Inhalt

Schätzung der festen Effekte

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

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

$$(\boldsymbol{y} - \boldsymbol{X}\beta)'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\beta) \to \min_{\boldsymbol{\beta}}$$

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

$$\widehat{\beta} = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}. \tag{10}$$

(Wir nehmen hier an, dass die Inversen $\boldsymbol{V}^{-1}$ und $\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}$ 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

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

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

Optimalität des Schätzers für β

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

Erwartungstreue $\mathrm{E}(\widehat{\beta}) = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}$ ist klar.

Sei $\widetilde{\boldsymbol{\beta}} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{y}$ ein anderer linearer erwartungstreuer Schätzer für $\boldsymbol{\beta}$.
$\Rightarrow \mathrm{E}(\widetilde{\beta}) = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{X}\boldsymbol{\beta} \overset{!}{=} \boldsymbol{\beta} \quad\Rightarrow\quad \boldsymbol{a} = \boldsymbol{0}$ und $\boldsymbol{B}\boldsymbol{X} = \boldsymbol{I_p}$.
$\Rightarrow$ Für alle $\boldsymbol{c} \in \mathbb{R}^p$ gilt

$$
\begin{aligned}
& \mathrm{Var}(\boldsymbol{c}'\widetilde{\beta}) - \mathrm{Var}(\boldsymbol{c}'\widehat{\beta}) \\
= \; & \boldsymbol{c}'\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}'\boldsymbol{c} - \boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{c} \\
= \; & \boldsymbol{c}'\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}'\boldsymbol{c} - \boldsymbol{c}'\boldsymbol{B}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{B}'\boldsymbol{c} \\
= \; & \boldsymbol{c}'\boldsymbol{B}\boldsymbol{V}^{1/2}[\boldsymbol{I_n} - \boldsymbol{V}^{-1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1/2}]\boldsymbol{V}^{1/2}\boldsymbol{B}'\boldsymbol{c} \geq 0,
\end{aligned}
$$

da die mittlere Matrix eine Projektionsmatrix und damit positiv semi-definit ist.
Damit ist $\widehat{\beta}$ der lineare erwartungstreue Schätzer mit der kleinsten Varianz. $\square$

Optimalität des Schätzers für β

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

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

  Die Optimalitätseigenschaften von $\widehat{\beta}$ 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 $\boldsymbol{b}$ zu erhalten. Dazu benötigen wir die hierarchische Modellformulierung (8).

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

- $\widehat{\boldsymbol{b}}$ ist wegen des Satzes vom iterierten Erwartungswert unverzerrt:

$$\mathrm{E}(\widehat{\boldsymbol{b}}) = \mathrm{E}(\mathrm{E}(\boldsymbol{b}|\boldsymbol{y})) = \mathrm{E}(\boldsymbol{b}) = \mathbf{0}.$$

  Für die zufälligen Effekte wird hierbei Unverzerrtheit als $\mathrm{E}(\widehat{\boldsymbol{b}}) = \mathrm{E}(\boldsymbol{b})$ definiert und <u>nicht</u> etwa als $\mathrm{E}(\widehat{\boldsymbol{b}}|\boldsymbol{b}) = \boldsymbol{b}$ für alle $\boldsymbol{b}$.

Prädiktion der zufälligen Effekte

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

$$
\begin{aligned}
\mathrm{E}[(\tilde{\boldsymbol{b}} - \boldsymbol{b})'(\tilde{\boldsymbol{b}} - \boldsymbol{b})] &= \mathrm{E}[(\tilde{\boldsymbol{b}} - \widehat{\boldsymbol{b}} + \widehat{\boldsymbol{b}} - \boldsymbol{b})'(\tilde{\boldsymbol{b}} - \widehat{\boldsymbol{b}} + \widehat{\boldsymbol{b}} - \boldsymbol{b})] \\
&= \mathrm{E}[(\tilde{\boldsymbol{b}} - \widehat{\boldsymbol{b}})'(\tilde{\boldsymbol{b}} - \widehat{\boldsymbol{b}})] + 2\,\mathrm{E}[(\tilde{\boldsymbol{b}} - \widehat{\boldsymbol{b}})'(\widehat{\boldsymbol{b}} - \boldsymbol{b})] \\
&\qquad + \mathrm{E}[(\widehat{\boldsymbol{b}} - \boldsymbol{b})'(\widehat{\boldsymbol{b}} - \boldsymbol{b})] \\
&= \mathrm{E}[(\tilde{\boldsymbol{b}} - \widehat{\boldsymbol{b}})'(\tilde{\boldsymbol{b}} - \widehat{\boldsymbol{b}})] + \mathrm{E}[(\widehat{\boldsymbol{b}} - \boldsymbol{b})'(\widehat{\boldsymbol{b}} - \boldsymbol{b})] \\
&\geq \mathrm{E}[(\widehat{\boldsymbol{b}} - \boldsymbol{b})'(\widehat{\boldsymbol{b}} - \boldsymbol{b})],
\end{aligned}
$$

da $\mathrm{E}[(\tilde{\boldsymbol{b}} - \widehat{\boldsymbol{b}})'(\widehat{\boldsymbol{b}} - \boldsymbol{b})] = \mathrm{E}_{\boldsymbol{y}}\, \mathrm{E}_{\boldsymbol{b}|\boldsymbol{y}}[(\tilde{\boldsymbol{b}} - \widehat{\boldsymbol{b}})'(\mathrm{E}(\boldsymbol{b}|\boldsymbol{y}) - \boldsymbol{b})] = \boldsymbol{0}$.

