

Inhalt

- 1 Das lineare gemischte Modell
- 2 Likelihood-Schätzung für lineare gemischte Modelle
- 3 Likelihood-Inferenz im linearen gemischten Modell
- 4 Bayes-Schätzung für lineare gemischte Modelle
- 5 Additive gemischte Modelle
 - Penalisierte Splines
 - Penalisierte Regression und LMMs
 - Bayesianische Sicht auf Penalisierungsansätze
 - Weitere Beispiele
- 6 Das generalisierte lineare gemischte Modell
- 7 Likelihood-Schätzung für generalisierte lineare gemischte Modelle

Motivation

Die gemeinsame log-Likelihood für \mathbf{y} und \mathbf{b} im LMM, aus der sich die BLUPs für $\boldsymbol{\beta}$ und \mathbf{b} ergeben, hat die Form (siehe (15))

$$l(\mathbf{y}, \mathbf{b}) = \log L(\mathbf{y}, \mathbf{b}) = -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})' \mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) - \frac{1}{2}\mathbf{b}' \mathbf{G}^{-1}\mathbf{b}.$$

Diese kann als **penalisierte Likelihood** aufgefasst werden mit **Strafterm** $\mathbf{b}' \mathbf{G}^{-1}\mathbf{b}$.

Dies öffnet eine Verbindung zu **penalisierten Regressionsmethoden** (mit quadratischen Straftermen) und damit zu einer sehr allgemeinen Modellklasse.

Wir beginnen mit dem ausführlichen Beispiel der penalisierten Splines; weitere knappe Beispiele zeigen dann die Größe der flexiblen Modellklasse auf.

Das nichtparametrische Regressionsproblem

Modell: Die metrische Zielvariable y läßt sich durch eine deterministische Funktion $f(z)$ der metrischen Kovariablen z und einen additiven Fehler ε erklären:

$$y_i = f(z_i) + \varepsilon_i, i = 1, \dots, n.$$

- **Annahmen über f :**

Glattheitsanforderungen, z.B. via Stetigkeit oder Differenzierbarkeit von f .
Deutlich flexibler als Linearitätsannahme $f(z) = \beta_0 + \beta_1 z$ (lineares Modell).

- **Annahmen über ε :**

Wie im linearen Modell $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2) \Rightarrow E(y_i) = f(z_i)$.

D.h. die Funktion f modelliert den Erwartungswert der Zielvariablen.

Splines

Idee: Modelliere f stückweise durch Polynome, ohne dass Sprungstellen / Kanten entstehen \rightarrow **Splines** auf dem Definitionsbereich $[a, b]$ von z .

Definition von Splines:

Wähle **Knoten** $a = \kappa_1 < \dots < \kappa_m = b$. f ist ein Spline vom Grad $\ell \geq 0$, falls gilt

- $f(z)$ ist auf den Intervallen $[\kappa_j, \kappa_{j+1})$ ein Polynom vom Grad ℓ .
- $f(z)$ ist stetig (falls $\ell > 0$) und $(\ell - 1)$ -mal stetig differenzierbar. (**Glattheit**)

Konstruktive Darstellung durch Basisfunktionen:

Jeder Spline f vom Grad ℓ mit Knoten $\kappa_1 < \dots < \kappa_m$ kann eindeutig als Linearkombination $f(z) = \sum_{j=1}^d \gamma_j B_j(z)$ von $d = \ell + m - 1$ Basisfunktionen B_j dargestellt werden, die den d dimensionalen Spliner Raum $S_\ell(\kappa_1, \dots, \kappa_m)$ aufspannen.

Bekannte Basen: **Truncated-Power-Series** (TP) und **Basic-Splines** (B-Splines).

