Package 'gsaot'

January 10, 2025

Title Compute Global Sensitivity Analysis Indices Using Optimal Transport

Version 0.1.0

Description Computing Global Sensitivity Indices from given data using Optimal Transport, as defined in Borgonovo et al (2024) <doi:10.1287/mnsc.2023.01796>. You provide an input sample, an output sample, decide the algorithm, and compute the indices.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.2

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

Imports boot, ggplot2, patchwork (>= 1.2.0), Rcpp, RcppEigen (>= 0.3.4.0.0), Rdpack (>= 2.4), stats, transport (>= 0.15.0)

URL https://github.com/pietrocipolla/gsaot,

https://pietrocipolla.github.io/gsaot/

BugReports https://github.com/pietrocipolla/gsaot/issues

LinkingTo Rcpp, RcppEigen

VignetteBuilder knitr

RdMacros Rdpack

NeedsCompilation yes

Author Leonardo Chiani [aut, cre, cph] (<https://orcid.org/0009-0007-2491-6290>),

> Emanuele Borgonovo [rev], Elmar Plischke [rev], Massimo Tavoni [rev]

Maintainer Leonardo Chiani <leonardo.chiani@polimi.it>

Repository CRAN

Date/Publication 2025-01-10 14:00:06 UTC

Contents

ower_bound
ot_indices
ot_indices_1d
ot_indices_smap
ot_indices_wb
blot.gsaot_indices
plot_inner_stats
print.gsaot_indices
16

Index

lower_bound Calculate lower bounds for Optimal Transport sensitivity indices

Description

Calculate lower bounds for Optimal Transport sensitivity indices

Usage

```
lower_bound(
   y,
   M,
   bound = "dummy",
   dummy_optns = NULL,
   cost = "L2",
   discrete_out = FALSE,
   solver = "sinkhorn",
   solver_optns = NULL,
   scaling = TRUE
)
```

Arguments

У	An array or a matrix containing the output values.
М	A scalar representing the number of partitions for continuous inputs.
bound	(default "dummy") A string defining the type of lower bound to compute. Should be "dummy" or "entropic". See details for more information.
dummy_optns	(default NULL) A list containing the options on the distribution of the dummy variable. See details for more information.
cost	(default "L2") A string or function defining the cost function of the Optimal Transport problem. It should be "L2" or a function taking as input y and returning a cost matrix. If cost="L2", ot_indices uses the squared Euclidean metric.

lower_bound

discrete_out	(default FALSE) Logical, by default the output sample in y are equally weighted.
	If discrete_out=TRUE, the function tries to create an histogram of the realiza-
	tions and to use the histogram as weights. It works if the output is discrete or
	mixed and the number of realizations is large. The advantage of this option is to
	reduce the dimension of the cost matrix.
solver	Solver for the Optimal Transport problem. Currently supported options are:
	• "sinkhorn" (default), the Sinkhorn's solver (Cuturi 2013).
	 "sinkhorn_log", the Sinkhorn's solver in log scale (Peyré et al. 2019).
	• "transport", a solver of the non regularized OT problem using transport::transport().
solver_optns	(optional) A list containing the options for the Optimal Transport solver. See
	details for allowed options and default ones.
scaling	(default TRUE) Logical that sets whether or not to scale the cost matrix.

Details

The function allows the computation of two different lower bounds. With bound="dummy", the function samples from a distribution defined in dummy_optns (by default a standard normal), independent from the output y and then computes the indices using the algorithm specified in solver. Under the hood, lower_bound calls the other available functions in the package:

- ot_indices_1d() (for solver="1d)
- ot_indices_wb() (for solver="wasserstein-bures")
- ot_indices() (for solver %in% c("sinkhorn", "sinkhorn_log", "wasserstein")) The user can choose the distribution of the dummy variable using the argument dummy_optns. dummy_optns should be a named list with at least a term called "distr" defining the sampling function. The other terms in the list are used as arguments to the sampling function. With bound="entropic", the function computes the lower bound of the entropic indices. In this case, solver should be either "sinkhorn" or "sinkhorn_log"

Value

An object of class gsaot_indices for bound="dummy" or a scalar for bound="entropic".

