Description of R file "MOI_NP.R"

Loyce Kayanula, Kristan A. Schneider

Availability and updates

The R-file "MOI_NP.R" is also available via GitHub. Updates of the code and this description will be made available there https://github.com/Maths-against-Malaria/Non-parametric-MOI-estimation.

The non-parametric maximum likelihood estimates (MLE)

All functions needed to calculate the non-parametric MLE of the MOI and lineage frequencies from molecular datasets based are described here. Also the plug-in estimates for the average MOI and the linage prevalences are provided. This description partly uses functions already described in Hashemi and Schneider(2023) (Meraj Hashemi, Kristan A. Schneider bioRxiv 2023.06.01.543300; doi: https://doi.org/10.1101/2023.06.01.543300). The respective parts of the description here are taken from there.

The first step is to load the R-file ""MOI_NP.R". The second step is to is to import the molecular data. The final step is to derive the MLE using he function `MOI.NP`.

Loading the R-file. Save the R-file "MOI_NP.R" in a directory `path` and load it using the function `source`. E.g., if the file is stored in source `"C:/Documents/backslash/Musterfrau"`, the file is loaded by running the following line.

```
source("C:/Documents/backslash/Musterfrau/MOI-MLE-IDM.R")
```

Several packages are required, which are installed by loading the source file. Usere might receive a promt asking from which server to load the required packages which are currently not installed.

Importing data using `DatImp`. Import molecular data using the function `DatImp(path)`. Here, `path` is the location where the molecular dataset is stored. Data needs to be stored in a standardized fashion (see section Data format) as either an ".xlsx"-, ".csv"- or ".txt"-file. If the data is stored in an ".xls"-file, it has to be converted into an ".xlsx"-file with appropriate spreadsheet software.

Code Example. The following code imports the file "Example_Data.xlsx" (see additional files), which is stored in "C:/Documents/backslash/Molecular Data/Example_Data.xlsx". The data contains molecular information of two microsatellite markers ("M1", "M2") from $N = 60$ sampes. Only the first 10 lines of output are shown.

```
path <- "C:/Documents/backslash/Molecular Data/Example_Data.xlsx"
```

```
DatImp(path)
```

```
##    SampleID   M1   M2
## 1     sam_1  201  175
## 2     sam_2  208 <NA>
## 3     sam_3  213  183
## 4     sam_4  201  153
## 5     sam_4  213  168
## 6     sam_4  217 <NA>
## 7     sam_5  213  168
```

```
## 8     sam_5  208  175
## 9     sam_6  201  153
```

To read the data into an array named `dat` use the following code.

```
dat <- DatImp(path)
```

Data format. Molecular data needs to be stored either as ".xlsx"-, ".csv"- or ".txt"-file in a specific format. An Example is provided as an additional file. The format for ".xlsx"-files is described. A data set consists of two or more columns. The first contains the sample IDs, the remaining columns molecular information from samples. One collum corresponds to one marker. Here we describe the simples case of a single molecular marker. Each sample is stored in a $k \times 2$ block (in case of n molecular markers as a $k \times (n+1)$ block). In the first column at least the first row must contain the sample ID. The lineages present in the sample are stored in the second column in consecutive rows in any arbitrary order. Below are four alternative schematic descriptions of a sample in which lineages 1, 2 and 4 were observed. Note that missing values can occur, that the same lineage might be entered multiple times for a sample (but it is counted only once) and that the sample ID has to occur only in the first row. Missing values must be left empty. Examples:

ID1	lineage 2
	lineage 4
	lineage 1

ID1	lineage 1
ID1	lineage 2
	lineage 4

ID1	lineage 1
	lineage 2
	lineage 4
	lineage 4

ID1	lineage 1
	lineage 2
	lineage 4
ID1	lineage 4

Sample IDs and lineages are entered as numbers or strings. See the file "Example_Data.xlsx" for an example of microsatellite data. The first row is reserved for column labels. It can be left empty, but this row must not be omitted. The table below shows again the first 10 rows of the example data set "Example_Data.xlsx", corresponding to the first 6 samples. The data has two molecular markers "M1" and "M2".

```
dat
```

```
##    SampleID   M1   M2
## 1     sam_1  201  175
## 2     sam_2  208 <NA>
## 3     sam_3  213  183
## 4     sam_4  201  153
## 5     sam_4  213  168
## 6     sam_4  217 <NA>
## 7     sam_5  213  168
## 8     sam_5  208  175
## 9     sam_6  201  153
## 10    sam_6  208 <NA>
```

