

# Datenglättung mit gemischten Regressionsmodellen

Florian Hiemeyer

1. Oktober 2003

## Zusammenfassung

Seminar Analyse komplexer Strukturen, bei Prof. Gerhard Tutz, SS2003:  
Die Methode gemischter Modelle und ihr Einsatz als Ersatz von P-Splines

## 1 Ausgangspunkt

Ein häufig auftauchendes Problem bei Glättungsverfahren ist, dass die einzelnen, per Basisfunktionen ermittelten Glättungsfunktionen, zwischen zwei Knoten vernünftig geglättet sind, zugleich aber zwischen den einzelnen Parametern  $\beta_j$  der Abschnitte große Unterschiede bestehen. Diese sorgen oftmals für kantige, unglatte Verbindungsstellen.

Ein häufig eingesetztes Werkzeug zur Vermeidung dieses Fehlverhaltens ist der Ersatz des Kleinste-Quadrate-Kriteriums durch die Minimierung einer penalisierten Residuenquadratsumme (SPP):

$$SPP(\boldsymbol{\beta}) = \sum_{i=1}^n \left( y_i - \sum_{j=0}^p \beta_j B_j(x_i) \right)^2 + \lambda \underbrace{\sum_{j=k}^p (\Delta^k \beta_j)^2}_{\text{sog. Penalisierung/Strafterm}}$$

Dabei stehen die  $x_j$  für die Meßstellen,  $y_j$  für die Meßwerte,  $B_j$  für die Basisfunktion zwischen den Knoten  $j$  und  $j + 1$ , und  $\Delta^k$  für den Differenzenoperator  $k$ -ter Ordnung.

Der Glättungsparameter  $\lambda$  bestimmt das Gewicht des Strafterms auf die penalisierte Summe, und damit das Verhältnis Datentreue/Glattheit. Problematisch dabei ist insbesondere der Differenzenoperator, der für größere  $k$  sehr unübersichtlich wird.

Dieses Verfahren ist aber - nach (Speed, 1991; Seite 43f) - äquivalent zu einem weiteren, echt statistischen, Verfahren: den gemischten Modellen (Mixed Models) in der Regression.

Von dieser Möglichkeit handelt diese Seminararbeit, in der zunächst die Technik der Mixed Models vorgestellt werden soll, und anschließend die Verwendung als Glättungsverfahren aufgezeigt werden soll.

## 2 Gemischte Regressionsmodelle

Im klassischen Regressionsmodell werden die erklärenden Variablen als fix und ohne Meßfehler gemessen betrachtet, der Zufallsfehler wird als  $\epsilon$  „zusätzlich“ eingeführt. Das Modell lässt sich allerdings auch auf Daten mit und ohne Meßfehler übertragen. Dann spricht man von gemischten Modellen (engl.: Mixed Models).

Der Vorteil der Anwendung der gemischten Modelle liegt vor allem darin, dass die Notation und das Grundprinzip der Parameterbestimmung (Maximum Likelihood) dem der generalisierten linearen Regression entspricht. Damit bleibt ein Großteil der Analysemethoden, die für generalisierte Regressionsmodelle vorgesehen sind, leicht umsetzbar. Dies betrifft vor allem auch Methoden zur Behandlung fehlender Daten und den Umgang mit Meßfehlern sowie korrelierten Fehlern (z. B. bei Zeitreihen).

### 2.1 Notation und Definitionen

#### 2.1.1 Das Gauß-Mixed Model

Zur Erläuterung der Grundprinzipien der Mixed Models zunächst der wichtigste Spezialfall, das Gauss'sche Mixed Model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \epsilon, \begin{bmatrix} \mathbf{u} \\ \epsilon \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

Hierbei ist dann  $\mathbf{G}$  die Kovarianzmatrix der zufälligen Effekte, und  $\mathbf{R}$  die Kovarianzmatrix des Zufallsfehlers. Sowohl der Zufallsfehler, als auch die zufälligen Effekte in diesem Modell sind also normalverteilt, wenn auch mit unterschiedlichen Varianzen.

