Evaluation und Metaanalyse

MSc Klinische Psychologie und Psychotherapie

SoSe 2024

Prof. Dr. Dirk Ostwald

(6) Linear Mixed Models Schätzung

Motivation

Schaeuffele et al. (2024)

“Building on previous work, we expected considerable variability and thus used a random-effects model to account for heterogeneity of included studies (...) Across settings, TD-CBT revealed significantly stronger symptom reduction in depression ($g = 0.74$, 95% CI = 0.57 - 0.92, $P < 0.001$) (...) than controls at posttreatment”

```r
library(metafor)                                               # metafor R Paket
D       = read.csv("./6_Data/TD-CBT.csv")                      # Präprozessierte Schaueffele et al. Daten
res     = rma(yi = D$yi, vi = D$vi)                            # metafor random effects
```

```
Random-Effects Model (k = 63; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.4015 (SE = 0.0875)
tau (square root of estimated tau^2 value):      0.6337
I^2 (total heterogeneity / total variability):   88.29%
H^2 (total variability / sampling variability):  8.54

Test for Heterogeneity:
Q(df = 62) = 344.9632, p-val < .0001

Model Results:

estimate      se    zval    pval   ci.lb   ci.ub
  0.7445  0.0885  8.4169  <.0001  0.5712  0.9179  ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Motivation

Schaeuffele et al. (2024)

```r
library(metafor)                                    # metafor R Paket
D       = read.csv("./6_Data/TD-CBT.csv")           # Präprozessierte Schaueffele et al. Daten
res     = rma(yi = D$yi, vi = D$vi)                 # metafor random effects
```

Viechtbauer (2010) Conducting Meta-Analyses in R with the `metafor` package

"Assuming the observed outcomes and corresponding sampling variances are supplied via `yi` and `vi`, the random-effects model is fitted with `rma(yi, vi, data = dat)`. Restricted maximum-likelihood estimation is used by default when estimating τ^2 (the REML estimator is approximately unbiased and quite efficient; see Viechtbauer (2005)). While the HS, HE, DL, and SJ estimators of $\hat{\tau}^2$ are based on closed-form solutions, the ML, REML, and EB estimators must be obtained numerically. For this, the `rma()` function makes use of the Fisher scoring algorithm, which is robust to poor starting values and usually converges quickly (Harville (1977); Jennrich and Sampson (1976)). By default, the starting value is set equal to the value of the Hedges estimator and the algorithm terminates when the change in the estimated value of τ^2 is smaller than 10—5 from one iteration to the next. The maximum number of iterations 100 by default (which should be sufficient in most case)."

Motivation

Problemstellung

Wir betrachten das Linear Mixed Model

$$y = X_f \beta_f + X_r \beta_r + \varepsilon \tag{1}$$

wobei

- y ein n-dimensionaler beobachtbarer Zufallsvektor ist, der Daten modelliert,
- $X_f \in \mathbb{R}^{n \times p}$ eine vorgegebene Fixed-Effects-Designmatrix ist,
- $X_r \in \mathbb{R}^{n \times q}$ eine vorgegebene Random-Effects-Designmatrix ist,
- $\beta_f \in \mathbb{R}^p$ ein fester, aber unbekannter, Vektor von Fixed-Effects ist,
- $\beta_r \sim N(0_q, \sigma^2_{\beta_r} I_q)$ ein q-dimensionaler nicht-beobachtbarer Zufallsvektor von Random-Effects ist,
- $\varepsilon \sim N(0_n, \sigma^2_\varepsilon I_n)$ ein n-dimensionaler nicht-beobachtbarer Zufallsvektor von Fehlertermen ist,
- β_r und ε unabhängige Zufallsvektoren sind,
- $\sigma^2_{\beta_r}, \sigma^2_\varepsilon > 0$ feste unbekannte Parameter, genannt *Varianzkomponenten*, sind.