The first sample (sam_1) contains two lineages, "201" at markers "M1" and "175" at marker "M2". The second sample (sam_2) has lineages "208" at marker "M1" but the information at marker "M2" is missing. The fourth sample (sam_4) contains lineages "201", "213", and "217" a marker "M1" (indicating at least MOI = 3), but only the two lineages "153" and "168" at marker "M2". For more information and examples see the description in https://github.com/Maths-against-Malaria/MOI—Incomplete-Data-Model.git

The function `MOI.NP`. The function `MOI.NP(data, markername, M, CI=FALSE, B=1000, alpha_level = 0.95, p = NULL, km=NULL, maxit =1000, eps=10^-8)` is used to calculate the MLE. The basic syntax is `MOI.NP(data, markername, M)` which contains the data from a single molecular marker, the name of the marker and the maximum MOI value M.

The output is a list, which contains a list called "estimates" in which the estimates are stored. This list has 9 elements: (1) the MLE of the lineages frequencies $(\hat{p}_1, \ldots, \hat{p}_n)$, (2) the MLE of the MOI distribution

$(\hat{\kappa}_1, \ldots, \hat{\kappa}_M)$, (3) the plug-in estimate for the average MOI $\hat{\psi} = \sum_{m=1}^{M} m\hat{\kappa}_m$, (4) the maximum MOI imposed, M, (5) the lineage counts, i.e., the number of samples in which each lineage was detected, (6) the sample size N, (7) the observed prevalences, (8) the plug in estimates for the prvalences, with the prevalence of linega k being $1 - \hat{G}(1 - p_k)$, where $\hat{G}(t) = \sum_{m=1}^{M} \hat{\kappa}_m t^m$ is the PGF of the estimated MOI distribution, and (9) the heterozygosity calculated as $1 - \sum_{k=1}^{n} \hat{p}_k^2$.

Code example. This code calculates the MLE for the second marker "M2" in the data "Example_Data.xlsx" imposing a maximum MOI of $M = 6$. Empty observations (in this case sam_2) will be disregarded.

```
MOI.NP(dat[,c("SampleID","M2")], "M2", 6)
```

```
## $estimates
## $estimates$`MLE of lineage freqs.`
##     M2_153     M2_168     M2_175     M2_183     M2_185
## 0.49559692 0.17354544 0.07394282 0.18182642 0.07508840
##
## $estimates$`MLE of MOI distribution`
##     P[MOI=1]     P[MOI=2]     P[MOI=3]     P[MOI=4]     P[MOI=5]     P[MOI=6]
## 4.572243e-01 4.321600e-01 1.106156e-01 2.760176e-08 7.220376e-18 1.857262e-29
##
## $estimates$`average MOI`
## [1] 1.653391
##
## $estimates$`maximum MOI for algorithm`
## [1] 6
##
## $estimates$`lineage counts`
##      M2_153 M2_168 M2_175 M2_183 M2_185
## [1,]     38     16      7     16      7
##
## $estimates$`sample size`
## [1] 59
##
## $estimates$`observed prevalence`
##    M2_153    M2_168    M2_175    M2_183    M2_185
## 0.6440678 0.2711864 0.1186441 0.2711864 0.1186441
##
## $estimates$`estimated prevalences`
##    M2_153    M2_168    M2_175    M2_183    M2_185
## 0.6452280 0.2645063 0.1181239 0.2760365 0.1198897
##
## $estimates$heterozygosity
## [1] 0.680099
```

The resulting estimates are $\hat{p}_1 = 0.49559692$, $\hat{p}_2 = 0.17354544$, $\hat{p}_3 = 0.07394282$, $\hat{p}_4 = 0.18182642$, and $\hat{p}_5 = 0.07508840$ for the lineage frequencies, $\hat{\kappa}_1 = 4.572243e-01$, $\hat{\kappa}_2 = 4.321600e-01$, $\hat{\kappa}_3 = 1.106156e-01$, $\hat{\kappa}_4 = 2.760176e-08$, $\hat{\kappa}_5 = 7.220376e-18$, and $\hat{\kappa}_6 = 1.857262e-29$. Furthermore, the average MOI is estimated as 1.653391, while the heterozygosity is 0.680099.

Several options can be specified. As optional arguments `km` and `p` initial frequency distributions for MOI and lineages can be specified for the numerical iteration. Here, `km` has to be a frequency vector of length `M` and `p` a frequency vector corresponding to the number of lineages in the data. In case the algorithm is not converging the maximum number of iterations can be set by specifying the argument `maxit` (default `maxit=1000`).

