Estimating the order k of the Lehmann's alternative from the data

2023-05-20

We consider Lehmann's alternative with $(1-\theta)F + \theta F^k$ and we estimate the value of k from the dat through Monte Carlo simulation.

```
library(foreign)
library(readr)
library(R.matlab)
library(doSNOW)
library(foreach)
library(nout)
library(isotree)
```

Digits dataset

```
data = readMat("~/nout/trials/RealData/Datasets/Dataset digits/pendigits.mat")

dataset = cbind(data$X, data$y); colnames(dataset)[ncol(dataset)] = "y"
in_ind = which(dataset[,ncol(dataset)]==0)
out_ind = which(dataset[,ncol(dataset)]==1)

# Initializing parameters
set.seed(321)

B=10^4

 ress = foreach(b = 1:B, .combine=c) %dopar% {
   inlier = sample(in_ind, size = 1)
   outlier = sample(out_ind, size = 1)

   greater.logi = inlier<outlier

   return(greater.logi)
}

greater.prob = mean(ress)

(k=greater.prob/(1-greater.prob))
```

```
## [1] 1.023472
```

Credit Card dataset

```
dataset = read_csv("~/nout/trials/RealData/Datasets/Dataset creditcard/creditcard.csv")
in_ind = which(dataset[,ncol(dataset)]==0)
out_ind = which(dataset[,ncol(dataset)]==1)
```

```r
# Initializing parameters
set.seed(321)

B=10^4

 ress = foreach(b = 1:B, .combine=c) %dopar% {
   inlier = sample(in_ind, size = 1)
   outlier = sample(out_ind, size = 1)

   greater.logi = inlier<outlier

   return(greater.logi)
}

greater.prob = mean(ress)

(k=greater.prob/(1-greater.prob))
```

```
## [1] 0.7176228
```

Shuttle (Statlog) dataset

```r
data = readMat("~/nout/trials/RealData/Datasets/Dataset shuttle/shuttle.mat")
dataset = cbind(data$X, data$y); colnames(dataset)[ncol(dataset)] = "y"
in_ind = which(dataset[,ncol(dataset)]==0)
out_ind = which(dataset[,ncol(dataset)]==1)

# Initializing parameters
set.seed(321)

B=10^4

 ress = foreach(b = 1:B, .combine=c) %dopar% {
   inlier = sample(in_ind, size = 1)
   outlier = sample(out_ind, size = 1)

   greater.logi = inlier<outlier

   return(greater.logi)
}

greater.prob = mean(ress)

(k=greater.prob/(1-greater.prob))
```

```
## [1] 1.009646
```

Cover type dataset

```r
data = readMat("~/nout/trials/RealData/Datasets/Dataset cover type/cover.mat")
dataset = cbind(data$X, data$y); colnames(dataset)[ncol(dataset)] = "y"
in_ind = which(dataset[,ncol(dataset)]==0)
out_ind = which(dataset[,ncol(dataset)]==1)
```

```
# Initializing parameters
set.seed(321)

B=10^4

 ress = foreach(b = 1:B, .combine=c) %dopar% {
   inlier = sample(in_ind, size = 1)
   outlier = sample(out_ind, size = 1)

   greater.logi = inlier<outlier

   return(greater.logi)
}

greater.prob = mean(ress)

(k=greater.prob/(1-greater.prob))
```

```
## [1] 0.1487651
```

Mammography dataset

```
data = readMat("~/nout/trials/RealData/Datasets/Dataset mammography/mammography.mat")
dataset = cbind(data$X, data$y); colnames(dataset)[ncol(dataset)] = "y"
in_ind = which(dataset[,ncol(dataset)]==0)
out_ind = which(dataset[,ncol(dataset)]==1)

# Initializing parameters
set.seed(321)

B=10^4

 ress = foreach(b = 1:B, .combine=c) %dopar% {
   inlier = sample(in_ind, size = 1)
   outlier = sample(out_ind, size = 1)

   greater.logi = inlier<outlier

   return(greater.logi)
}

greater.prob = mean(ress)

(k=greater.prob/(1-greater.prob))
```

```
## [1] 1.289902
```