

PSIMEX: Simex algorithm for errors on pedigree structures

Erica Ponzi

August 28, 2017

1 Introduction

Pedigree SIMEX (P-SIMEX) is a generalization of the SIMEX algorithm used to correct for measurement error in regression models. P-SIMEX applies to error in the pedigree structures, which can affect different estimates in quantitative genetics, as for example inbreeding regression or heritability. The algorithm starts from a given initial pedigree error and calculates the error-prone estimate starting from it. It then simulates progressively increasing errors into the pedigree and calculates the estimate for each of them. By repeating this procedure for a given number of simulations for each error level, it computes the average across simulations and calculates a value for the estimate for each error level. These values are used to fit a function of the estimates in terms of the error proportion in the pedigree and a correct value is extrapolated for the null error rate. Error-free values are given together with their standard errors. The fitting function can be chosen among linear and quadratic by the user.

Here we apply P-SIMEX to the calculation of inbreeding depression and heritability. The error structures modelled by P-SIMEX are the missing paternities across the pedigree and the misassignment of paternities, that is the presence of wrong relationships among individuals in the examined population. The pedigree information are contained in the file 'pedigree', where for each individual the ids of its father and mother are reported, as well its sex and generation number.

```
pedigree0 <- pedigree
head(pedigree0)

##   id father mother generation sex
## 1  1     NA     NA           1  1
## 2  2     NA     NA           1  1
## 3  3     NA     NA           1  1
## 4  4     NA     NA           1  1
## 5  5     NA     NA           1  1
```

```
## 6 6 NA NA 1 1
```

The trait data are reported in the file 'data', where we have information about the covariates included in the model we want to use for the estimation of the parameter of interest. The trait is reported here as 'y'. Any name for the trait is admitted provided coherence is kept in the model specification. In the MCMCglmm case, we recommend not to use the name 'trait', since it can cause some inconsistencies with the package internal functions.

```
head(data)
##   sex f_inb animal.id id year animal      y
## 1  1    0      1  1    1      1 12.99242
## 2  1    0      2  2    1      2 11.59230
## 3  1    0      3  3    1      3 14.02854
## 4  1    0      4  4    1      4 12.35644
## 5  1    0      5  5    1      5 12.82242
## 6  1    0      6  6    1      6 10.12616
```

In case we use a MCMCglmm for the calculation of heritability and we include repeated measurement in it, we need to include in this dataset a copy of the variable 'id' and call it 'animal', as requested by the package 'MCMCglmm'. When the parameter of interest is the inbreeding depression, we can already include the inbreeding coefficient in the dataset or the algorithm will anyway calculate it from the pedigree data and include it in the model covariates.

2 P-SIMEX on misassigned paternities

In this section we will see how to use PSIMEX in the calculation of a parameter of interest when the pedigree is affected by misassignment of paternities. We can model this case by simulating wrong paternities uniformly across the whole pedigree or by replacing true fathers with similar individuals and we can apply this either to the calculation of inbreeding depression or to the heritability. The main function we will use is **Psimex** and the specification of its parameters will allow us to model each different case.

