Package 'pedprobr'

April 13, 2023

Type Package

Title Probability Computations on Pedigrees

Version 0.8.0

Description An implementation of the Elston-Stewart algorithm for calculating pedigree likelihoods given genetic marker data (Elston and Stewart (1971) <doi:10.1159/000152448>). The standard algorithm is extended to allow inbred founders. 'pedprobr' is part of the 'ped suite', a collection of packages for pedigree analysis in R. In particular, 'pedprobr' depends on 'pedtools' for pedigree manipulations and 'pedmut' for mutation modelling. For more information, see 'Pedigree Analysis in R' (Vigeland, 2021, ISBN:9780128244302).

License GPL (>= 2)

URL https://github.com/magnusdv/pedprobr

BugReports https://github.com/magnusdv/pedprobr/issues

Depends R (>= 4.1.0), pedtools (>= 1.1.0)

Imports pedmut (>= 0.5.0)

Suggests testthat

Encoding UTF-8

Language en-GB

RoxygenNote 7.2.3

SystemRequirements MERLIN (https://csg.sph.umich.edu/abecasis/merlin/) for calculations involving multiple linked markers.

NeedsCompilation no

Author Magnus Dehli Vigeland [aut, cre] (<https://orcid.org/0000-0002-9134-4962>)

Maintainer Magnus Dehli Vigeland <m.d.vigeland@medisin.uio.no>

Repository CRAN

Date/Publication 2023-04-13 15:40:09 UTC

R topics documented:

`allGenotypes` *Genotype matrix*

Description

An autosomal marker with `n` alleles has `choose(n+1, 2)` possible unordered genotypes. This function returns these as rows in a matrix.

Usage

```
allGenotypes(n)
```

Arguments

`n`	A positive integer.

Value

An integer matrix with two columns and `choose(n+1, 2)` rows.

Examples

```
allGenotypes(3)
```

genoCombinations *Genotype combinations*

Description

Returns the possible genotype combinations in a pedigree, given partial marker data. This function is mainly for internal use.

Usage

```
genoCombinations(x, partialmarker, ids, make.grid = TRUE)
```

Arguments

x	a ped() object.
partialmarker	a marker() object compatible with x.
ids	a vector with ID labels of one or more pedigree members.
make.grid	a logical indicating if the result should be simplified to a matrix.

Value

If make.grid = FALSE (the default) the function returns a list of integer vectors, one vector for each element of ids. Each integer represents a genotype, in the form of a row number of the matrix allGenotypes(n), where n is the number of alleles of the marker.

If make.grid = TRUE, the cartesian product of the vectors is taken, resulting in a matrix with one column for each element of ids.

haldane *Genetic map functions*

Description

Simple implementations of the classical map functions of Haldane and Kosambi, relating the genetic distance and the recombination rate between two linked loci.

Usage

```
haldane(cM = NULL, rho = NULL)

kosambi(cM = NULL, rho = NULL)
```

Arguments

cM	A numeric vector with genetic distances in centiMorgan, or NULL.
rho	A numeric vector with recombination rates, or NULL.

Value

A numeric of the same length as the input.

Examples

```
cM = 0:200
dat = cbind(Haldane = haldane(cM = cM),
            Kosambi = kosambi(cM = cM))
matplot(cM, dat, ylab = "Recombination rate", type = "l")
legend("topleft", legend = colnames(dat), col = 1:2, lty = 1:2)


rho = seq(0, 0.49, length = 50)
dat2 = cbind(Haldane = haldane(rho = rho),
             Kosambi = kosambi(rho = rho))
matplot(rho, dat2, xlab = "Recombination rate", ylab = "cM", type = "l")
legend("topleft", legend = colnames(dat), col = 1:2, lty = 1:2)
```

HWprob *Hardy-Weinberg probabilities*

Description

Hardy-Weinberg probabilities

Usage

```
HWprob(allele1, allele2, afreq, f = 0)
```

Arguments

allele1, allele2	Vectors of equal length, containing alleles in the form of indices of afreq
afreq	A numeric vector with allele frequencies
f	A single number in [0, 1]; the inbreeding coefficient

Value

A numeric vector of the same length as allele1 and allele2

Examples

```
p = 0.1; q = 1-p
hw = HWprob(c(1,1,2), c(1,2,2), c(p, q))
stopifnot(all.equal(hw, c(p^2, 2*p*q, q^2)))
```

`likelihood` *Pedigree likelihood*

Description

The `likelihood()` and `likelihood2()` functions constitute the heart of **pedprobr**. The former computes the pedigree likelihood for each indicated marker. The latter computes the likelihood for a pair of linked markers separated by a given recombination rate.

