Package ‘oncomodel’

February 20, 2015

Type Package

Title Maximum likelihood tree models for oncogenesis

Version 1.0

Date 2008-01-25

Author Anja von Heydebreck <anja.von.heydebreck@merck.de>,
contribution from Christiane Heiss <christiane.heiss@web.de>

Maintainer Anja von Heydebreck <anja.von.heydebreck@merck.de>

Description Computing probabilistic tree models for oncogenesis based
on genetic data using maximum likelihood.

Depends R (>= 2.5.1), ade4

License GPL (>= 2)

Repository CRAN

Date/Publication 2012-10-29 08:59:19

NeedsCompilation no

R topics documented:

oncomodel-package ........................................... 2
boot.conf.values ............................................. 3
comp.freq .................................................. 4
is.tree ...................................................... 4
kidney ......................................................... 5
leafset.prob .................................................. 6
leafset.prob2 ................................................ 6
MLparameters ................................................. 7
MLtopology ................................................... 8
mrca ........................................................ 9
subtree ....................................................... 9

Index 11
oncomodel-package

Maximum likelihood tree models for oncogenesis

Description

Computing probabilistic tree models for oncogenesis based on genetic data using maximum likelihood.

Details

Package: oncomodel
Type: Package
Version: 1.0
Date: 2008-01-18
License: GPL version 2 or newer

Author(s)

original by Anja von Heydebreck <anja.von.heydebreck@merck.de>, contributions from Christiane Heiss <christiane.heiss@web.de> Maintainer: Anja von Heydebreck <anja.von.heydebreck@merck.de>

References


Examples

```r
## NOT RUN
## The example needs longer run time.
data(kidney)
## Maximum likelihood tree model
#y <- MLtopology(kidney$x)

## Graphical presentation
#y.phyl <- newick2phylog(y$newick)
#plot.phyl(y.phyl, cnodes =1, clabel.n=0.6, f=0.75, sub="Oncogenic tree of given aberrations")

## Bootstrap confidence values (in percent) and the splits occurring in
## more than 10 percent of the bootstrap data sets
#boot.conf.values(kidney$x, nrep=2)

## Probability for aberration -3|-3p
#leafset.prob(c("-3|-3p", "+5|+5q"), kidney$res)
```
## Probability for aberration -3|-3p
leafset.prob2(c("-3|-3p", "+5|+5q"), kidney$res)
## END(NOT RUN)

### boot.conf.values

**Bootstrap Confidence Values**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compute the bootstrap confidence values (in percent) for the inner edges and display the splits occurring in &gt; 10 percent of the bootstrap data sets (the splits are characterized by one of the two subsets of leaves).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>boot.conf.values(data, random.seed = 12345, nrep = 500)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arguments</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
</tr>
<tr>
<td>random.seed</td>
</tr>
<tr>
<td>nrep</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>We use the nonparametric bootstrap (Felsenstein, 1985) to assess the uncertainty of properties of the estimated tree model. The proposed tree structure has to be interpreted with caution. Nevertheless we think that the model can at least serve an exploratory purpose, allowing us to formulate hypotheses about the evolution of karyotypes in the data set.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequencies</td>
</tr>
<tr>
<td>confidence values</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>
Examples

```r
## NOT RUN
## The calculation of bootstrap confidence values needs longer run time.
#data(kidney)
#boot.conf.values(kidney$x, nrep=2)
## END(NOT RUN)
```

---

**comp.freq**  
*Compare Model Probabilities to Frequencies*

**Description**

Compares the model probabilities of single alterations and pairs to the observed frequencies and shows scatterplots for the comparisons.

**Usage**

```r
comp.freq(x, tree, p)
```

**Arguments**

- `x`  
a binary data matrix with rows representing tumors and columns representing genetic alterations.
- `tree`  
the tree in matrix format.
- `p`  
a vector of edge parameters (model probabilities).

**Examples**

```r
data(kidney)
comp.freq(kidney$x, kidney$res$tree, kidney$res$p)
```

---

**is.tree**  
*Compare with Tree Format*

**Description**

Tests whether a 2 x n matrix represents a rooted tree in the format accepted by the oncomodel package.

**Usage**

```r
is.tree(tree)
```

**Arguments**

- `tree`  
a 2 x n matrix.
Details

In the accepted format, the columns of the integer matrix represent the edges of the tree, with the entry in the first row being closer to the root. The leaves have to be the smallest integers of the matrix. If an edge has a smaller column index than a second one, it may not be on the path from the second edge to the root (the order of the columns has to be compatible with the partial order of the edges of the tree).

Examples

```r
data(kidney)
is.tree(kidney$res$tree)
```

Description

This data set contains cytogenetic data from 173 cases of renal clear cell carcinoma, covering 7 frequent chromosomal aberrations, as well as the corresponding maximum likelihood tree model.