Motivation

Problemstellung

Wie bereits gesehen, impliziert das Linear Mixed Model insbesondere die gemeinsame Verteilung

$$\begin{pmatrix} \beta_r \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0_q \\ X_f\beta_f \end{pmatrix}, \begin{pmatrix} \sigma^2_{\beta_r} I_q & \sigma^2_{\beta_r} X_r^T \\ \sigma^2_{\beta_r} X_r & \sigma^2_{\beta_r} X_r X_r^T + \sigma^2_\varepsilon I_n \end{pmatrix}\right) \tag{2}$$

von Datenvektor und Random-Effects-Vektor, sowie die marginale Datenverteilung

$$y \sim N(X_f\beta_f, \sigma^2_{\beta_r} X_r X_r^T + \sigma^2_\varepsilon I_n). \tag{3}$$

Mithilfe der Definitionen des *Varianzkomponentenvektors* θ und des marginalen Datenkovarianzmatrixparameters V_θ

$$\theta := (\sigma^2_{\beta_r}, \sigma^2_\varepsilon) \text{ und } V_\theta := \sigma^2_{\beta_r} X_r X_r^T + \sigma^2_\varepsilon I_n, \tag{4}$$

wird die marginale Datenverteilung des Linear Mixed Models häufig auch als

$$y \sim N(X_f\beta_f, V_\theta) \tag{5}$$

geschrieben.

Das Punktschätzerproblem für ein Linear Mixed Model hat dann drei zentrale Aspekte (vgl. Searle (1988)):

(1) Die Angabe eines Schätzers $\hat{\beta}_f$ für den Fixed-Effects-Parameter β_f.
(2) Die Angabe eines Schätzers $\hat{\beta}_r$ für den Random-Effects-Parameter β_r.
(3) Die Angabe eines Schätzers $\hat{\theta}$ für den Varianzkomponentenparameters θ.

Motivation

Problemstellung

Die Lösung dieses Problems ist nicht trivial, bleibt Gegenstand aktueller Forschung und nutzt im Allgmeinen iterative Methoden. Generell werden meist folgende Ansätze verfolgt.

(1) Schätzung von β_f basierend auf der geschätzten Marginalverteilung von y

$\Rightarrow$ V_θ wird durch $V_{\hat{\theta}}$ ersetzt und β_f durch den *Generalisierten-Kleinste-Quadrate Schätzer* geschätzt.

(2) Schätzung von β_r basierend auf der geschätzten gemeinsamen Verteilung von β_r und y

$\Rightarrow$ V_θ wird durch $V_{\hat{\theta}}$ und β_f durch $\hat{\beta}_f$ ersetzt und β_r durch seinen *bedingten Erwartungswert* geschätzt.

(3) Schätzung von θ basierend auf der geschätzten Marginalverteilung von y

$\Rightarrow$ Die Varianzkomponenten θ werden iterativ mithilfe des *Restricted Maximum-Likelihood* Verfahrens geschätzt.

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Bedingter Erwartungswert der Random-Effects

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Selbstkontrollfragen

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Generalisierte Kleinste-Quadrate Schätzung

- Kleinste-Quadrate Schätzung heißt auf English "Ordinary Least Squares" (OLS).
- Zur Abgrenzung nennen wir den bekannten Betaparameterschätzer im Folgenden "OLS-Schätzer".
- Generalisierte Kleinste-Quadrate Schätzung heißt auf English "Generalized Least Squares" (GLS).
- Der GLS-Schätzer ist ein Betaparameterschätzer für das ALM

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N\left(0_n, \sigma^2 V\right) \text{ mit } V \neq I_n \tag{6}$$

- Der GLS-Schätzer ist also im Fall nicht-sphärischer Fehlerkovarianzmatrixparameter angezeigt.
- Der GLS-Schätzer stellt sicher, dass T-Statistiken auch im Fall $V \neq I_n$ t-verteilt sind.
- Im Kontext der Fixed-Effects-Schätzung eines Linear Mixed Models gilt in Hinblick auf obiges ALM speziell

$$X := X_f, \beta := \beta_f, \sigma^2 := \sigma_\varepsilon^2 \sigma_{\beta_r}^2, V := \frac{1}{\sigma_\varepsilon^2} X_r X_r^T + \frac{1}{\sigma_{\beta_r}^2} I_n \text{ und somit } \sigma^2 V = V_\theta. \tag{7}$$

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Definition (Generalisierte Kleinste-Quadrate Schätzer)

Gegeben sei ein Allgemeines Lineares Modell der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N\left(0_n, \sigma^2 V\right) \text{ mit} \tag{8}$$

mit $\sigma^2 > 0$ und einer positiv-definiten Matrix $V \in \mathbb{R}^{n \times n}$. Dann heißt

$$\hat{\beta}_{\text{GLS}} := \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y \tag{9}$$

der *Generalisierte-Kleinste-Quadrate-Schätzer von* β und

$$\hat{\sigma}^2_{\text{GLS}} := \frac{\left(y - X\hat{\beta}_{\text{GLS}}\right)^T V^{-1} \left(y - X\hat{\beta}_{\text{GLS}}\right)}{n-p} \tag{10}$$

der *Generalisierte-Kleinste-Quadrate-Schätzer von* σ^2.