Usage

```
likelihood(x, ...)

## S3 method for class 'ped'
likelihood(
  x,
  markers = NULL,
  peelOrder = NULL,
  lump = TRUE,
  eliminate = 0,
  logbase = NULL,
  loopBreakers = NULL,
  verbose = FALSE,
  theta = 0,
  ...
)

## S3 method for class 'list'
likelihood(x, markers = NULL, logbase = NULL, ...)

likelihood2(x, ...)

## S3 method for class 'ped'
likelihood2(
  x,
  marker1,
  marker2,
  rho,
  peelOrder = NULL,
  eliminate = 0,
  logbase = NULL,
  loopBreakers = NULL,
  verbose = FALSE,
  ...
)

## S3 method for class 'list'
```

```
likelihood2(x, marker1, marker2, logbase = NULL, ...)
```

Arguments

x	A ped object, a singleton object, or a list of such objects.
...	Further arguments.
markers	One or several markers compatible with x. Several input forms are possible: • A marker() object compatible with x. • A list of marker objects. • A vector of names or indices of markers attached to x. If x is a list, this is the only valid input.
peelOrder	For internal use.
lump	Activate allele lumping, i.e., merging unobserved alleles. This is an important time saver, and should be applied in nearly all cases. (The parameter exists mainly for debugging purposes.) The lumping algorithm will detect (and complain) if any markers use a non-lumpable mutation model. Default: TRUE.
eliminate	Mostly for internal use: a non-negative integer indicating the number of iterations in the internal genotype-compatibility algorithm. Positive values can save time if the number of alleles is large.
logbase	Either NULL (default) or a positive number indicating the basis for logarithmic output. Typical values are exp(1) and 10.
loopBreakers	A vector of ID labels indicating loop breakers. If NULL (default), automatic selection of loop breakers will be performed. See breakLoops().
verbose	A logical.
theta	Theta correction.
marker1, marker2	Single markers compatible with x.
rho	The recombination rate between marker1 and marker2. To make biological sense rho should be between 0 and 0.5.

Details

The implementation is based on the peeling algorithm of Elston and Stewart (1971). A variety of situations are covered; see the Examples section for some demonstrations.

- autosomal and X-linked markers
- 1 marker or 2 linked markers
- complex inbred pedigrees
- markers with mutation models
- pedigrees with inbred founders

For more than two linked markers, see likelihoodMerlin().

Value

A numeric with the same length as the number of markers indicated by `markers`. If `logbase` is a positive number, the output is `log(likelihood, logbase)`.

Author(s)

Magnus Dehli Vigeland

References

Elston and Stewart (1971). *A General Model for the Genetic Analysis of Pedigree Data.* doi:10.1159/000152448

See Also

`likelihoodMerlin()`, for likelihoods involving more than 2 linked markers.

Examples

```
### Simple likelihood ###
p = 0.1
q = 1 - p

# Singleton
s = singleton() |> addMarker(geno = "1/2", afreq = c("1" = p, "2" = q))

stopifnot(all.equal(likelihood(s), 2*p*q))

# Trio
t = nuclearPed() |>
  addMarker(geno = c("1/1", "1/2", "1/1"), afreq = c("1" = p, "2" = q))

stopifnot(all.equal(likelihood(t), p^2 * 2*p*q * 0.5))


### Example of calculation with inbred founders ###

### Case 1: Trio with inbred father
x = cousinPed(0, child = TRUE)
x = addSon(x, 5)
x = relabel(x, old = 5:7, new = c("father", "mother", "child"))

# Add equifrequent SNP; father homozygous, child heterozygous
x = addMarker(x, father = "1/1", child = "1/2")

# Plot with genotypes
plot(x, marker = 1)

# Compute the likelihood
lik1 = likelihood(x, markers = 1)
```

```
### Case 2: Using founder inbreeding
# Remove ancestry of father
y = subset(x, c("father", "mother", "child"))

# Indicate that the father has inbreeding coefficient 1/4
founderInbreeding(y, "father") = 1/4

# Plot (notice the inbreeding coefficient)
plot(y, marker = 1)

# Likelihood should be the same as above
lik2 = likelihood(y, markers = 1)

stopifnot(all.equal(lik1, lik2))
```

`lumpAlleles` *Allele lumping*

Description

Perform allele lumping (i.e., merging unobserved alleles) for all markers attached to the input pedigree.