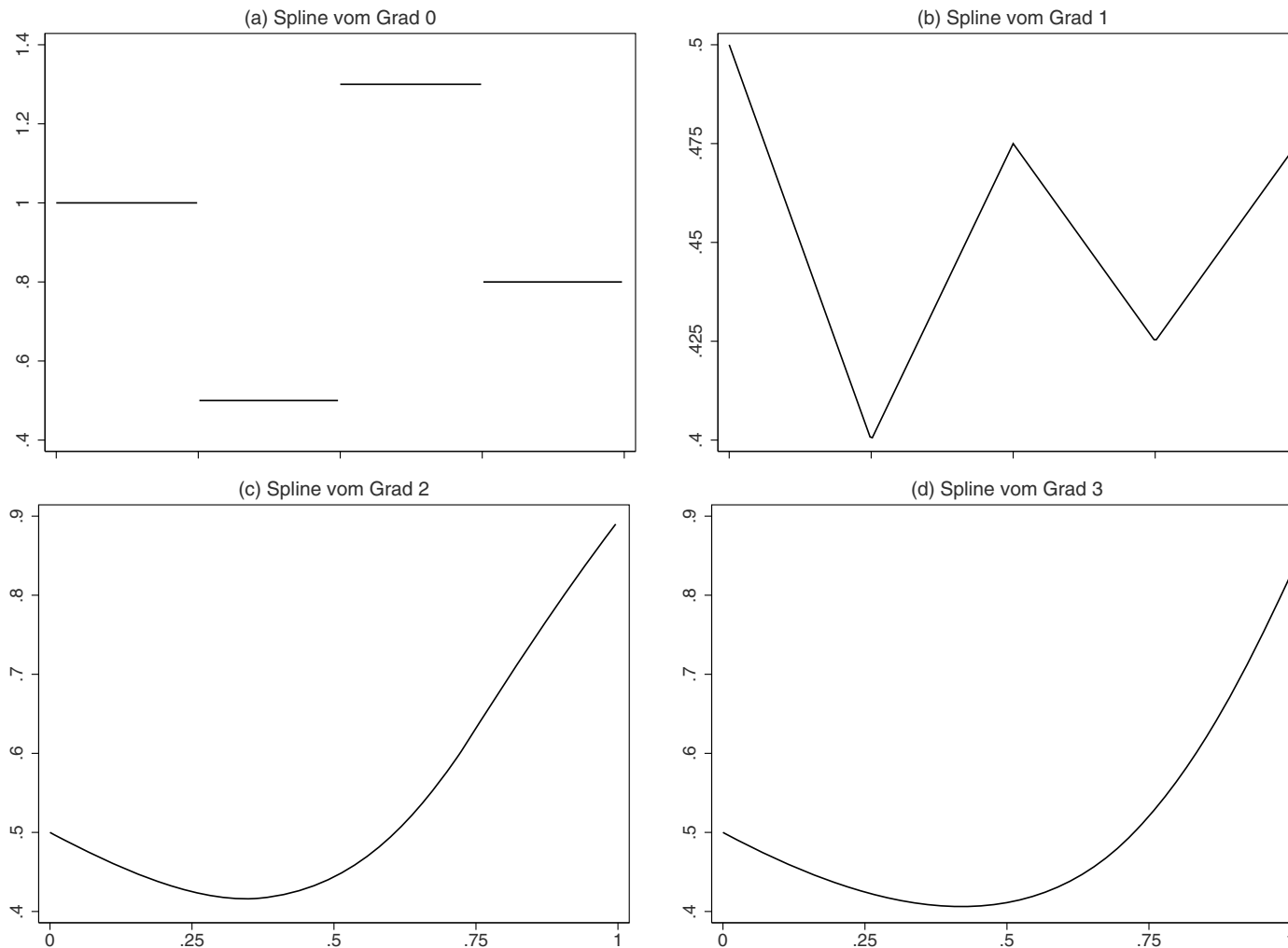


Abbildung 1: Beispiele für Splines vom Grad 0, 1, 2 und 3 zu den Knoten $\kappa_1 = 0, \kappa_2 = 0.25, \kappa_3 = 0.5, \kappa_4 = 0.75$ und $\kappa_5 = 1$.
Quelle: Fahrmeir et al (2009, Abb. 7.5).

Definition der d Basisfunktionen

Für eine Knotenmenge $\{\kappa_1, \dots, \kappa_m\}$ und einen Grad ℓ :

TP-Basisfunktionen

$$B_j(z) = z^{j-1}, \quad j = 1, \dots, \ell + 1 \quad (\text{Polynome})$$

$$B_{\ell+j}(z) = (z - \kappa_j)_+^\ell, \quad j = 2, \dots, m - 1 \quad (\text{abgeschnittene Potenzen}).$$

mit $(u)_+^\ell = [\max(u, 0)]^\ell$.

B-Spline Basisfunktionen (rekursiv)

$$B_j^0(z) = \mathbb{1}_{[\kappa_j, \kappa_{j+1})}(z), \quad j = 1, \dots, d,$$

$$B_j^\ell(z) = \frac{z - \kappa_{j-\ell}}{\kappa_j - \kappa_{j-\ell}} B_{j-1}^{\ell-1}(z) + \frac{\kappa_{j+1} - z}{\kappa_{j+1} - \kappa_{j+1-\ell}} B_j^{\ell-1}(z), \quad j = 1, \dots, d$$

mit zusätzlichen Knoten $\kappa_{1-\ell}, \dots, \kappa_0 < a$ und $b < \kappa_{m+1}, \dots, \kappa_{m+\ell}$.

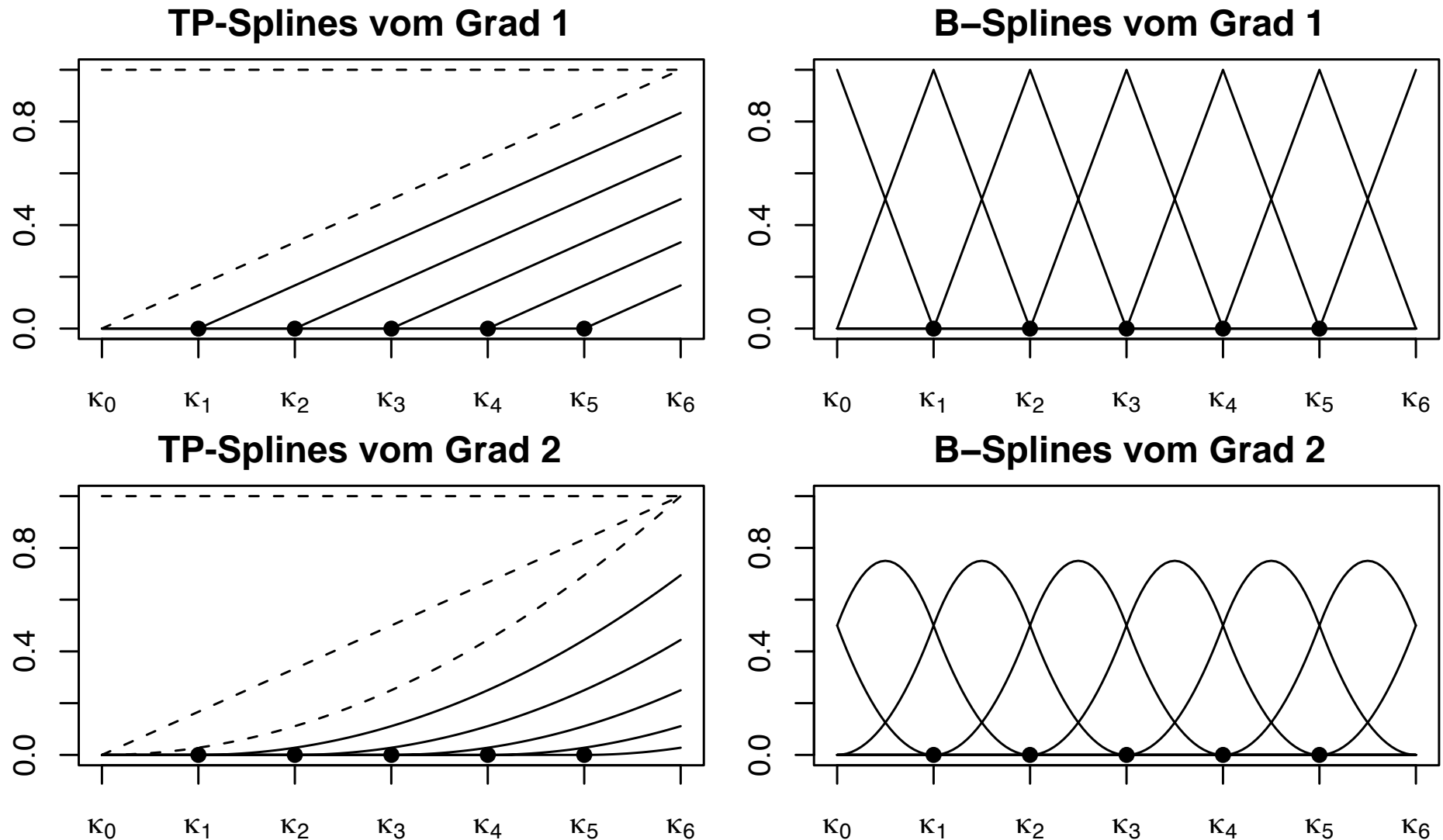


Abbildung 2: Lineare (oben) und quadratische (unten) TP- (links) und B-Spline-Basifunktionen (rechts) für $K = 5$ äquidistante Knoten mit $\kappa_0 = 0$. Für die TP-Basis sind die Polynomterme gestrichelt dargestellt.