Gleichheit gilt für $\tilde{\boldsymbol{b}} = \widehat{\boldsymbol{b}}$.

Prädiktion der zufälligen Effekte

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

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

Herleitung für Interessierte: → **Extrablatt**.

Klar: Aus (12) folgt wiederum (Satz B.4, Fahrmeir et al., 2007)

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

Aus (12) folgt (Satz B.4, Fahrmeir et al., 2007) der bedingten Erwartungswert

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

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

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

Henderson's mixed model equations

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

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

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

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

Gemeinsame / penalisierte Likelihood

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

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

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

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

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

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

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

$$\boldsymbol{A} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G}^{-1} \end{pmatrix}, \tag{17}$$

so lässt sich die Lösung der Schätzgleichung (14) auch in der folgenden kompakten Form schreiben $\rightarrow$ **Übung**

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

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

Shrinkage

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

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

$$\mathrm{E}(\widehat{\boldsymbol{b}}) = \mathrm{E}(\boldsymbol{b}) = \mathbf{0},$$

und nicht

$$\mathrm{E}(\widehat{\boldsymbol{b}}|\boldsymbol{b}) = \boldsymbol{b} \;\text{ für alle }\; \boldsymbol{b}.$$

Shrinkage

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

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

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

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

Shrinkage - Varianz-Bias-Tradeoff

Beispiel Random-Intercept-Modell (ohne Kovariablen)

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

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

	$\hat{\boldsymbol{b}}^{BLUP}$	$\hat{\boldsymbol{b}}^{KQ}$	
$\hat{b}_i$			
$\mathrm{E}(\hat{b}_i \mid b_i)$			
$\mathrm{Var}(\hat{b}_i \mid b_i)$			
$\mathrm{MSE}(\hat{b}_i)$			

$$\mathrm{MSE}(\hat{b}_i) = \mathrm{E}_b\, \mathrm{E}_{y|b}[(\hat{b}_i - b_i)^2] = \mathrm{E}_b[\mathrm{Var}(\hat{b}_i|b_i) + (E(\hat{b}_i|b_i) - b_i)^2].$$

Schätzung der Kovarianzstruktur

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

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

$$\widehat{\boldsymbol{R}} = \boldsymbol{R}(\widehat{\boldsymbol{\vartheta}}), \quad \widehat{\boldsymbol{G}} = \boldsymbol{G}(\widehat{\boldsymbol{\vartheta}}), \quad \widehat{\boldsymbol{V}} = \boldsymbol{V}(\widehat{\boldsymbol{\vartheta}}),$$

die für die Berechnung der BLUPs von $\boldsymbol{\beta}$ und $\boldsymbol{b}$ verwendet werden können.

Maximum-Likelihood (ML)-Schätzung von ϑ

Die **ML-Schätzung** von $\boldsymbol{\vartheta}$ basiert auf der Likelihood des marginalen Modells

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

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

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

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

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

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

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

Maximierung von $l_P(\boldsymbol{\vartheta})$ bezüglich $\boldsymbol{\vartheta}$ liefert den ML-Schätzer $\widehat{\boldsymbol{\vartheta}}_{ML}$.

Motivation REML-Schätzung

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

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

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

  Dies ist der restringierte Maximum Likelihood-Schätzer. Er berücksichtigt den Verlust an Freiheitsgraden durch Schätzung von $\boldsymbol{\beta}$.

REstringierte Maximum-Likelihood (REML) Schätzung von ϑ

<u>Grundidee</u>: Basiere Likelihood nicht auf $\boldsymbol{y}$, sondern auf $(n-p)$ linear unabhängige Fehlerkontraste $\boldsymbol{Ay}$, deren Verteilung von $\boldsymbol{\beta}$ unabhängig ist.

Teile die log-Likelihood in zwei Teile auf für $\boldsymbol{Ay}$ (für ϑ) und $\boldsymbol{By}$ (für $\boldsymbol{\beta}$) mit

1. $\boldsymbol{A}$ von Rang $n-p$ und $\boldsymbol{B}$ von Rang p.
2. Die zwei Teile sind statistisch unabhängig, also hier $\text{Cov}(\boldsymbol{Ay}, \boldsymbol{By}) = \boldsymbol{0}$.
3. $\boldsymbol{Ay}$ sind <u>Fehlerkontraste</u>, d.h. $\text{E}(\boldsymbol{Ay}) = \boldsymbol{0}$.
4. $\boldsymbol{BX}$ hat Rang p ($\Rightarrow$ $\boldsymbol{By}$ schätzt eine eineindeutige Funktion von $\boldsymbol{\beta}$).