Wegen dem Additionstheorem für die Normalverteilung folgt daraus die rechenstechnisch vereinfachende Tatsache, dass  $\mathbf{y}$  selbst auch wieder normalverteilt ist, wobei gilt

$$\begin{aligned} \mathbb{E}[\mathbf{y}] &= \mathbb{E}[\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \epsilon] = \\ &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbb{E}[\mathbf{u}] + \mathbb{E}[\epsilon] = \\ &= \mathbf{X}\boldsymbol{\beta} \\ \mathbb{V}_{\mathbf{y}} = \text{Cov}(\mathbf{y}) &= \text{Cov}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \epsilon) = \\ &= \text{Cov}(\mathbf{X}\boldsymbol{\beta}) + \text{Cov}(\mathbf{Z}\mathbf{u}) + \text{Cov}(\epsilon) = \\ &= \mathbf{Z}\text{Cov}(\mathbf{u})\mathbf{Z}^t + \text{Cov}(\epsilon) = \\ &= \mathbf{Z}\mathbf{G}\mathbf{Z}^t + \mathbf{R} \end{aligned}$$

#### 2.1.2 Generalisierte gemischte Modelle

Eine verallgemeinerte Form des gemischten Modells, mit geeigneter Linkfunktion  $\mathbf{g}$  bietet die Formel:

$$\mathbb{E}[\mathbf{y}|\mathbf{u}] = \mathbf{g}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}), \mathbf{u} \sim (\mathbf{0}, \mathbf{G})$$

### 3 Methodik der Gemischten Modelle

#### 3.1 Parameterbestimmung mit BLUP

Die Parameter eines gemischten Modells werden nach der Methode der **Best Linear Unbiased Predictor (BLUP)**, wenn die Varianzmatrizen  $\mathbf{G}$ ,  $\mathbf{R}$  bekannt sind:

$$\begin{aligned} \mathbf{X}^t \mathbf{R}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{X}^t \mathbf{R}^{-1} \mathbf{Z} \hat{\mathbf{u}} &= \mathbf{X}^t \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^t \mathbf{R}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} + (\mathbf{Z}^t \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1}) \hat{\mathbf{u}} &= \mathbf{Z}^t \mathbf{R}^{-1} \mathbf{y} \end{aligned}$$

#### 3.2 Anwendungsbeispiele zu gemischten Modellen

##### 3.2.1 Ertrag von Milchkühen, bei bekanntem $\mathbf{G}$

Es wird die Menge des ersten Milchertrags bei Kühen betrachtet (,yield'), die Prädiktoren sind das Vatertier (,sire') der Kuh und die Herde des Tieres (,herd'). Dabei wird angenommen, dass der Parameter des ,sire'- wegen des Prinzips der genetischen Vererbung - auch bei gleichem Vatertier zufällig schwankt. Der Wert für ,herd' wird dagegen als nicht-zufällig, also fix, angenommen. Der sire-Effekt wird - z. B. aufgrund früherer Erkenntnisse - als  $u \sim \mathcal{N}(\mathbf{0}, 0.1\mathbf{I} * \sigma^2)$  beschrieben: Damit gilt für die gemeinsame Verteilung der  $\mathbf{u}$ ,  $\boldsymbol{\epsilon}$ :

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} 0.1 * \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \sigma^2 \right)$$

Beispielhaft soll nun die Berechnung des gemischten Modells an diesen Daten berechnet werden.

$$\text{Daten} = \begin{pmatrix} \text{Herd} & \text{Sire} & \text{Yield} \\ 1 & A & 110 \\ 1 & D & 100 \\ 2 & B & 110 \\ 2 & D & 100 \\ 2 & D & 100 \\ 3 & C & 110 \\ 3 & C & 110 \\ 3 & D & 100 \\ 3 & D & 100 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{y} = (110, 100, 110, 100, 100, 110, 110, 100, 100)^T$$

$$\boldsymbol{\beta} = (h_1, h_2, h_3)^T, \mathbf{u} = (s_A, s_B, s_C, s_D)^T$$

$$\mathbf{R} = \sigma^2 \mathbf{I}$$

$$\mathbf{G} = 0.1 * \sigma^2 \mathbf{I}$$

Lösungsgleichungen:

$$\begin{aligned} \mathbf{X}^t \mathbf{R}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{X}^t \mathbf{R}^{-1} \mathbf{Z} \hat{\mathbf{u}} &= \mathbf{X}^t \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^t \mathbf{R}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} + (\mathbf{Z}^t \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1}) \hat{\mathbf{u}} &= \mathbf{Z}^t \mathbf{R}^{-1} \mathbf{y} \end{aligned}$$