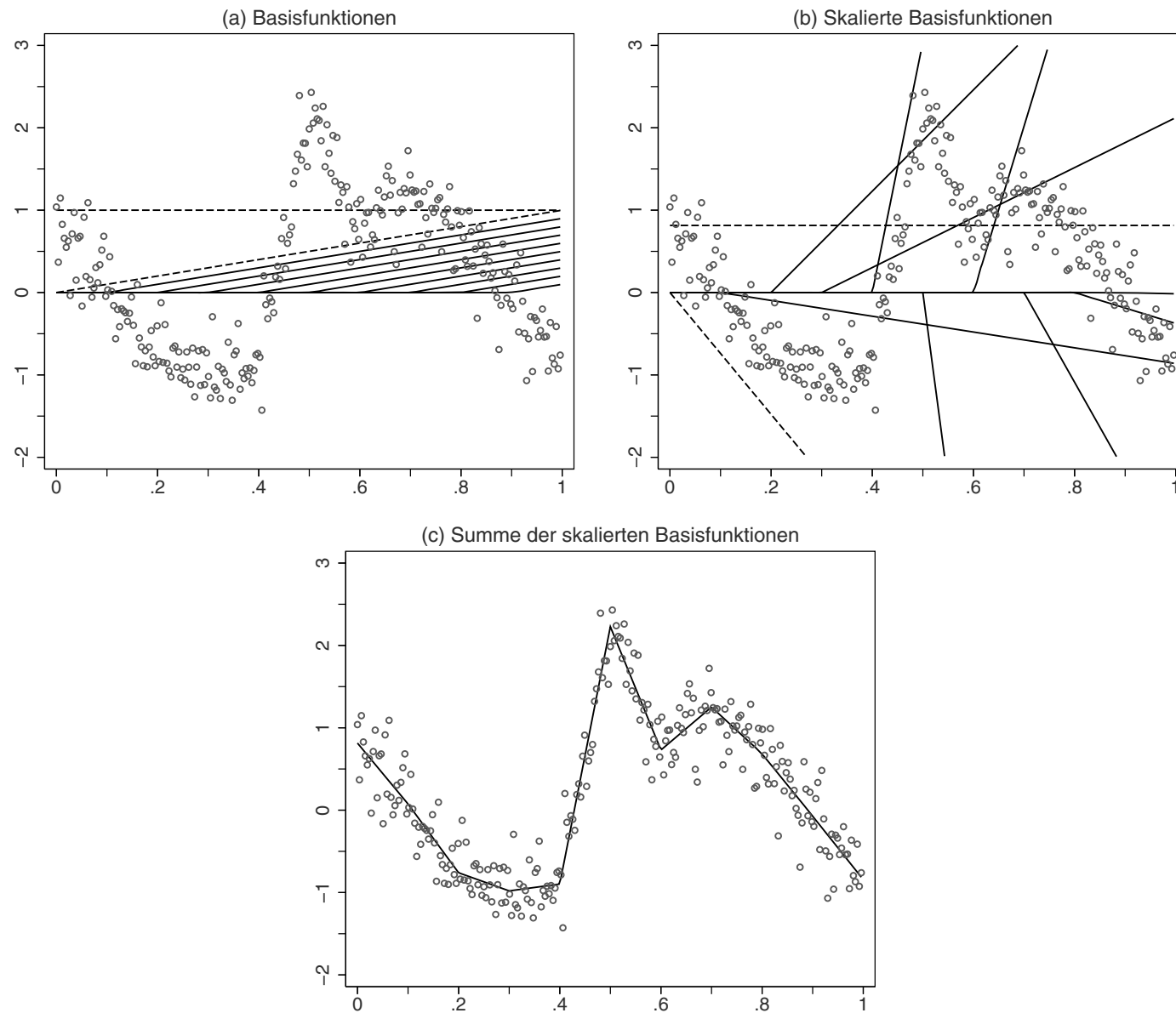


Abbildung 3: Konstruktion einer Funktion aus skalierten TP-Basisfunktionen für lineare Splines. Quelle: Fahrmeir et al (2009, Abb. 7.6).

Basisdarstellung

Damit ergibt sich die Darstellung des nichtparametrischen Regressionsproblems als

$$y_i = f(z_i) + \varepsilon_i \approx \sum_{j=1}^d \gamma_j B_j(z_i) + \varepsilon_i.$$

Man erhält ein **lineares Modell** in dem Parametervektor $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d)'$

$$\mathbf{y} = \mathbf{B}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

mit Designmatrix aus TP- bzw. B-Spline-Basis

$$\mathbf{B} = \begin{pmatrix} B_1(z_1) & \dots & B_d(z_1) \\ \vdots & & \vdots \\ B_1(z_n) & \dots & B_d(z_n) \end{pmatrix}.$$

Mit KQ-Schätzer ergibt sich $\hat{f}(z) = (B_1(z), \dots, B_d(z))(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{y}$.

Einfluss der Knotenzahl

Es gilt der **Varianz-Bias-Trade Off**:

- große Knotenzahl: flexible Funktionsschätzung, aber große Variabilität (Overfitting)
- kleine Knotenzahl: glattere Schätzung, aber möglicher Bias