Examples

```
N <- 1000
mx <- c(1, 1, 1)
Sigmax <- matrix(data = c(1, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 1), nrow = 3)
x1 <- rnorm(N)
x2 <- rnorm(N)
x3 <- rnorm(N)
x <- cbind(x1, x2, x3)
x <- mx + x %*% chol(Sigmax)
A <- matrix(data = c(4, -2, 1, 2, 5, -1), nrow = 2, byrow = TRUE)
y <- t(A %*% t(x))</pre>
```

```
ot_indices
```

Calculate Optimal Transport sensitivity indices for multivariate y

Description

ot_indices calculates sensitivity indices using Optimal Transport (OT) for a multivariate output sample y with respect to input data x. Sensitivity indices measure the influence of inputs on outputs, with values ranging between 0 and 1.

Usage

```
ot_indices(
 х,
 у,
 Μ,
  cost = "L2",
  discrete_out = FALSE,
  solver = "sinkhorn",
  solver_optns = NULL,
  scaling = TRUE,
  boot = FALSE,
  stratified_boot = FALSE,
 R = NULL,
  parallel = "no",
  ncpus = 1,
  conf = 0.95,
  type = "norm"
)
```

Arguments

```
Х
```

A matrix or data.frame containing the input(s) values. The values can be numeric, factors, or strings. The type of data changes the partitioning. If the values are continuous (double), the function partitions the data into M sets. If the values are discrete (integers, strings, factors), the number of partitioning sets is data-driven.

У	A matrix containing the output values. Each column represents a different output variable, and each row represents a different observation. Only numeric values are allowed.
М	A scalar representing the number of partitions for continuous inputs.
cost	(default "L2") A string or function defining the cost function of the Optimal Transport problem. It should be "L2" or a function taking as input y and re- turning a cost matrix. If cost="L2", ot_indices uses the squared Euclidean metric.
discrete_out	(default FALSE) Logical, by default the output sample in y are equally weighted. If discrete_out=TRUE, the function tries to create an histogram of the realiza- tions and to use the histogram as weights. It works if the output is discrete or mixed and the number of realizations is large. The advantage of this option is to reduce the dimension of the cost matrix.
solver	Solver for the Optimal Transport problem. Currently supported options are:
	• "sinkhorn" (default), the Sinkhorn's solver (Cuturi 2013).
	• "sinkhorn_log", the Sinkhorn's solver in log scale (Peyré et al. 2019).
	• "transport", a solver of the non regularized OT problem using transport::transport().
solver_optns	(optional) A list containing the options for the Optimal Transport solver. See details for allowed options and default ones.
scaling	(default TRUE) Logical that sets whether or not to scale the cost matrix.
boot	(default FALSE) Logical that sets whether or not to perform bootstrapping of the OT indices.
stratified_boo	
	(default FALSE) Logical that sets the type of resampling performed. With stratified_boot=FALSE, the function resamples the dataset and then creates the partitions. Otherwise, first, it creates the partitions and then it performs stratified bootstrapping with strata being the partitions.
R	(default NULL) Positive integer, number of bootstrap replicas.
parallel	(default "no") The type of parallel operation to be used (if any). If miss- ing, the default is taken from the option boot.parallel (and if that is not set, "no"). Only considered if boot = TRUE. For more information, check the boot::boot() function.
ncpus	(default 1) Positive integer: number of processes to be used in parallel operation: typically one would chose this to the number of available CPUs. Check the ncpus option in the boot::boot() function of the boot package.
conf	(default 0.95) Number between 0 and 1 representing the confidence level. Only considered if boot = TRUE.
type	(default "norm") Method to compute the confidence interval. Only considered if boot = TRUE. For more information, check the type option of boot::boot.ci().

