

Package ‘bayestools’

March 30, 2018

Title Tools for Pre- and Post-processing in Phylogenetic Bayesian Inference.

Version 0.0.0.9000

Description Package for bayesian inference in phylogenetics and evolution.

It provides functions for prior specification in divergence time estimation using fossils as well as other kinds of data. The package provides tools for interacting with the input and output of bayesian platforms in evolutionary biology such as BEAST2. The package implements a way to measure interdependence between probability density functions in the context of comparisons between prior and posterior bayesian densities.

Depends R (>= 3.4.2)

License GNU General Public License v3.0

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

BugReports <https://github.com/gaballench/bayestools/issues>

Suggests knitr,
rmarkdown

VignetteBuilder knitr

R topics documented:

findParams	2
laventa	3
lognormalBeast	4
measureSensit	5
mswd.test	6

Index	8
--------------	----------

findParams	<i>Function for estimation of probability density function parameters through quadratic optimization</i>
------------	--

Description

Function for estimation of probability density function parameters through quadratic optimization

Usage

```
findParams(q, p, output = "complete", pdffunction, params)
```

Arguments

q	A numeric vector of observed quantiles, might come from a HPD from a previous study (along with a median), or from other sources of prior information. See Notes .
p	A numeric vector of quantile probabilities, 0.05, 0.50 and 0.95 by default
output	One of two possible values: "complete" and "parameters". For the latter the complete output of the <code>optim</code> function is returned with information on convergence and squared errors (that might be useless for simple cases) or just the parameters.
pdffunction	A character vector (of length one) with the name of the PDF function of interest. Technically this argument supports any PDF function of the form <code>pDIST</code> (e.g., <code>pnorm</code> , <code>ppois</code> , <code>pexp</code>).
params	A character vector with the name of the parameter(s) to optimize in the probability density function. These should match the parameter names of the respective PDF function, e.g., "lambda" in the function <code>ppois</code>

Value

Either a list with the complete output of convergence, squared errors and parameter values, or just a vector of parameter values. Depends on the value of output. Warnings may be triggered by the function `optim` since the optimization is a heuristic process, whenever a given iteration results in an invalid value for a given combination of parameters, the `optim` function tries another combination of values but inform the user about the problem through a warning. In general these can be safely disregarded.

Note

Notes on distributions

Author(s)

Main code by Gustavo A. Ballen with important contributions in expression call structure and vectorized design by Klaus Schliep (<Klaus.Schliep@umb.edu>).

Examples

```
# Find the best parameters for a standard normal density that fit the observed quantiles -1.644854, 0, and 1.644854
findParams(q = c(-1.644854, 0, 1.644854), p = c(0.05, 0.50, 0.95), output = "complete", pdffunction = "pnorm", par

# Given that we have prior on the age of a fossil to be 1 - 10 Ma and that we want to model it with a lognormal dist
findParams(q = c(1, 5.5, 10), p = c(0.05, 0.50, 0.95), output = "complete", pdffunction = "plnorm", params = c("me
```

laventa

Geochronology samples from the Honda Group in Colombia

Description

A dataset containing geochronology data from several samples along the stratigraphic column of the Honda and Huila groups in the Tatacoa Desert area. The dataset was compiled from the Table 3.2 in Flynn et al. (1997).

Usage

```
data(laventa)
```

Format

A data frame with 87 rows and 7 variables:

age Estimated age (in Ma) from a given rock sample

one_sigma Standard deviation of the age estimate

sample Sample code as in Table 3.2

unit Stratigraphic unit in either the Honda Group or the Huila Group

elevation Position in the stratigraphic column, in meters

mineral The mineral used for dating the sample

comments Comments from footnotes in the original table ...

References

Flynn, J.J., Guerrero, J. & Swisher III, C.C. (1997) Geochronology of the Honda Group. In: R. F. Kay, R. H. Madden, R. L. Cifelli, and J. J. Flynn (Eds), Vertebrate Paleontology in the Neotropics: the Miocene Fauna of La Venta, Colombia. Smithsonian Institution Press, pp. 44–60.

lognormalBeast	<i>Constructing a curve for the user-specified lognormal prior using Beast2 parameters</i>
----------------	--

Description

Constructing a curve for the user-specified lognormal prior using Beast2 parameters

Usage

```
lognormalBeast(M, S, meanInRealSpace = TRUE, offset = 0, from, to,
  by = 0.05, ...)
```

Arguments

M	Mean of the lognormal density in Beast2.
S	Standard deviation of the lognormal density in Beast2.
meanInRealSpace	Whether to plot the mean on the real- or log-space (i.e., apply log(M) before plotting). Please see under details.
offset	Hard lower bound.
from, to	Starting and ending point to calculate considering the offset as zero. That is, from will affect produce a starting point of (offset + from) and an ending point of (offset + to)

Details

This function creates a matrix of x-y values given parameters of a lognormal density as specified in the program Beast2. It's main purpose is for plotting but other uses such as sensitivity quantification are available. Please note that the value of mean depends on whether we expect it to be in real or log space. Please refer to Heath (2015) for more info: [Heath, T. A. \(2015\). Divergence Time Estimation using BEAST v2.](#)

Value

A matrix of two columns consisting of the x and y values of the lognormal density.