Rechnen ergibt:

$$\begin{pmatrix} 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 3 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 4 & 0 & 0 & 2 & 2 \\ 1 & 0 & 0 & 11 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 11 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 12 & 0 \\ 1 & 2 & 2 & 0 & 0 & 0 & 15 \end{pmatrix} \begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \\ \hat{s}_A \\ \hat{s}_B \\ \hat{s}_C \\ \hat{s}_D \end{pmatrix} = \begin{pmatrix} 210 \\ 310 \\ 420 \\ 110 \\ 110 \\ 220 \\ 500 \end{pmatrix}$$

Nach Gaußelimination erhält man:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (105.64, 104.28, 105.46)^T \\ \hat{\mathbf{u}} &= (0.40, 0.52, 0.76, -1.76)^T \end{aligned}$$

Zum Vergleich: mit festen Effekten (wegen Singularität von  $\mathbf{X}$  ist  $s_D = 0$ ):

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (100, 100, 100)^t \\ \hat{\mathbf{u}} &= (10, 10, 10, 0)^t \end{aligned}$$

### 3.2.2 Mixed Models bei der Analyse gruppierter Daten (bei unbek. $\mathbf{G}$ )

Eine der wichtigsten Verwendungszwecke Gemischter Modelle stellt die Analyse gruppierter Daten, zum Beispiel zeitlich wiederholte Messungen an verschiedenen Subjekten, dar. Als Beispiel gehe ich hier auf die Messung des Gewichts von 48 Schweinen über 9 Wochen (Wand,2002;1f) ein. Dabei gilt das Modell:

$$weight_{ij} = \beta_0 + U_i + \beta_1 week_j + \epsilon_{ij}, 1 \leq i \leq 48, 1 \leq j \leq 9 \text{ mit } U_i \sim \mathcal{N}(0, \sigma_U^2)$$

Der Parameter  $\beta_0$  ist hierbei die Baseline im Regressionsmodell, während die  $U_i$  den Faktor-Intercept der einzelnen Schweine erklären. Dabei sind sowohl die Zufälligkeit des Faktor-Intercept für jedes einzelne Schwein enthalten, als auch – aufgrund der Zufälligkeit des Parameters – auch Korrelationen zwischen den Gewichtsmessungen eines Tieres.

Hierbei sind nun nicht nur die Parameter  $U_i$  aus den Daten zu schätzen, sondern auch deren, nun als unbekannt angesehene, Kovarianzmatrix  $\mathbf{G}$ . Die Lösung für diese Problemstellung erfolgt mittels der **Maximum-Likelihood** oder **Restricted-Maximum-Likelihood** Technik.

## 3.3 Die ML-Methode bei gemischten Modellen

### 3.3.1 Klassische ML-Methode: verzerrt

Aufgrund der Normalverteilung der  $\mathbf{y}$  im gemischten Modell, mit  $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$  und  $\mathbf{V}_y = Cov(\mathbf{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^t + \mathbf{R}$  ergibt sich zunächst

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}_y).$$

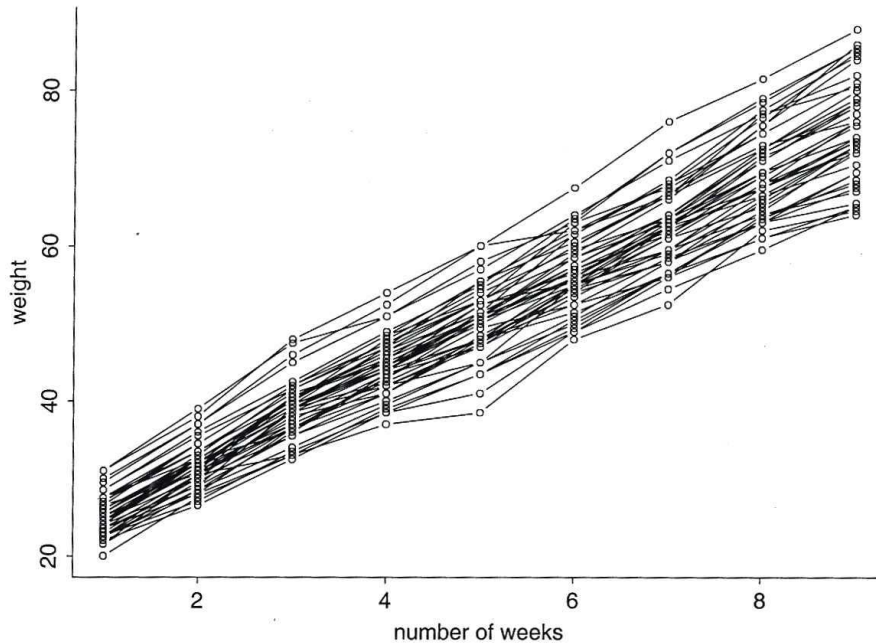


Abbildung 1: Verbundene Linien zeigen die Messungen am selben Tier.

