Compiled: September 14, 2020

Notes 1: BLUP of Composite Likelihood

By: Eric Morenz

1 Notation and References

Our goal is to estimate $\boldsymbol{b}$ the vector of cluster predictions in mixed effect models. We begin in the simplest case of a random intercept linear mixed model. We really heavily on the notes *Fitting Linear Mixed-Effects Models Using* `lme4` by Bates, Machler, Bolker, and Walker. We write the model as

$$Y = X\beta + Z\boldsymbol{b} + \epsilon,$$

where $b = G\boldsymbol{u}$ denotes the random effects with G a block diagonal positive semi-definte matrix and $\boldsymbol{u} \sim \mathcal{N}(0, I)$. The coefficients β denotes the fixed effects and (X, Z) denote the covariates for the fixed and random effects. ϵ is the independent random error term. We denote the covariance matrix of Y, $\Xi = \sigma^2(I + Z^T G^T G Z)$,the notation is similar to Lumley doc.

2 Pairwise approximation

We now suppose that the subscript p denotes that data is stacked in pairs. For example if we had one cluster of k observations, then $Y_p = [Y_{12}, Y_{13}, \dots, Y_{1k}, Y_{23}, \dots Y_{k-1k}]$, where $Y_{ij} = [Y_i, Y_j]$, the pairs of observations in order.

We make an deduction as to the form of the population estimator of the BLUP

$$\tilde{\boldsymbol{b}} = G Z_P^T \Xi_P^{-1}(\theta)(Y_P - X_P\tilde{\beta}_\theta) \tag{1}$$

and we posit as a sample version of Equation **??**

$$\hat{\boldsymbol{b}} = G Z_S^T W^{1/2} \Xi_S^{-1}(\theta) W^{1/2}(Y_S - X_S\tilde{\beta}_\theta) \tag{2}$$

using the same notation as in the paper where $W^{1/2}$ is an $N_P \times N_P$ diagonal matrix with entry k being $\pi_k^{-1/2}$. We next rewrite the first equation as a vector of sums. Consider first that there is only one cluster and index pairs from a cluster as i, j. Then we may rewrite the residual vector as follows

$$Y_P - X_P\tilde{\beta}_\theta = (e_k)_{i=1}^{2N_P}$$

with $e_k = y_k - \sum_{\ell=1}^{P} x_{k\ell}\tilde{\beta}_\ell$ denoting the k^{th} population level residuals.

Next we remark that we can relabel the observations so that every odd entry is the observation from the first index of the pair and every even entry is the observation from the second index. Then we sort these entries in ascending order of pairs, first by the first index then by the second index. For example with 3 observations

$$Y_P = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_1 \\ Y_3 \\ Y_2 \\ Y_3 \end{pmatrix}$$

We may then rewrite the residual vector as

$$(e_k)_{i=1}^{2N_P} = (e_i)_{i<j}$$

using the reorderd version.

Next if we denote $\Sigma_{ij} = VCOV(Y_i, Y_j)$ a 2 x 2 matrix then Ξ_P^{-1} is block diagonal with $k = 1, \ldots, N_P$ entries Σ_{ij}^{-1} and the ordering $i < j$. For now we will label the entries of

$$\Sigma_{ij}^{-1} = \begin{pmatrix} \sigma^{ii} & \sigma^{ij} \\ \sigma^{ij} & \sigma^{jj} \end{pmatrix}$$

Leading to

$$\Xi_P^{-1}(\theta)(Y_P - X_P\tilde{\beta}_\theta) = \begin{pmatrix} \sigma^{ii}e_i + \sigma^{ij}e_j \\ \sigma^{jj}e_j + \sigma^{ij}e_i \end{pmatrix}_{i<j}$$

where we have a vector of length $2N_P$ defined in these pairs, which we can rewrite as

$$(\sigma^{ii}e_i + \sigma^{ij}e_j)_{i\neq j}$$

Our last task is to now multiply this vector by GZ_P^T. G is a model parameter which defines $Var(\boldsymbol{b})$ and is a $p \times p$ symmetric positive definite matrix. Z_P is the covariate matrix index in the same was as the rest of our data was. Therefore we for now ignore G and focus on the vector

$$Z_P^T\Xi_P^{-1}(\theta)(Y_P - X_P\tilde{\beta}_\theta) = \left(\textstyle\sum_{i<j} z_{ip}(\sigma^{ii}e_i + \sigma^{ij}e_j) + z_{jp}(\sigma^{jj}e_j + \sigma^{ij}e_i)\right)_{p=1}^{d}$$

where d is the number of dimensions. This can also be simplified as

$$Z_P^T\Xi_P^{-1}(\theta)(Y_P - X_P\tilde{\beta}_\theta) = \left(\textstyle\sum_{i\neq j} z_{ip}[\sigma^{ii}e_i + \sigma^{ij}e_j]\right)_{p=1}^{d}$$

