

Inhalt

- 1 Das lineare gemischte Modell
- 2 Likelihood-Schätzung für lineare gemischte Modelle
 - Schätzung der festen und Vorhersage der zufälligen Effekte
 - Schätzung der Kovarianzstruktur
 - Numerische Berechnung der Schätzer
- 3 Likelihood-Inferenz im linearen gemischten Modell
- 4 Bayes-Schätzung für lineare gemischte Modelle
- 5 Additive gemischte Modelle
- 6 Das generalisierte lineare gemischte Modell
- 7 Likelihood-Schätzung für generalisierte lineare gemischte Modelle

Schätzung der festen Effekte

Für die Schätzung der festen Effekte β verwenden wir die marginale Verteilung (9), $\mathbf{y} \sim N(\mathbf{X}\beta, \mathbf{V})$.

Wir nehmen zunächst an, dass die Kovarianzen $\text{Cov}(\mathbf{b}) = \mathbf{G}$ und $\text{Cov}(\varepsilon) = \mathbf{R}$, und damit $\text{Cov}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$, bekannt sind. Dann ist (9) ein allgemeines lineares Modell. Das verallgemeinerte Kleinste-Quadrate (KQ)-Kriterium

$$(\mathbf{y} - \mathbf{X}\beta)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) \rightarrow \min_{\beta}$$

für β ergibt den verallgemeinerten KQ-Schätzer (Aitken-Schätzer)

$$\hat{\beta} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}. \quad (10)$$

(Wir nehmen hier an, dass die Inversen \mathbf{V}^{-1} und $(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1}$ existieren. Verallgemeinerungen mit generalisierten Inversen existieren.)

Schätzung der festen Effekte

Dies ist gleichzeitig der Maximum-Likelihood (ML)-Schätzer, wenn wir die Normalverteilungsannahme treffen: Die log-Likelihood für β aus dem marginalen Modell (9) ist

$$l(\beta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\beta)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta). \quad (11)$$

$$\frac{d}{d\beta} l(\beta) = \mathbf{X}' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) \stackrel{!}{=} \mathbf{0} \quad \Rightarrow \quad \hat{\beta} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}.$$

Optimalität des Schätzers für β

$\hat{\beta}$ ist der BLUE, der beste lineare erwartungstreue Schätzer (best linear unbiased estimator) für β . **Beweis:**

Erwartungstreue $E(\hat{\beta}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta = \beta$ ist klar.

Sei $\tilde{\beta} = \mathbf{a} + \mathbf{B}\mathbf{y}$ ein anderer linearer erwartungstreuer Schätzer für β .

$\Rightarrow E(\tilde{\beta}) = \mathbf{a} + \mathbf{B}\mathbf{X}\beta \stackrel{!}{=} \beta \quad \forall \beta \Rightarrow \mathbf{a} = \mathbf{0}$ und $\mathbf{B}\mathbf{X} = \mathbf{I}_p$.

\Rightarrow Für alle $\mathbf{c} \in \mathbb{R}^p$ gilt

$$\begin{aligned} & \text{Var}(\mathbf{c}'\tilde{\beta}) - \text{Var}(\mathbf{c}'\hat{\beta}) \\ &= \mathbf{c}'\mathbf{B}\mathbf{V}\mathbf{B}'\mathbf{c} - \mathbf{c}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{c} \\ &= \mathbf{c}'\mathbf{B}\mathbf{V}\mathbf{B}'\mathbf{c} - \mathbf{c}'\mathbf{B}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}'\mathbf{c} \\ &= \mathbf{c}'\mathbf{B}\mathbf{V}^{1/2}[\mathbf{I}_n - \mathbf{V}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1/2}]\mathbf{V}^{1/2}\mathbf{B}'\mathbf{c} \geq 0, \end{aligned}$$

da die mittlere Matrix eine Projektionsmatrix (idempotent) und damit positiv semi-definit ist. Damit ist $\hat{\beta}$ der lineare erwartungstreue Schätzer mit der kleinsten Varianz. □

