

Inhalt

- 1 Das lineare gemischte Modell
- 2 Likelihood-Schätzung für lineare gemischte Modelle
- 3 Likelihood-Inferenz im linearen gemischten Modell**
 - Kovarianzmatrizen für die Schätzer
 - Tests für die festen Effekte
 - Tests für die zufälligen Effekte oder Varianzkomponenten
 - Modellselektion
- 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 Kovarianzmatrix der Schätzer

Bei bekannter Kovarianzstruktur ist wegen (18) → **Übung**

$$\text{Cov} \begin{pmatrix} \hat{\beta} \\ \hat{\mathbf{b}} - \mathbf{b} \end{pmatrix} = (\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{A})^{-1}. \quad (21)$$

Beachte, dass $\text{Cov}(\hat{\mathbf{b}} - \mathbf{b})$ statt $\text{Cov}(\hat{\mathbf{b}})$ verwendet wird, was den Bias durch Shrinkage und die Variabilität in \mathbf{b} berücksichtigt.

Geschätzte Versionen mit eingesetztem $\hat{\mathbf{R}}$ und $\hat{\mathbf{A}}$ können zur Konstruktion approximativer Konfidenz- oder Vorhersageintervalle für Ausdrücke in β und \mathbf{b} verwendet werden (z.B. für neue Beobachtungen \mathbf{y}_i); die Diagonalelemente ergeben geschätzte Varianzen für die Komponenten von $\hat{\beta}$ bzw. $\hat{\mathbf{b}}$.

Beachte, dass die Variabilität durch die Schätzung der Kovarianzparameter nicht berücksichtigt wird und die tatsächlichen Varianzen daher unterschätzt werden.

Schätzung der Kovarianzmatrix der Schätzer

Manchmal wird auch eine Alternative verwendet, bei der \mathbf{b} wie ein fester Effekt behandelt wird, indem man auf \mathbf{b} bedingt. Dann erhält man \rightarrow **Übung**

$$\text{Cov} \left(\left(\begin{array}{c} \hat{\beta} \\ \hat{\mathbf{b}} \end{array} \right) \middle| \mathbf{b} \right) = (\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{A})^{-1} \mathbf{C}'\mathbf{R}^{-1}\mathbf{C} (\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{A})^{-1}. \quad (22)$$

Diese Kovarianzmatrix ist von der **Sandwich-Matrix**-Form, da sie wie ein Sandwich aus $(\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{A})^{-1}$ und $\mathbf{C}'\mathbf{R}^{-1}\mathbf{C}$ zusammengesetzt ist.

Nach Einsetzen der Schätzer $\hat{\mathbf{R}}$, $\hat{\mathbf{A}}$ erhält man die geschätzte Kovarianzmatrix.

Wir präferieren (21). (22) führt zu kleineren Standardfehlern,

$$\begin{aligned} \text{Cov} \left(\begin{array}{c} \hat{\beta} \\ \hat{\mathbf{b}} - \mathbf{b} \end{array} \right) &= \text{Cov} \left(\left(\begin{array}{c} \hat{\beta} \\ \hat{\mathbf{b}} \end{array} \right) \middle| \mathbf{b} \right) \\ &= (\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{A})^{-1} \mathbf{A} (\mathbf{C}'\mathbf{R}^{-1}\mathbf{C} + \mathbf{A})^{-1} \geq \mathbf{0}. \end{aligned}$$

Tests für die festen Effekte

Mit den geschätzten Kovarianzmatrizen lassen sich **Wald-Tests** für

$$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0} \text{ gegen } H_A : \mathbf{L}\boldsymbol{\beta} \neq \mathbf{0} \quad (23)$$

oder Konfidenzintervalle konstruieren. Dazu benutzen wir, dass

$$T_W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{L}' [\mathbf{L} \text{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{L}']^{-1} \mathbf{L} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

asymptotisch $\chi^2_{\text{Rang}(\mathbf{L})}$ -verteilt ist. Es können auch robuste Kovarianzmatrix-Schätzer verwendet werden, siehe z.B. Liang & Zeger (1986). *Longitudinal Data Analysis Using Generalized Linear Models*. Biometrika, 73 (1), 13-22.