Bemerkungen

- Es muss nicht notwendigerweise $V = I_n$ gelten.
- Die Fehlerkomponenten in ε können unterschiedliche Varianzen haben oder korreliert sein.
- Im Fall $V = I_n$ gilt weiterhin

$$\hat{\beta}_{\text{GLS}} = \left(X^T I_n^{-1} X\right)^{-1} X^T I_n^{-1} y = \left(X^T X\right)^{-1} X^T y =: \hat{\beta}_{\text{OLS}}. \tag{11}$$

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Beispiel

Gegeben sei das Fixed-Effects-Modell der Metaanalyse,

$$y = 1_n \delta + \varepsilon \text{ mit } \varepsilon \sim N(0, V) \tag{12}$$

wobei

- y der Vektor von n studienspezifischen Effektstärkeschätzern ist
- $1_n \in \mathbb{R}^{n \times 1}$ die Fixed-Effects-Designmatrix bezeichnet,
- δ die wahre, aber unbekannte, Effektstärke bezeichnet und
- ε der Zufallsfehler ist, für den angenommen wird, dass

$$V := \text{diag}(\sigma_1^2, ..., \sigma_n^2) \tag{13}$$

  mit bekannten studienspezifischen Varianzschätzern $\sigma_i^2, i = 1, ..., n$ ist.

Offenbar gilt hier in Hinblick auf die Definition des GLS-Schätzers, dass $\sigma^2 := 1$ und $V \neq I_n$ ist. Weiterhin gilt

$$V^{-1} = \text{diag}(w_1, ..., w_n) \text{ mit } w_i := \frac{1}{\sigma_i^2} \text{ für } i = 1, ..., n. \tag{14}$$

Beispiel (fortgeführt)

Damit ergibt sich aber

$$
\begin{aligned}
\hat{\beta}_{\mathsf{GLS}} &= \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y \\
&= \left(1_n^T \mathsf{diag}(w_1, ..., w_n) 1_n\right)^{-1} 1_n^T \mathsf{diag}(w_1, ..., w_n) y \\
&= \left(\begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\
&= \left(\begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}\right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} w_1 y_1 \\ \vdots \\ w_n y_n \end{pmatrix} \\
&= \left(\sum_{i=1}^n w_i\right)^{-1} \sum_{i=1}^n w_i y_i \\
&= \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i y_i \\
&= \hat{\delta}
\end{aligned}
\tag{15}
$$

Wir sehen also, dass der Fixed-Effects-Modell Effektstärkeschätzer der GLS-Schätzer dieses Modells ist.

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Theorem (GLS-Schätzer und OLS-Schätzer)

Gegeben sei ein *untransformiertes ALM* der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N\left(0_n, \sigma^2 V\right) \tag{16}$$

mit $\sigma^2 > 0$ und einer positiv-definiten Matrix $V \in \mathbb{R}^{n \times n}$ und es sei $\hat{\beta}_{\text{GLS}}$ der Generalisierte-Kleinste-Quadrate-Schätzer von β. Weiterhin sei $K \in \mathbb{R}^{n \times n}$ eine Matrix mit den Eigenschaften

$$KK^T = V \text{ und } \left(K^{-1}\right)^T K^{-1} = V^{-1} \tag{17}$$

Schließlich sei

$$y^* = X^*\beta + \varepsilon^* \text{ mit } y^* := K^{-1}y, X^* := K^{-1}X, \varepsilon^* := K^{-1}\varepsilon \tag{18}$$

das *transformierte ALM*. Dann gelten

(1) Der GLS-Schätzer des untransformierten ALMs ist der OLS-Schätzer des transformierten ALMs.

(2) Für den Zufallsfehler im transformierten ALM gilt $\varepsilon^* \sim N(0_n, \sigma^2 I_n)$.

Bemerkungen

- Der zu schätzende wahre, aber unbekannte, Betaparameter ist in beiden ALMs identisch.
- Im transformierten ALM ist der Fehlerkovarianzmatrixparameter sphärisch, also T-Statistiken t-verteilt.
- Man nennt die Transformation des ALMs durch K auch eine "Whitening-Transformation".