Optimalität des Schätzers für β

- Unter der Normalverteilungsannahme kann man sogar zeigen, dass $\hat{\beta}$ der beste erwartungstreue Schätzer ist.
- Mit der Transformation $\mathbf{X}^* = \mathbf{V}^{-1/2}\mathbf{X}$, $\mathbf{y}^* = \mathbf{V}^{-1/2}\mathbf{y}$, $\boldsymbol{\varepsilon}^* = \mathbf{V}^{-1/2}\boldsymbol{\varepsilon}$ gilt $\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$ mit $\boldsymbol{\varepsilon}^* \sim N(\mathbf{0}, \mathbf{I}_n)$.

Somit kann das marginale Modell $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ auf das gewöhnliche lineare Modell $\mathbf{y}^* \sim N(\mathbf{X}^*\boldsymbol{\beta}, \mathbf{I}_n)$ zurückgeführt werden.

Die Optimalitätseigenschaften von $\hat{\beta}$ ergeben sich dann unmittelbar aus dem Gauß-Markov-Theorem für das gewöhnliche lineare Modell.

Prädiktion der zufälligen Effekte

Oft ist es von Interesse, auch Vorhersagen für die zufälligen Effekte \mathbf{b} zu erhalten. Dazu benötigen wir die hierarchische Modellformulierung (8).

Unter der Normalverteilungsannahme ist der bedingte Erwartungswert $\hat{\mathbf{b}} := E(\mathbf{b}|\mathbf{y})$ von \mathbf{b} , gegeben die Daten \mathbf{y} , die **beste lineare erwartungstreue Vorhersage** (BLUP, best linear unbiased prediction) für \mathbf{b} .

- $\hat{\mathbf{b}}$ ist wegen des Satzes vom iterierten Erwartungswert **unverzerrt**:

$$E(\hat{\mathbf{b}}) = E(E(\mathbf{b}|\mathbf{y})) = E(\mathbf{b}) = \mathbf{0}.$$

Für die zufälligen Effekte wird hierbei Unverzerrtheit als $E(\hat{\mathbf{b}}) = E(\mathbf{b})$ definiert und nicht etwa als $E(\hat{\mathbf{b}}|\mathbf{b}) = \mathbf{b}$ für alle \mathbf{b} .

Prädiktion der zufälligen Effekte

- $\hat{\mathbf{b}} = E(\mathbf{b}|\mathbf{y})$ minimiert $E[(\tilde{\mathbf{b}} - \mathbf{b})'(\tilde{\mathbf{b}} - \mathbf{b})]$ in der Klasse der unverzerrten, linearen Schätzer $\tilde{\mathbf{b}}$:

$$\begin{aligned}
 E[(\tilde{\mathbf{b}} - \mathbf{b})'(\tilde{\mathbf{b}} - \mathbf{b})] &= E[(\tilde{\mathbf{b}} - \hat{\mathbf{b}} + \hat{\mathbf{b}} - \mathbf{b})'(\tilde{\mathbf{b}} - \hat{\mathbf{b}} + \hat{\mathbf{b}} - \mathbf{b})] \\
 &= E[(\tilde{\mathbf{b}} - \hat{\mathbf{b}})'(\tilde{\mathbf{b}} - \hat{\mathbf{b}})] + 2E[(\tilde{\mathbf{b}} - \hat{\mathbf{b}})'(\hat{\mathbf{b}} - \mathbf{b})] \\
 &\quad + E[(\hat{\mathbf{b}} - \mathbf{b})'(\hat{\mathbf{b}} - \mathbf{b})] \\
 &= E[(\tilde{\mathbf{b}} - \hat{\mathbf{b}})'(\tilde{\mathbf{b}} - \hat{\mathbf{b}})] + E[(\hat{\mathbf{b}} - \mathbf{b})'(\hat{\mathbf{b}} - \mathbf{b})] \\
 &\geq E[(\hat{\mathbf{b}} - \mathbf{b})'(\hat{\mathbf{b}} - \mathbf{b})],
 \end{aligned}$$

