Evaluation und Metaanalyse

MSc Klinische Psychologie und Psychotherapie

SoSe 2024

Prof. Dr. Dirk Ostwald

(5) Linear Mixed Models Formulierung

Überblick

- Linear Mixed Models sind eine Weiterentwicklung des Allgemeinen Linearen Modells (ALM).
- Durch iterative Schätzverfahren sind Linear Mixed Models in den letzten 50 Jahren sehr populär geworden.
- In **R** sind Linear Mixed Models durch die Verfügbarkeit der Pakete `nlme` und `lme4` sehr verbreitet.
- Klassische Anwendungen von Linear Mixed Models sind Mehrebenen- und Longitudinalanalysen.
- Fixed- und Random-Effects Modelle der Metaanalyse sind spezielle Linear Mixed Models.
- Die Restricted Maximum-Likelihood-Schätzung ist eng mit der Theorie der Linear Mixed Models verwoben.
- Viele weitere Modelle sind Spezialfälle von Linear Mixed Models, z.B. die Bayesianische ALM Schätzung.
- Linear Mixed Models sind die state-of-the-art Inferenzmodelle in vielen Anwendungsfeldern

Wir formulieren zunächst ein allgemeines Linear Mixed Model. Zur Parameterinferenz eines Linear Mixed Models betrachten wir dann

- die Generalisierte Kleinste-Quadrate Schätzung der Fixed Effects und ihre Konfidenzintervalle,
- den bedingten Erwartungswert der Random-Effects, sowie
- die Varianzkomponentenschätzung mit Restricted Maximum-Likelihood.

Wir betrachten schließlich Anwendungsbeispiele

(1) Anwendung von Linear Mixed Models im Bereich Metaanalyse
(2) Anwendung von Linear Mixed Models im Bereich von Multizentren-Psychotherapiestudien
(3) Anwendung von Linear Mixed Models im Bereich Veränderungsmessung bei Psychotherapie

Modellformulierung

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Bedingter Erwartungswert der Random-Effects

Varianzkomponentenschätzung mit Restricted Maximum Likelihood

Linear Mixed Model Inferenz mit `nlme`

Selbstkontrollfragen

Modellformulierung

Definition (Linear Mixed Model)

Es sei

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

wobei

- y ein n-dimensionaler beobachtbarer Zufallsvektor ist, der *Daten* genannt wird,
- $X_f \in \mathbb{R}^{n\times p}$ eine vorgegeben Matrix ist, die *Fixed-Effects-Designmatrix* genannt wird,
- $\beta_f \in \mathbb{R}^p$ ein unbekannter fester Vektor ist, der *Fixed Effects* genannt wird,
- $X_r \in \mathbb{R}^{n\times q}$ eine vorgegeben Matrix ist, die *Random-Effects-Designmatrix* genannt wird,
- β_r ein q-dimensionaler latenter Zufallsvektor ist der *Random Effects* genannt wird und für den gilt, dass

$$\beta_r \sim N(0_q, \Sigma_{\beta_r}) \text{ mit } \Sigma_{\beta_r} \in \mathbb{R}^{q\times q} \text{ p.d.}, \tag{2}$$

- ε ein n-dimensionaler latenter Zufallsvektor ist, der *Zufallsfehler* genannt wird und für den gilt, dass

$$\varepsilon \sim N(0_n, \Sigma_\varepsilon) \text{ mit } \Sigma_\varepsilon \in \mathbb{R}^{n\times n} \text{ p.d. und unabhängig von } \beta_r. \tag{3}$$

Dann heißt (1) *Linear Mixed Model*.

Bemerkungen

- Man bezeichnet Darstellung des Linear Mixed Models in dieser Definition auch als *strukturelle Form* bezeichnen.
- Häufig gelten $\Sigma_{\beta_r} := \sigma^2_{\beta_r} I_q$ mit $\sigma^2_{\beta_r} > 0$ und $\Sigma_\varepsilon := \sigma^2_\varepsilon I_n$ mit $\sigma^2_\varepsilon > 0$.

Modellformulierung

Definition (Verteilungsdarstellung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model

$$y = X_f\beta_f + X_r\beta_r + \varepsilon \text{ mit } \beta_r \sim N(0_q, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon). \tag{4}$$

Dann nennt man die äquivalente Darstellung dieses Modells mit der marginalen Verteilung

$$\beta_r \sim N(0_q, \Sigma_{\beta_r}) \tag{5}$$

und der bedingten Verteilung

$$y \,|\, \beta_r \sim N(X_f\beta_f + X_r\beta_r, \Sigma_\varepsilon) \tag{6}$$

die Verteilungsdarstellung des Linear Mixed Models

Bemerkungen

- Die Äquivalenz folgt mit dem Theorem zur linear-affinen Transformation multivariate Normalverteilungen.
- Intuitiv beschreibt der Ausdruck $y = X_f\beta_f + X_r\beta_r + \varepsilon$ vor allem eine bedingte Verteilung.
- Die Fehlerkovarianzmatrix Σ_ε ist die Kovarianzmatrix dieser bedingten Verteilung.