⇒ Wesentliche Frage nach der **Anzahl der zu verwendenden Knoten!**

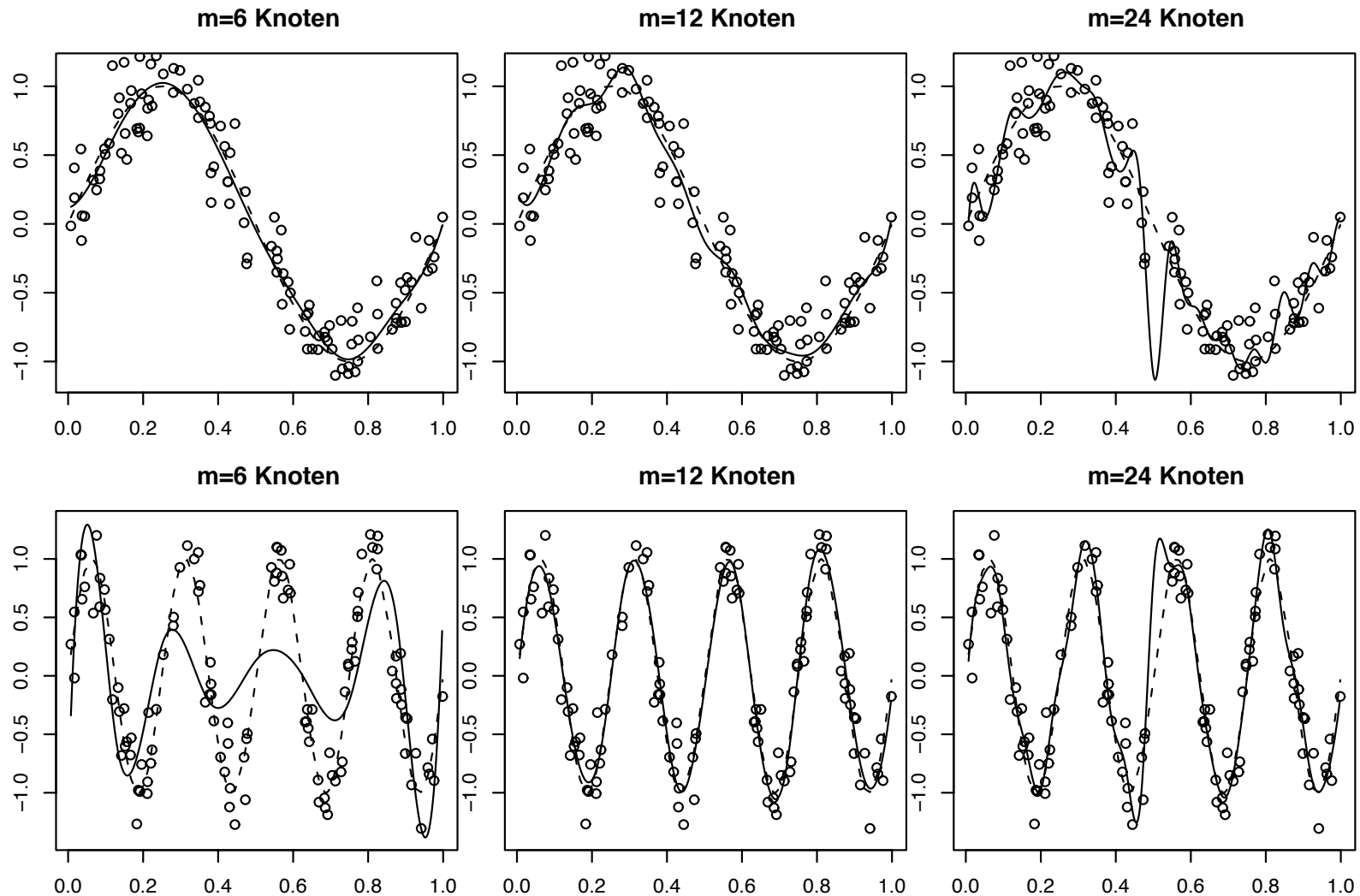


Abbildung 4: Abhängigkeit der Glättung mit Splines von der Knotenwahl. Die gestrichelte Linie gibt jeweils die wahre Funktion wieder, die durchgezogene Linie die Schätzung mit B-Splines mit 6, 12 bzw. 24 Knoten.

Grundidee penalisierte Splines

- Approximiere $f(z)$ durch einen Spline mit einer **großen Zahl von Knoten** (üblicherweise ca. 20 bis 40) \Rightarrow prinzipiell flexibel genug, auch stark variierende Funktionen $f(z)$ darzustellen.
- Führe zusätzlich einen **Strafterm** ein, der eine zu große **Variabilität der Schätzung bestraft**. (**Regularisierung**)
- Minimiere anstelle des üblichen KQ-Kriteriums ein um diesen Strafterm erweitertes **penalisiertes KQ-Kriterium** (äquivalent: Maximierung einer penalisierten log-Likelihood).

Penalisierte Splines auf TP-Basis

$$f(z) = \underbrace{\gamma_1 + \gamma_2 z + \dots + \gamma_{\ell+1} z^\ell}_{\text{1. Teil: globale polynomiale Form}} + \underbrace{\gamma_{\ell+2} (z - \kappa_2)_+^\ell + \dots + \gamma_d (z - \kappa_{m-1})_+^\ell}_{\text{2. Teil: Abweichung von globaler polynomialer Form}} .$$

Rauhe Funktionsschätzungen bei großer Variabilität des 2. Modellteils.

Da die ℓ te Ableitung von $(z - \kappa_j)_+^\ell$ gleich $\ell!$ für $z > \kappa_j$ und 0 sonst ist, $j \geq \ell + 2$, sind die γ_j proportional zu **Sprüngen in der ℓ ten Ableitung**.

⇒ Strafterm für betragsmäßig große Koeffizienten der abgeschnittenen Potenzen:

$$\text{pen}(\gamma, \mathbf{K}) = \lambda \sum_{j=\ell+2}^d \gamma_j^2 = \lambda \gamma' \mathbf{K} \gamma$$

mit $\gamma = (\gamma_1, \dots, \gamma_d)'$ und $\mathbf{K} = \text{diag}(\underbrace{0, \dots, 0}_{(\ell+1)}, \underbrace{1, \dots, 1}_{(m-2)})$.

Penalisierte Splines auf B-Spline-Basis: P-Splines

- Ein geeigneter Strafterm lässt sich über das Integral der (quadrierten) k -ten Ableitung definieren. Häufig wird dabei $k = 2$ verwendet:

$$\int (f''(z))^2 dz = \sum_{i=1}^d \sum_{j=1}^d \gamma_i \gamma_j \int B_i''(z) B_j''(z) dz = \gamma' \mathbf{K} \gamma$$

mit Strafmatrix $\mathbf{K} = (\int B_i''(z) B_j''(z) dz)_{i,j}$.

- Für die k -ten Ableitungen lässt sich der Strafterm über **Differenzen k -ter Ordnung** Δ^k der Parameter γ approximieren. Diese sind rekursiv definiert

$$\begin{aligned} \Delta^1 \gamma_j &= \gamma_j - \gamma_{j-1}, \\ \Delta^2 \gamma_j &= \Delta^1 \Delta^1 \gamma_j = \Delta^1 \gamma_j - \Delta^1 \gamma_{j-1} = \gamma_j - 2\gamma_{j-1} + \gamma_{j-2}, \\ &\vdots \\ \Delta^k \gamma_j &= \Delta^{k-1} \gamma_j - \Delta^{k-1} \gamma_{j-1}. \end{aligned}$$

Penalisiertes KQ-Kriterium

Minimiere das **penalisierte KQ-Kriterium** über γ in Abhängigkeit vom $\lambda \geq 0$,

$$PKQ(\lambda, \gamma) = (\mathbf{y} - \mathbf{B}\gamma)'(\mathbf{y} - \mathbf{B}\gamma) + \lambda\gamma' \mathbf{K}\gamma.$$

- Der **Glättungsparameter** $\lambda \geq 0$ steuert dabei den **Kompromiss zwischen Datentreue und Glattheit** der Funktionsschätzung.
- $\lambda \rightarrow 0$: Geringes Gewicht für den Strafterm. $PKQ(\lambda, \gamma) \approx KQ(\gamma)$ und $\hat{\gamma}$ nahe beim KQ-Schätzer.
- $\lambda \rightarrow \infty$: Schätzung vollständig durch den Strafterm dominiert. Resultat:
 - Bei der TP-Basis ein Polynom vom Grad ℓ als Schätzung für $f(z)$.
 - Bei P-Splines des Grades $\ell \geq k - 1$, k te Differenzen, ein Polynom vom Grad $k - 1$.
- Durch Wahl von $\lambda \geq 0$ Kontinuum zwischen diesen beiden Extremen.