da $E[(\tilde{\mathbf{b}} - \hat{\mathbf{b}})'(\hat{\mathbf{b}} - \mathbf{b})] = E_{\mathbf{y}} E_{\mathbf{b}|\mathbf{y}}[(\tilde{\mathbf{b}} - \hat{\mathbf{b}})'(E(\mathbf{b}|\mathbf{y}) - \mathbf{b})] = \mathbf{0}$.

Gleichheit gilt für $\tilde{\mathbf{b}} = \hat{\mathbf{b}}$.

Prädiktion der zufälligen Effekte

Um $\hat{\mathbf{b}}$ explizit zu bestimmen, betrachte die gemeinsame Verteilung von \mathbf{y} und \mathbf{b} . Diese ergibt sich aus dem zweistufigen hierarchischen Modell für $\mathbf{y}|\mathbf{b}$ und \mathbf{b} zu

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{b} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V} & \mathbf{Z}\mathbf{G} \\ \mathbf{G}\mathbf{Z}' & \mathbf{G} \end{pmatrix} \right). \quad (12)$$

Herleitung für Interessierte: \rightarrow **Extrablatt**.

Klar: Aus (12) folgt wiederum (Satz B.4, Fahrmeir et al., 2009)

- $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$
- $\mathbf{y}|\mathbf{b} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{V} - \mathbf{Z}\mathbf{G}\mathbf{G}^{-1}\mathbf{G}\mathbf{Z}' = \mathbf{R})$.

Aus (12) folgt (Satz B.4, Fahrmeir et al., 2009) der bedingte Erwartungswert

$$E(\mathbf{b}|\mathbf{y}) = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (13)$$

In der Praxis wird der unbekanntes Vektor $\boldsymbol{\beta}$ in (13) durch den verallgemeinerten KQ-Schätzer $\hat{\boldsymbol{\beta}}$ aus (10) ersetzt,

$$\hat{\mathbf{b}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

Henderson's mixed model equations

BLUE und BLUP (gemeinsam auch mit BLUP bezeichnet) $\hat{\beta}$ und $\hat{\mathbf{b}}$ sind die Lösung der sogenannten Henderson's mixed model equations

$$\begin{aligned} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\beta} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{b}} &= \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\beta} + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{b}} &= \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{aligned} \quad (14)$$

Für diese Schätzgleichungen gibt es mehrere Herleitungen, u.a. folgende drei:

- 1 $\hat{\beta}$ und $\hat{\mathbf{b}}$ haben minimale Varianz unter den linearen erwartungstreuen Schätzern/Vorhersagen. (Analog: Linearkombinationen, z.B. $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\mathbf{b}}$.)
- 2 Eine bayesianische Herleitung (mehr in Kapitel 4).
- 3 Die Schätzgleichungen ergeben sich bei Maximierung der **gemeinsamen Likelihood** $L(\mathbf{y}, \mathbf{b})$ von \mathbf{y} und \mathbf{b} simultan bezüglich β und \mathbf{b} .

Gemeinsame / penalisierte Likelihood

Wegen $p(\mathbf{y}, \mathbf{b}) = p(\mathbf{y}|\mathbf{b})p(\mathbf{b})$ ergibt sich (bis auf additive Konstanten)

$$l(\mathbf{y}, \mathbf{b}) = \log L(\mathbf{y}, \mathbf{b}) = -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})' \mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) - \frac{1}{2}\mathbf{b}' \mathbf{G}^{-1}\mathbf{b}. \quad (15)$$

