögliche Wahl von λ
- Wahl der Knotenzahl bei P-Splines
- Auswahl der Messpunkte beim Kriging
- Wahl der Kovarianzfunktion beim Kriging
- Güte des Modells \rightarrow statistische Tests

Literatur



L. Fahrmeir, T. Kneib und S. Lang

Regression - Modelle, Methoden und Anwendungen

Springer, Berlin, 2007



H. Wackernagel

Multivariate Geostatistics

Springer, Berlin, 1998



L. Fahrmeir, F. Sagerer und G. Sussmann

Geoadditive regression for analyzing small-scale geographical variability
in car insurance

Blätter der Deutschen Gesellschaft für Versicherungsmathematik,
28:47–65, 2007.