This is a sum of terms depending only pairs of observations. We note that for now we may ignore the G matrix as this is just a simple multivariate delta method of our original vector of interest. The sample form of this estimator for any choice p can clearly be written as

$$\sum_{i\neq j} \frac{R_{ij}}{\pi_{ij}} z_{ip}[\sigma^{ii}e_i + \sigma^{ij}e_j]$$

We then conclude that the postulated sample form is indeed correct for the population approximation.

3 Pairwise Approximation Simulation Results

We evaluate the performance of the pairwise approximation derived in Equation ??. The R file `PairwiseBLUPs.R` evaluates the performance of the approximation for a single cluster for $N = 10, 25, 50, 100, 500, 1000$. Fifty clusters are simulated in total, however this first approximation, does not borrow information between clusters, but `lme4::lmer` does.

The R code `PairwiseBLUPs.R` on Github describes the simulation we use to evaluate the performance of the pairwise likelihood approach to approximate model BLUPS. The model is correctly specified, low dimensional and simple in this case, I doubt in more complicated settings the approximations will be as close.

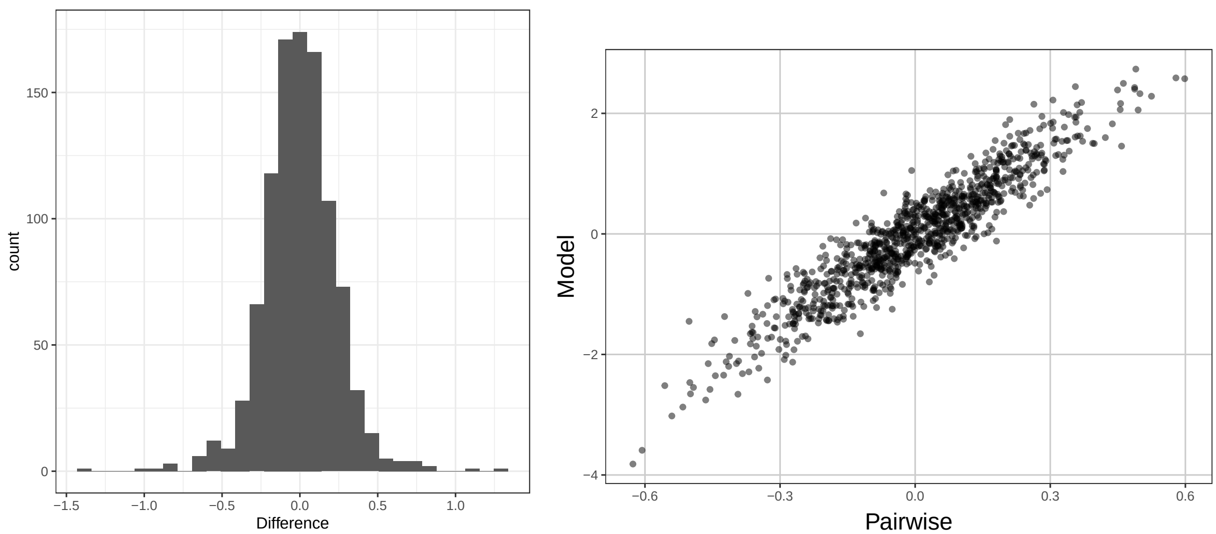

Figure 1: 200 replications of a dataset with 50 clusters and 50 observations per cluster.

We observe in Figure ?? that the full pairwise approach *appears* approximately unbiased. And strongly correlated with the model BLUPS. The approximations appear to be $O_P\left(n^{-1/2}\right)$ at first glance, and the standard error of the difference appears to be quite small. However, each observed point is downward biased towards 0, this is likely due to the approximation not being correctly implemented.