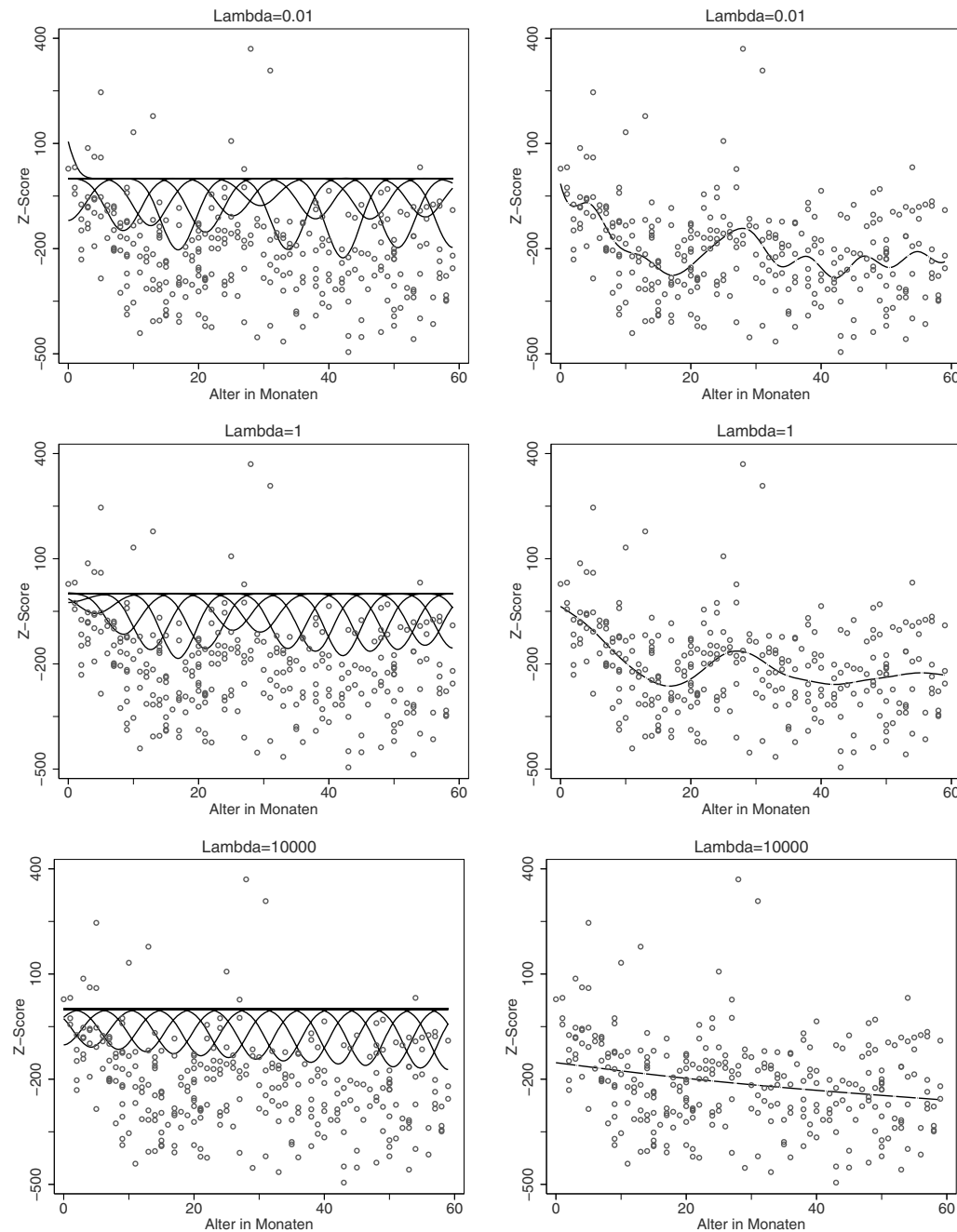


Abbildung 5: Einfluss des Glättungsparameters auf die Schätzung eines P-Splines basierend auf zweiten Differenzen. Quelle: Fahrmeir (2009, Abb. 7.15).

Die Rolle des Glättungsparameters

- Vorteil: Die **Glattheit** der Schätzung wird nicht mehr über Anzahl und Position der Knoten, sondern nur durch den Glättungsparameter λ gesteuert.
⇒ Frage nach der Wahl **geeigneter Werte für den Glättungsparameter λ !**

Optimalitätskriterien zur Bestimmung von **Glättungsparametern**, z.B.:

- Überlegungen zum **Bias-Varianz-Trade Off**. Einen solchen Kompromiss erhält man durch die Betrachtung des mittleren quadratischen Fehlers.
- Approximation des quadratischen Prognosefehlers durch **Kreuzvalidierung**.
- Verwendung von anderen aus der **Modellwahl** bekannten Kriterien, z.B. AIC.
- Hier: **Darstellung der Penalisierungsansätze als gemischtes Modell**.

TP-Penalierungsansatz als gemischtes Modell

Für penalisierte Splines mit TP-Basis ist das penalisierte KQ-Kriterium

$$PKQ(\lambda, \gamma) = (\mathbf{y} - \mathbf{B}\gamma)'(\mathbf{y} - \mathbf{B}\gamma) + \lambda \sum_{j=l+2}^d \gamma_j^2.$$

Zerlege die Designmatrix $\mathbf{B} = (\mathbf{X} | \mathbf{Z})$ und die Koeffizienten γ in

- $\boldsymbol{\beta} = (\gamma_1, \dots, \gamma_{l+1})'$, die **nicht penalisierten** Polynomkoeffizienten,
- $\mathbf{b} = (\gamma_{l+2}, \dots, \gamma_d)'$, die **penalisierten** Koeff. der abgeschnittenen Potenzen.

$$\Rightarrow PKQ(\lambda, \gamma)/\sigma^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})' \left(\frac{1}{\sigma^2} \mathbf{I}_n \right) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) + \mathbf{b}' \left(\frac{\lambda}{\sigma^2} \mathbf{I}_{m-2} \right) \mathbf{b}.$$

Dies ist das penalisierte KQ-Kriterium (16) für ein LMM

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad \mathbf{b} \sim N(\mathbf{0}, \frac{\sigma^2}{\lambda} \mathbf{I}_{m-2}). \quad (25)$$

- $\boldsymbol{\beta}$ und \mathbf{b} können als Vektoren **fester** / **zufälliger Effekte** aufgefasst werden.
- Die **Varianz** τ^2 der zufälligen Effekte ist dabei $\tau^2 = \sigma^2/\lambda$.