- K mit den geforderten Eigenschaften kann durch die *Cholesky-Zerlegung* von V gewonnen werden.

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Beweis

(1) Für den GLS-Schätzer im untransformierten Modell gilt

$$
\begin{aligned}
\hat{\beta}_{\mathsf{GLS}} &= \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y \\
&= \left(X^T \left(K^{-1}\right)^T K^{-1} X\right)^{-1} X^T \left(K^{-1}\right)^T K^{-1} y \\
&= \left(\left(K^{-1} X\right)^T K^{-1} X\right)^{-1} \left(K^{-1} X\right)^T K^{-1} y \\
&= \left({X^*}^T X^*\right)^{-1} {X^*}^T y^*.
\end{aligned} \tag{19}
$$

Dies aber entspricht dem OLS-Schätzer im transformierten Modell.

(2) Mit der Tatsache, dass für eine invertierbare Matrix A immer gilt, dass $\left(A^{-1}\right)^T = \left(A^T\right)^{-1}$ und dem Theorem zur linear-affinen Transformation multivariater Normalverteilungen ergibt sich

$$
\begin{aligned}
\varepsilon^* &\sim N\left(K^{-1} 0_n, K^{-1}(\sigma^2 V){K^{-1}}^T\right) \\
&= N\left(0_n, \sigma^2 K^{-1} V {K^{-1}}^T\right) \\
&= N\left(0_n, \sigma^2 K^{-1} K K^T {K^{-1}}^T\right) \\
&= N\left(0_n, \sigma^2 K^{-1} K K^T {K^T}^{-1}\right) \\
&= N\left(0_n, \sigma^2 I_n\right).
\end{aligned} \tag{20}
$$

$\square$

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Anwendung zur Schätzung der Fixed-Effects in einem Linear Mixed Model

Gegeben sei die marginale Datenverteilung eines Linear Mixed Models basierend auf einem Schätzer $\hat{\theta}$,

$$y \sim N(X_f\beta_f, V_{\hat{\theta}}) \tag{21}$$

Dann ist der GLS-Schätzer in diesem Modell

$$\hat{\beta}_f = \left(X_f^T V_{\hat{\theta}}^{-1} X_f\right)^{-1} X_f^T V_{\hat{\theta}}^{-1} y \tag{22}$$

ein populärer Schätzer für β_f.

Für Eigenschaften von $\hat{\beta}_f$ siehe zum Beispiel Harville (1977) und Searle, Casella, and McCulloch (1992).

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Theorem (Frequentistische Verteilung des GLS-Betaparameterschätzers)

Gegeben sei ein Allgemeines Lineares Modell der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 V) \tag{23}$$

mit $\sigma^2 > 0$ und einer positiv-definiten Matrix $V \in \mathbb{R}^{n \times n}$. Weiterhin sei

$$\hat{\beta}_{\text{GLS}} := \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y \tag{24}$$

der Generalisierte-Kleinste-Quadrate-Schätzer von β. Dann gilt

$$\hat{\beta}_{\text{GLS}} \sim N\left(\beta, \sigma^2 (X^T V^{-1} X)^{-1}\right). \tag{25}$$

Bemerkung

- Das Theorem ist eine Generalisierung des entsprechenden Theorems für den OLS-Betaparameterschätzer.

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

<u>Beweis</u>

Das Theorem folgt mit dem Theorem zur linear-affinen Transformation von multivariaten Normalverteilungen. Speziell gilt hier

$$\hat{\beta}_{\mathsf{GLS}} \sim N(\mu, \Sigma) \tag{26}$$

mit

$$\mu = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} X \beta = \beta \tag{27}$$

und

$$\begin{aligned}
\Sigma &= \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} \left(\sigma^2 V\right) \left(\left(X^T V^{-1} X\right)^{-1} X^T V^{-1}\right) \\
&= \sigma^2 \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} V \left(\left(X^T V^{-1} X\right)^{-1} X^T V^{-1}\right)^T \\
&= \sigma^2 \left(X^T V^{-1} X\right)^{-1} X^T \left(\left(X^T V^{-1} X\right)^{-1} X^T V^{-1}\right)^T \\
&= \sigma^2 \left(X^T V^{-1} X\right)^{-1} X^T \left(V^{-1}\right)^T X \left(\left(X^T V^{-1} X\right)^{-1}\right)^T \\
&= \sigma^2 \left(X^T V^{-1} X\right)^{-1} \left(X^T V^{-1} X\right) \left(X^T V^{-1} X\right)^{-1} \\
&= \sigma^2 \left(X^T V^{-1} X\right)^{-1}
\end{aligned} \tag{28}$$

$\square$

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Bedingter Erwartungswert der Random-Effects

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Selbstkontrollfragen

Bedingter Erwartungswert der Random Effects

Das Linear Mixed Model impliziert wie gesehen eine gemeinsame Verteilung von Daten und unbeobachtbarem Random-Effects-Vektor β_r. Ein Standardvorgehen im Bereich der Linear Mixed Model Schätzung ist es, β_r durch den Erwartungswert der auf den Daten bedingten Verteilung von β_r zu schätzen, wobei unbekannte Parameterwerte wiederrum durch ihre Schätzer ersetzt werden.