Wichtig: Alle asymptotischen Ergebnisse in diesem Kapitel gelten in LMMs für **Longitudinal- oder Cluster-Daten**, in denen \mathbf{y} aus N unabhängigen Teilvektoren \mathbf{y}_i besteht, für $N \rightarrow \infty$. Resultate für allgemeine LMMs sind schwierig zu zeigen (vgl. die Diskussion in Ruppert, Wand & Carroll (2003), Kapitel 4.8).

Die Wald-Teststatistik basiert auf Standardfehlern (21), die die wahre Variabilität in $\hat{\beta}$ durch Vernachlässigung der Schätzung von ϑ unterschätzen.

Dieses Problem wird häufig durch approximative t - oder F -Statistiken verringert:

- Für einen einzelnen Parameter β_j in β wird die Verteilung von $(\hat{\beta}_j - \beta_j) / \widehat{SE}(\hat{\beta}_j)$ durch eine t -Verteilung approximiert.
- Für generelle Hypothesen (23) wird die Verteilung von

$$F = \frac{(\hat{\beta} - \beta)' \mathbf{L}' [\mathbf{L} \text{Cov}(\hat{\beta}) \mathbf{L}']^{-1} \mathbf{L} (\hat{\beta} - \beta)}{\text{Rang}(\mathbf{L})}$$

durch eine F -Verteilung mit $\text{Rang}(\mathbf{L})$ Zähler-Freiheitsgraden approximiert.

Die Freiheitsgrade für die t -Verteilung bzw. die Nenner-Freiheitsgrade für die F -Verteilung werden aus den Daten geschätzt. Häufig wird die Satterthwaite-Approximation verwendet, die die Momente der Verteilungen matcht.

Für kleine Stichproben gibt die Methode nach Kenward und Rogers bessere Ergebnisse. Diese berücksichtigt die zusätzliche Unsicherheit durch Schätzung von ϑ .

Die Berechnung von Satterthwaite-artigen Freiheitsgraden ist recht rechenintensiv.

In R: R-Paket `pbkrtest`, siehe Halekoh & Højsgaard (2014), A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models – The R Package `pbkrtest`, *Journal of Statistical Software*, 59 (9).

Alternativ können wir den **Likelihood-Quotienten-Test** (LQT) verwenden. Bei nicht so großem n gibt der LQT meist genauere Ergebnisse als der Wald-Test.

Die Teststatistik des LQT

$$T_{LQT} = 2 \sup_{H_A} l(\beta, \vartheta) - 2 \sup_{H_0} l(\beta, \vartheta)$$

ist asymptotisch unter der Nullhypothese $\chi^2_{\text{Rang}(L)}$ -verteilt. Voraussetzungen:

- genestete Modelle
- gleiche Kovarianzstruktur in jedem Modell
- **keine REML-Schätzung!** (Bei verschiedenen festen Effekten unter H_0 und H_A unterscheiden sich die Fehlerkontraste $\mathbf{A}_0 \mathbf{y}$ und $\mathbf{A}_A \mathbf{y} \Rightarrow$ die Likelihoods für $\mathbf{A}_0 \mathbf{y}$ und $\mathbf{A}_A \mathbf{y}$ sind nicht vergleichbar.)

Beachte außerdem, dass die Approximation bei kleinem $n - p$ schlecht sein kann.

Tests für die (Ko-)Varianzkomponenten in ϑ

Bei Longitudinal- oder Clusterdaten mit

$$\mathbf{D} = \begin{pmatrix} d_{11} & \dots & d_{1q} \\ \vdots & \ddots & \vdots \\ d_{1q} & \dots & d_{qq} \end{pmatrix} = \left(\begin{array}{c|c} \mathbf{D}_1 & \begin{matrix} d_{1q} \\ \vdots \\ d_{qq} \end{matrix} \\ \hline d_{1q} & \dots & d_{qq} \end{array} \right)$$

könnte z.B. ein Test für

$$H_0 : \mathbf{D} = \left(\begin{array}{c|c} \mathbf{D}_1 & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline 0 & \dots & 0 \end{array} \right) \text{ mit } \mathbf{D}_1 \text{ positiv-definit } (q-1 \times q-1) \text{ gegen}$$

$$H_A : \mathbf{D} \text{ positiv-semidefinit } (q \times q) \quad (24)$$

von Interesse sein. Dies entspricht einem Test für den q ten zufälligen Effekt, z.B.