2.1 Uniform replacement

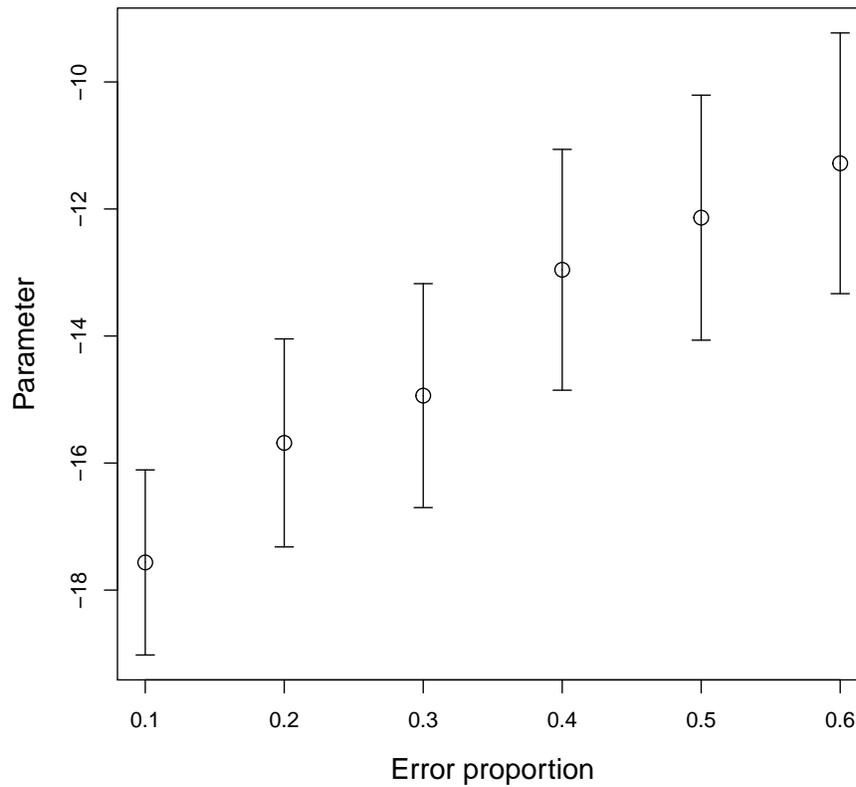
In this case, the error is simulated by replacing fathers with random male individuals taken from the same generations.

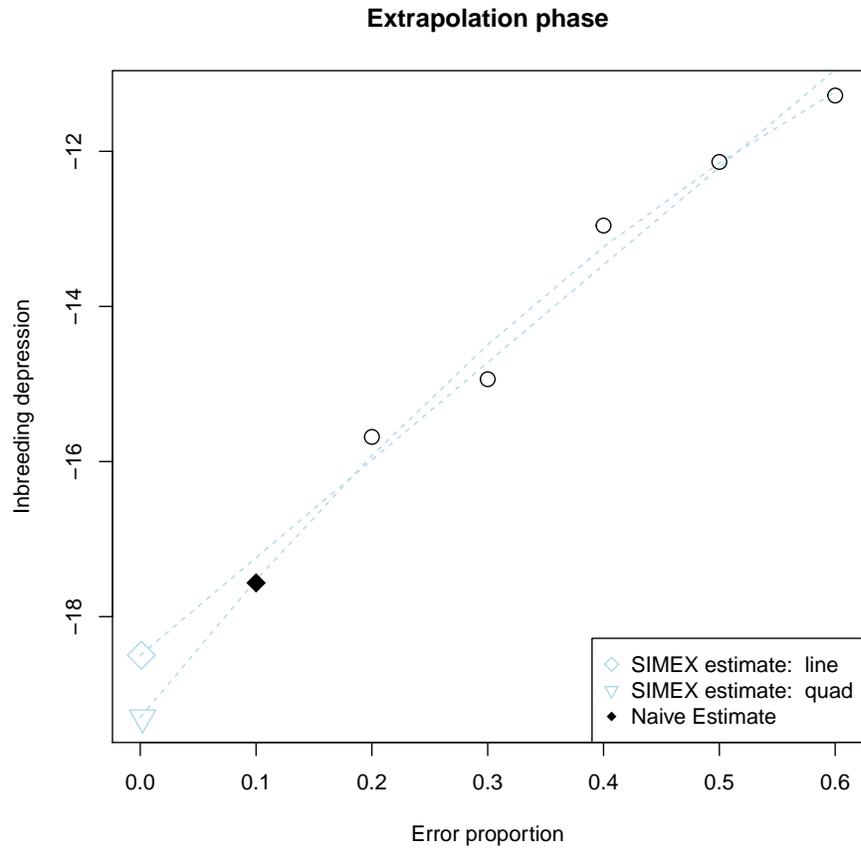
2.1.1 Inbreeding depression

```
lambda <- c(0.2, 0.3, 0.4, 0.5, 0.6)
lambda0 <- 0.1
model <- lm(y~sex+f_inb, data = data)

results_mis <- Psimex(pedigree0, data,
  lambda, lambda0, B = 10,
  model, parameter = "inbreeding",
  error = "misassignment", way = "uniform",
  fitting.method = c("linear", "quadratic"),
  ntop = NA, nbottom = NA,
  prior, nitt, thin, burnin)
```