Dieses Vorgehen entspricht damit letztlich einer Bayesianischen Punktschätzung von β_r mit marginaler Verteilung ("Prior distribution")

$$\beta_r \sim N\left(0_q, \hat{\sigma}^2_{\beta_r} I_q\right) \tag{29}$$

und bedingter Verteilung ("Likelihood")

$$y \,|\, \beta_r \sim N\left(X_f \hat{\beta}_f + X_r \beta_r, \hat{\sigma}^2_{\varepsilon} I_n\right) \tag{30}$$

Anwendung des Theorems zur bedingten Normalverteilungen auf die hier relevante gemeinsame Verteilung

$$\begin{pmatrix} \beta_r \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0_q \\ X_f \hat{\beta}_f \end{pmatrix}, \begin{pmatrix} \hat{\sigma}^2_{\beta_r} I_q & \hat{\sigma}^2_{\beta_r} X_r^T \\ \hat{\sigma}^2_{\beta_r} X_r & V_{\hat{\theta}} \end{pmatrix}\right) \tag{31}$$

ergibt dann als Schätzer für den Random-Effects Parameter den Erwartungswertparameter der Verteilung $\beta_r \,|\, y$

$$\hat{\beta}_r = \mu_{\beta_r|y} = \hat{\sigma}^2_{\beta_r} X_r^T V_{\hat{\theta}}^{-1}(y - X_f \hat{\beta}_f). \tag{32}$$

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Bedingter Erwartungswert der Random-Effects

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Selbstkontrollfragen

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Zentrale Referenzen

Patterson and Thompson (1971)

- Fehlerkontrastmotivation der Restricted Maximum-Likelihood Zielfunktion

Harville (1977)

- Übersicht zu Restricted Maximum-Likelihood Methoden und numerischer Auswertung

Searle, Casella, and McCulloch (1992)

- Ausführliche Übersicht zum Problem der Varianzkomponentenschätzung

Bates and DebRoy (2004)

- Integration von Restricted Maximum-Likelihood in Penalized Least Squares

Starke and Ostwald (2017)

- Expectation-Maximization und Restricted Maximum-Likelihood aus der Perspektive von Variational Inference

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Motivation

Die Maximum-Likelihood Methode kann auf verzerrte Varianzschätzer führen

Zum Beispiel ist der Maximum-Likelihood-Schätzer des Varianzparameters des Normalverteilungsmodells

$$\hat{\sigma}^2_{\text{ML}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_{\text{ML}})^2 \text{ mit } \hat{\mu}_{\text{ML}} := \frac{1}{n}\sum_{i=1}^{n} y_i =: \bar{y}. \tag{33}$$

verzerrt und nur asymptotisch erwartungstreu. Speziell gilt mit der Erwartungstreue der Stichprobenvarianz

$$\mathbb{E}\left(\hat{\sigma}^2\right) = \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2\right) = \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right) = \sigma^2, \tag{34}$$

dass

$$\mathbb{E}\left(\hat{\sigma}^2_{\text{ML}}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_{\text{ML}})^2\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n}(y_i - \bar{y}_n)^2\right) = \frac{n-1}{n}\sigma^2 \tag{35}$$

Da $(n-1)/n < 1$ insbesondere bei kleinem n, unterschätzt der Maximum-Likelihood-Schätzer σ^2.

Patterson and Thompson (1971) schreiben "The difference between the two methods [ML und ReML] is analogous to the well-known difference between two methods of estimating the variance σ^2 of a normal distribution [wie oben] (...)." und Harville (1977) merkt an "One criticism of the ML approach to the estimation of [σ^2] is that the ML estimator (...) takes no account of the loss in degrees of freedom that results from estimating [μ] (...) These"deficiencies" are eliminated in the restricted maximum likelihood (REML) approach (...)"

