Robust mixed effects models: simulations and limitations

M. Koller[1], W. A. Stahel[1]

[1] Seminar für Statistik, Department of Mathematics, ETH Zürich, Switzerland

Keywords: Robust, Mixed Models, Variance Components.

Abstract

The standard linear mixed effects model is

$$\underline{y} = \boldsymbol{X}\underline{\beta} + \boldsymbol{Z}\underline{b} + \underline{\varepsilon}, \quad \underline{b} \sim \mathcal{N}(\underline{0}, \sigma^2 \boldsymbol{V}_b(\underline{\theta})) \quad \text{and} \quad \underline{\varepsilon} \sim \mathcal{N}(\underline{0}, \sigma^2 \boldsymbol{I}),$$

where $\underline{\beta}$ is the vector of the fixed effects, $\underline{b}$ is the vector of the random effects, with a covariance matrix parameterized by $\underline{\theta}$, and $\underline{\varepsilon}$ collects the errors. We assume the random effects to be independent of the errors. Let $\boldsymbol{U}_b(\underline{\theta})$ be any square root of $\boldsymbol{V}_b(\underline{\theta})$. In the following we replace $\underline{b}$ by $\boldsymbol{U}_b(\underline{\theta})\,\underline{b}^*$, $\underline{b}^*$ being the spherical random effects, to avoid the inversion of $\boldsymbol{U}_b(\underline{\theta})$.

We robustify the scoring equations for $\underline{\beta}$ and $\underline{b}^*$ as follows.

$$\boldsymbol{X}^\intercal \psi_e\big(\big(\underline{y} - \boldsymbol{X}\underline{\beta} - \boldsymbol{Z}\boldsymbol{U}_b(\underline{\theta})\,\underline{b}^*\big)/\sigma\big) = 0,$$

$$\boldsymbol{U}_b(\underline{\theta})^\intercal\,\boldsymbol{Z}^\intercal \psi_e\big(\big(\underline{y} - \boldsymbol{X}\underline{\beta} - \boldsymbol{Z}\boldsymbol{U}_b(\underline{\theta})\,\underline{b}^*\big)/\sigma\big) - \frac{\lambda_e}{\lambda_b}\psi_b(\underline{b}^*/\sigma) = 0,$$

where the ψ_e and ψ_b functions are applied componentwise and λ_e and λ_b are correction factors depending on ψ_e and ψ_b. This formulation of the scoring equations is only valid for diagonal $\boldsymbol{V}_b(\underline{\theta})$ but can be generalized to the non-diagonal case.

Note that the scoring equations for $\underline{\beta}$ could also be written in a form that does not involve the random effects. But then we would need to be able to partition $\underline{y}$ into independent sub-vectors to robustify them by applying a ψ-function. This is not always possible, e.g., for crossed random effects. Considering the random effects b_j and the residuals ε_i, a ψ-function can be applied to all these independent quantities individually. This leads to a more effective robustification if these quantities are prone to outliers independently.

As an example, consider a two-way ANOVA with crossed random effects. In this case, the correlation structure of $\underline{y}$ cannot be arranged into independent blocks thus thwarting any robustification on this level. But since the crossed nature of the random effects is encoded only in $\boldsymbol{Z}$ and not in $\boldsymbol{V}_b(\underline{\theta})$, which in this example is a simple diagonal matrix, this is not a problem when using the above formulation. By splitting the robustification of the residuals and the random effects, we gain the ability to down-weight just one (or some) of the observations or the random effects.

The robustified estimating equations for σ and $\underline{\theta}$ are derived from the classical ones. The robustification follows the lines of the Design Adaptive Scale estimate in the robust linear regression case, i.e., we use a linear approximation to the distribution of the residuals and the spherical random effects to calculate the normalization constants (cf. reference). It is noteworthy that in the classical case we recover the REML estimates.

We explore for which data this method has its limitations, e.g., pedigree-based mixed effects models, and where this is not an issue, e.g., crossed random effects. The robustness properties of the method are evaluated by a couple of simulation studies. The linear approximations suffice to ensure consistent results for Huber functions down to quite small tuning constants k. For various scenarios we demonstrate both bounded influence and breakdown properties.

References

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