Theorem (Gemeinsame Verteilung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model. Dann gilt für die gemeinsame Verteilung von Daten und Random Effects, dass

$$\begin{pmatrix} \beta_r \\ y \end{pmatrix} \sim N\left(\mu_{\beta_r,y}, \Sigma_{\beta_r,y}\right) \tag{7}$$

mit

$$\mu_{\beta_r,y} := \begin{pmatrix} 0_q \\ X_f\beta_f \end{pmatrix} \in \mathbb{R}^{q+n} \text{ und } \Sigma_{\beta_r,y} := \begin{pmatrix} \Sigma_{\beta_r} & \Sigma_{\beta_r}X_r^T \\ X_r\Sigma_{\beta_r} & X_r\Sigma_{\beta_r}X_r^T + \Sigma_\varepsilon \end{pmatrix} \in \mathbb{R}^{(q+n)\times(q+n)} \tag{8}$$

<u>Beweis</u>

Die gemeinsame Verteilung des Linear Mixed Models ergibt sich direkt durch Anwendung des Theorems zu Gemeinsamen Normalverteilungen auf die Verteilungsdarstellung des Linear Mixed Models.

□

Theorem (Marginale Datenverteilung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model. Dann gilt für die marginale Verteilung der Daten, dass

$$y \sim N\left(\mu_y, \Sigma_y\right) \tag{9}$$

mit

$$\mu_y := X_f \beta_f \in \mathbb{R}^n \text{ und } \Sigma_y := X_r \Sigma_{\beta_r} X_r^T + \Sigma_\varepsilon \in \mathbb{R}^{n \times n} \tag{10}$$

Beweis

Die Aussage ergibt sich direkt aus dem Theorem zur Gemeinsamen Verteilung des Linear Mixed Models und dem Theorem zu Marginalen Normalverteilungen.

$\square$

Bemerkungen

- Linear Mixed Models erlauben es, nicht-sphärische Kovarianzmatrixstrukturen zu modellieren.
- Gilt speziell $\Sigma_{\beta_r} := \sigma^2_{\beta_r} I_q, \sigma^2_{\beta_r} > 0$ und $\Sigma_\varepsilon := \sigma^2_\varepsilon I_n, \sigma^2_\varepsilon > 0$, so folgt

$$y \sim N\left(X_f \beta_f, \sigma^2_{\beta_r} X_r X_r^T + \sigma^2_\varepsilon I_n\right) \tag{11}$$

- Parameter wie $\sigma^2_{\beta_r}$ und σ^2_ε nennt man *Kovarianzkomponenten*.

Definition (Hierarchische Darstellung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model

$$y = X_f\beta_f + X_r\beta_r + \varepsilon \text{ mit } \beta_r \sim N(0_q, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon). \quad (12)$$

Dann nennt man die äquivalente Darstellung dieses Modells in der Form

$$\begin{aligned} \beta_r &= 0_q + \eta && \text{mit } \eta \sim N(0_q, \Sigma_{\beta_r}) \Leftrightarrow \ \beta_r \sim N(0_q, \Sigma_{\beta_r}) \\ y &= X_f\beta_f + X_r\beta_r + \varepsilon && \text{mit } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \ \Leftrightarrow y\,|\,\beta_r \sim N(X_f\beta_f + X_r\beta_r, \Sigma_\varepsilon) \end{aligned} \quad (13)$$

die *hierarchische Darstellung des Linear Mixed Models*

Bemerkung

- Man nennt diese Darstellung auch ein *Mehrebenenmodell*.
- Es ist leicht, sich Linear Mixed Models mit mehr als den hier spezifizierten zwei Ebenen vorzustellen.
- Die obigen Verteilungsaussagen gelten natürlich auch für die Hierarchische Form des Linear Mixed Models.

Allgemeines Lineares Modell

Linear Mixed Models

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition des Linear Mixed Models wieder.
2. Geben Sie die Definition der Verteilungsdarstellung des Linear Mixed Models wieder.
3. Geben Sie das Theorem zur Gemeinsamen Verteilung des Linear Mixed Models wieder.
4. Geben Sie das Theorem zur Marginalen Datenverteilung des Linear Mixed Models wieder.
5. Geben Sie die Definition der Hierarchischen Darstellung des Linear Mixed Models wieder.