Review Patterson and Thompson (1971) - REML

* This paper considers the estimation of weights to be used in the recovery of inter-block information in incomplete block designs with possibly unequal block sizes.
* This method consists of maximising the likelihood, not all the data, but of a set of selected error contrasts.
* Nelder’s (1968) method can only be used when the design are generally balanced, that is when the block sizes are equal.
* The maximum likelihood procedure described by Harlty and Roa (1967) can be used buy does not give the same estimates as Nelder’s method in the balanced case.
* Both methods are using the same total sum of squares of deviations, but whereas one methods equates the sum of squares to (n-1)\*sigma^2 and the other equates the sum of squares to n\*sigma^2. The former method gives an unbiased estimate of sigma^2, the latter maximizes the likelihood of the sample.
* A different method for unbalanced design has been descried by Cunningham & Henderson (1968) and modified by Thompson (1969). This method allows for error in estimation of treatment effects but in general the estimates are not efficient.
* The contrasts among yields are divided into two sets: 1) contrasts between treatment totals. 2) Error contrasts.
* The summary of the method is the present paper consists of maximising the joint likelihood of all possible contrasts in set 2). Contrasts in set 1) are excluded from the likelihood function.
* *Asymptotic variance* is the variances of from the asymptotically normal estimator. An asymptotically normal estimator is a consistent estimator whose distribution around the true parameter θ approaches a normal distribution with standard deviation shrinking in proportion to 1/(n)^(-1) as the sample size n grows.
* Inversion of the matrix H has been shown in the paper by Henderson and Searle (1981).
* Author uses
* H^(-1) exists if (Z`Z + Γ^(-1)) ^(-1) exists. Z’Z gives a diagonal matrix with elements k\_j, where k\_j is the number of plots in block j. Hence, Z`Z + Γ^(-1) can only be singular if \gamma equals –k\_j^(-1), this means the variance components of between blocks becomes negative. Therefore, the author of the paper impose the condition that \gamma > -1/k\_max, where k-max is the largest number of plots in a single blocks.
* REML by also break down if the fisher’s information matrix is singular. The author distinguished three cases.
  + First case is when the number of observation equals the number of treatments, i.e. n=t., because all treatment contrast accounts for all the n-1 degrees of freedom, i.e. there will be no error contrast for construct the restricted likelihood function.
  + Second case is when some treatment comparisons are totally confounded with every block contrasts. i.e. SZ = 0 and W = Γ and hence E = 0.
  + The last case is when all the intra-block comparisons are confounded with the treatment contrasts. Hence, we cannot estimate the \gamma and \sigma individually.
* X’X can be singular, this can be resolved by replacing it by X’H^{-1}X and apply the generalized inverse.

Comparing different REML

There are four different algorithms have to developed based on REML.

Patterson and Thompson introduced REML with Fisher’s scoring algorithm. However, this method can be computationally intensive for large data sets because it requires the calculation of an expected information matrix for the variances components.

This computational issue leads to the development of average information (AI) algorithm by Gilmour et al. The AI algorithm is obtaining a new information matrix by approximating the averages of the observed and expected information matrices. Since, AI algorithm requires calculation of the second derivatives of the restrictive likelihood function, AI algorithm is also known as second order method. There are several disadvantages of AI algorithm, these are AI algorithm does not guarantee to have monotonic convergence sequence, and the variance parameters are not ensured to remain within the parameter space. These two disadvantages are strongly disturbed by the choices of starting values.

Further development leads to the expectation – maximisation (EM) algorithm by Dempster et al. This method, being a first order method, guarantees to have monotonic convergence sequence and is insensitive to starting values, but it may require large number of iterations for convergence.

An improved EM algorithm has introduced by Liu, namely Parameter Expanded EM (PXEM) algorithm. This method reduces the number of iterations.

Some improved methods had described by combining the EM algorithm or PXEM algorithm with the AI algorithm