Diese kann als **penalisierte Likelihood** aufgefasst werden mit **Strafterm** $\mathbf{b}' \mathbf{G}^{-1}\mathbf{b}$.
Die Maximierung ist äquivalent zur Minimierung des **penalisierten KQ-Kriteriums**

$$KQ_{pen}(\boldsymbol{\beta}, \mathbf{b}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})' \mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) + \mathbf{b}' \mathbf{G}^{-1}\mathbf{b}. \quad (16)$$

- Der erste Term entspricht einem verallgemeinerten KQ-Kriterium.
- Der zweite Term berücksichtigt die **Verteilungsannahme** für \mathbf{b} und bestraft ($\mathbf{b}' \mathbf{G}^{-1}\mathbf{b} \geq 0$ wegen $\mathbf{G} \geq 0$ als Kovarianz) Abweichungen von $E(\mathbf{b}) = \mathbf{0}$.
 - Für $\mathbf{G} = \mathbf{0}$ gilt $\hat{\mathbf{b}} = \mathbf{0}$.
 - Für $\mathbf{G}^{-1} \rightarrow \mathbf{0}$ geht $\mathbf{b}' \mathbf{G}^{-1}\mathbf{b} \rightarrow 0$ und \mathbf{b} wird wie ein fester Effekt geschätzt.

Differenzieren von $KQ_{pen}(\boldsymbol{\beta}, \mathbf{b})$ und Nullsetzen der Ableitungen ergibt (14).

Definiert man $\mathbf{C} = (\mathbf{X}|\mathbf{Z})$ und die partitionierte Matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix}, \quad (17)$$

so lässt sich die Lösung der Schätzgleichung (14) auch in der folgenden kompakten Form schreiben

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{pmatrix} = (\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{A})^{-1}\mathbf{C}'\mathbf{R}^{-1}\mathbf{y}. \quad (18)$$

In dieser Form erkennt man die enge Beziehung zur [Ridge-Schätzung](#). Ebenso besteht ein enger Zusammenhang zur (empirischen) Bayes-Schätzung (mehr dazu später).

Shrinkage

$\hat{\mathbf{b}}$ ist ein **Shrinkage-Schätzer** für \mathbf{b} , d.h. seine Komponenten haben eine geringere Streuung als sie hätten, wenn \mathbf{b} als feste Effekte behandelt würden.

Dies scheint auf den ersten Blick der Eigenschaft der Unverzerrtheit des BLUP zu widersprechen. Erwartungstreue bedeutet hier jedoch

$$E(\hat{\mathbf{b}}) = E(\mathbf{b}) = \mathbf{0},$$

und nicht

$$E(\hat{\mathbf{b}}|\mathbf{b}) = \mathbf{b} \text{ für alle } \mathbf{b}.$$

Shrinkage

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

Der BLUP für \mathbf{y} ist also ein **gewichtetes Mittel** des Populationsmittels $\mathbf{X}\hat{\boldsymbol{\beta}}$ und der individuellen Daten \mathbf{y} . Die beobachteten Daten werden gegen das Populationsmittel „geschrumpft“ (“borrowing of strength”).

Beachte, dass $\mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}'$.

- Wenn die Varianz der Residuen \mathbf{R} im Verhältnis groß ist, wird viel Gewicht auf das Populationsmittel $\mathbf{X}\hat{\boldsymbol{\beta}}$ gelegt.
- Wenn \mathbf{R} klein ist, gilt das Gegenteil.

Shrinkage - Varianz-Bias-Tradeoff

Beispiel Random-Intercept-Modell (ohne Kovariablen)

$$y_{ij} = b_i + \varepsilon_{ij}, \quad b_i \stackrel{iid}{\sim} N(0, \tau^2), \quad \varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2), \\ j = 1, \dots, n_i, i = 1, \dots, N.$$

Für den BLUP $\hat{\mathbf{b}}^{BLUP}$ und den KQ-Schätzer $\hat{\mathbf{b}}^{KQ}$, der \mathbf{b} als feste Effekte schätzt, gilt

