mR90iod – program running method ℜ with iteration on data
Tom Druet
druet.t@fsagx.ac.be
Apr 21,2001

Introduction

Method ℜ is a method for variance components estimations. The method should be used only for computational reasons. Indeed, it is less demanding than a program running REML or MCMC. However, the theoretical properties of the method are not perfectly known.

Basically, the method works by comparing two sets of solutions: one estimated on the whole data set and the second estimated on a reduced data set (partial data set) based on a sample. After computing both set of solutions, regression factor have to be computed. When correct (co)variances are used all regression factors should be equal to 1. There are as many regression factors as variances and covariances.

How to run this program

To prepare the data and the parameter file, rules are the same as for programs BLUPf90 and REMLf90.
The program asks five questions:
Use fixed effects of complete data set or not: yes –no?
The program computes first solutions for the complete data set. It can use the solution found for the fixed effects for computing solution for the partial data. The solution of fixed effects are then blocked. Then aim of this is to reduce sampling variance and also to eliminate bias resulting from selection on fixed effects.
You have to type "yes" or "no" (not "y" or "n").

Limits of selection (2), random seed
To select the sample for the partial data sets, records are selected randomly. The program proceeds by generating while reading each record a uniform random number comprised between 0 and 1. If the number generated is comprised between the first limit of selection and the second one, then the record is kept (else the record is skipped). To generate the random number, the program need a seed. Each time the same limits and the same seed are used, the same records are sampled.
You have to type three number, tow reals (r1 and r2) comprised between 0 and 1 (r1 must be smaller than r2) and an integer as seed. r2-r1 gives the percentage of records selected.

Acceleration factor?
(co)variances are updated after each computation of regression factors until (co)variances are similar to those obtained 10 rounds earlier. Sometimes, updates are rather small and regression factors move very slowly towards 1. This can be accelerated by making the updates bigger by using an "acceleration factor". The optimal acceleration factor is different for each data set. Generally, an acceleration factor of "10" works well in simple models.
When updates are too big, and regression factors are getting less and less close to 1, acceleration factor as to be reduced. In some extreme case, it might be that a regression factor

lower than 1 should be used (when number of records per level increases: sire model, random regressions).

PCG convergence criteria?
This is the convergence criteria for the preconditioned conjugate gradient algorithm applied to both data sets.
Name of parameter file?
Simply the name of the same parameter file as for running BLUPf90 or REMLf90.

Some limits of application of this program

When a REML program is easily implemented it should be preferred to mRiod. Method ℜ doesn't work very well on small data sets. At least, ten thousands of records should be used. The advantage of the method is only to be able to handle larger data sets with large number of equations (we do not recommend mRiod with a sire model for instance).

Method ℜ can not compute residual (co)variances. In single trait analysis, the method can only estimate variance ratios. You can do this by using the estimate and the residual variance you choose in the parameter file (this value is kept constant). In multitrait analysis, the residual covariance matrix in the parameter file is also kept constant.

Example

The example results from a simulated data set (ex.dat). Records were generated with two fixed effects (sex (row2) and 50 contemporary groups (row3)) and one random genetic effect (row 1). There are 861 records and 991 animals in pedigree file (ex.ped).
The parameter file is ex.par:

```
DATAFILE
 ex.dat
 NUMBER_OF_TRAITS
          1
 NUMBER_OF_EFFECTS
          3
 OBSERVATION(S)
          4
 WEIGHT(S)

 EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
          2             2  cross
          3            50  cross
          1           991  cross
 RANDOM_RESIDUAL VALUES
   90.00000
 RANDOM_GROUP
          3
 RANDOM_TYPE
 add_animal
 FILE
 ex.ped
 (CO)VARIANCES
   10.00000
```

The program asks:

```
use fixed effects of complete data set or not: yes - no?
```

```
yes
limits of selection (2),random seed
0 0.5 26
acceleration factor?
10
pcg convergence criteria?
1ᵉ-12
name of parameter file ?
ex.par
```

We select 50 % of the records with seed = 26 and acceleration factor is 10.
The program runs then a total of 473 pcg rounds, makes 18 updates of variance and converges to the value of 16.989. Residual variance is 90 (see parameter file). We can estimate heritability: h² = 16.989/(16.989+90.000) = 0.159.

If an acceleration factor of 1 is used, the programs runs 2046 pcg runs and makes 137 updates. Estimated value is very close: 16.971 (h² = 0.159).