Examples

```
# Generate a matrix for the lognormal density with mean 1 and standard deviation 1, with mean in real space, and
lognormalBeast(M = 1, S = 1, meanInRealSpace = TRUE, from = 0, to = 10)
# The same as above but with an offset of 10, that is, the curve starts at 10 as if it was 0 to values will start in
lognormalBeast(M = 1, S = 1, meanInRealSpace = TRUE, offset = 10, from = 0, to = 10)
```

measureSensit

Calculate the Intersection Between Two Densities

Description

Calculate the Intersection Between Two Densities

Usage

```
measureSensit(d1, d2, splits = 500, rawData = c(TRUE, TRUE), plot = TRUE,
  x_limit = "auto", colors = c("red", "blue", "gray"), ...)
```

Arguments

d1, d2	Either two vectors of empirical (i.e., MCMC-produced) values OR a data frame/matrix with columns x and y for values fitted to a density from which to calculate areas. If rawData is set to TRUE in any instance, the data must be place in vectors and not multidimensional objects.
splits	A numerical argument controlling the number of subdivisions of the intersection area for numerical integration
rawData	Are d1 and/or d2 raw data for which a density should be calculated? a vector of length two containing logical values indicating whether any of the arguments d1 or d2 are raw data or whether the user is inputting already calculated densities (e.g., the output from the density, curve, or dDIST functions, or any two-dimension object with x and y values)
plot	Should a plot be produced?
colors	A vector of three colors, namely, color of the d1 density (e.g., the prior), color of the d2 density e.g., the posterior), and color of the intersection.
...	Further arguments to pass to the graphical functions such as lines and plot internally, such as main, xlim, ylim, xlab, ylab, etc.

Details

Sensitivity is measured as the overlapping portion between two densities following Ballen (in prep). It has a value between 0 and 1. The values of the vector rawData determine the behavior of the function and therefore attention must be paid to their consistence with the nature of arguments d1 and d2. Despite the function was designed in order to allow to quantify posterior sensitivity to the prior, this can be used to quantify any overlapping between two given densities and for any other purpose.

Value

A numeric vector with the value of the intersection between two densities. As a side effect, a plot is produced to an active (or new) graphical device.

Examples

```
## Not run:
# Set seed and colors to use in plots in the order: Prior, posterior, and intersection
set.seed(1985)
colors <- c("red", "blue", "lightgray")
# Sensitivity in two identical distributions
below <- measureSensit(d1 = rnorm(1000000, mean = 0, 1), d2 = rnorm(1000000, mean = 0, 1), main = "Comp. dependence",
  legend(x = "topright", legend = round(below, digits = 2)))
# Sensitivity in two distributions partially overlapping
below <- measureSensit(d1 = rnorm(1000000, mean = 3, 1), d2 = rnorm(1000000, mean = 0, 1), main = "Partial dependence",
  legend(x = "topright", legend = round(below, digits = 2)))
# Sensitivity in two completely-different distributions
below <- measureSensit(d1 = rnorm(1000000, mean = 8, 1), d2 = rnorm(1000000, mean = 0, 1), main = "Complete independence",
  legend(x = "topright", legend = round(below, digits = 2)))
# Don't plot, just return the intersection
measureSensit(d1 = rnorm(1000000, mean = 3, 1), d2 = rnorm(1000000, mean = 0, 1), plot = FALSE)

## End(Not run)
```

mswd.test	<i>Reduced chi-square test or mean square weighted deviation (mswd) test</i>
-----------	--

Description

Reduced chi-square test or mean square weighted deviation (mswd) test

Usage

```
mswd.test(age, sd)
```

Arguments

age	A vector of age radiometric age estimates
sd	A vector of the standard deviation corresponding to each element in age

Details

From Ludwig (2003:646): "By convention, probabilities of fit greater than 0.05 are generally considered as arguably satisfying the mathematical assumptions of an isochron, while lower probabilities are generally taken as indicating the presence of "geological" scatter, and hence a significant possibility of bias in the isochron age." The term $mswd \cdot d.f.$ is used because the term $1/d.f.$ was involved in the calculation of $mswd$, and we need the chi-sq. The null hypothesis is that the isochron conditions hold

Value

A numeric vector of length one with the p-value corresponding to the test.

Examples

```
data(laventa)
# Do the age estimates for the Neiva Fm. conform to the isochron hypothesis?
oneLevelIndex <- which(laventa$unit == "Neiva Fm.")
mswd.test(age = laventa$age[oneLevelIndex], sd = laventa$one_sigma[oneLevelIndex])
# The p-value is smaller than the nominal alpha of 0.05, so we can reject the null hypothesis of isochron condition

# Do the age estimates for the samples JG-R 88-2 and JG-R 89-2 conform to the isochron hypothesis?
twoLevelsIndex <- which(laventa$sample == "JG-R 89-2" | laventa$sample == "JG-R 88-2")
dataset <- laventa[twoLevelsIndex, ]
remove the values 21 and 23 because of their abnormally large standard deviations
mswd.test(age = dataset$age[c(-21, -23)], sd = dataset$one_sigma[c(-21, -23)])
# The p-value is larger than the nominal alpha of 0.05, so we can not reject the null hypothesis of isochron condition
```

Index

*Topic **datasets**

lamenta, [3](#)

findParams, [2](#)

lamenta, [3](#)

lognormalBeast, [4](#)

measureSensit, [5](#)

mswd.test, [6](#)

Notes, [2](#)