Application of maximum likelihood theory to estimation of variance components

Let

\[ \mathbf{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_p) = (\log \sigma_1^2, \log \sigma_2^2, \ldots, \log \sigma_p^2) \]

be the vector of log variances in the mixed linear model.

For \( \hat{\sigma}^2 \), the MLE of \( \sigma^2 \) let

\[ \hat{\gamma} = (\log \hat{\sigma}_1^2, \log \hat{\sigma}_2^2, \ldots, \log \hat{\sigma}_p^2) \]

Consider a full rank version of the fixed effects part of the mixed linear model let

\[ l(\mathbf{\beta}, \sigma^2) = \log L(\mathbf{\beta}, \sigma^2) \]

be the (unrestricted) log-likelihood and

\[ l_r(\sigma^2) \]

be the restricted log-likelihood
Define 
\[
\ell^R (\beta, \Sigma) = \ell (\beta, (e^1, e^2, \ldots, e^n))
\]
and 
\[
\ell^X (\Sigma) = \ell (e^1, e^2, \ldots, e^n)
\]
These are a loglikelihood and a restricted loglikelihood for an alternative (from \(\xi^2\)) parameterization in terms of log variances (i.e. components of \(\Sigma\)).

Now think of \(\Sigma\) or \((\beta, \Sigma)\) playing the role of \(\Theta\) in the theory of maximum likelihood — see what we get.

Taking first the ML case: let
\[
M = \begin{pmatrix}
M_{11} & M_{12} \\
M_{21} & M_{22}
\end{pmatrix}
\]

where
\[
M_{11} = \left( \frac{\partial^2 \ell^X}{\partial \beta_i \partial \beta_j} \right)_{\text{MLE}}
\]
\[
M_{22} = \left( \frac{\partial^2 \ell}{\partial \beta_i \partial \beta_j} \right)_{\text{MLE}}
\]
\[
M_{12} = \left( \frac{\partial \ell^X}{\partial \beta_i} \left( \frac{\partial \ell}{\partial \beta_j} \right) \right)_{\text{MLE}}
\]
\[
M_{21} = \left( \frac{\partial \ell}{\partial \beta_i} \left( \frac{\partial \ell^X}{\partial \beta_j} \right) \right)_{\text{MLE}}
\]
Then \(-M^{-1} = Q\) functions as an estimated variance-covariance matrix for \((\hat{Y}, \hat{\beta})\) which large sample theory says is approximately MVN and so approximate confidence limits for \(Y_i\) are

\[ \hat{Y}_i \pm Z \sqrt{Q_{ii}} \]

where \(Q_{ii}\) is the \(i\)th diagonal entry of \(Q = -M^{-1}\)

So approximate confidence limits for \(\tilde{\sigma}_i^2\) are

\[ (\hat{\tilde{\sigma}}_i - Z \sqrt{Q_{ii}}, \hat{\tilde{\sigma}}_i + Z \sqrt{Q_{ii}}) \]

A similar thing can be done based on REML estimates - let

\[ M_\text{REML} = \left( \begin{array}{c|c} \frac{\partial^2 \ell}{\partial \hat{\beta}_j \partial \hat{\beta}_k} & \hat{\beta}_j \hat{\beta}_k \text{REML} \\ \hline \frac{\partial^2 \ell}{\partial \hat{Y}_i \partial \hat{Y}_j} \end{array} \right) \]

have entries that are logs of REML estimates of variance components

and \(Q = -M_\text{REML}^{-1}\) functions as an estimated variance-covariance matrix for \(\hat{\beta}_j \hat{\beta}_k \text{REML}\), and we proceed as above with \(Q_{ii}\) the \(i\)th diagonal entry of \(Q_\text{REML}\).
Note

1. This is what R does.
2. The algorithm seems to help in small samples.

Another (old-fashioned) approach to inference for variance components based on ANOVA-type MS's:

Suppose $MS_1$, $MS_2$, ..., $MS_k$ are independent r.v.'s and

$$ \frac{MS_i}{EMS_i} \sim \chi^2_{df_i} $$

It is common to build estimates of variance components using linear combinations of MS's:

$$ S^2 = a_1 MS_1 + a_2 MS_2 + \ldots + a_k MS_k $$

and consider this r.v. and approximation to its distribution:

$$ E S^2 = a_1 EMS_1 + a_2 EMS_2 + \ldots + a_k EMS_k $$

$$ Var S^2 = \sum a_i^2 Var MS_i $$
\[
= \sum a_i^2 \text{Var} \left[ \frac{\text{EMS}_i}{df_i}, \frac{(df_j)}{\text{EMS}_i} \right]
\]

\[
= \sum a_i^2 \left[ \frac{\text{EMS}_i}{df_i} \right]^2 \text{Var} \left( \frac{a \chi^2_{df_i}}{df_i} \text{ r.v.} \right)
\]

\[
= \sum a_i^2 \left( \frac{\text{EMS}_i}{df_i} \right)^2 2 \frac{df_i}{df_i} \text{r.v.}
\]

\[
= 2 \sum a_i^2 \frac{(\text{EMS}_i)^2}{df_i}.
\]

And a natural (plug-in) estimate is \( \hat{\text{Var}} S^2 = 2 \sum a_i^2 \frac{(\text{EMS}_i)^2}{df_i} \).