Schätzung über gemischte Modelle

- Im gemischten Modell (25) können die Varianzen σ^2 und τ^2 über **ML-** oder **REML-Schätzung** bestimmt werden.
- Den **optimalen Glättungsparameter** erhält man dann als $\hat{\lambda} = \hat{\sigma}^2 / \hat{\tau}^2$.
- Die Hypothese

$H_0 : f$ Polynom vom Grad ℓ bzw. $k - 1$ gegen $H_A : f$ flexiblere Funktion

lässt sich durch einen Test überprüfen für

$$H_0 : \tau^2 = 0 \quad \text{gegen} \quad H_A : \tau^2 > 0.$$

(Beachte die Hinweise zum Testen von Varianzkomponenten, Folien 66 ff.! Hier lässt sich \mathbf{y} nicht in N unabhängige Teilvektoren zerlegen, sodass die asymptotischen Mischungsverteilungen für T_{LQT} nicht gültig sind. Die exakte Verteilung oder Bootstrap-Verfahren bieten Alternativen.)

Punktweise Konfidenzbänder

Der BLUP für $f(z) = \mathbf{C}_z(\boldsymbol{\beta}', \mathbf{b}')'$ mit $\mathbf{C}_z = (B_1(z), \dots, B_d(z))$ ergibt sich aus (18) mit $\mathbf{C} = \mathbf{B} = [\mathbf{X}|\mathbf{Z}]$ als

$$\hat{f}(z) = \mathbf{C}_z(\hat{\boldsymbol{\beta}}', \hat{\mathbf{b}}')' = \mathbf{C}_z(\mathbf{C}'\mathbf{C} + \lambda \text{diag}(\mathbf{0}, \mathbf{I}_{m-2}))^{-1} \mathbf{C}'\mathbf{y},$$

mit der (bayesianischen) Kovarianz aus (21)

$$\text{Cov}(\hat{f}(z) - f(z)) = \sigma^2 \mathbf{C}_z(\mathbf{C}'\mathbf{C} + \lambda \text{diag}(\mathbf{0}, \mathbf{I}_{m-2}))^{-1} \mathbf{C}'_z.$$

Diese Kovarianz berücksichtigt möglichen Bias in $\hat{f}(z)$. → Übung

Unter der Annahme asymptotischer Normalverteilung können hiermit approximative punktweise Konfidenzbänder konstruiert werden. → Übung

Für simultane Konfidenzbänder siehe Ruppert, Wand und Carroll (2003, 6.5).

Penalisierte Regression und LMMs

Für allgemeine Penaliserungsansätze mit penalisierter Residuenquadratsumme

$$PKQ(\lambda) = (\mathbf{y} - \mathbf{B}\boldsymbol{\gamma})'(\mathbf{y} - \mathbf{B}\boldsymbol{\gamma}) + \lambda\boldsymbol{\gamma}'\mathbf{K}\boldsymbol{\gamma} \quad (26)$$

könnte man nun versuchen, mit $\tau^2 = \sigma^2/\lambda$ ein LMM zu definieren der Form

$$\mathbf{y} = \mathbf{B}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad \boldsymbol{\gamma} \sim N(\mathbf{0}, \tau^2 \mathbf{K}^{-1}).$$

Problem: Die Inverse \mathbf{K}^{-1} existiert nicht immer, $0 < \text{Rang}(\mathbf{K}) < \dim(\boldsymbol{\gamma})$.

- **Beispiel:** Für penalisierte B-Splines hat $\mathbf{K}_k = \mathbf{D}'_k \mathbf{D}_k$ keinen vollen Rang.

Konsequenz: Die aus (26) resultierende Dichte der zufälligen Effekte

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

ist damit **teilweise uneigentlich**, d.h. sie lässt sich nicht normieren.

Konstruktion einer geeigneten Zerlegung

Sei $h = \text{Rang}(\mathbf{K})$ und $d = \dim(\gamma)$. Zerlege γ in Teilvektoren β und \mathbf{b} , so dass

$$\gamma = \underbrace{\tilde{\mathbf{X}}}_{d \times (d-h)} \underbrace{\beta}_{(d-h) \times 1} + \underbrace{\tilde{\mathbf{Z}}}_{d \times h} \underbrace{\mathbf{b}}_{h \times 1}.$$

Wähle $\tilde{\mathbf{X}}$ und $\tilde{\mathbf{Z}}$, so dass

- $\mathbf{K}\tilde{\mathbf{X}} = \mathbf{0}$, so dass β nicht durch \mathbf{K} penalisiert wird,
- $\tilde{\mathbf{Z}}' \mathbf{K} \tilde{\mathbf{Z}} = \mathbf{I}_h$, so dass sich der Strafterm als $\lambda \mathbf{b}' \mathbf{b}$ schreiben lässt:

$$\gamma' \mathbf{K} \gamma = (\tilde{\mathbf{X}} \beta + \tilde{\mathbf{Z}} \mathbf{b})' \underbrace{\mathbf{K}(\tilde{\mathbf{X}} \beta + \tilde{\mathbf{Z}} \mathbf{b})}_{\mathbf{0}} = \mathbf{b}' \underbrace{\tilde{\mathbf{Z}}' \mathbf{K} \tilde{\mathbf{Z}}}_{\mathbf{I}_h} \mathbf{b} = \mathbf{b}' \mathbf{b}.$$

β , \mathbf{b} können dann als feste und (i.i.d.) zufällige Effekte aufgefasst werden im LMM

$$\mathbf{y} = \mathbf{B}\gamma + \varepsilon = \mathbf{B}\tilde{\mathbf{X}}\beta + \mathbf{B}\tilde{\mathbf{Z}}\mathbf{b} + \varepsilon =: \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \varepsilon, \quad \mathbf{b} \sim N(\mathbf{0}, \tau^2 \mathbf{I}_h).$$

$\lambda = \sigma^2 / \tau^2$ lässt sich wieder über REML oder ML schätzen.

Details zur Konstruktion

- Verwende eine Basis des Nullraums von \mathbf{K} als Spalten von $\tilde{\mathbf{X}} \Rightarrow \mathbf{K}\tilde{\mathbf{X}} = \mathbf{0}$.
Z.B. Polynome vom Grad 0 bis $k - 1$ für B-Splines mit k ten Differenzen.
- Für $\tilde{\mathbf{Z}}$, verwende Spektralzerlegung $\mathbf{K} = \underset{d \times h}{\mathbf{\Gamma}} \underset{h \times h}{\mathbf{\Omega}_+} \underset{h \times d}{\mathbf{\Gamma}'}$ mit $\mathbf{\Omega}_+$ Diagonalmatrix der positiven Eigenwerte, $\mathbf{\Gamma}$ orthonormale Matrix der Eigenvektoren.
Definiere $\tilde{\mathbf{Z}} = \mathbf{L}(\mathbf{L}'\mathbf{L})^{-1}$ mit $\mathbf{L} = \mathbf{\Gamma}\mathbf{\Omega}_+^{1/2}$ (also $\mathbf{K} = \mathbf{L}\mathbf{L}'$).
 $\Rightarrow \tilde{\mathbf{Z}}'\mathbf{K}\tilde{\mathbf{Z}} = (\mathbf{L}'\mathbf{L})^{-1} \mathbf{L}'\mathbf{L}\mathbf{L}'\mathbf{L} (\mathbf{L}'\mathbf{L})^{-1} = \mathbf{I}_h$.
- Die Spektralzerlegung ist nicht immer notwendig. So kann beispielsweise für P-Splines auch $\mathbf{L} = \mathbf{D}'$ mit der Differenzenmatrix \mathbf{D} gewählt werden.