Furthermore, the numerical threshold for convergence can be adjusted by the optional argument `eps` (default `eps=10^-8`).

Code example. The following code uses $p_1 = 0.5$, $p_2 = 0.15$, $p_3 = 0.1$, $p_4 = 0.2$, and $p_5 = 0.05$ as initial frequencies, $\kappa_1 = 0.5$, $\kappa_2 = 0.4$, $\kappa_3 = 0.05$, $\kappa_4 = 0.048$, $\kappa_5 = 0.001$, and $\kappa_6 = 0.001$ for the initial frequency and MOI distributions to calculate the MOI of marker "M2" in the data "Example_Data.xlsx". Furthermore it uses a maximum of `maxit=500` iterations and a numerical threashold of convergence of `eps=10^-6`.

```
MOI.NP(dat[,c("SampleID","M2")], "M2", 6, p= c(0.5,0.15,0.1,0.2,0.05),
            km = c(0.5,0.4,0.05,0.048,0.001,0.001), maxit=500, eps=10^-6)
```

```
## $estimates
## $estimates$`MLE of lineage freqs.`
##     M2_153     M2_168     M2_175     M2_183     M2_185
## 0.49559710 0.17354527 0.07394287 0.18182632 0.07508844
##
## $estimates$`MLE of MOI distribution`
##     P[MOI=1]     P[MOI=2]     P[MOI=3]     P[MOI=4]     P[MOI=5]     P[MOI=6]
## 4.572236e-01 4.321640e-01 1.106097e-01 2.748504e-06 5.870681e-15 1.502510e-23
##
## $estimates$`average MOI`
## [1] 1.653392
##
## $estimates$`maximum MOI for algorithm`
## [1] 6
##
## $estimates$`lineage counts`
##      M2_153 M2_168 M2_175 M2_183 M2_185
## [1,]     38     16      7     16      7
##
## $estimates$`sample size`
## [1] 59
##
## $estimates$`observed prevalence`
##    M2_153    M2_168    M2_175    M2_183    M2_185
## 0.6440678 0.2711864 0.1186441 0.2711864 0.1186441
##
## $estimates$`estimated prevalences`
##    M2_153    M2_168    M2_175    M2_183    M2_185
## 0.6452281 0.2645061 0.1181240 0.2760363 0.1198897
##
## $estimates$heterozygosity
## [1] 0.6800989
```

Adding confidence intervalls (CIs) to the estimates. The function `MOI.NP` also allows to calculate bootstrap percentile confidence intervals (CIs). These are calculated if the option `CI = TRUE` is specified (default `CI = FALSE`). This will result $\alpha \times 100\%$ confidence interval. The coverage α is specified by the option `alpha_level` default (`alpha_level = 0.95`). The CIs are by default calculated based on `B` bootstrap repeats (default `B= 1000`). The output of the function contain a list with two elements. The first element contains the estimates as described above. The second are the element is a list called `CIs` with 5 elements: (1) the CI of the average MOI, (2) the CI for the heterozygosity, (3) the CI for each component of the MOI distribution, (4) the CI for each lineage frequency, and (5) th CI for each lineage prevalence.