Details

Solvers:

OT is a widely studied topic in Operational Research and Calculus. The reference for the OT solvers in this package is Peyré et al. (2019). The default solver is "sinkhorn", the Sinkhorn's

solver introduced in Cuturi (2013). It solves the entropic-regularized version of the OT problem. The "sinkhorn_log" solves the same OT problem but in log scale. It is more stable for low values of the regularization parameter but slower to converge. The option "transport" is used to choose a solver for the non-regularized OT problem. Under the hood, the function calls transport::transport() from package transport. This option does not define the solver per se, but the solver should be defined with the argument solver_optns. See the next section for more information.

Solver options:

The argument solver_optns should be empty (for default options) or a list with all or some of the required solver parameters. All the parameters not included in the list will be set to default values. The solvers "sinkhorn" and "sinkhorn_log" have the same options:

- numIterations (default 1e3): a positive integer defining the maximum number of Sinkhorn's iterations allowed. If the solver does not converge in the number of iterations set, the solver will throw an error.
- epsilon (default 0.01): a positive real number defining the regularization coefficient. If the value is too low, the solver may return NA.
- maxErr (default 1e-9): a positive real number defining the approximation error threshold between the marginal histogram of the partition and the one computed by the solver. The solver may fail to converge in numIterations if this value is too low.

The solver "transport" has the parameters:

- method (default "networkflow): string defining the solver of the OT problem.
- control: a named list of parameters for the chosen method or the result of a call to transport::trcontrol().
- threads (default 1): an Integer specifying the number of threads used in parallel computing.

For details regarding this solver, check the transport::transport() help page.

Value

A gsaot_indices object containing:

- method: a string that identifies the type of indices computed.
- indices: a names array containing the sensitivity indices between 0 and 1 for each column in x, indicating the influence of each input variable on the output variables.
- bound: a double representing the upper bound of the separation measure or an array representing the mean of the separation for each input according to the bootstrap replicas.
- x, y: input and output data provided as arguments of the function.
- inner_statistic: a list of matrices containing the values of the inner statistics for the partitions defined by partitions. If method = wasserstein-bures, each matrix has three rows containing the Wasserstein-Bures indices, the Advective, and the Diffusive components.
- partitions: a matrix containing the partitions built to calculate the sensitivity indices. Each column contains the partition associated to the same column in x. If boot = TRUE, the object contains also:
- indices_ci: a data.frame with first column the input, second and third columns the lower and upper bound of the confidence interval.
- inner_statistic_ci: a list of matrices. Each element of the list contains the lower and upper confidence bounds for the partition defined by the row.

ot_indices_1d

- bound_ci: a list containing the lower and upper bounds of the confidence intervals of the separation measure bound.
- type, conf: type of confidence interval and confidence level, provided as arguments.

References

Cuturi M (2013). "Sinkhorn distances: Lightspeed computation of optimal transport." Advances in neural information processing systems, **26**.

Peyré G, Cuturi M, others (2019). "Computational optimal transport: With applications to data science." *Foundations and Trends*® *in Machine Learning*, **11**(5-6), 355–607.

See Also

ot_indices_1d(), ot_indices_wb()

Examples

```
N <- 1000
mx <- c(1, 1, 1)
Sigmax <- matrix(data = c(1, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 1), nrow = 3)
x1 <- rnorm(N)
x2 <- rnorm(N)
x3 <- rnorm(N)
x <- cbind(x1, x2, x3)
x <- mx + x %*% chol(Sigmax)
A <- matrix(data = c(4, -2, 1, 2, 5, -1), nrow = 2, byrow = TRUE)
y <- t(A %*% t(x))
x <- data.frame(x)
M <- 25
# Calculate sensitivity indices
sensitivity_indices <- ot_indices(x, y, M)
sensitivity_indices</pre>
```

ot_indices_1d Evaluate Optimal Transport indices on one dimensional outputs

Description

Evaluate Optimal Transport indices on one dimensional outputs

Usage

```
ot_indices_1d(
    x,
    y,
    M,
    boot = FALSE,
    R = NULL,
    parallel = "no",
    ncpus = 1,
    conf = 0.95,
    type = "norm"
)
```