Compiled: September 14, 2020

Notes 2: Sample BLUPs via Risk Minimization

By: Eric Morenz

In a simulation setting, we want to evaluate the properties of the pairwise BLUPS. The first step is to show that pairwise likelihood BLUPS given in Notes (1) is an appropriate approximation of the model predictions. Second, we aim to show that the weighted estimator from Equation **??** from Notes (1) is unbiased for the population level pairwise BLUPs. And finally we want to evaluate alternative biased estimators for the population level BLUPs that produce more conservative estimates when sample sizes are small.

As an alternative approach to a sampling unbiased estimator of the estimand derived from the pariwise approximation, we consider a pairwise approximation of the penalized loss function. In Section **??** we construct a pairwise approximation of the loss function for LMM with arbitrary weight matrices. And we evaluate properties of the general weighted penalized problem.

In the subsequent sectoin wish to compare choices of sampling schemes and whether post-hoc adjustments (e.g. down-weighting) agree with different well defined estimators. This theory will be backed by the same simulations in this chapter.

4 Sampling Simulation Results

We compare three sampling schemes to estimate the pairwise BLUPS. We describe the simulation here, and the code can be found on Github under `BLUPSimS1.R` and the functions used for estimation in `HelpFunctions.R`. The sampling scheme is messy and we only aim to measure the samping variability of these approaches. The steps of simulation are the following.

1) Generate a dataset with $N = 200$ clusters, of size $M = 50$.

2) Fit a linear mixed model b and a pairwise approach $b_p$to estimate the BLUPs.

3) Sample a collection of clusters

4) Sample within each cluster using three different apporaches

5) Estimate b_P with three different estimands

6) Repeat (3-5) 100 times.

The three sampling schemes are 1) a simple random sample 2) Random sample within cluster of random size 3) sampling proportional to size (not entierly comparable). The three estimators are 1) Unbiased estimation of b_p using cluster weights 2) Unbiased estimation of b_p excluding cluster weights 3) unbiased estimation of the penalized loss function and the resulting estimate.

5 Penalized Least Squares

We begin with the beautiful result that L^2 penalized least squares regression can be viewed as a ordinary least squares regression with "pseudo-data". We may then write the optimization problem as

$$\begin{pmatrix} \hat{u} \\ \hat{\beta} \end{pmatrix} = \arg\min_{\beta,u} \left|\left| \begin{bmatrix} W_1^{1/2}\tilde{Y} \\ 0 \end{bmatrix} - \begin{bmatrix} W_1^{1/2}ZG & W_1^{1/2}X \\ W_2^{1/2} & 0 \end{bmatrix} \begin{bmatrix} u \\ \beta \end{bmatrix} \right|\right|^2 \tag{3}$$

where W_1, W_2 are known fixed weight matrices. And we may view them as weights for the residuals and predictions respectively. Is Ξ correct? double check notation. This problem has as solution

$$\begin{pmatrix} \hat{u} \\ \hat{\beta} \end{pmatrix} = \begin{bmatrix} GZ^TW_1ZG + W_2 & GZ^TW_1X \\ X^TW_1ZG & X^TW_1X \end{bmatrix}^{-1} \begin{bmatrix} GZ^TW_1\tilde{Y} \\ X^TW_1\tilde{Y} \end{bmatrix} \tag{4}$$

We notice then that there is only one big difference in prediction when we include weights on the second term. The Identity matirx becomes a weight matrix. We can get the form of the predictions by

$$\hat{u} = V^{-1}GZ^TW_1[Y - X\hat{\beta}], \tag{5}$$

wtih

$$V = GZ^TW_1^{1/2}\Big(I - (X^TW_1X)^{-1}\Big)W_1^{1/2}ZG + W_2 \tag{6}$$

We can see that we don't believe that this solution would agree with the one given in Notes (1) if we provided the same weigths to the observations. Except maybe only in the random intercept model.

6 Results

We observe that we haven't quite fixed the kinks. We appear to be underestimating the BLUPs when using the first estimator. Fortunately, the second approach appears to work well. It is also easier to implement using exsiting code.