\Rightarrow Die Zerlegung von γ ist nicht eindeutig.

Frequentistische und Bayesianische Sicht

- **Frequentistisch** betrachtet sind die Parameter γ feste, unbekannte Parameter.
- Im umformulierten gemischten Modell enthält der **b** -Teil von γ zufällige Effekte - keine (festen) Parameter, sondern Zufallsgrößen!
- Streng genommen wäre die Darstellung als LMM nur als algorithmischer Trick zu betrachten, nicht als tatsächliche Umformulierung des Modells.
- **Bayesianisch** betrachtet ist dies kein Problem, da eh alle Parameter Zufallsgrößen sind. Die zwei Darstellungen sind äquivalente Formulierungen der gleichen Priori-Annahmen.

Bayesianische Sicht auf Penalisierungsansätze

Beobachtungsmodell

$$\mathbf{y} = \mathbf{B}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

mit Designmatrix \mathbf{B} , die entsprechende Basisfunktionen enthält.

Penalisierung der Form $pen(\lambda, \mathbf{K}) = \lambda \boldsymbol{\gamma}' \mathbf{K} \boldsymbol{\gamma}$ entspricht (improperer) multivariater **Priori-Normalverteilung** $p(\boldsymbol{\gamma} | \mathbf{K}, \tau^2) \propto \exp\left(-\frac{1}{2\tau^2} \boldsymbol{\gamma}' \mathbf{K} \boldsymbol{\gamma}\right)$ für $\tau^2 = \sigma^2 / \lambda$ oder

$$\boldsymbol{\gamma} \sim N(\mathbf{0}, \tau^2 \mathbf{K}^{-1}).$$

Beispiel: TP-Basis

- **Priori-Normalverteilung** für die Koeffizienten der abgeschnittenen Polynome: $\gamma_{l+2}, \dots, \gamma_d \stackrel{iid}{\sim} N(0, \tau^2)$.
- **Nichtinformative Priori** für die Koeffizienten der globalen Polynome: $p(\gamma_j) \propto \text{const}, j = 1, \dots, l + 1$.

Die beiden verschiedenen Typen von Priori-Verteilungen spiegeln die Unterscheidung in unpenalisiert und penalisiert zu schätzende Parameter wieder.

Beispiel: B-Spline-Basis

- Priori-Annahme für γ : **Random Walks** (Irrfahrten) **der Ordnung k** als stochastisches Analogon zur Differenzenbestrafung. Z.B. $k = 1$ (RW1):

$$\gamma_j = \gamma_{j-1} + u_j, \quad u_j \sim N(0, \tau^2), \quad j = 2, \dots, d.$$

- **nichtinformative** Priori-Verteilung $p(\gamma_1) \propto \text{const}$
- \Rightarrow **bedingte Verteilungen** $\gamma_j | \gamma_{j-1}, \dots, \gamma_1 \sim N(\gamma_{j-1}, \tau^2)$
- **gemeinsame Verteilung** für γ : NV mit EW $\mathbf{0}$ und Präzisionsmatrix \mathbf{K}/τ^2 , $\mathbf{K} = \mathbf{D}'_1 \mathbf{D}_1$
- **Präzisionsmatrix** hat **keinen vollen Rang** \Rightarrow Kovarianzmatrix $\tau^2 \mathbf{K}^{-1}$ existiert nicht bzw. gemeinsame Priori ist eine **teilweise uneigentlichen Verteilung**
- trotzdem **eigentliche Posteriori-Verteilung** (NV) mit

$$E(\gamma | \mathbf{y}) = \left(\mathbf{B}' \mathbf{B} + \frac{\sigma^2}{\tau^2} \mathbf{K} \right)^{-1} \mathbf{B}' \mathbf{y}, \quad \text{Cov}(\gamma | \mathbf{y}) = \left(\mathbf{B}' \mathbf{B} + \frac{\sigma^2}{\tau^2} \mathbf{K} \right)^{-1}.$$

Variierende Koeffizienten

- Glatte Interaktion $g(z)x$ zwischen metrischer Variable z , binärer Variable x .
- Das Modell sollte dann auch die Haupteffekte $x\beta_1$ und $f(z)$ enthalten,

$$y_i = \beta_0 + x_i\beta_1 + f(z_i) + g(z_i)x_i + \varepsilon_i.$$

- $f(z)$ ist der nichtlineare Effekt von z , falls $x = 0$,
- $f(z) + g(z) + \beta_1$ ist der nichtlineare Effekt von z , falls $x = 1$.

Zentriere $g(z)$ um Null, damit das Modell identifizierbar ist:

$$g(z)x + \beta_1x = (g(z) + c)x + (\beta_1 - c)x.$$

- $g(z)$ lässt sich wieder über Basisfunktionen approximieren. Die Designmatrix enthält das Produkt der Werte der Basisfunktionen und der x -Werte.
- Eine geeignete Strafmatrix ist wie zuvor z.B. $\mathbf{K}_k = \mathbf{D}'_k \mathbf{D}_k$ für B-Splines.
- Schätzung über das lineare gemischte Modell analog zu penalisierten Splines. Analog für x stetig oder kategorial. Details in Fahrmeir et al (2009, 8.3-8.4).

