Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

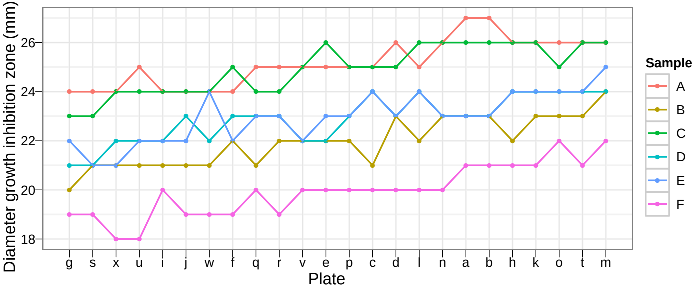
D MATH
Departement of Mathematics
Seminar für Statistik

Robust Estimation of Mixed Models

Summer school of the International Association for Statistical Computing
Katholieke Universiteit Leuven

1 Motivation: Penicillin Data

Data from an experiment *to assess the variability between samples of penicillin by the* B. subtilis *method*.

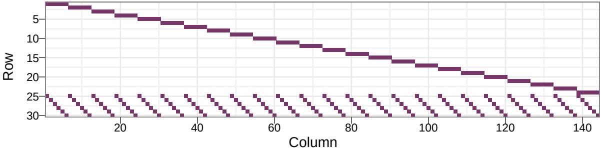


Model: $y_{ij} = \mu + plate_i + sample_j + e_{ij} \quad i = 1, \ldots, 24,\ j = 1, \ldots, 6.$

(Data and plots taken from Bates, 2011.)

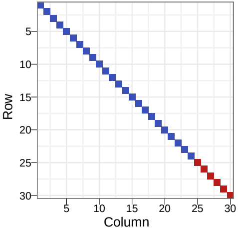
Matrix formulation: $\underline{Y} = \boldsymbol{X}\underline{\beta} + \boldsymbol{Z}\underline{B} + \underline{E}$.

Matrix $\boldsymbol{Z}^\mathsf{T}$:

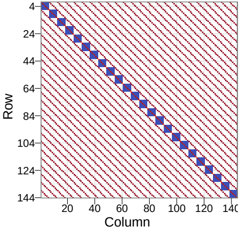


$$\underline{b} = \begin{bmatrix} plate_1 \\ \vdots \\ plate_{24} \\ sample_1 \\ \vdots \\ sample_6 \end{bmatrix}$$

$\mathbf{Cov}(\underline{B}) = \boldsymbol{\Sigma}_{b,\theta}$:



$\mathbf{Cov}(\underline{Y}) = \boldsymbol{Z}\boldsymbol{\Sigma}_{b,\theta}\boldsymbol{Z}^\mathsf{T} + \sigma_e^2\boldsymbol{I}$:



Outline

1 Motivation: Penicillin Data

2 The Linear Mixed Effects Model

3 Estimation in the classical case

4 Robustification

4.1 Fixed and Random Effects

4.2 Variance Components

5 Tools

2 The Linear Mixed Effects Model

$$\underline{Y} = \boldsymbol{X}\underline{\beta} + \boldsymbol{Z}\underline{B} + \underline{E},$$

$$\underline{B} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{b,\theta}), \quad \underline{E} \sim \mathcal{N}\left(0, \sigma_e^2 \boldsymbol{I}_n\right), \quad \underline{B} \perp \underline{E},$$

$$\underline{r}(\underline{\beta}, \underline{b}) = \underline{y} - \boldsymbol{X}\underline{\beta} - \boldsymbol{Z}\underline{b}.$$

$\boldsymbol{\Sigma}_{b,\theta}$ is a parametrized $q \times q$ matrix with parameter vector $\underline{\theta}$.

The log of the density is

$$-2\mathrm{d}(\underline{\beta}, \underline{b}, \underline{\theta}, \sigma_e | \underline{y}) = (n+q)\log 2\pi + n \log \sigma_e + \log|\boldsymbol{\Sigma}_{b,\theta}| + \frac{1}{\sigma_e^2} \underline{r}\left(\underline{\beta}, \underline{b}\right)^{\mathsf{T}} \underline{r}\left(\underline{\beta}, \underline{b}\right) + \underline{b}^{\mathsf{T}} \boldsymbol{\Sigma}_{b,\theta}^{-1} \underline{b}.$$

Usually, to get the likelihood:
Integrate the density over the random effects $\underline{b}$ first.