Arguments

x	A matrix or data.frame containing the input(s) values. The values can be nu- meric, factors, or strings. The type of data changes the partitioning. If the values are continuous (double), the function partitions the data into M sets. If the values are discrete (integers, strings, factors), the number of partitioning sets is data-driven.
У	An array containing the output values.
М	A scalar representing the number of partitions for continuous inputs.
boot	(default FALSE) Logical that sets whether or not to perform bootstrapping of the OT indices.
R	(default NULL) Positive integer, number of bootstrap replicas.
parallel	(default "no") The type of parallel operation to be used (if any). If miss- ing, the default is taken from the option boot.parallel (and if that is not set, "no"). Only considered if boot = TRUE. For more information, check the boot::boot() function.
ncpus	(default 1) Positive integer: number of processes to be used in parallel operation: typically one would chose this to the number of available CPUs. Check the ncpus option in the boot::boot() function of the boot package.
conf	(default 0.95) Number between 0 and 1 representing the confidence level. Only considered if boot = TRUE.
type	(default "norm") Method to compute the confidence interval. Only considered if boot = TRUE. For more information, check the type option of boot::boot.ci().

Value

A gsaot_indices object containing:

- method: a string that identifies the type of indices computed.
- indices: a names array containing the sensitivity indices between 0 and 1 for each column in x, indicating the influence of each input variable on the output variables.
- bound: a double representing the upper bound of the separation measure or an array representing the mean of the separation for each input according to the bootstrap replicas.

- x, y: input and output data provided as arguments of the function.
- inner_statistic: a list of matrices containing the values of the inner statistics for the partitions defined by partitions. If method = wasserstein-bures, each matrix has three rows containing the Wasserstein-Bures indices, the Advective, and the Diffusive components.
- partitions: a matrix containing the partitions built to calculate the sensitivity indices. Each column contains the partition associated to the same column in x. If boot = TRUE, the object contains also:
- indices_ci: a data.frame with first column the input, second and third columns the lower and upper bound of the confidence interval.
- inner_statistic_ci: a list of matrices. Each element of the list contains the lower and upper confidence bounds for the partition defined by the row.
- bound_ci: a list containing the lower and upper bounds of the confidence intervals of the separation measure bound.
- type, conf: type of confidence interval and confidence level, provided as arguments.

See Also

ot_indices(), ot_indices_wb()

Examples

```
x <- rnorm(1000)
y <- 10 * x
ot_indices_1d(data.frame(x), y, 30)</pre>
```

ot_indices_smap Evaluate sensitivity maps using Optimal Transport indices

Description

Evaluate sensitivity maps using Optimal Transport indices

Usage

```
ot_indices_smap(x, y, M)
```

Arguments

x	A matrix or data.frame containing the input(s) values. The values can be numeric, factors, or strings. The type of data changes the partitioning. If the values are continuous (double), the function partitions the data into M sets. If the values are discrete (integers, strings, factors), the number of partitioning sets is data-driven.
У	A matrix containing the output values. Each column is interpreted as a different output.
М	A scalar representing the number of partitions for continuous inputs.

A matrix where each column represents an input and each row represents an output. The values are indices between 0 and 1 computed using ot_indices_1d().

Examples

```
N <- 1000
x1 <- rnorm(N)
x2 <- rnorm(N)
x <- cbind(x1, x2)
y1 <- 10 * x1
y2 <- x1 + x2
y <- cbind(y1, y2)
ot_indices_smap(data.frame(x), y, 30)</pre>
```

ot_indices_wb	Evaluate Wasserstein-Bures approximation of the Optimal Transport
	solution

Description

Evaluate Wasserstein-Bures approximation of the Optimal Transport solution

Usage

```
ot_indices_wb(
    x,
    y,
    M,
    boot = FALSE,
    R = NULL,
    parallel = "no",
    ncpus = 1,
    conf = 0.95,
    type = "norm"
)
```

Arguments

A matrix or data.frame containing the input(s) values. The values can be numeric, factors, or strings. The type of data changes the partitioning. If the values are continuous (double), the function partitions the data into M sets. If the values are discrete (integers, strings, factors), the number of partitioning sets is data-driven.