Bivariate Glättung und räumliche Effekte

Für die Darstellung von räumlichen Effekten oder nichtparametrischen Interaktionsoberflächen lässt sich die Methodik der P-Splines verallgemeinern:

$$y_i = f(z_{i1}, z_{i2}) + \varepsilon_i \approx \sum_{j=1}^{d_1} \sum_{k=1}^{d_2} \gamma_{jk} B_{jk}(z_{i1}, z_{i2}) + \varepsilon_i$$

mit z.B.

$$B_{jk}(z_1, z_2) = B_j^{(1)}(z_1) B_k^{(2)}(z_2), \quad j = 1, \dots, d_1, \quad k = 1, \dots, d_2$$

eine **Tensorprodukt-Basis** (Produkte aller univariater B-Spline-Basisfunktionen). Ein geeigneter Strafterm ist hier z.B. (mit Kronecker-Produkt \otimes , analog für \mathbf{D}_k)

$$\begin{aligned} \lambda \boldsymbol{\gamma}' \mathbf{K} \boldsymbol{\gamma} &= \lambda \left[\sum_{j=1}^{d_1} \sum_{k=2}^{d_2} (\gamma_{jk} - \gamma_{j,k-1})^2 + \sum_{k=1}^{d_2} \sum_{j=2}^{d_1} (\gamma_{jk} - \gamma_{j-1,k})^2 \right] \\ &= \lambda \boldsymbol{\gamma}' [(\mathbf{I}_{d_2} \otimes \mathbf{D}_1)' (\mathbf{I}_{d_2} \otimes \mathbf{D}_1) + (\mathbf{D}_1 \otimes \mathbf{I}_{d_1})' (\mathbf{D}_1 \otimes \mathbf{I}_{d_1})] \boldsymbol{\gamma}. \end{aligned}$$

Die Schätzung über gemischte Modelle läuft wie für allgemeine Penaliserungsansätze beschrieben. Für Details, weitere (z.B. radiale) Basisfunktionen und Beziehungen zum Kriging in der räumlichen Statistik, siehe Fahrmeir et al (2009, Abschnitt 7.2).

Für höherdimensionale Glättung, siehe Fahrmeir et al (2009, Abschnitt 7.3).

Markov-Zufallsfelder

Bei räumlichen Daten in Regionen oder auf einem Gitter wird räumliche Glättung meist über Nachbarschaften umgesetzt:

$$r \in N(s) \Leftrightarrow r \text{ ist zu } s \text{ benachbart.}$$

Verwenden wir ein Intercept γ_s pro Region s , so ist ein möglicher Strafterm

$$\lambda \sum_{s=1}^d \sum_{r \in N(s), r < s} (\gamma_r - \gamma_s)^2 = \lambda \gamma' \mathbf{K} \gamma$$

mit $\mathbf{K} = (K_{sr})_{sr}$, $K_{sr} = -1$ für $r \in N(s)$, $K_{sr} = |N(s)|$ für $r = s$, $K_{sr} = 0$, sonst. Eine Schätzung über gemischte Modelle ist wieder möglich, für Details und die Beziehung zu Markov-Zufallsfeldern siehe Fahrmeir et al (2009, Abschnitt 7.2.4).

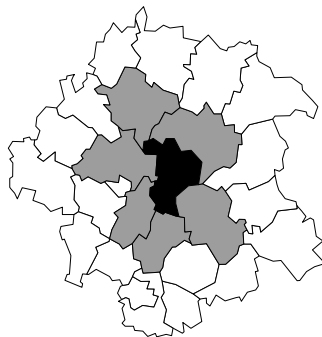
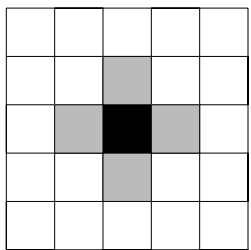


Abbildung 6: Nachbarschaften 1. Ordnung für reguläres Gitter und irreguläre Regionen. Die Nachbarn der schwarz gekennzeichneten Region sind in grau wiedergegeben. Quelle: Fahrmeir et al (2009, Abb. 7.45).

Ordinale Kovariablen

Bei einer Kovariablen mit geordneten Kategorien $k = 0, \dots, K$, bei der ein glatter Anstieg/Abfall der Koeffizienten $\gamma_0 = 0, \gamma_1, \dots, \gamma_K$ angenommen wird, schlagen

- Gertheiss, J. and G. Tutz (2009). *Penalized regression with ordinal predictors*. International Statistical Review 77, 345–365.

den Strafterm

$$\lambda \boldsymbol{\gamma}' \mathbf{K} \boldsymbol{\gamma} = \lambda \sum_{j=1}^K (\gamma_j - \gamma_{j-1})^2 = \lambda \boldsymbol{\gamma}' \mathbf{D}'_1 \mathbf{D}_1 \boldsymbol{\gamma}$$

vor. Die Schätzung über ein LMM läuft analog zu penalisierteren Splines. Der Test

$$H_0 : \gamma_1 = \dots = \gamma_K = \gamma_0 = 0 \quad \text{gegen} \quad H_A : \gamma_k \neq 0 \text{ für mindestens ein } k \quad (27)$$

auf Einfluss der Kovariablen über einen (R)LRT

$$H_0 : \tau^2 = 0 \quad \text{gegen} \quad H_A : \tau^2 > 0$$

im zugehörigen LMM hat größere Power als ein F-Test für (27) im linearen Modell:

- Gertheiss, J. and F. Oehrlein (2011). *Testing linearity and relevance of ordinal predictors*. Electronic Journal of Statistics 5, 1935-1959.

Lineare Gemischte Modelle als Modulare Modelle

Die Beispiele zeigen, dass LMMs ein mächtiges Werkzeug sind, mit dem flexibel Modellbausteine in **strukturiert-additiven Regressionsmodellen** modular kombiniert werden können:

- zufällige Effekte für z.B. Longitudinal- oder Clusterdaten
- räumliche Terme
- glatte Terme oder Interaktionen, variierende Koeffizienten
- ...

Damit sind z.B. Modelle der Form

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + b_{i0} + u_{ij1}b_{i1} + f_1(z_{ij1}) + f_2(z_{ij2}, z_{ij3}) + f_3(\mathbf{s}_{ij}) + f_4(z_{ij4})x_{ij,p+1} + \varepsilon_{ij}$$

mit Kovariablen \mathbf{x} und \mathbf{u} , metrischen Kovariablen \mathbf{z} , räumlichen Koordinaten \mathbf{s} , festen Effekten $\boldsymbol{\beta}$ und zufälligen Effekten \mathbf{b} möglich.

Software

Penalisierungsansätze können formal als LMM aufgeschrieben werden, weisen jedoch nicht die sonst typische Gruppierungsstruktur auf.

⇒ Verwende zur Schätzung spezialisierte Software, die an diese Struktur angepasst ist (und die nötigen Designmatrizen \mathbf{X} und \mathbf{Z} automatisch konstruiert).

Hier ist insbesondere das Paket `mgcv` von Simon Wood zu nennen.

Die `gamm()`-Funktion oder `gam()` mit Option `method = "(RE)ML"` fiten (generalisierte) additive gemischte Modelle mit (RE)ML-Schätzung.

- Feste und zufällige Effekte werden wie in `lme()` spezifiziert.
- Glatte Effekte $f(z)$ über `s(z)`.
- Glatte räumliche oder Interaktions-Oberflächen über `s(z1, z2)` oder `te(z1, z2)`.
- Für variierende Koeffizienten ist das `by`-Argument vorgesehen, `s(z, by=x)`.
- Weitere Möglichkeiten und Optionen, siehe die Manual sowie Wood (2006).