Usage

```
lumpAlleles(x, markers = NULL, always = FALSE, verbose = FALSE)
```

Arguments

`x`	A ped object or a list of such.
`markers`	A vector of names or indices referring to markers attached to `x`. (Default: All markers.)
`always`	A logical. If TRUE, lumping is always attempted. By default (FALSE) lumping is skipped for markers where all individuals are genotyped.
`verbose`	A logical.

Value

An object similar to `x`, but whose attached markers have reduced allele set.

Examples

```
x = nuclearPed() |> addMarker(geno = c("1/1", NA, NA), alleles = 1:4)

# Before lumping
afreq(x, 1)

# Lump
y = lumpAlleles(x, verbose = TRUE)
afreq(y, 1)
```

---

merlin *Pedigree likelihoods computed by MERLIN*

Description

These functions enable users to call MERLIN (Abecasis et al., 2002) from within R.

Usage

```
merlin(
  x,
  options,
  markers = NULL,
  linkageMap = NULL,
  verbose = TRUE,
  generateFiles = TRUE,
  cleanup = TRUE,
  dir = tempdir(),
  logfile = NULL,
  merlinpath = NULL,
  checkpath = TRUE
)

likelihoodMerlin(
  x,
  markers = NULL,
  linkageMap = NULL,
  rho = NULL,
  logbase = NULL,
  perChrom = FALSE,
  options = "--likelihood --bits:100 --megabytes:4000 --quiet",
  ...
)

checkMerlin(program = NULL, version = TRUE, error = FALSE)
```

Arguments

x	A ped object.
options	A single string containing all arguments to merlin except for the input file indications.
markers	A vector of names or indices of markers attached to x. (Default: all markers).
linkageMap	A data frame with three columns (chromosome; marker name; centiMorgan position) to be used as the marker map by MERLIN.
verbose	A logical.
generateFiles	A logical. If TRUE (default), input files to MERLIN named '_merlin.ped', '_merlin.dat', '_merlin.map', and '_merlin.freq' are created in the directory indicated by dir. If FALSE, no files are created.
cleanup	A logical. If TRUE (default), the MERLIN input files are deleted after the call to MERLIN.
dir	The name of the directory where input files should be written.
logfile	A character. If this is given, the MERLIN screen output will be dumped to a file with this name.
merlinpath	The path to the folder containing the merlin executables. If the executables are on the system's search path, this can be left as NULL (default).
checkpath	A logical indicating whether to check that the merlin executable is found.
rho	A vector of length one less than the number of markers, specifying the recombination rate between each consecutive pair.
logbase	Either NULL (default) or a positive number indicating the basis for logarithmic output. Typical values are exp(1) and 10.
perChrom	A logical; if TRUE, likelihoods are reported per chromosome.
...	Further arguments passed on to merlin().
program	A character containing "merlin", "minx" or both (default), optionally including full paths.
version	A logical. If TRUE (default), it is checked that running program produces a printout starting with "MERLIN 1.1.2".
error	A logical, indicating if an error should be raised if program is not found. Default: FALSE.

Details

For these functions to work, the program MERLIN must be installed (see link in the Reference section below) and correctly pointed to in the PATH variable. The merlin() function is a general wrapper which runs MERLIN with the indicated options, after creating the appropriate input files. For convenience, MERLIN's "–likelihood" functionality is wrapped in a separate function.

The merlin() function creates input files "_merlin.ped", "_merlin.dat", "_merlin.map" and "_merlin.freq" in the dir directory, and then runs the following command through a call to system():

```
merlin -p _merlin.ped -d _merlin.dat -m _merlin.map -f
_merlin.freq  <options>
```

likelihoodMerlin() first runs merlin() with options = "--likelihood --bits:100 --megabytes:4000 --quiet", and then extracts the likelihood values from the MERLIN output. Note that the output is the *total* likelihood including all markers.

For likelihood computations with linked markers, the argument rho should indicate the recombination fractions between each consecutive pair of markers (i.e., rho[i] is the recombination rate between markers i-1 and i). These will be converted to centiMorgan distances using Haldane's map function, and used to create genetic marker map in a MERLIN-friendly format.

Value

merlin() returns the screen output of MERLIN invisibly.

likelihoodMerlin() returns a single number; the total likelihood using all indicated markers.

checkMerlin() returns TRUE if the MERLIN executable indicated by program is found on the system. Otherwise FALSE, or (if error = TRUE) an error is raised.

Author(s)

Magnus Dehli Vigeland

References

Abecasis et al. (2002) Nat Gen 30:97-101. https://csg.sph.umich.edu/abecasis/merlin/.