Nun ist es das Ziel, einen Schätzer für  $\beta$  zu finden. Im Falle einer bekannten Varianz  $\mathbf{V}_y$ , läßt sich ein Schätzer für  $\beta$  mittels der Maximierung der Likelihoodfunktion von  $\mathbf{y}$  finden:

$$L(\beta, \mathbf{V}_y) = (2\pi)^{-\frac{n}{2}} |\mathbf{V}_y|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^t \mathbf{V}_y^{-1} (\mathbf{y} - \mathbf{X}\beta)\right\}$$

und  $l(\beta, \mathbf{V}_y) = -\frac{1}{2}\{n \log(2\pi) + \log |\mathbf{V}_y| + (\mathbf{y} - \mathbf{X}\beta)^t \mathbf{V}_y^{-1} (\mathbf{y} - \mathbf{X}\beta)\}$

Eine Maximum-Likelihood-Schätzung kann auch über die einzelnen Parameter optimiert werden. Da  $\beta$  nur im letzten Term der log-Likelihood  $l(\beta, \mathbf{V}_y)$  auftaucht, findet sich so

$$\hat{\beta} = (\mathbf{X}^t \mathbf{V}_y^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}_y^{-1} \mathbf{y}$$

Dies ist die Anwendung der **generalisierten least square (GLS)** Methode, wobei  $\mathbf{V}_y$  weiterhin unbekannt ist.

Nun ergibt aber das Einsetzen des GLS-Schätzers für  $\hat{\beta}$  die sogenannte **Profile-log-Likelihood**

$$l_P(\mathbf{V}_y) = -\frac{1}{2}[\log |\mathbf{V}_y| + \mathbf{y}^t \mathbf{V}_y^{-1} \{\mathbf{I} - \mathbf{X}(\mathbf{X}^t \mathbf{V}_y^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}_y^{-1}\} \mathbf{y}] - \frac{n}{2} \log(2\pi)$$

bzw.

$$l_P(\mathbf{V}_y) = -\frac{1}{2}[\log |\mathbf{V}_y| + \mathbf{y}^t \mathbf{V}_y^{-1} \mathbf{y} - \mathbf{y}^t \mathbf{P}^* \mathbf{y}] - \frac{n}{2} \log(2\pi)$$

Durch Maximierung von  $l_P(\mathbf{V}_y)$  mittels iterativer, numerischer Methoden, wie dem Fisher-Scoring-Algorithmus findet sich ein Schätzer für die Kovarianzmatrix  $\mathbf{V}_y$ . Über die obige **GLS**-Formel für  $\hat{\beta}$  lassen sich dann die weiteren (fixen) Parameter bestimmen.

Bei dieser klassischen Form der ML-Schätzung ist allerdings mit einer nicht unerheblichen Verzerrung zu rechnen, die aus den verlorenen Freiheitsgraden durch die Schätzung von  $\hat{\beta}$  resultiert.

### 3.3.2 Der unverzerzte REML-Schätzer

Diese Verzerrung lässt sich bei Verwendung der sogenannten “Restricted Maximum Likelihood” (**REML**) vermeiden. Hierbei wird die (modifizierte) Verteilung der  $\hat{\epsilon}$  der fixen Effekte. Damit ergibt sich die **restricted-log-likelihood**  $l_R(\mathbb{V}_y)$ :

$$l_R(\mathbb{V}_y) = l_P(\mathbb{V}_y) - \frac{1}{2} \log |\mathbf{X}^t \mathbb{V}_y \mathbf{X}|$$

Mit dieser Formel lassen sich anschliessend z. B. mit dem Fisher-Scoring-Algorithmus die Kovarianzmatrix  $\mathbb{V}_y$  und die Parameter  $\hat{\beta}$  bestimmen.