3 Estimation in the classical case

Given $\underline{\widetilde{\theta}}$, we can uniquely estimate $\underline{\widetilde{\beta}}$, $\underline{\widetilde{b}}$ and $\tilde{\sigma}_e$.

The objective function that has to be minimized depends only on $\underline{\theta}$.

Solve using a general purpose optimizer allowing boundaries.

4 Robustification

Again: the log of the density is

$$-2d(\underline{\beta}, \underline{b}, \underline{\theta}, \sigma_e | \underline{y}) = (n+q)\log 2\pi + n\log\sigma_e + \log|\boldsymbol{\Sigma}_{b,\theta}| \\ + \frac{1}{\sigma_e^2}\underline{r}(\underline{\beta}, \underline{b})^\intercal \underline{r}(\underline{\beta}, \underline{b}) + \underline{b}^\intercal \boldsymbol{\Sigma}_{b,\theta}^{-1}\underline{b}.$$

4.1 $\underline{\beta}$ and $\underline{b}$

Replace the second line by

$$\frac{1}{\lambda_e}\underline{1}^\intercal \rho_e\left(\underline{r}(\underline{\beta}, \underline{b})/\sigma_e\right) + \frac{1}{\lambda_b}\underline{1}^\intercal \rho_b\left(\boldsymbol{\Sigma}_{b,\theta}^{-1/2}\underline{b}\right).$$

4.2 σ_e and $\underline{\theta}$

Estimate directly from $\underline{r}$ and $\underline{\widetilde{b}}$ and correct to get unbiased estimates.
To estimate σ_e we use Huber's Proposal II,

$$\hat{\sigma}_e^2 = \frac{\sum_i w_{e,i}^2 r_i^2}{\sum_i \mathbf{E}\left[w_{e,i}^2 r_i^2\right]},$$

where $w_{e,i}$ are the robustness weights $w_{e,i} = \rho'_e(r_i/\hat{\sigma}_e)/(r_i/\hat{\sigma}_e)$.
Calculation of $\mathbf{E}\left[w_{e,i}^2 r_i^2\right]$: Approximate r_i linearly in terms of $\underline{e}$ and $\underline{b}$.

$$\begin{aligned}\underline{r} &= \underline{y} - \boldsymbol{X}\underline{\widetilde{\beta}} - \boldsymbol{Z}\underline{\widetilde{b}} \\ &= \boldsymbol{X}\underline{\beta} + \boldsymbol{Z}\underline{b} + \underline{e} - \boldsymbol{X}\underline{\widetilde{\beta}} - \boldsymbol{Z}\underline{\widetilde{b}} \\ &= \underline{e} + \boldsymbol{X}(\underline{\beta} - \underline{\widetilde{\beta}}) + \boldsymbol{Z}(\underline{b} - \underline{\widetilde{b}})\end{aligned}$$

Use a *von Mises expansion* for $\underline{\widetilde{\beta}}$, $\underline{\widetilde{b}}$. (DAS-scale uses a similar approach.)

For diagonal $\mathbf{\Sigma}_{b,\theta}$, e.g., simple uncorrelated random effects, a similar approach can be used to estimate $\underline{\theta}$.

For correlated random effects, componentwise weights might break the correlation structure. We use blockwise weights based on the Mahalanobis distance instead.

The estimation of $\underline{\theta}$ is then considerably more complicated and the details have not been fully worked out yet.

5 Tools

Helpful tools for developing a robust method:

- Study of a simple example.
- Estimates should be unbiased for all sensible tuning constants and converge to the classical solution for tuning constants approaching ∞.
- Sensitivity curves for various types of outliers (e.g., for a one-way anova: single observations, group, and spread of a group).

Open Problems

- Implementation of non-diagonal case.
- Treatment of special correlation structures?
- Theoretical robustness properties.

References

Koller, M. and W. A. Stahel (2011). Sharpening wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis 55*(8), 2504–2515.

Bates, D. M. (2011). *lme4: Mixed-effects modeling with R (http://lme4.r-forge.r-project.org/lMMwR/)*.