Х

У	A matrix containing the output values. Each column represents a different output variable, and each row represents a different observation. Only numeric values are allowed.
М	A scalar representing the number of partitions for continuous inputs.
boot	(default FALSE) Logical that sets whether or not to perform bootstrapping of the OT indices.
R	(default NULL) Positive integer, number of bootstrap replicas.
parallel	(default "no") The type of parallel operation to be used (if any). If miss- ing, the default is taken from the option boot.parallel (and if that is not set, "no"). Only considered if boot = TRUE. For more information, check the boot::boot() function.
ncpus	(default 1) Positive integer: number of processes to be used in parallel operation: typically one would chose this to the number of available CPUs. Check the ncpus option in the boot::boot() function of the boot package.
conf	(default 0.95) Number between 0 and 1 representing the confidence level. Only considered if boot = TRUE.
type	(default "norm") Method to compute the confidence interval. Only considered if boot = TRUE. For more information, check the type option of boot::boot.ci().

A gsaot_indices object containing:

- method: a string that identifies the type of indices computed.
- indices: a names array containing the sensitivity indices between 0 and 1 for each column in x, indicating the influence of each input variable on the output variables.
- bound: a double representing the upper bound of the separation measure or an array representing the mean of the separation for each input according to the bootstrap replicas.
- x, y: input and output data provided as arguments of the function.
- inner_statistic: a list of matrices containing the values of the inner statistics for the partitions defined by partitions. If method = wasserstein-bures, each matrix has three rows containing the Wasserstein-Bures indices, the Advective, and the Diffusive components.
- partitions: a matrix containing the partitions built to calculate the sensitivity indices. Each column contains the partition associated to the same column in x. If boot = TRUE, the object contains also:
- indices_ci: a data.frame with first column the input, second and third columns the lower and upper bound of the confidence interval.
- inner_statistic_ci: a list of matrices. Each element of the list contains the lower and upper confidence bounds for the partition defined by the row.
- bound_ci: a list containing the lower and upper bounds of the confidence intervals of the separation measure bound.
- type, conf: type of confidence interval and confidence level, provided as arguments.

See Also

ot_indices(), ot_indices_1d()

Examples

```
N <- 1000
mx <- c(1, 1, 1)
Sigmax <- matrix(data = c(1, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 1), nrow = 3)
x1 <- rnorm(N)
x2 <- rnorm(N)
x3 <- rnorm(N)
x <- cbind(x1, x2, x3)
x <- mx + x %*% chol(Sigmax)
A <- matrix(data = c(4, -2, 1, 2, 5, -1), nrow = 2, byrow = TRUE)
y <- t(A %*% t(x))
x <- data.frame(x)
y <- y
ot_indices_wb(x, y, 100)</pre>
```

plot.gsaot_indices Plot Optimal Transport sensitivity indices

Description

Plot Optimal Transport based sensitivity indices using ggplot2 package.

Usage

```
## S3 method for class 'gsaot_indices'
plot(x, ranking = NULL, wb_all = FALSE, dummy = NULL, ...)
```

Arguments

х	An object generated by ot_indices, ot_indices_1d, or ot_indices_wb.
ranking	An integer with absolute value less or equal than the number of inputs. If pos- itive, select the first ranking inputs per importance. If negative, select the last ranking inputs per importance.
wb_all	(default FALSE) Logical that defines whether or not to plot the Advective and Diffusive components of the Wasserstein-Bures indices
dummy	(default NULL) A double or and object of class gsaot_indices that represents a lower bound.
	Further arguments passed to or from other methods.

```
plot_inner_stats
```

