0.1 Cook's Distance for LMEs

Cook's Distance is a well known diagnostic technique used in classical linear models, extended to LME models. For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ.

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on 'one-step' methods. *Cook (1986)* gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

0.1.1 Change in the precision of estimates

The effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.

Cook's distance

In the study of Linear model diagnostics, Cook proposed a measure that combines the information of leverage and residual of the observation, now known simply as the

Cook's Distance. **?** would later adapt the Cook's distance measure for the analysis of LME models.

0.2 Cook's Distance

0.2.1 Cook's Distance

Cooks Distance (D_i) is an overall measure of the combined impact of the ith case of all estimated regression coefficients. It uses the same structure for measuring the combined impact of the differences in the estimated regression coefficients when the kth case is deleted. $D_{(k)}$ can be calculated without fitting a new regression coefficient each time an observation is deleted. Cook's D statistics (i.e. colloquially Cook's Distance) is a measure of the influence of observations in subset U on a vector of parameter estimates (**?**).

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)}$$

If V is known, Cook's D can be calibrated according to a chi-square distribution with degrees of freedom equal to the rank of $\boldsymbol{X}$ (**?**).

0.2.2 Cook's Distance

- For variance components γ: $CD(\gamma)_i$,
- For fixed effect parameters β: $CD(\beta)_i$,
- For random effect parameters $\boldsymbol{u}$: $CD(u)_i$,
- For linear functions of $b\hat{e}ta$: $CD(\psi)_i$

Random Effects

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects.

linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

0.2.3 Information Ratio

0.3 Computation and Notation

with $\boldsymbol{V}$ unknown, a standard practice for estimating $\boldsymbol{X\beta}$ is the estime the variance components σ_j^2, compute an estimate for $\boldsymbol{V}$ and then compute the projector matrix A, $\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{AY}$.

? remarks that $\boldsymbol{D}$ is a block diagonal with the $i-$th block being $u\boldsymbol{I}$

0.4 Exention of Cook's Distance methodology to LME models

Cook's Distance is extended to LME models. For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ.

Diagnostic methods for variance components are based on 'one-step' methods. **?** gives a completely general method for assessing the influence of local departures from assumptions in statistical models. For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

Cook's Distance was extended from classical linear models to LME models. For linear mixed effects models, Cook's distance can be extended to model influence diagnostics by definining.

$$CD_{\beta i} = \frac{(\hat{\beta} - \hat{\beta}_{[i]})^T(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})(\hat{\beta} - \hat{\beta}_{[i]})}{p}$$

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

0.4.1 Cook's Distance

- For variance components γ: $CD(\gamma)_i$,
- For fixed effect parameters β: $CD(\beta)_i$,
- For random effect parameters $\boldsymbol{u}$: $CD(u)_i$,
- For linear functions of $\hat{beta}$: $CD(\psi)_i$

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.

0.5 Cook's Distance for LMEs

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on 'one-step' methods. **?** gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

Random Effects

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects.

linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ.

0.6 Cook's Distance for LMEs

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on 'one-step' methods. **?** gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$\mathrm{CD}_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$\mathrm{CD}_i(b) = g\prime_{(i)}(I_r + \mathrm{var}(\hat{b})D)^{-2}\mathrm{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

¿¿¿¿¿¿¿ origin/master