Code example. The following code uses calculates the MLE and 95% bootstrap CIs for the first marker ("M1") in the data set "Example_Data.xlsx".

```r
MOI.NP(dat[,c("SampleID","M1")], "M1", 6, CI=TRUE)
```

```
## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
```

```
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations

## Warning in NonparametricEst(newdtB, M, p, km, maxit = maxit, eps = eps): EM
## algorithm failed to converge within 1000 iterations
## $estimates
## $estimates$`MLE of lineage freqs.`
##     M1_201     M1_208     M1_213     M1_217
## 0.42497262 0.24939515 0.24832971 0.07730252
##
## $estimates$`MLE of MOI distribution`
##     P[MOI=1]     P[MOI=2]     P[MOI=3]     P[MOI=4]     P[MOI=5]     P[MOI=6]
## 3.879155e-01 4.478972e-01 1.641870e-01 2.399172e-07 1.775955e-20 2.847995e-39
##
## $estimates$`average MOI`
## [1] 1.776272
##
## $estimates$`maximum MOI for algorithm`
## [1] 6
##
## $estimates$`lineage counts`
##      M1_201 M1_208 M1_213 M1_217
## [1,]     36     23     23      8
##
## $estimates$`sample size`
## [1] 60
##
```

```
## $estimates$`observed prevalence`
##    M1_201    M1_208    M1_213    M1_217
## 0.6000000 0.3833333 0.3833333 0.1333333
##
## $estimates$`estimated prevalences`
##    M1_201    M1_208    M1_213    M1_217
## 0.5976199 0.3870458 0.3856195 0.1317663
##
## $estimates$heterozygosity
## [1] 0.689557
##
##
## $CIs
## $CIs$`CI average MOI`
##     2.5%    97.5%
## 1.528420 2.046494
##
## $CIs$`CI heterozygosity`
##      2.5%     97.5%
## 0.6297055 0.7165304
##
## $CIs$`CIs MOI distribution`
##                   2.5%        97.5%
## P[MOI=1] 1.932377e-01 5.646405e-01
## P[MOI=2] 1.783444e-01 7.029715e-01
## P[MOI=3] 7.044128e-09 3.682168e-01
## P[MOI=4] 2.374164e-23 1.279552e-01
## P[MOI=5] 1.804015e-45 1.405623e-05
## P[MOI=6] 2.284360e-73 3.813399e-21
##
## $CIs$`CIs lineage frequencies`
##               2.5%     97.5%
## M1_201 0.32481308 0.5303872
## M1_208 0.16550252 0.3517229
## M1_213 0.15928161 0.3478949
## M1_217 0.02961551 0.1270035
##
## $CIs$prevalences
##               2.5%     97.5%
## M1_201 0.46788474 0.7184190
## M1_208 0.26780822 0.5187909
## M1_213 0.26176566 0.5066676
## M1_217 0.04958672 0.2136543
```

Code example. The following code uses calculates the MLE and 90% bootstrap CIs for the first marker ("M2") in the data set "Example_Data.xlsx" based on 200 bootstrap repeats.

```
MOI.NP(dat[,c("SampleID","M2")], "M2", 6, CI=TRUE, alpha_level=0.90, B=200)
```

```
## $estimates
## $estimates$`MLE of lineage freqs.`
##     M2_153     M2_168     M2_175     M2_183     M2_185
## 0.49559692 0.17354544 0.07394282 0.18182642 0.07508840
##
## $estimates$`MLE of MOI distribution`
```

```
##     P[MOI=1]     P[MOI=2]     P[MOI=3]     P[MOI=4]     P[MOI=5]     P[MOI=6]
## 4.572243e-01 4.321600e-01 1.106156e-01 2.760176e-08 7.220376e-18 1.857262e-29
##
## $estimates$`average MOI`
## [1] 1.653391
##
## $estimates$`maximum MOI for algorithm`
## [1] 6
##
## $estimates$`lineage counts`
##      M2_153 M2_168 M2_175 M2_183 M2_185
## [1,]     38     16      7     16      7
##
## $estimates$`sample size`
## [1] 59
##
## $estimates$`observed prevalence`
##    M2_153    M2_168    M2_175    M2_183    M2_185
## 0.6440678 0.2711864 0.1186441 0.2711864 0.1186441
##
## $estimates$`estimated prevalences`
##    M2_153    M2_168    M2_175    M2_183    M2_185
## 0.6452280 0.2645063 0.1181239 0.2760365 0.1198897
##
## $estimates$heterozygosity
## [1] 0.680099
##
##
## $CIs
## $CIs$`CI average MOI`
##       5%      95%
## 1.477895 1.857157
##
## $CIs$`CI heterozygosity`
##        5%       95%
## 0.5905313 0.7347975
##
## $CIs$`CIs MOI distribution`
##                      5%          95%
## P[MOI=1] 3.112001e-01 5.988579e-01
## P[MOI=2] 2.651939e-01 6.444930e-01
## P[MOI=3] 9.565355e-09 2.217114e-01
## P[MOI=4] 2.598664e-23 3.166252e-09
## P[MOI=5] 5.088532e-41 6.674495e-21
## P[MOI=6] 1.852306e-60 4.279466e-35
##
## $CIs$`CIs lineage frequencies`
##                5%       95%
## M2_153 0.41147953 0.5981572
## M2_168 0.12373800 0.2161989
## M2_175 0.03145297 0.1226993
## M2_183 0.10715120 0.2474272
## M2_185 0.03255743 0.1283070
##
```

```
## $CIs$prevalences
##                5%       95%
## M2_153 0.53789104 0.7613870
## M2_168 0.18405844 0.3489824
## M2_175 0.05056268 0.1993348
## M2_183 0.17223018 0.3646635
## M2_185 0.05179261 0.2052741
```