	$\hat{\mathbf{b}}^{BLUP}$	$\hat{\mathbf{b}}^{KQ}$	
\hat{b}_i			
$E(\hat{b}_i b_i)$			
$\text{Var}(\hat{b}_i b_i)$			
$\text{MSE}(\hat{b}_i)$			

$$\text{MSE}(\hat{b}_i) = E_b E_{y|b} [(\hat{b}_i - b_i)^2] = E_b [\text{Var}(\hat{b}_i | b_i) + (E(\hat{b}_i | b_i) - b_i)^2].$$

Schätzung der Kovarianzstruktur

Sei ϑ der Parametervektor, der alle unbekannt Parameter in den Kovarianzmatrizen $\mathbf{R} = \mathbf{R}(\vartheta)$, $\mathbf{G} = \mathbf{G}(\vartheta)$ und $\mathbf{V} = \mathbf{V}(\vartheta)$ enthält.

Indem man einen (konsistenten) Schätzer $\hat{\vartheta}$ einsetzt, erhält man die geschätzten Kovarianzmatrizen

$$\hat{\mathbf{R}} = \mathbf{R}(\hat{\vartheta}), \quad \hat{\mathbf{G}} = \mathbf{G}(\hat{\vartheta}), \quad \hat{\mathbf{V}} = \mathbf{V}(\hat{\vartheta}),$$

die für die Berechnung der BLUPs von β und \mathbf{b} verwendet werden können.

Maximum-Likelihood (ML)-Schätzung von ϑ

Die **ML-Schätzung** von ϑ basiert auf der **Likelihood** des marginalen Modells

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

Die log-Likelihood für $\boldsymbol{\beta}$ und ϑ ist bis auf additive Konstanten gegeben durch

$$l(\boldsymbol{\beta}, \vartheta) = -\frac{1}{2} \left\{ \log |\mathbf{V}(\vartheta)| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}(\vartheta)^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}.$$

Maximiert man $l(\boldsymbol{\beta}, \vartheta)$ für festes ϑ bezüglich $\boldsymbol{\beta}$, so erhält man (wieder)

$$\hat{\boldsymbol{\beta}}(\vartheta) = \arg \max_{\boldsymbol{\beta}} (l(\boldsymbol{\beta}, \vartheta)) = (\mathbf{X}' \mathbf{V}(\vartheta)^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}(\vartheta)^{-1} \mathbf{y}.$$

Einsetzen von $\hat{\boldsymbol{\beta}}(\vartheta)$ in $l(\boldsymbol{\beta}, \vartheta)$ liefert die nur von ϑ abhängige **Profil-log-Likelihood**

$$l_P(\vartheta) = l(\hat{\boldsymbol{\beta}}(\vartheta), \vartheta) = -\frac{1}{2} \left\{ \log |\mathbf{V}(\vartheta)| + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\vartheta))' \mathbf{V}(\vartheta)^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\vartheta)) \right\}.$$

Maximierung von $l_P(\vartheta)$ bezüglich ϑ liefert den **ML-Schätzer** $\hat{\vartheta}_{ML}$.

Motivation REML-Schätzung

ML-Schätzer für Varianzen sind (nach unten) verzerrt, da der Verlust von Freiheitsgraden durch die Schätzung von $\hat{\beta}$ unberücksichtigt bleibt.

Beispiel lineares Modell $Y \sim N(\mathbf{X}\beta, \sigma^2 I_n)$

- ML-Schätzer ist $\hat{\sigma}^2 = \frac{1}{n} \hat{\epsilon}' \hat{\epsilon}$, aber $E(\hat{\sigma}^2) = \frac{n-p}{n} \sigma^2$.
- Benutzt wird üblicherweise der erwartungstreue Schätzer $\hat{\sigma}^2 = \frac{1}{n-p} \hat{\epsilon}' \hat{\epsilon}$.

Dies ist der **restringierte Maximum Likelihood**-Schätzer. Er berücksichtigt den Verlust an Freiheitsgraden durch Schätzung von β .