Simulation results





```

results_mis$description
## [1] "P SIMEX for inbreeding : misassignment paternities in pedigree"

results_mis$error
## [1] "misassignment"

results_mis$fitting.method
## [1] "line" "quad"

results_mis$way
## [1] "uniform"

results_mis$extrapolated_data

```

```

## $fitting.method
## [1] "line" "quad"
##
## $inb_dep
## [1] -18.49745 -19.30173
##
## $se
## [1] 0.7095429 0.6539443
##
## $sampling_se
## [1] 4.821621 4.399151
##
## $var
## [1] 23.75148 19.78018
##
## $AIC
## [1] 9.354286 7.754176

results_mis$lambda
## [1] 0.2 0.3 0.4 0.5 0.6

results_mis$lambda0
## [1] 0.1

results_mis$starting.value

## $model
##
## Call:
## lm(formula = y ~ sex + f_inb, data = data)
##
## Coefficients:
## (Intercept)      sex      f_inb
##      10.082      2.034     -17.565
##
##
## $inb0
## [1] -17.56494
##
## $se0
## [1] 0.7433484
##
## $pval0
## [1] 1.777324e-89
##

```

```
## $mean_inb0
## [1] 0.06351693
##
## $median_inb0
## [1] 0.05380249
##
## $var_inb0
## [1] 0.003530739
```

2.2 Replacement with similar individuals

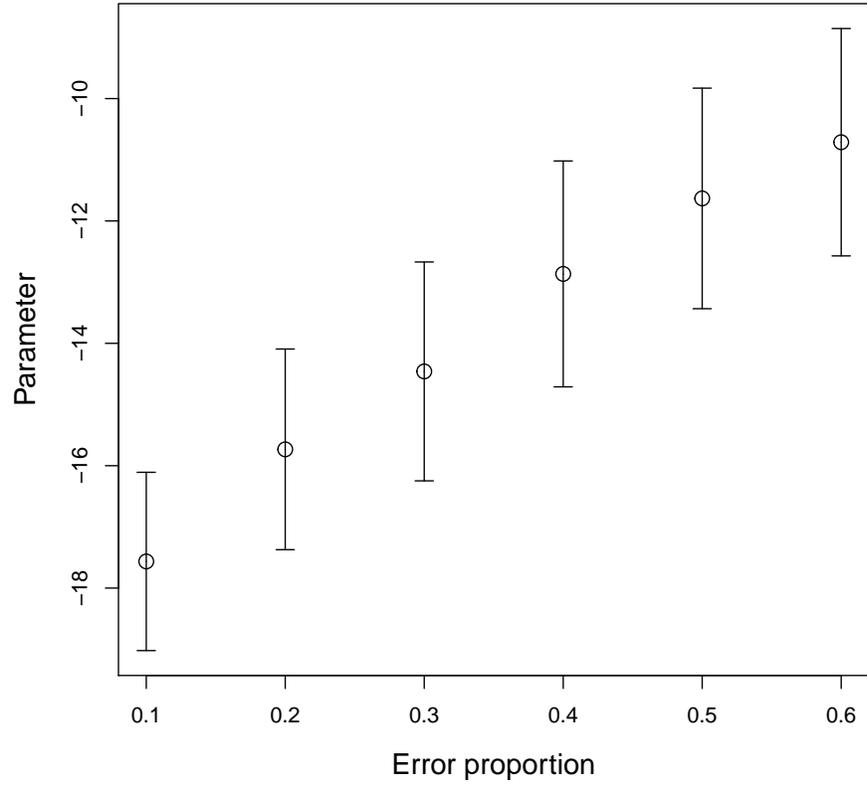
In this case, the error is simulated by replacing fathers with the most similar male individuals taken from the same generations.