$\Rightarrow$ ReML für Varianzparameterschätzung scheint eine gute Idee zu sein.

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Maximum-Likelihood und Restricted Maximum-Likelihood

Betrachtet man die marginale Datenverteilung des Linear Mixed Models

$$y \sim N(X_f\beta_f, V_\theta) \tag{36}$$

so ergibt sich für die Log-Likelihood Funktion der Varianzkomponenten θ für einen Schätzer $\hat{\beta}_f$ von β_f

$$\begin{aligned}\ell(\theta) &= \ln\left((2\pi)^{-\frac{n}{2}}\,|V_\theta|^{-\frac{1}{2}}\,\exp\left(-\frac{1}{2}(y - X_f\hat{\beta}_f)^T V_\theta^{-1}(y - X_f\hat{\beta}_f)\right)\right) \\ &= -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|V_\theta| - \frac{1}{2}(y - X_f\hat{\beta}_f)^T V_\theta^{-1}(y - X_f\hat{\beta}_f)\end{aligned} \tag{37}$$

Die (numerische) Maximierung dieser Funktion hinsichtlich θ führt zu einem Maximum-Likelihood-Schätzer von θ,

$$\hat{\theta}_{\mathsf{ML}} := \mathsf{argmax}_\theta\left(-\frac{1}{2}\ln|V_\theta| - \frac{1}{2}(y - X_f\hat{\beta}_f)^T V_\theta^{-1}(y - X_f\hat{\beta}_f)\right). \tag{38}$$

Ein zentrales Resultat ist, dass ein Restricted Maximum-Likelihood-Schätzer von θ gegeben ist durch

$$\hat{\theta}_{\mathsf{ReML}} := \mathsf{argmax}_\theta\left(-\frac{1}{2}\ln|V_\theta| - \frac{1}{2}\ln|X_f^T V_\theta^{-1} X_f| - \frac{1}{2}(y - X_f\hat{\beta}_f)^T V_\theta^{-1}(y - X_f\hat{\beta}_f)\right). \tag{39}$$

Die Zielfunktion der ReML Methode und der ML Methode unterscheiden sich also nur hinsichtlich eines Terms.

Im Folgenden wollen wir die Motivation für die Einführung des Terms $-\frac{1}{2}\ln|X_f^T V_\theta^{-1} X_f|$ (sehr) grob skizzieren.

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Motivation der Restricted Maximum-Likelihood Funktion durch Fehlerkontraste

Grundidee des von Patterson and Thompson (1971) formulierten Ansatzes ist es, den Effekt von β_f aus y herauszurechnen und dann die Likelihood-Funktion der so transformierten Daten hinsichtlich von θ zu maximieren.

Genauer ist das Ziel den Datenvektor durch eine lineare Transformation mit einer Matrix $M \in \mathbb{R}^{m \times n}$ in einen anderen Vektor z zu transformieren, dessen Erwartungswert für jeden möglichen Wert von β_f der Nullvektor ist, also

$$z = My \text{ mit } \mathbb{E}(z) = 0_m \text{ für alle } \beta_f \in \mathbb{R}^p \tag{40}$$

und dann die Log-Likelihood-Funktion von z zu maximieren.

Eine solche Matrix M muss insbesondere die Bedingung

$$MX_f = 0_m \tag{41}$$

erfüllen, denn dann gilt

$$\mathbb{E}(z) = \mathbb{E}(My) = \mathbb{E}(MX_f\beta_f) = \mathbb{E}(0_{mp}\beta_f) = 0_m \text{ für alle } \beta_f \in \mathbb{R}^p. \tag{42}$$

Eine prinzipielle Möglichkeit für die Wahl von M ist die $n \times n$ Matrix

$$M = I_n - P_n \text{ mit } P_n := X_f(X_f^T X_f)^{-1} X_f^T \in \mathbb{R}^{n \times n} \tag{43}$$

mit der sogenannten *Projektionsmatrix* P.

