

Package ‘lorad’

December 17, 2023

Title Lowest Radial Distance Method of Marginal Likelihood Estimation

Version 0.0.1.0

Description Estimates marginal likelihood from a posterior sample using the method described in Wang et al. (2023) <[doi:10.1093/sysbio/syad007](https://doi.org/10.1093/sysbio/syad007)>, which does not require evaluation of any additional points and requires only the log of the unnormalized posterior density for each sampled parameter vector.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Suggests knitr, rmarkdown, testthat (>= 3.0), rstan (>= 2.32),
bridgesampling (>= 1.1)

Imports stats

Maintainer Analisa Milkey <analisa.milkey@uconn.edu>

Author Analisa Milkey [aut, cre] (<<https://orcid.org/0000-0002-1157-4363>>),
Elena Korte [aut],
Paul O. Lewis [aut] (<<https://orcid.org/0000-0001-9852-8759>>)

Config/testthat/edition 3

VignetteBuilder knitr, rmarkdown

Depends R (>= 3.0)

LazyData true

NeedsCompilation no

Repository CRAN

Date/Publication 2023-12-17 08:00:05 UTC

R topics documented:

gtrigsamples	2
k80samples	3
lorad_calc_log_sum	4
lorad_estimate	4

lorad_standardize	5
lorad_standardize_estimation_sample	6
lorad_summary	7
lorad_transform	7

Index	9
--------------	----------

gtrigsamples	<i>Sequence data used in gtrig vignette</i>
--------------	---

Description

Sequence data used in gtrig vignette

Usage

gtrigsamples

Format

gtrigsamples:

A data frame with 10,001 rows and 35 columns:

Iteration MCMC iteration

Posterior Log of the unnormalized posterior density

Likelihood Log likelihood

Prior Log of the prior density

alpha Shape parameter of the (mean=1) Gamma distribution of among-site rate heterogeneity

edge_length_proportions.1. Proportion of total tree length used by edge 1

edge_length_proportions.2. Proportion of total tree length used by edge 2

edge_length_proportions.3. Proportion of total tree length used by edge 3

edge_length_proportions.4. Proportion of total tree length used by edge 4

edge_length_proportions.5. Proportion of total tree length used by edge 5

edge_length_proportions.6. Proportion of total tree length used by edge 6

edge_length_proportions.7. Proportion of total tree length used by edge 7

edgelens.1. Edge length 1

edgelens.2. Edge length 2

edgelens.3. Edge length 3

edgelens.4. Edge length 4

edgelens.5. Edge length 5

edgelens.6. Edge length 6

edgelens.7. Edge length 7

er.1. Exchangeability parameter for A to C

er.2. Exchangeability parameter for A to G

er.3. Exchangeability parameter for A to T

er.4. Exchangeability parameter for C to G
er.5. Exchangeability parameter for C to T
er.6. Exchangeability parameter for G to T
pi.1. Nucleotide relative frequency for A
pi.2. Nucleotide relative frequency for C
pi.3. Nucleotide relative frequency for G
pi.4. Nucleotide relative frequency for t
pinvar Proportion of invariable sites
site_rates.1. Rate for site category 1
site_rates.2. Rate for site category 1
site_rates.3. Rate for site category 1
site_rates.4. Rate for site category 1
tree_length Tree length (sum of all edge lengths) in substitutions per site

Source

The program RevBayes (version 1.2.1) was used to obtain a sample from the Bayesian posterior distribution for 5 green plant rbcL sequences under a GTR+I+G model.

k80samples

Sequence data used in k80 vignette

Description

Sequence data used in k80 vignette

Usage

k80samples

Format

k80samples:
 A data frame with 10,000 rows and 4 columns:
iter Iteration
log.kernel Log unnormalized posterior
edgelen Edge length in substitutions per site
kappa Transition transversion rate ratio

Source

doi: [10.1093/sysbio/syad007](https://doi.org/10.1093/sysbio/syad007)

lorad_calc_log_sum *Calculate a sum on log scale*

Description

Calculates the (natural) log of a sum without leaving the log scale by factoring out the largest element.