- H_0 : Random Intercept gegen H_A : Random Intercept und Slope ($q = 2$)
- H_0 : keine zufälligen Effekte gegen H_A : Random Intercept ($q = 1$).

Asymptotische Verteilung des LQT

(24) verletzt die Standardannahme der getesteten Parameter im **Innern des Parameterraums**. Z.B. $d_{qq} \geq 0$ als Varianz und $d_{qq} = 0$ (unter H_0) daher am **Rand des Parameterraums**. (Generell: komplizierte Restriktionen an $\mathbf{D} \geq 0$ und $\mathbf{D}_1 \geq 0$.)

Für Longitudinal- oder Cluster-Daten mit $N \rightarrow \infty$ hat T_{LQT} für (24) asymptotisch unter H_0 eine 0.5:0.5-Mischungsverteilung aus zwei χ^2_{q-1} - und χ^2_q -Verteilungen (χ^2_0 -Verteilung = Punktmasse an der Null).

Für komplexere Hypothesen für \mathbf{D} können komplexere χ^2 -Mischungen entstehen.

Diese Mischungsverteilung gilt nicht, wenn ein Nuisance-Parameter (Störparameter) am Rand des Parameterraums liegt, also z.B. die Varianz der Random Intercepts (nahe) Null ist, während die Varianz der Random Slopes getestet wird.

Exakte Verteilung des LQT

Kann \mathbf{y} nicht in viele unabhängige Teilvektoren $\mathbf{y}_i, i = 1, \dots, N$ unterteilt werden, so approximiert die Mischungsverteilung die Verteilung von T_{LQT} meist schlecht.

Die exakte Verteilung von T_{LRT} für $H_0 : \tau_1^2 = 0$ im allgemeinen LMM mit $\mathbf{G} = \tau_1^2 \mathbf{I}_{q_1}$ und $\mathbf{R} = \sigma^2 \mathbf{I}_n$ ist im R-Paket `RLRsim` (F. Scheipl) implementiert. Approximationen für $\mathbf{G} = \text{diag}(\tau_1^2 \mathbf{I}_{q_1}, \dots, \tau_s^2 \mathbf{I}_{q_s})$, $s > 1$, beruhen auf

- Greven, S., Crainiceanu, C., Küchenhoff, H. & Peters, A. (2008). *Restricted Likelihood Ratio Testing for Zero Variance Components in Linear Mixed Models*. JCGS, 17 (4): 870-891.
- Scheipl, F., Greven, S. & Küchenhoff, H. (2008). *Size and Power of Tests for a Zero Random Effect Variance or Polynomial Regression in Additive and Linear Mixed Models*. CSDA, 52 (7): 3283-3299.

Wenn nur Parameter in ϑ getestet werden, kann auch ein restringierter LQT basierend auf der REML-Schätzung verwendet werden. Die Asymptotik ist entsprechend, die exakte Verteilung ist ebenfalls in `RLRsim` implementiert.

Modellselektion

AIC- und BIC-Werte in Modelloutputs, z.B. von `lme()` in R, basieren meist auf der marginalen Likelihood (9), z.B. das Akaike Informationskriterium (AIC)

$$AIC = -2 \log l(\hat{\beta}, \hat{\vartheta}) + 2(p + d),$$

mit d die Anzahl der Parameter in ϑ . Ein AIC basierend auf der marginalen restringierten log-Likelihood kann analog definiert werden.

Die Herleitung des AIC nimmt unabhängig und identisch verteilte Beobachtungen sowie einen Parameterraum \mathbb{R}^{p+d} an. Bei der Selektion von Parametern in ϑ (mit Restriktionen, z.B. Varianzen aus $[0, \infty)$) führt dies zu einem Bias und tendenziell zu kleineren Modellen ohne zufällige Effekte.

Modellselektion

Neuere Ansätze verwenden für die Selektion von zufälligen Effekten ein AIC basierend auf der konditionalen log-Likelihood für $\mathbf{y}|\mathbf{b} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{R})$. Die korrekten Freiheitsgrade werden hergeleitet in

- Greven, S. und Kneib, T. (2010). *On the Behaviour of Marginal and Conditional AIC in Linear Mixed Models*. Biometrika, 97(4): 773-789

und das konditionale AIC ist implementiert im R-Paket cAIC4.