Motivation der Restricted Maximum-Likelihood Funktion durch Fehlerkontraste

Es gilt dann nämlich

$$MX_f = (I_n - P_n)X_f = X_f - P_nX_f = X_f - X_f(X_f^TX_f)^{-1}X_f^TX_f = X_f - X_f = 0_{nn}. \tag{44}$$

Nutzt man also diese Matrix M zur Transformation der Daten, ergibt sich

$$z = My = (I_n - P_n)y = y - X_f(X_f^TX_f)^{-1}X_f^Ty = y - X_f\hat{\beta}_f = \hat{\varepsilon} \tag{45}$$

und wir sehen, dass eine solche Matrix M die Daten auf die Residuals, also die Differenz zwischen Daten und Modellvorhersage nach Schätzung der Fixed-Effects projiziert. Die Matrix P_n nennt man dementsprechend auch *Residual-forming matrix* oder *Projektionsmatrix* und die Matrix M *Fehlerkontrastmatrix*. Der Vektor z sind dann die Residuals und ReML wird auch häufig als *Residual Maximum-Likelihood* bezeichnet. Eine Zeile einer solchen Matrix M nennt man auch *Fehlerkontrast*, die Matrix M daher eine *Fehlerkontrastmatrix*.

Prinzipiell würde man nun die Log-Likelihood Funktion von $z \in \mathbb{R}^n$, das aufgrund des Theorems zur linear-affinen Transformation multivariater Normalverteilungen die Verteilung

$$z \sim N(MX\beta_f, MV_\theta M^T) \tag{46}$$

hat, also

$$\ell(\theta) = \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|MV_\theta M^T| - \frac{1}{2}(My)^T(MV_\theta M^T)^{-1}My. \tag{47}$$

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Motivation der Restricted Maximum-Likelihood Funktion durch Fehlerkontraste

Leider funktioniert die vorgeschlagene Wahl von M in dieser Form nicht, "da $\text{rg}(M) = m < n$".

Man wählt daher die ersten $n-p$ Zeilen von M und erhält eine Matrix $K \in \mathbb{R}^{m \times n}$ mit vollem Spaltenrang $m = n-p$.

Dabei gilt weiterhin $\mathbb{E}(z) = \mathbb{E}(Ky) = 0_m$ und man möchte

$$\ell(\theta) = \frac{m}{2}\ln 2\pi - \frac{1}{2}\ln|KV_\theta K^T| - \frac{1}{2}(Ky)^T(KV_\theta K^T)^{-1}Ky \tag{48}$$

maximieren.

Searle, Casella, and McCulloch (1992) beweisen nun, dass

$$\ln|KV_\theta K^T| = \ln|V_\theta| + \ln|X_f V_\theta^{-1} X_f| \tag{49}$$

und

$$(Ky)^T(KV_\theta K^T)^{-1}Ky = y^T P_n y = (y - X_f\hat{\beta}_f)^T V_\theta^{-1}(y - X_f\hat{\beta}_f) \tag{50}$$

Dies ist intuitiv zumindest unter dem Aspekt, dass K Teil von P_n ist, einsichtig. Damit ergibt sich für die Log-Likelihood-Funktion von $z = Ky$ aber, dass

$$\ell(\theta) = \frac{m}{2}\ln 2\pi - \frac{1}{2}\ln|V_\theta| - \frac{1}{2}\ln|X_f V_\theta^{-1} X_f| - \frac{1}{2}(y - X_f\hat{\beta}_f)^T V_\theta^{-1}(y - X_f\hat{\beta}_f) \tag{51}$$

also identisch mit der ReML Zielfunktion ist.

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Die hier gegebene Darstellung lässt allerdings viele Fragen offen

- Warum sind ReML Schätzer der Varianzkomponenten unverzerrt (vgl. Foulley (1993))?
- Was sind weitere generelle Eigenschaften der ReML Schätzer (vgl. Harville (1977))
- Was genau ist das Problem bei Residualprojektion mit $M = (I_n - P_n)$?
- Wie und wann funktionieren die Beweise von Searle, Casella, and McCulloch (1992)?

Darüber hinaus ergeben sich zumindest folgende Fragen

- Was verhält sich die Fehlerkontrastmotivation zur Expectation-Maximization Motivation (vgl. Laird (1982))?
- Wie verhält sich die Fehlerkontrastmotivation zur bedingten Verteilungsmotivation (vgl. Verbyla (1990))?
- Welche Algorithmen eignen sich zur Maximierung der ReML Zielfunktion (vgl. Lindstrom and Bates (1990))?
- Wie verhalten sich ReML und Penalized-Least-Squares (vgl. Bates and DebRoy (2004))?