### 3.3.3 Bestimmung der zufälligen Parameter

Während in vielen Fragestellungen, in denen gemischte Modelle verwendet werden, vor allem die Kovarianzmatrix  $G$  interessiert, benötigt man für die Verwendung als Glättungsverfahren auch einen Schätzer für die zufälligen Werte. Die Formel aus Kapitel (3.1) erweist sich dabei häufig als sehr aufwändig. Eine Alternative dazu ist die Regression von  $\mathbf{u}$  auf  $\mathbf{y}$ . Die zugehörige Formel lautet:

$$\hat{\mathbf{u}} = \mathbf{GZ}^t \mathbb{V}_y^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta})$$

## 4 Regressionsmodelle als Smoother

### 4.1 Beispiel mit fixen und gemischten Effekten

Sei  $y_i = \sin(3x_i\pi) + 0.4\epsilon_i$ , mit  $x_i \sim \mathcal{U}[0, 1]$ ,  $\epsilon_i \sim \mathcal{N}(0, 1)$ . Dann lassen sich die Werte mittels eines linearen Splines glätten (Beispiel nach: Wand, 2002, Seite 6):

$$y_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^K u_k (x_i - \kappa_k)_+ + \epsilon_i$$

Zum Vergleich werden nun folgende Modelle verwendet:

- nur fixe Effekte: die  $u_k$  werden wie die  $\beta$ 's mit der kleinsten Quadrate Methode geschätzt.
- mit Zufallseffekten in den Splines:  $u_i \sim \mathcal{N}(0, \sigma_U^2)$  unabhängig.

In Abbildung 2 ist bei a) die Periodizität der Sinusfunktion noch grob zu erkennen, die Approximation der Kurve ist allerdings sehr “zackig”. Dagegen sind die einzelnen Splines bei der Approximation mit dem gemischten Modell wesentlich besser angepasst, die Kurve liegt ohne große Sprünge sehr nahe an der Originalkurve.

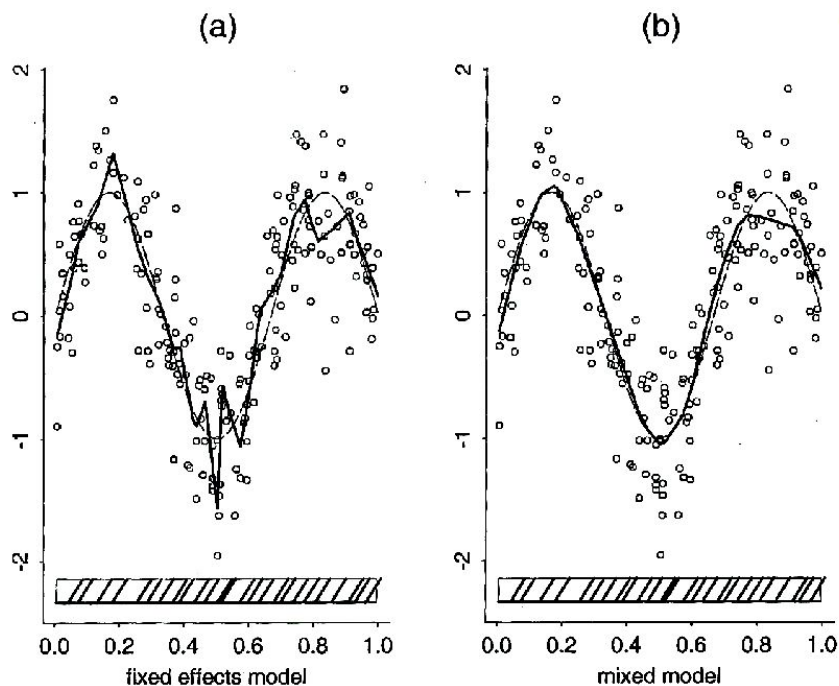


Abbildung 2: Vergleich zwischen fixem (a) und gemischtem (b) Regressionsmodell

## 4.2 Gründe für die Überlegenheit des gemischten Modells

Wegen  $u_i \sim \mathcal{N}(0, \sigma_u^2)$  sind "große" Parameter unwahrscheinlich. Betragsmäßig hohe Steigungsänderungen werden also ebenfalls penalisiert, die Kurve wird glatter. Zudem wird die Varianz der Zufallsparameter aus den Daten geschätzt (bei großer Streuung sind höhere Parameter wahrscheinlicher). Die Äquivalenz zum P-spline ist beweisbar (Speed, 1991; Seite 43f).