REstringierte Maximum-Likelihood (REML) Schätzung von ϑ

Grundidee: Basiere Likelihood nicht auf \mathbf{y} , sondern auf $(n - p)$ linear unabhängigen Fehlerkontrasten $\mathbf{A}\mathbf{y}$, deren Verteilung von β unabhängig ist.

Teile die log-Likelihood in zwei Teile auf für $\mathbf{A}\mathbf{y}$ (für ϑ) und $\mathbf{B}\mathbf{y}$ (für β) mit

- 1 \mathbf{A} von Rang $n - p$ und \mathbf{B} von Rang p .
- 2 Die zwei Teile sind statistisch unabhängig, also hier $\text{Cov}(\mathbf{A}\mathbf{y}, \mathbf{B}\mathbf{y}) = \mathbf{0}$.
- 3 $\mathbf{A}\mathbf{y}$ sind Fehlerkontraste, d.h. $E(\mathbf{A}\mathbf{y}) = \mathbf{0}$.
- 4 $\mathbf{B}\mathbf{X}$ hat Rang p ($\Rightarrow \mathbf{B}\mathbf{y}$ schätzt eine eindeutige Funktion von β).

Geeignete Matrizen: $\mathbf{A} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$, $\mathbf{B} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$.

→ Englischer Name **residual** oder **restricted maximum likelihood estimation**.

Aus den ersten zwei Punkten folgt, dass die log-Likelihood $l = l' + l^*$ ist (bis auf additive Konstanten), l' für $\mathbf{A}\mathbf{y}$ und l^* für $\mathbf{B}\mathbf{y}$.

Maximierung von l^* (aus $\mathbf{B}\mathbf{y}$) bzgl. β ergibt den MLE/BLUE $\hat{\beta}$ (abhängig von ϑ).

Maximierung von l' (aus $\mathbf{A}\mathbf{y}$) bzgl. ϑ : Die restringierte log-Likelihood ist (bis auf additive Konstanten) unabhängig von dem gewählten Fehlerkontrast

$$\begin{aligned} l_R(\vartheta) &= -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}| - \frac{1}{2} (\mathbf{y} - \mathbf{X} \hat{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}) \\ &= l(\hat{\beta}(\vartheta), \vartheta) - \frac{1}{2} \log |\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}| \end{aligned} \quad (19)$$

Herleitung für Interessierte: → **Extrablatt**

Maximierung ergibt den restringierten ML- (REML)-Schätzer $\hat{\vartheta}_{REML}$ für ϑ .

Alternative Herleitung des REML-Schätzers

Alternativ maximiert der REML-Schätzer $\hat{\vartheta}_{REML}$ die **marginale log-Likelihood** für ϑ , bei der β aus der Likelihood $L(\beta, \vartheta)$ des marginalen Modells herausintegriert wird

$$l_R(\vartheta) = \log \left(\int L(\beta, \vartheta) d\beta \right). \quad (20)$$

Diese ergibt sich bis auf additive Konstanten wieder als (19). → **Übung**

Auch beim LMM ist die **Reduktion der Verzerrung** von $\hat{\vartheta}_{ML}$ der Hauptgrund für die Verwendung von $\hat{\vartheta}_{REML}$ als Schätzer für ϑ . Es ist jedoch allgemein nicht gesichert, ob auch der mean squared error (MSE) geringer wird.