Usage

```
lorad_calc_log_sum(logx)
```

Arguments

logx Numeric vector in which elements are on log scale

Value

The log of the sum of the (exponentiated) elements supplied in logx

lorad_estimate *Calculates the LoRaD estimate of the marginal likelihood*

Description

Provided with a data frame containing sampled parameter vectors and a dictionary relating column names to parameter types, returns a named character vector containing the following quantities:

- logML (the estimated log marginal likelihood)
- nsamples (number of samples)
- nparams (length of each parameter vector)
- training_frac (fraction of samples used for training)
- tsamples (number of samples used for training)
- esamples (number of sampled used for etimation)
- coverage (nominal fraction of the estimation sampled used)
- esamplesused (number of estimation samples actually used for estimation)
- realized_coverage (actual fraction of estimation sample used)
- rmax (lowest radial distance: defines boundary of working parameter space)
- log_delta (volume under the unnormalized posterior inside working parameter space)

Usage

```
lorad_estimate(params, colspec, training_frac, training_mode, coverage)
```

Arguments

params	Data frame in which rows are sample points and columns are parameters, except that last column holds the log posterior kernel
colspec	Named character vector associating column names in params with column specifications
training_frac	Number between 0 and 1 specifying the training fraction
training_mode	One of random, left, or right, specifying how training fraction is chosen
coverage	Number between 0 and 1 specifying fraction of training sample used to compute working parameter space

Value

Named character vector of length 11.

Examples

```
normals <- rnorm(1000000,0,10)
prob_normals <- dnorm(normals,0,10,log=TRUE)
proportions <- rbeta(1000000,1,2)
prob_proportions <- dbeta(proportions,1,2,log=TRUE)
lengths <- rgamma(1000000, 10, 1)
prob_lengths <- dgamma(lengths,10,1,log=TRUE)
paramsdf <- data.frame(
  normals,prob_normals,
  proportions, prob_proportions,
  lengths, prob_lengths)
columnkey <- c(
  "normals"="unconstrained",
  "prob_normals"="posterior",
  "proportions"="proportion",
  "prob_proportions"="posterior",
  "lengths"="positive",
  "prob_lengths"="posterior")
results <- lorad_estimate(paramsdf, columnkey, 0.5, 'random', 0.1)
lorad_summary(results)
```

lorad_standardize	<i>Transforms unconstrained parameters to have the same location and scale</i>
-------------------	--

Description

Standardizes parameters that have already been transformed (if necessary) to have unconstrained support. Standardization involves subtracting the sample mean and dividing by the sample standard deviation. Assumes that the log posterior kernel (i.e. the log of the unnormalized posterior) is the last column in the supplied data frame.

Usage

```
lorad_standardize(df, coverage)
```

Arguments

df	Data frame containing a column for each model parameter sampled and a final column of log posterior kernel values
coverage	Fraction of the training sample used to compute working parameter space

Value

List containing the log-Jacobian of the standardization transformation, the inverse square root matrix, a vector of column means, and rmax (radial distance to furthest point in working parameter space)

```
lorad_standardize_estimation_sample
```

Transforms training sample using training sample means and standard deviations

Description

Transforms training sample using training sample means and standard deviations

Usage

```
lorad_standardize_estimation_sample(standardinfo, y)
```

Arguments

standardinfo	List containing the log Jacobian of the standardization transformation, the inverse square root matrix, the column means, and rmax (the radial distance representing the edge of the working parameter space)
y	Data frame containing a column for each transformed model parameter in the estimation sample, with last column being the log kernel values

Value

A new data frame consisting of the standardized estimation sample with log kernel in last column

lorad_summary	<i>Summarize output from <code>lorad_estimate()</code></i>
---------------	--

Description

Summarize output from `lorad_estimate()`

Usage

```
lorad_summary(results)
```

Arguments

`results` Named character vector returned from `lorad_estimate()`

Value

String containing a summary of the supplied results object

Examples

```
normals <- rnorm(1000000,0,10)
prob_normals <- dnorm(normals,0,10,log=TRUE)
paramsdf <- data.frame(normals,prob_normals)
columnkey <- c("normals"="unconstrained", "prob_normals"="posterior")
results <- lorad_estimate(paramsdf, columnkey, 0.5, 'left', 0.1)
lorad_summary(results)
```

lorad_transform	<i>Log (or log-ratio) transform parameters having constrained support</i>
-----------------	---

Description

Log-transforms parameters with support (0,infinity), log-ratio transforms K-dimensional parameters with support a (K-1)-simplex, logit transforms parameters with support [0,1], and leaves unchanged parameters with unconstrained support (-infinity, infinity).

Usage

```
lorad_transform(params, colspec)
```

Arguments

`params` Data frame containing a column for each model parameter sampled as well as one or more columns that, when summed, constitute the log joint posterior kernel

`colspec` Named character vector matching each column name in `params` with a column specification

Value

A new data frame comprising transformed parameter values with a final column holding the log joint posterior kernel

Index

* datasets

gtrigsamples, 2

k80samples, 3

gtrigsamples, 2

k80samples, 3

lorad_calc_log_sum, 4

lorad_estimate, 4

lorad_estimate(), 7

lorad_standardize, 5

lorad_standardize_estimation_sample, 6

lorad_summary, 7

lorad_transform, 7