HiLMM: Package for heritability estimation

A. Bonnet

April 13, 2015

This vignette explains how to use the HiLMM package which is dedicated to the estimation of heritability in linear mixed models. For further details on the statistical model and on the implementation, we refer the reader to [1].

After installing the package in R, the package has to be loaded by using the following instruction:

```
> library(HiLMM)
```

The package HiLMM contains two functions called `estim_herit` and `data_simu`.

The `estim_herit` function provides an estimation of the heritability and its usage is described as follows:

```
res_herit=estim_herit(Y,W)
```

where the arguments are:

- `Y`: Vector of observations of size n

- `W`: Matrix which contains the genetic information of size $n \times N$

and where the output `res_herit` is a list of four attributes:

- `heritability`: estimated value of the heritability

- `CI_low`: lower bound of the confidence interval for the heritability

- `CI_up`: upper bound of the confidence interval for the heritability

- `standard deviation`: standard deviation associated to the estimation of the heritability

The `data_simu` function generates datasets which can be used as arguments of the function `estim_herit`, which are a vector of observations and a matrix which has the features of a genetic information matrix. This function allows the user to test the function `res_herit` with a dataset of any size with given values of `eta_star` and `q`.These parameters are defined in Equations (4) and (5) of [1].

The usage of this function is as follows:

```
data_sim=data_simu(n,N,eta_star,q)
```

where the arguments are:

- `n`: number of observations

- `N`: number of SNPs

- `eta_star`: value of the heritability

- `q`: proportion of SNPs which are associated to non null effects

and where the output `data_sim` is identical to the arguments of `estim_herit`, that is:

- `Y`: Vector of observations of size n

- `W`: Matrix which contains the genetic information of size $n \times N$

We can for example generate a dataset with the function `data_simu` and then apply the function `estim_herit`.

```
> data_sim= data_simu(500,1000,0.6,0.5)
> res_herit=estim_herit(data_sim$Y,data_sim$W)

> data_sim$W[1:10,1:10]

       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,]    1    0    0    1    0    1    1    0    2    0
 [2,]    0    1    2    1    1    1    0    0    2    2
 [3,]    0    0    2    0    2    0    0    1    2    0
 [4,]    1    0    1    1    1    1    0    1    2    1
 [5,]    1    1    0    1    2    1    0    1    0    1
 [6,]    0    0    1    1    1    0    0    1    2    0
 [7,]    1    0    0    0    0    2    0    1    1    0
 [8,]    1    1    0    0    1    1    1    1    2    0
 [9,]    0    0    0    1    0    1    1    1    2    1
[10,]    0    0    2    0    1    0    2    1    2    1

> data_sim$Y[1:10]

 [1]   6.612510 -36.649905  -4.904274 -13.695514 -32.967129  16.195379
 [7]  14.335360  39.472756  56.666535  34.280563
```

The generated matrix `W` contains 0,1,2 at position (i, j) according to the version of the j-th SNP of the i-th individual. More precisely, 0,2,1 means respectively that the individual has two copies of the less frequent SNP, two copies of the more frequent SNP and one copy of each. In practice, the j-th column of W is generated according to a binomial distribution with parameters n and p_j, p_j representing the frequency of the less frequent allele and being uniformly chosen between 0.1 and 0.5.

For this dataset we get the following results.

```
> res_herit$heritability
```

```
[1] 0.4933943
```

```
> res_herit$CI_low
```

```
[1] 0.3237532
```

```
> res_herit$CI_up
```

```
[1] 0.6630354
```

```
> res_herit$standard_deviation
```

```
[1] 0.08655158
```

estim_herit provides an estimation for the heritability, the bounds of the 95%
confidence interval and the standard deviation.
Hereafter, we also provide some information about the R session

```
> sessionInfo()
```

```
R version 3.1.1 (2014-07-10)
Platform: x86_64-pc-linux-gnu (64-bit)

locale:
 [1] LC_CTYPE=fr_FR.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=fr_FR.UTF-8        LC_COLLATE=C
 [5] LC_MONETARY=fr_FR.UTF-8    LC_MESSAGES=fr_FR.UTF-8
 [7] LC_PAPER=fr_FR.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=fr_FR.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] HiLMM_1.1

loaded via a namespace (and not attached):
[1] tools_3.1.1
```

References

[1] Anna Bonnet, Elisabeth Gassiat, and Céline Lévy-Leduc. Heritability estimation in
high-dimensional sparse linear mixed models, 2014. http://arxiv.org/abs/1404.3397.