Examples

```
if(checkMerlin()) {

### Trivial example for validation
x = nuclearPed(1) |>
  addMarker("1" = "1/2") |>            # likelihood = 1/2
  addMarker("1" = "1/1", "3" = "1/2")  # likelihood = 1/8

# MERLIN likelihoods
lik1 = likelihoodMerlin(x, markers = 1, verbose = FALSE)
lik2 = likelihoodMerlin(x, markers = 2, verbose = FALSE)
likTot = likelihoodMerlin(x, verbose = FALSE)
stopifnot(all.equal(
  round(c(lik1, lik2, likTot), c(3,3,4)), c(1/2, 1/8, 1/16)))

# Example with ped lists
y = list(singleton(1), singleton(2))
y = setMarkers(y, locus = list(alleles = 1:2))
genotype(y[[1]], marker = 1, id = "1") = "1/2"
genotype(y[[2]], marker = 1, id = "2") = "1/1"
lik = likelihoodMerlin(y, verbose = FALSE)
stopifnot(all.equal(round(lik, 3), 1/8))

### Linked markers
z = nuclearPed(2)
m = marker(z, geno = c("1/1", "1/2", "1/2", "1/2"))
```

```
z = setMarkers(z, list(m, m))

# By MERLIN...
L1 = likelihoodMerlin(z, markers = 1:2, rho = 0.25, verbose = FALSE)

# ...and by pedprobr
L2 = likelihood2(z, marker1 = 1, marker2 = 2, rho = 0.25)

# stopifnot(all.equal(signif(L1, 3), signif(L2, 3)))
}
```

`oneMarkerDistribution` *Genotype distribution for a single marker*

Description

Computes the genotype probability distribution of one or several pedigree members, possibly conditional on known genotypes for the marker.

Usage

```
oneMarkerDistribution(
  x,
  ids,
  partialmarker,
  loopBreakers = NULL,
  eliminate = 0,
  grid.subset = NULL,
  verbose = TRUE
)
```

Arguments

`x`	A ped object or a list of such.
`ids`	A numeric with ID labels of one or more pedigree members.
`partialmarker`	Either a `marker` object or the name (or index) of a marker attached to `x`. If `x` has multiple components, only the latter is allowed.
`loopBreakers`	(Only relevant if the pedigree has loops). A vector with ID labels of individuals to be used as loop breakers. If NULL (default) loop breakers are selected automatically. See `breakLoops()`.
`eliminate`	A non-negative integer, indicating the number of iterations in the internal genotype-compatibility algorithm. Positive values can save time if `partialmarker` has many alleles.

`grid.subset` (Optional; not relevant for most users.) A numeric matrix describing a subset of all marker genotype combinations for the `ids` individuals. The matrix should have one column for each of the `ids` individuals, and one row for each combination: The genotypes are described in terms of the matrix `M = allGenotypes(n)`, where `n` is the number of alleles for the marker. If the entry in column `j` is the integer `k`, this means that the genotype of individual `ids[j]` is row `k` of `M`.

`verbose` A logical.

Value

A named k-dimensional array, where `k = length(ids)`, with the joint genotype distribution for the `ids` individuals. The probabilities are conditional on the known genotypes and the allele frequencies of `partialmarker`.

Author(s)

Magnus Dehli Vigeland

See Also

`twoMarkerDistribution()`

Examples

```
# Trivial example giving Hardy-Weinberg probabilities
s = singleton(id = 1)
m = marker(s, alleles = 1:2) # equifrequent SNP
oneMarkerDistribution(s, ids = 1, partialmarker = m)

# Conditioning on a partial genotype
genotype(m, id = 1) = "1/-"
oneMarkerDistribution(s, ids = 1, partialmarker = m)

# Genotype distribution for a child of heterozygous parents
trio = nuclearPed(father = "fa", mother = "mo", child = "ch")
m1 = marker(trio, fa = "1/2", mo = "1/2")
oneMarkerDistribution(trio, ids = "ch", partialmarker = m1)

# Joint distribution of the parents, given that the child is heterozygous
m2 = marker(trio, ch = "1/2", afreq = c("1" = 0.5, "2" = 0.5))
oneMarkerDistribution(trio, ids = c("fa", "mo"), partialmarker = m2)

# A different example: The genotype distribution of an individual (id = 8)
# whose half cousin (id = 9) is homozygous for a rare allele.
y = halfCousinPed(degree = 1) |>
  addMarker("9" = "a/a", afreq = c(a = 0.01, b = 0.99))

oneMarkerDistribution(y, ids = 8, partialmarker = 1)
```

`setMutationModel` *Set a mutation model*

Description

This function offers a convenient way to attach mutation models to a pedigree with marker data. It wraps `pedmut::mutationModel()`, which does the main work of creating the models, but relieves the user from having to loop through the markers in order to supply the correct alleles and frequencies for each marker.