Usage

```r
data(kidney)
```

Format

A list with the following components:

- x  
  the binary data matrix

- res  
  the ML tree model, a list with components:
   - treethe tree in matrix format
   - pthe conditional probabilities associated with the edges
   - var.namescharacter vector of variable names
   - totloglikthe log-likelihood of the tree model.
   - newickthe tree model in Newick format
leafset.prob

*Probability of a Set of Leaves*

**Description**
Computes the probability of a set of leaves (chromosomal aberrations) in an oncogene tree model.

**Usage**
```
leafset.prob(leafset, y)
```

**Arguments**
- **leafset**: a character vector of one or more leaves (chromosomal aberrations).
- **y**: the tree model as obtained from MLtopology.

**Value**
The probability of exactly a given set of leaves (chromosomal aberrations) in the tree.

**References**

**Examples**
```
data(kidney)
leafset.prob(c("-3|-3p", "+5|+5q"), kidney$res)
```

leafset.prob2

*Probability of a Set of Leaves*

**Description**
Computes the probability of a set of leaves (chromosomal aberrations) in an oncogene tree model.

**Usage**
```
leafset.prob2(leafset, y)
```

**Arguments**
- **leafset**: a character vector of one or more leaves (chromosomal aberrations).
- **y**: the tree model as obtained from MLtopology.
Value

The probability of exactly the given set of leaves (chromosomal aberrations) of the tree.

References


Examples

data(kidney)
leafset.prob2(c("-3|-3p", "+5|+5q"), kidney$res)

MLparameters

Compute Maximum Likelihood Parameters

Description

Computes the maximum likelihood parameters for a given tree topology.

Usage

MLparameters(x, tree, freq = NULL)

Arguments

x          a binary matrix whose rows are the (preferably unique) genetic profiles.
tree       the tree in matrix format.
freq       a vector whose length equals the number of rows of x, giving the frequency of each profile in the data.

Value

p          a vector of the maximum likelihood edge parameters (model probabilities).
totloglik  the log-likelihood at the ML parameters.

Examples

data(kidney)
MLparameters(kidney$x, kidney$res$tree, freq = NULL)
MLtopology

Compute Maximum Likelihood Tree Topology

Description

Tries to compute the maximum likelihood tree model for a given data set through stepwise leaf insertion and rearrangements.

Usage

MLtopology(x, verbose = FALSE)

Arguments

x

a binary matrix with rows representing tumors and columns representing genetic alterations.

verbose

a Boolean value indicating whether intermediate results of the algorithm are to be printed.

Value

A list with the following components:

tree

the resulting tree in matrix format.

p

a vector of the maximum likelihood edge parameters (model probabilities).

totloglik

the log-likelihood of the tree model.

var.names

the character vector with the names of alterations.

newick

the tree model in Newick format.

References


Examples

```r
## NOT RUN
## The computation of the maximum likelihood tree model needs longer run time.
data(kidney)
y <- MLtopology(kidney$x)
## END(NOT RUN)
```
**mrca**

*Most Recent Common Ancestor*

**Description**

Computes the most recent common ancestor node for a pair of nodes of a tree.

**Usage**

```r
mrca(x, y, tree)
```

**Arguments**

- `x` a node of the tree.
- `y` a node of the tree.
- `tree` a tree in matrix format.

**Value**

the most recent common ancestor node of `x` and `y`.

**Examples**

```r
data(kidney)
mrca(1,2, kidney$res$tree)
```

**subtree**

*Subtree*

**Description**

Computes the subtree rooted at a given node.

**Usage**

```r
subtree(node, tree)
```

**Arguments**

- `tree` a tree in matrix format.
- `node` a node of the tree.

**Value**

The edge indices of the subtree rooted at the given node.
Examples

```r
data(kidney)
subtree(14, kidney$res$tree)
```
Index

*Topic attribute
  is.tree, 4
  leafset.prob, 6
  leafset.prob2, 6
  mrca, 9
*Topic datasets
  kidney, 5
*Topic distribution
  leafset.prob, 6
  leafset.prob2, 6
*Topic hplot
  comp.freq, 4
*Topic manip
  subtree, 9
*Topic models
  MLparameters, 7
  MLtopology, 8
*Topic nonparametric
  boot.conf.values, 3
*Topic package
  oncomodel-package, 2
*Topic tree
  MLtopology, 8
boot.conf.values, 3
comp.freq, 4
is.tree, 4
kidney, 5
leafset.prob, 6
leafset.prob2, 6
MLparameters, 7
MLtopology, 8
mrca, 9
oncomodel (oncomodel-package), 2
oncomodel-package, 2
subtree, 9