Empirische Versionen von BLUP und BLUE

Mit den durch **(RE)ML** geschätzten Kovarianzmatrizen $\hat{\mathbf{R}}$, $\hat{\mathbf{G}}$ und $\hat{\mathbf{V}}$ ergeben sich empirische Versionen von BLUP und BLUE, **EBLUP** und **EBLUE**, für \mathbf{b} und β ,

$$\hat{\beta} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}, \quad \hat{\mathbf{b}} = \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}).$$

- Im Gegensatz zum linearen Modell sind die Schätzer der festen Effekte β im LMM abhängig von der Kovarianzmatrix \mathbf{V} . Daher sind die Schätzer $\hat{\beta}(\hat{\vartheta}_{REML})$ und $\hat{\beta}(\hat{\vartheta}_{ML})$ **nicht identisch**.
- Nach Einsetzen von $\hat{\vartheta}$ gelten die Optimalitätseigenschaften nicht mehr exakt.
- Ebenso sind die Varianzen bzw. Kovarianzmatrizen der Schätzer analytisch nicht exakt zugänglich (mehr später).

Numerische Berechnung der Schätzer

- Die Maximierung von $l_R(\vartheta)$ [$l_P(\vartheta)$] zur Berechnung von $\hat{\vartheta}_{REML}$ [$\hat{\vartheta}_{ML}$] erfolgt wegen Nichtlinearität in ϑ meist per [Newton-Raphson](#) oder [Fisher-Scoring](#).
- Dabei können diese Algorithmen keine Restriktion von Parametern auf $[0, \infty)$ berücksichtigen, sodass negative Schätzer für Varianzen entstehen könnten.
- Daher maximiert `lme()` in R-Paket `nlme` die (restringierte) log-Likelihood bzgl. der skalierten log-Varianzen. Allerdings kann diese Methode ein Maximum in der Null (kein zufälliger Effekt) nicht finden.
- `lmer()` in R-Paket `lme4` verwendet eine Parametrisierung, die über eine Nebenbedingung positiv (semi-)definites \mathbf{G} sicherstellt und singuläres \mathbf{G} erlaubt. Die Funktion verwendet Optimierung mit Nebenbedingungen statt Newton-Raphson.

Numerische Berechnung der Schätzer

- Die Bedingungen an ϑ unterscheiden sich zwischen marginalem (\mathbf{V} positiv (semi-)definit) und konditionalem Modell (\mathbf{G} und \mathbf{R} positiv (semi-)definit). Softwarepakete maximieren bzgl. ϑ zum Teil über einen größeren Parameterraum als das konditionale Modell implizieren würde.
- Die historisch erste Alternative ist der EM-Algorithmus, der die zufälligen Effekte als fehlende Daten behandelt. Die Konvergenz kann langsam sein.

Likelihood-Schätzung in R

Das klassische Paket zur LMM-Schätzung in R ist `nlme`. Eine ausführliche Beschreibung findet man im Buch der Autoren, Pinheiro & Bates (2000).

Der iterative Optimierungsalgorithmus ist eine Hybridversion aus zunächst **EM**- und im Anschluss **Newton-Raphson(NR)-Algorithmus**.

- **EM-Algorithmus**: Iterationen schnell und einfach zu berechnen. Schnell nahe ans Optimum, Konvergenz nahe des Optimums potentiell sehr langsam.
- **NR-Algorithmus**: Iterationen sehr rechenintensiv. (Berechnung 1. und 2. Ableitung, Scorefunktion, Hessematrix, in jedem Schritt.) Möglicherweise instabil entfernt vom Optimum. Sehr schnelle Konvergenz nahe des Optimums.

Alternativen

- Für additive und generalisierte additive gemischte Modelle (mehr dazu in Kapitel 5) eignet sich das R-Paket `mgcv`, siehe auch Wood (2006).
- `lme4` ist eine neuere Weiterentwicklung von Douglas Bates und Koautoren, das lineare und generalisierte lineare gemischte Modelle fitten kann (unter Benutzung von S4-Klassen). Siehe auch Bates, Maechler, Bolker & Walker (2015). *Fitting Linear Mixed-Effects Models using lme4*. Journal of Statistical Software, angenommen. <http://arxiv.org/abs/1406.5823>.
- Viele Beispiele für lineare gemischte Modelle mit SAS `proc mixed` finden sich in Verbeke & Molenberghs, 2000.