Usage

```
setMutationModel(x, model, markers = NULL, ...)
```

Arguments

`x`	A ped object or a list of such.
`model`	A model name implemented by `pedmut::mutationModel()` (see Details), or NULL.
`markers`	A vector of names or indices referring to markers attached to `x`. (Default: All markers.)
`...`	Arguments forwarded to `pedmut::mutationModel()`, e.g., `rate`.

Details

Currently, the following models are implemented in the `pedmut` package:

- `equal` : All mutations equally likely; probability $1 - rate$ of no mutation
- `proportional` : Mutation probabilities are proportional to the target allele frequencies
- `onestep`: A mutation model for microsatellite markers, allowing mutations only to the nearest neighbours in the allelic ladder. For example, '10' may mutate to either '9' or '11', unless '10' is the lowest allele, in which case '11' is the only option. This model is not applicable to loci with non-integral microvariants.
- `stepwise`: A common model in forensic genetics, allowing different mutation rates between integer alleles (like '16') and non-integer "microvariants" like '9.3'). Mutations also depend on the size of the mutation if the parameter 'range' differs from 1.
- `custom` : Allows any mutation matrix to be provided by the user, in the `matrix` parameter
- `random` : This produces a matrix of random numbers, where each row is normalised so that it sums to 1
- `trivial` : The identity matrix; i.e. no mutations are possible.

Value

An object similar to `x`.

Examples

```
### Example requires the pedmut package ###
if (requireNamespace("pedmut", quietly = TRUE)){

# A pedigree with data from a single marker
x = nuclearPed(1) |>
  addMarker(geno = c("a/a", NA, "b/b")) # mutation!

# Set `equal` model
y = setMutationModel(x, marker = 1, model = "equal", rate = 0.01)

# Inspect model
mutmod(y, 1)

# Likelihood
likelihood(y, 1)

# Remove mutation model
z = setMutationModel(y, model = NULL)
stopifnot(identical(z, x))
}
```

---

twoMarkerDistribution *Genotype distribution for two linked markers*

Description

Computes the joint genotype distribution of two markers for a specified pedigree member, conditional on known genotypes and the recombination rate between the markers.

Usage

```
twoMarkerDistribution(
  x,
  id,
  partialmarker1,
  partialmarker2,
  rho,
  loopBreakers = NULL,
  eliminate = 99,
  verbose = TRUE
)
```

Arguments

x	A ped object or a list of such.
id	A single ID label.

partialmarker1, partialmarker2
: Either a marker object, or the name (or index) of a marker attached to x.

rho
: A single numeric in the interval [0, 0.5]: the recombination fraction between the two markers.

loopBreakers
: (Only relevant if the pedigree has loops). A vector with ID labels of individuals to be used as loop breakers. If NULL (default) loop breakers are selected automatically. See breakLoops().

eliminate
: A non-negative integer, indicating the number of iterations in the internal algorithm for reducing the genotype space. Positive values can save time if partialmarker1 and/or partialmarker2 have many alleles.

verbose
: A logical.

Value

A named matrix giving the joint genotype distribution.

Author(s)

Magnus Dehli Vigeland

See Also

oneMarkerDistribution()

Examples

```
# A sib-pair pedigree
x = nuclearPed(children = c("bro1", "bro2"))

# Two SNP markers; first brother homozygous for the `1` allele
SNP1 = SNP2 = marker(x, bro1 = "1/1", afreq = c("1" = 0.5, "2" = 0.5))

plot(x, marker = list(SNP1, SNP2))

# Genotype distribution for the brother depends on linkage
twoMarkerDistribution(x, id = "bro2", SNP1, SNP2, rho = 0)
twoMarkerDistribution(x, id = "bro2", SNP1, SNP2, rho = 0.5)

# X-linked
chrom(SNP1) = chrom(SNP2) = "X"

plot(x, marker = list(SNP1, SNP2))

twoMarkerDistribution(x, id = "bro2", SNP1, SNP2, rho = 0)
twoMarkerDistribution(x, id = "bro2", SNP1, SNP2, rho = 0.5)
```

Index