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Anwendung zur Schätzung der Fixed-Effects in einem Linear Mixed Model

Gegeben sei die marginale Datenverteilung eines Linear Mixed Models basierend auf einem Schätzer $\hat{\beta}_f$,

$$y \sim N(X_f \hat{\beta}_f, V_\theta) \tag{52}$$

Dann ist der ReML Schätzer in diesem Modell,

$$\hat{\theta}_{\text{ReML}} := \text{argmax}_\theta \left(-\frac{1}{2} \ln |V_\theta| - \frac{1}{2} \ln |X_f^T V_\theta^{-1} X_f| - \frac{1}{2} (y - X_f \hat{\beta}_f)^T V_\theta^{-1} (y - X_f \hat{\beta}_f) \right), \tag{53}$$

ein populärer Schätzer für θ.

Zusammenfassung

Algorithmus zur Schätzung der Parameter eines Linear Mixed Models

(0) Initialisierung

- Wahl eines geeigneten Startwerts $\hat{\beta}_f^{(0)}$

(1) Für $k = 1, ..., K$

- ReML-Schätzung $\hat{\theta}^{(k)}$ basierend auf $\hat{\beta}_f^{(k-1)}$
- GLS-Schätzung $\hat{\beta}_f^{(k)}$ basierend auf $\hat{\theta}^{(k)}$

(2) Schätzung von $\hat{\beta}_r$ basierend auf $\hat{\theta}^{(K)}$ und $\hat{\theta}^{(K)}$

Referenzen I

Bates, Douglas M, and Saikat DebRoy. 2004. "Linear Mixed Models and Penalized Least Squares." *Journal of Multivariate Analysis* 91 (1): 1–17. https://doi.org/10.1016/j.jmva.2004.04.013.

Foulley, JL. 1993. "A Simple Argument Showing How to Derive Restricted Maximum Likelihood."

Harville, David A. 1977. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association* 72 (358): 320. https://doi.org/10.2307/2286796.

Jennrich, R. I., and P. F. Sampson. 1976. "Newton-Raphson and Related Algorithms for Maximum Likelihood Variance Component Estimation." *Technometrics* 18 (1): 11. https://doi.org/10.2307/1267911.

Laird, Nan M. 1982. "Computation of Variance Components Using the Em Algorithm." *Journal of Statistical Computation and Simulation* 14 (3-4): 295–303. https://doi.org/10.1080/00949658208810550.

Lindstrom, Mary J., and Douglas M. Bates. 1990. "Nonlinear Mixed Effects Models for Repeated Measures Data." *Biometrics* 46 (3): 673. https://doi.org/10.2307/2532087.

Patterson, H. D., and R. Thompson. 1971. "Recovery of Inter-Block Information When Block Sizes Are Unequal." *Biometrika* 58 (3): 545–54. https://doi.org/10.1093/biomet/58.3.545.

Schaeuffele, Carmen, Laura E. Meine, Ava Schulz, Maxi C. Weber, Angela Moser, Christina Paersch, Dominique Recher, et al. 2024. "A Systematic Review and Meta-Analysis of Transdiagnostic Cognitive Behavioural Therapies for Emotional Disorders." *Nature Human Behaviour* 8 (3): 493–509. https://doi.org/10.1038/s41562-023-01787-3.

Searle, Shayle R. 1988. "Mixed Models and Unbalanced Data: Wherefrom, Whereat and Whereto?" *Communications in Statistics - Theory and Methods* 17 (4): 935–68. https://doi.org/10.1080/03610928808829667.

Searle, Shayle R., George Casella, and Charles E. McCulloch. 1992. *Variance Components*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

Referenzen II

Starke, Ludger, and Dirk Ostwald. 2017. "Variational Bayesian Parameter Estimation Techniques for the General Linear Model." *Frontiers in Neuroscience* 11 (September). https://doi.org/10.3389/fnins.2017.00504.

Verbyla, A. P. 1990. "A Conditional Derivation of Residual Maximum Likelihood." *Australian Journal of Statistics* 32 (2): 227–30. https://doi.org/10.1111/j.1467-842X.1990.tb01015.x.

Viechtbauer, Wolfgang. 2005. "Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model." *Journal of Educational and Behavioral Statistics* 30 (3): 261–93. https://doi.org/10.3102/10769986030003261.

———. 2010. "Conducting Meta-Analyses in *R* with the **Metafor** Package." *Journal of Statistical Software* 36 (3). https://doi.org/10.18637/jss.v036.i03.