2.2.1 Inbreeding depression

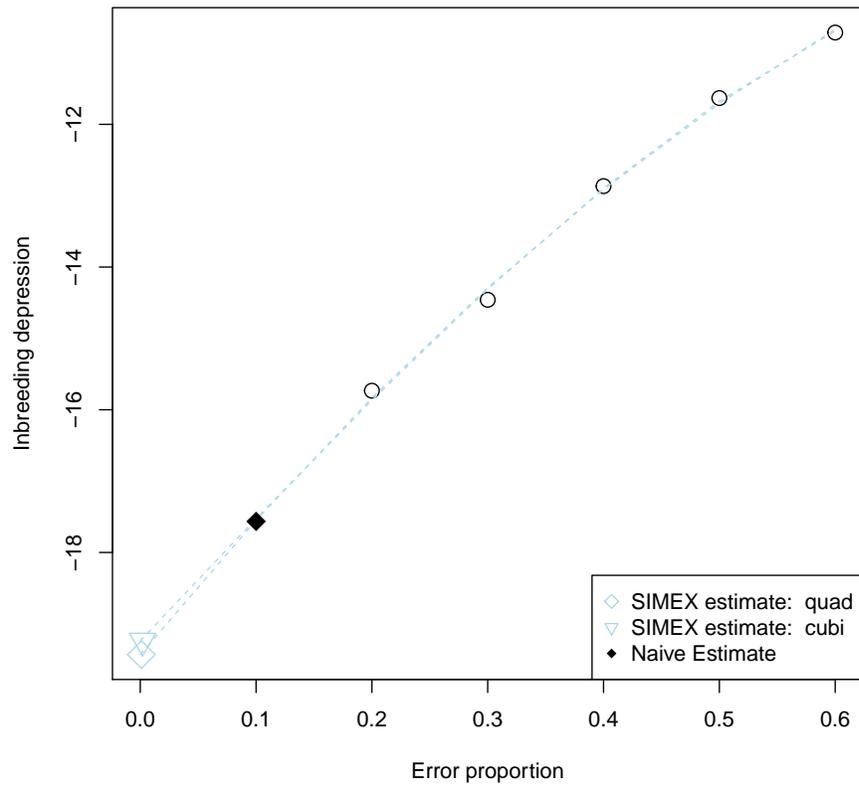
```
lambda <- c(0.2, 0.3, 0.4, 0.5, 0.6)
lambda0 <- 0.1
model <- lm(y~sex+f_inb, data = data)

results_mis1 <- Psimex(pedigree0, data,
                      lambda, lambda0, B = 10,
                      model, parameter = "inbreeding",
                      error = "misassignment", way = "similar",
                      fitting.method = c("quadratic", "cubic"),
                      ntop = NA, nbottom = NA,
                      prior, nitt, thin, burnin)
```

Simulation results



Extrapolation phase



```
results_mis1$description
## [1] "P SIMEX for inbreeding : misassignment paternities in pedigree"
results_mis1$error
## [1] "misassignment"
results_mis1$fitting.method
## [1] "quad" "cubi"
results_mis1$way
## [1] "similar"
results_mis1$extrapolated_data
```

```

## $fitting.method
## [1] "quad" "cubi"
##
## $inb_dep
## [1] -19.43139 -19.23879
##
## $se
## [1] 0.6340407 0.5582182
##
## $sampling_se
## [1] 4.282213 18.696869
##
## $var
## [1] 18.73936 349.88450
##
## $AIC
## [1] -3.789985 -2.269927

results_mis1$lambda

## [1] 0.2 0.3 0.4 0.5 0.6

results_mis1$lambda0

## [1] 0.1

results_mis1$starting.value

## $model
##
## Call:
## lm(formula = y ~ sex + f_inb, data = data)
##
## Coefficients:
## (Intercept)      sex      f_inb
##      10.082      2.034     -17.565
##
##
## $inb0
## [1] -17.56494
##
## $se0
## [1] 0.7433484
##
## $pval0
## [1] 1.777324e-89
##

```

```
## $mean_inb0
## [1] 0.06351693
##
## $median_inb0
## [1] 0.05380249
##
## $var_inb0
## [1] 0.003530739
```

All the described cases can be applied to the calculation of heritability as well. One will need to use a model specification as the one we illustrated in the first section and specify all parameters according to each specific case, as we did in the estimation of inbreeding depression. PSIMEX also allows to simulate missing paternities instead of misassigned paternities. This can be done by setting the 'error' parameter to 'missing', and can be applied to each case and parameter.