A ggplot object that, if called, will print

Examples

```
N <- 1000
mx <- c(1, 1, 1)
Sigmax <- matrix(data = c(1, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 1), nrow = 3)
x1 <- rnorm(N)</pre>
x2 <- rnorm(N)</pre>
x3 <- rnorm(N)
x <- cbind(x1, x2, x3)</pre>
x <- mx + x %*% chol(Sigmax)</pre>
A <- matrix(data = c(4, -2, 1, 2, 5, -1), nrow = 2, byrow = TRUE)
y <- t(A %*% t(x))
x <- data.frame(x)</pre>
M <- 25
# Calculate sensitivity indices
sensitivity_indices <- ot_indices_wb(x, y, M)</pre>
sensitivity_indices
plot(sensitivity_indices)
```

plot_inner_stats Plot Optimal Transport inner statistics

Description

Plot Optimal Transport based inner statistics for each partition using ggplot2 package. If provided, it plots also the uncertainty estimates.

Usage

```
plot_inner_stats(x, ranking = NULL, wb_all = FALSE, ...)
```

Arguments

х	An object generated by ot_indices, ot_indices_1d, or ot_indices_wb.
ranking	An integer with absolute value less or equal than the number of inputs. If pos- itive, select the first ranking inputs per importance. If negative, select the last ranking inputs per importance.

wb_all	(default FALSE) Logical that defines whether or not to plot the Advective and Diffusive components of the Wasserstein-Bures indices
	Further arguments passed to or from other methods.

A patchwork object that, if called, will print.

Examples

```
N <- 1000
mx <- c(1, 1, 1)
Sigmax <- matrix(data = c(1, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 1), nrow = 3)
x1 <- rnorm(N)
x2 <- rnorm(N)
x3 <- rnorm(N)
x <- cbind(x1, x2, x3)
x <- mx + x %*% chol(Sigmax)
A <- matrix(data = c(4, -2, 1, 2, 5, -1), nrow = 2, byrow = TRUE)
y <- t(A %*% t(x))
x <- data.frame(x)
M <- 25
# Get sensitivity indices
sensitivity_indices <- ot_indices(x, y, M)
plot_inner_stats(sensitivity_indices)
```

print.gsaot_indices Print Optimal Transport Sensitivity indices information

Description

Print Optimal Transport Sensitivity indices information

Usage

```
## S3 method for class 'gsaot_indices'
print(x, data = FALSE, ...)
```

Arguments

х	An object generated by ot_indices, ot_indices_1d, or ot_indices_wb.
data	Logical, indicating whether or not the input and output data should be printed.
	Further arguments passed to or from other methods.

Value

The information contained in argument x

Examples

```
N <- 1000
mx <- c(1, 1, 1)
Sigmax <- matrix(data = c(1, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 1), nrow = 3)
x1 <- rnorm(N)
x2 <- rnorm(N)
x3 <- rnorm(N)
x <- cbind(x1, x2, x3)
x <- mx + x %*% chol(Sigmax)
A <- matrix(data = c(4, -2, 1, 2, 5, -1), nrow = 2, byrow = TRUE)
y <- t(A %*% t(x))
x <- data.frame(x)
M <- 25
# Calculate sensitivity indices
sensitivity_indices <- ot_indices(x, y, M)
print(sensitivity_indices)</pre>
```

Index

boot::boot(), 5, 8, 11
boot::boot.ci(), 5, 8, 11

 $\texttt{lower_bound, 2}$

ot_indices, 4, *12*, *13*, ot_indices(), *3*, *9*, ot_indices_1d, *7*, *12*, *13*, ot_indices_1d(), *3*, *7*, *10*, ot_indices_smap, 9 ot_indices_wb, 10, *12*, *13*, ot_indices_wb(), *3*, *7*, 9

plot.gsaot_indices, 12
plot_inner_stats, 13
print.gsaot_indices, 14

transport::transport(), 3, 5, 6
transport::trcontrol(), 6