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

$\rightarrow$ 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 $\boldsymbol{Ay}$ und l^* für $\boldsymbol{By}$.

Maximierung von l^* (aus $\boldsymbol{By}$) bzgl. $\boldsymbol{\beta}$ ergibt den MLE/BLUE $\widehat{\boldsymbol{\beta}}$ (abhängig von $\boldsymbol{\vartheta}$).

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

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

Herleitung für Interessierte: $\rightarrow$ **Extrablatt**
Maximierung ergibt den restringierten ML- (REML)-Schätzer $\widehat{\boldsymbol{\vartheta}}_{REML}$ für $\boldsymbol{\vartheta}$.

Alternative Herleitung des REML-Schätzers

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

$$l_R(\boldsymbol{\vartheta}) = \log\left(\int L(\boldsymbol{\beta}, \boldsymbol{\vartheta}) d\boldsymbol{\beta}\right). \tag{20}$$

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

Auch beim LMM ist die Reduktion der Verzerrung von $\widehat{\boldsymbol{\vartheta}}_{ML}$ der Hauptgrund für die Verwendung von $\hat{\boldsymbol{\vartheta}}_{REML}$ als Schätzer für $\boldsymbol{\vartheta}$. 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 $\widehat{\boldsymbol{R}}$, $\widehat{\boldsymbol{G}}$ und $\widehat{\boldsymbol{V}}$ ergeben sich empirische Versionen von BLUP und BLUE, EBLUP und EBLUE, für $\boldsymbol{b}$ und $\boldsymbol{\beta}$,

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

- Im Gegensatz zum linearen Modell sind die Schätzer der festen Effekte $\boldsymbol{\beta}$ im LMM abhängig von der Kovarianzmatrix der Residuen $\boldsymbol{V}$. Daher sind die Schätzer $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}}_{REML})$ und $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}}_{ML})$ **nicht identisch**.
- Nach Einsetzen von $\widehat{\boldsymbol{\vartheta}}$ 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 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 z.B. `lme()` in R die (restringierte) log-Likelihood bzgl. der skalierten log-Varianzen (Parameterraum $\mathbb{R} = \log(0, \infty)$). Allerdings kann diese Methode ein Maximum in der Null (kein zufälliger Effekt) nicht finden.
- Die Bedingungen an ϑ unterscheiden sich zwischen marginalem ($\boldsymbol{V}$ positiv (semi-)definit) und konditionalem Modell ($\boldsymbol{G}$ und $\boldsymbol{R}$ positiv (semi-)definit). Softwarepakete maximieren bzgl. ϑ oft über einen größeren Parameterraum als das konditionale Modell implizieren würde.

Alternativen

- Die ML-Schätzer $\widehat{\boldsymbol{\beta}}_{ML}$ und $\widehat{\boldsymbol{\vartheta}}_{ML}$ kann man auch simultan durch die Maximierung der log-Likelihood $l(\beta, \boldsymbol{\vartheta})$ des marginalen Modells erhalten.
- Da laut (19)

$$l_R(\boldsymbol{\vartheta}) \quad = \quad l(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}), \boldsymbol{\vartheta}) - \frac{1}{2} \log |\boldsymbol{X}' \boldsymbol{V}^{-1} \boldsymbol{X}|,$$

mit $-\frac{1}{2} \log |\boldsymbol{X} \boldsymbol{V}^{-1} \boldsymbol{X}|$ unabhängig von $\boldsymbol{\beta}$, kann man $\widehat{\boldsymbol{\beta}}_{REML}$ und $\widehat{\boldsymbol{\vartheta}}_{REML}$ auch simultan durch die Maximierung der restringierten log-Likelihood erhalten,

$$l_R(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = l(\boldsymbol{\beta}, \boldsymbol{\vartheta}) - \frac{1}{2} \log |\boldsymbol{X}' \boldsymbol{V}(\boldsymbol{\vartheta})^{-1} \boldsymbol{X}|.$$

- Auch Schaukelalgorithmen sind möglich, die abwechselnd Iterationswerte $\widehat{\boldsymbol{\vartheta}}^{(k)}$ und $\widehat{\boldsymbol{\beta}}^{(k+1)} = \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}}^{(k)})$ sowie $\widehat{\boldsymbol{b}}^{(k+1)} = \widehat{\boldsymbol{b}}(\widehat{\boldsymbol{\vartheta}}^{(k)})$ aus (14) bestimmen.
- Eine Alternative ist der EM-Algorithmus, der die zufälligen Effekte als fehlende Daten behandelt. Dieser kann jedoch langsam konvergieren.

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 Paket `mgcv` in R von Simon Wood.
  Literatur dazu: Wood, S. N. (2006). *Generalized Additive Models. An Introduction with R.* Chapman & Hall/CRC, Boca Raton.

- `lme4` ist eine neuere Weiterentwicklung von Douglas Bates und Koautoren, das lineare und generalisierte lineare gemischte Modelle fitten kann (unter Benutzung von S4-Klassen).

- Viele Beispiele für lineare gemischte Modelle mit `SAS proc mixed` finden sich in Verbeke & Molenberghs, 2000.