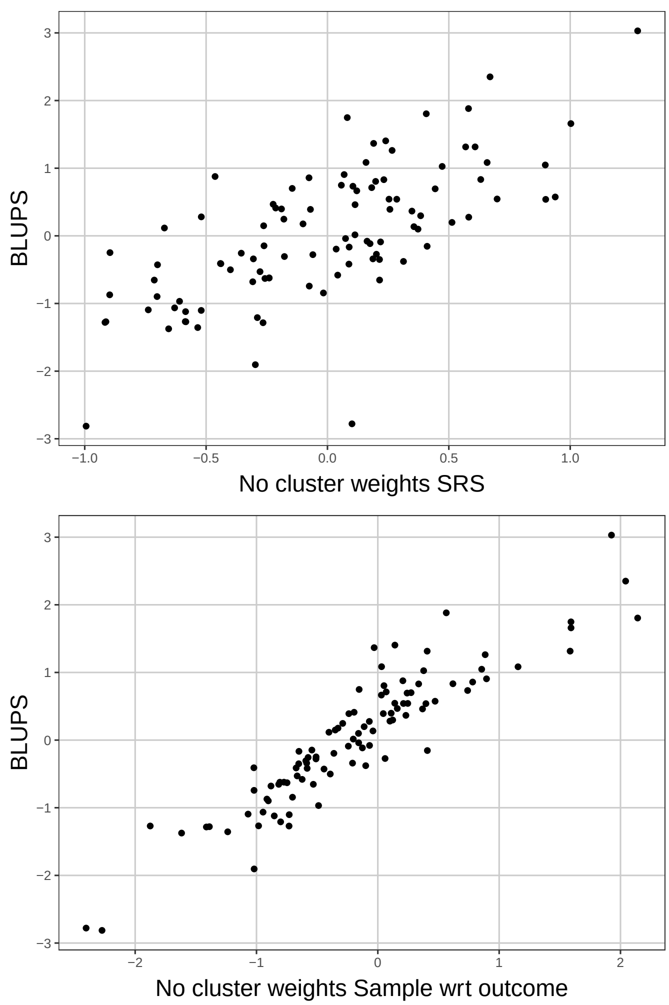

Figure 2: Estimation of BLUPs using approach **one**. 1000 replications of a dataset with 50 clusters and 50 observations per cluster.

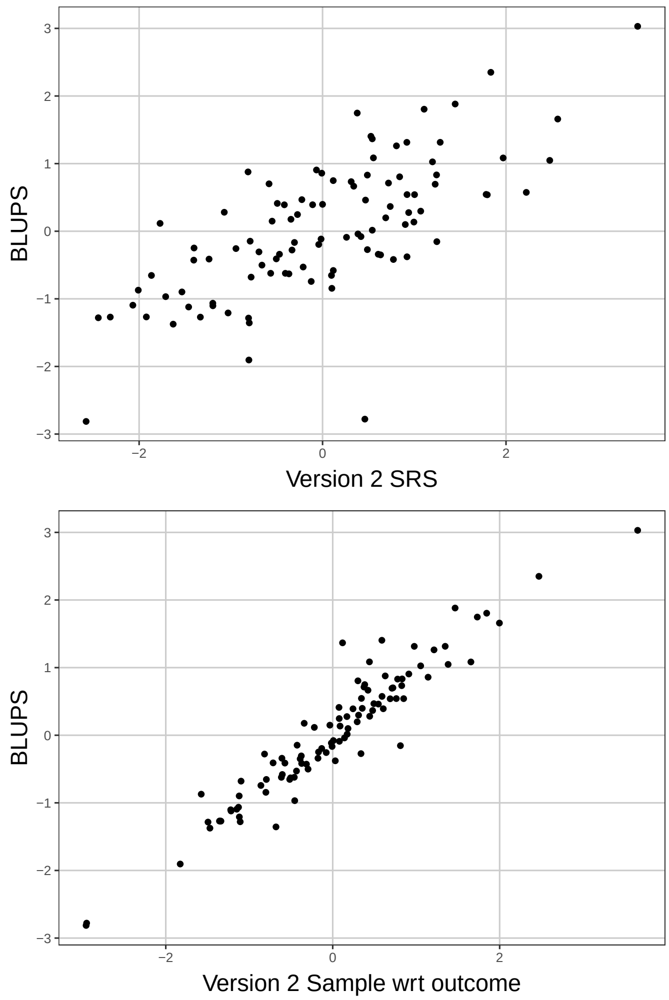

Figure 3: Estimation of BLUPs using approach **two**. 1000 replications of a dataset with 50 clusters and 50 observations per cluster.