Wasserstein tests

July 28, 2021

R topics documented:

`dist.torus.mod`

Description

Computes distance between two points on the two-dimensional flat torus, conceived as a periodic [0,1]x[0,1] space.

Usage

```
dist.torus.mod(coor)
```

Arguments

`coor`	A vector of coordinates of two points on the 2-dimensional flat torus. It must have the form (x1,y1,x2,y2).

Value

The distance between (x1,y1) and (x2,y2).

stat_1d

Description

Computes the distribution-free one-dimensional statistic defined in [1] for two samples on the real line.

Usage

```
stat_1d(x,y)
```

Arguments

x	A sample on the real line.
y	A sample on the real line.

Value

The statistic value corresponding to (x,y).

References

[1] RAMDAS, A., GARCIA, N. and CUTURI, M. (2015). On Wasserstein Two Sample Testing and Related Familites of Nonparametric Tests. Entropy 19.

test_1d

Description

Computes the p-value corresponding to the one-dimensional Wasserstein test defined in [1] for two samples on the real line.

Usage

```
test_1d(x,y,NR=1000,NS=1000)
```

Arguments

x	A sample on the real line.
y	A sample on the real line.
NR	The number of replications for the statistic distribution simulation under the null. The default is NR=1000.
NS	The size of the samples used to simulate the statistic distribution. The default is NS=1000.

Details

The statistic distribution is simulated by computing NR times the statistic value for two uniform laws of size NS in [0,1].

Value

The p-value corresponding to the sample (x,y).

References

[1] RAMDAS, A., GARCIA, N. and CUTURI, M. (2015). On Wasserstein Two Sample Testing and Related Familites of Nonparametric Tests. Entropy 19.

---

test_1d_circ

Description

This function takes two samples on the circle (conceived as a periodic [0,1]), finds the cutpoint defined in [1], sends the samples to the real line and perform the one-dimensional test defined in [2]. This procedure corresponds to perform a goodness-of-fit test based on Wasserstein distance directly on the circle.

Usage

```
test_1d_circ(x,y,by_seq=0.01,NR=1000,NS=1000)
```

Arguments

x	A sample on the circle.
y	A sample on the circle.
by_seq=0.01	The parameter by_seq to be passed to the transport_circ function.
NR=1000	The parameter NR to be passed to the test_1d function.
NS=1000	The parameter NS to be passed to the test_1d function.

Details

In order to perform the Wasserstein test on the two-dimensional flat torus, this function must be applied to both pair of marginals and the minimum of both p-values has to be kept.

Value

The p-value corresponding to the Wasserstein test between both measures (x,y) on the circle.

References

[1] RABIN, J., DELON, J. and GOUSSEAU, Y. (2009). Transportation Distances on the Circle. Journal of Mathematical Imaging and Vision 41. [2] RAMDAS, A., GARCIA, N. and CUTURI, M. (2015). On Wasserstein Two Sample Testing and Related Familites of Nonparametric Tests. Entropy 19.

`transport_circ`

Description

Finds the cut-point on the circle defined in [1] in order to send the problem to the real line.

Usage

```
transport_circ(x,y,by_seq=0.01)
```

Arguments

`x`	A sample on the periodic [0,1].
`y`	A sample on the periodic [0,1].
`by_seq`	The grid precision on the circle where cutpoints are looked for. The default is by_seq=0.01

Details

This function computes the Optimal Transportation problem for two measures (x,y) on the circle (conceived as a periodic [0,1]). Then, the cutpoint defined in [1] is searched among a set of candidates on a grid of precision by_seq on the circle. If the function fails to find a cutpoint the precision by_seq should be thinned.

Value

`opt_plan`	The optimal transportation plan between both measures on the circle.
`cut_point`	The cutpoint.
`x`	The original sample x.
`y`	The original sample y.

References

[1] RABIN, J., DELON, J. and GOUSSEAU, Y. (2009). Transportation Distances on the Circle. Journal of Mathematical Imaging and Vision 41.

wd_torus

Description

Computes empirical Wasserstein distance between two samples on the two-dimensional flat torus, conceived as a periodic [0,1]x[0,1] space, using the R package transport [1].

Usage

```
wd_torus(x1,y1,x2,y2,p=2,ot_method="networkflow")
```

Arguments

x1	A vector corresponding to the first coordinate of the first sample.
y1	A vector corresponding to the second coordinate of the first sample.
x2	A vector corresponding to the first coordinate of the second sample.
y2	A vector corresponding to the second coordinate of the second sample.
p	the power to which the torus distance between points is taken in order to compute transportation costs.
method	The algorithm to compute the optimal transportation problem. See [1].

Details

This function implements the function wasserstein from the R package transport [1], for the cost derived of the euclidean distance on the two-dimensional flat torus. For details on the wasserstein distance computation, see [1].

Value

The p-power of the empirical Wasserstein distance between both samples.

References

[1] Schuhmacher D, Bähre B, Gottschlich C, Hartmann V, Heinemann F, Schmitzer B (2020). transport: Computation of Optimal Transport Plans and Wasserstein Distances. R package version 0.12-2, https://cran.r-project.org/package=transport.

Index