## 4.3 Glätten additiver Modelle

Das Glätten mittels gemischten Modellen lässt sich allgemein für stetige Funktionen  $f$  und  $g$  mit  $y_i = \beta_0 + f(s_i) + g(t_i) + \epsilon_i$  anwenden.

Zunächst können die Funktionen wieder durch Regressions-splines gefittet werden:

$$y_i = \beta_0 + \beta_s s_i + \sum_{k=1}^{K_s} u_k^s (s_i - \kappa_k^s)_+ + \beta_t t_i + \sum_{k=1}^{K_t} u_k^t (t_i - \kappa_k^t)_+ + \epsilon_i$$

mit  $u_k^s \sim \mathcal{N}(0, \sigma_s^2)$ ,  $u_k^t \sim \mathcal{N}(0, \sigma_t^2)$  unabhängig

Damit ergibt sich für die Designmatrizen

$$\mathbf{X} = [1 \ s_i \ t_i]_{1 \leq i \leq n}, \mathbf{Z} = \left[ \underbrace{(s_i - \kappa_k^s)_+}_{1 \leq k \leq K^s}, \underbrace{(t_i - \kappa_k^t)_+}_{1 \leq k \leq K^t} \right]_{1 \leq i \leq n}$$

Das gemischte Modell lautet dann:  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$  mit

$$E \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \mathbf{0}, Cov \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \sigma_s^2 \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_t^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma^2 \mathbf{I} \end{bmatrix}$$

Die Knoten  $\kappa_k^s, \kappa_k^t$  können mit relativ einfachen Methoden (z. B. Quantil-Knoten) bestimmt werden (siehe beispielsweise das Seminar-Referat von Gerhardinger, 2003, S. 6).

## 5 Ausblicke

### 5.1 Einige Ergänzungen

Anstelle von truncated power series  $(x_i - \kappa_k)_+^q$  sind beliebige Basisfunktionen  $\phi_j(x_i)$  einsetzbar.

Zudem läßt sich das Glättungsergebnis (bei etwa identischen Knotenabständen) durch korrelierte zufällige Effekte verbessern.

$$-1 \leq \rho \leq 1, \mathbf{G} = \sigma_U^2 \begin{bmatrix} 1 & \rho^1 & \rho^2 & \dots & \rho^{n-1} \\ \rho^1 & 1 & \rho^1 & \dots & \rho^{n-2} \\ \rho^2 & \rho^1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & \dots & 1 \end{bmatrix}$$

### 5.2 Radiale Basen

Radiale Basen (univariat) lassen sich wie folgt einbinden:  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_k \mathbf{u} + \boldsymbol{\epsilon}$ ,  $Cov(\mathbf{u}) = \sigma_u^2 (\boldsymbol{\Omega}_k^{-1/2}) (\boldsymbol{\Omega}_k^{-1/2})^t$  mit  $\mathbf{Z}_k = \left[ |x_i - \kappa_k|^{2m-1} \right]_{1 \leq i \leq n}$  und  $\boldsymbol{\Omega}_K = \left[ |\kappa_k - \kappa_{k'}|^{2m-1} \right]_{1 \leq k, k' \leq K}$

Durch die Transformation  $\mathbf{Z} = \mathbf{Z}_k \boldsymbol{\Omega}_K^{-1/2}$  läßt sich die klassische MM-Theorie auf  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$  anwenden.

Wenn man statt der euklidischen Norm statt dem Betrag nutzt, läßt sich das Konzept der radialen Basen sehr einfach auf mehrdimensionale Daten übertragen.

## 6 Literaturverzeichnis

Gerhardinger, U. (2003) Entwicklung in Basisfunktionen, [http://www.stat.uni-muenchen.de/krause/COURSES/Seminar\\_SS03/materialien/gerhardinger.pdf](http://www.stat.uni-muenchen.de/krause/COURSES/Seminar_SS03/materialien/gerhardinger.pdf)

Robinson, G.K. (1991) That BLUP is a good thing: The estimation of random effects (with discussion). *Statistical Science* **6**, 15-51

Speed, Terry (1991) Comment on Robinson (1991)

Wand, M.P. (2002) Smoothing and mixed models, <http://www.maths.unsw.edu.au/wand/papers.html>.