Package 'torustest'

January 24, 2023

Type Package

Title Two-sample goodness-of-fit tests on the two-dimensional flat torus based on Wasserstein distance.

Version 0.1.0

Author Javier González-Delgado

Maintainer Javier González-Delgado <javier.gonzalez-delgado@math.univ-toulouse.fr>

Description The goal of torustest is to provide some practical approaches to perform two-sample goodness-of-fit tests for measures supported on the two-dimensional flat torus based on Wasserstein distance. These techniques have been introduced in González-Delgado, J., González-Sanz, A., Cortés, J., & Neuvial, P. (2021). Two-sample goodness-of-fit tests on the flat torus based on Wasserstein distance and their relevance to structural biology. arXiv:2108.00165.

Encoding UTF-8

Imports BAMBI, circular, doParallel, FRACTION, ggplot2, pracma, proxy, transport, uniformly

RoxygenNote 7.1.2

NeedsCompilation no

License MIT + file LICENSE

R topics documented:

dist.torus *Distance on the two-dimensional torus*

Description

Distance between two points on the two-dimensional torus (periodic [0,1] x [0,1].)

Usage

```
dist.torus(x, y)
```

Arguments

x	A vector of two coordinates in the two-dimensional flat torus, parameterized as the periodic [0,1) x [0,1).
y	A vector of two coordinates in the two-dimensional flat torus, parameterized as the periodic [0,1) x [0,1).

Value

The distance on the torus between x and y.

Examples

```
set.seed(10)
# Uniform distribution on [0,1] x [0,1]
x <- uniformly::runif_in_cube(1, 2, c(0.5, 0.5), 0.5)
y <- uniformly::runif_in_cube(1, 2, c(0.5, 0.5), 0.5)
dist.torus(x, y)

# Bivariate von Mises distribution
x <- uniformly::runif_in_cube(1, 2, c(0.5, 0.5), 0.5)
y <- BAMBI::rvmcos(1, kappa1 = 1, kappa2 = 1, mu1 = 0, mu2 = 0)/(2*pi)
dist.torus(x, y)
```

geodesic.projection *Geodesic projection of a pair of samples on the torus*

Description

Given a vector of integers u=(a,b), two samples on the torus are projected to the closed geodesic given by the canonical projection of the straight line with director vector u and containing the origin.

Usage

```
geodesic.projection(u, data_1, data_2, do_plots = FALSE, size_points = 1)
```

Arguments

u	A vector of two integers (a,b), with b non-zero.
data_1	n x 2 matrix containing n observations in the two-dimensional flat torus, parameterized as the periodic [0,1) x [0,1).
data_2	m x 2 matrix containing m observations in the two-dimensional flat torus, parameterized as the periodic [0,1) x [0,1).
do_plots	Whether to produce plots illustrating the steps of the algorithm for the given geodesic and samples.
size_points	If do_plots is TRUE, the size of the points to plot (to choose according to n,m).

Value

- plot_list - If do_plots is TRUE, a list with the corresponding plots.
- circle_one - The projected sample data_1 parameterized on the circle (periodic [0,1)).
- circle_two - The projected sample data_2 parameterized on the circle (periodic [0,1)).
- proj_data - A data frame containing the coordinates on [0,1]x[0,1] of both samples and the corresponding projections.

Examples

```
data_1 <- BAMBI::rvmcos(30) / (2 * pi)
data_2 <- BAMBI::rvmcos(30) / (2 * pi)
pr <- geodesic.projection(u = c(-1, -2), data_1, data_2, do_plots = TRUE)
pr$plot_list[[1]]
pr$plot_list[[2]]
pr$plot_list[[3]]
pr$plot_list[[4]]
pr$plot_list[[5]]
pr$circle_one; pr$circle_two
```

---

rgeodesic *Closed geodesics sampling*

Description

Samples a number of different closed geodesics on the torus. Each geodesic is represented by a vector of integers (B, A), where B follows a geometric distribution of parameter p and, for a given B=b, A is uniform in (0,1,...,b).

Usage

```
rgeodesic(ng, p = 0.1)
```

Arguments

ng	The size of the geodesics sample.
p	The parameter p of the geometric distribution of B.

Details

For big sample sizes, the value of p must be reduced in order to get a sample of different geodesics. This yields to less simple geodesics, with more revolutions over the torus.

Value

A ng x 2 matrix containing the vectors corresponding to the sampled closed geodesics.

References

[1] Petroni, N. C.(2019). Taking rational numbers at random. arXiv:1908.06944v1.

Examples

```
rgeodesic(5, p = 0.1)

# For bigger sample sizes, p must be reduced.
rgeodesic(100, p = 0.01)
```

---

sim.null.stat *Wasserstein statistic null distribution*

Description

Simulates Wasserstein statistic null distribution.

Usage

```
sim.null.stat(NR, NC = 1, n = 30)
```

Arguments

NR	Number of replicas to simulate.
NC	Number of cores for parallel computation.
n	Sample size for the simulated samples.

Value

A sample of size NR simulating the null distribution of the statistic.

References

[1] Delon, J., Salomon, J., Sobolevskii, A.: Fast transport optimization for Monge costs on the circle. SIAM J. Appl. Math. 70(7), 2239–2258 (2010).

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

Examples

```
sim.null.stat(100)
```

`stat.s1` *Wasserstein distribution-free statistic on S^1*

Description

Computes Wasserstein statistic between two samples on S^1, parameterized as a periodic [0,1).

Usage

```
stat.s1(x, y)
```

Arguments

`x`	numerical vector of sample in [0,1).
`y`	numerical vector of sample in [0,1).

Value

The statistic realization of samples x, y.

References

[1] Delon, J., Salomon, J., Sobolevskii, A.: Fast transport optimization for Monge costs on the circle. SIAM J. Appl. Math. 70(7), 2239–2258 (2010).

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

Examples

```
set.seed(10)
stat.s1(runif(50), runif(50))
stat.s1(runif(50), as.numeric(circular::rvonmises(50, pi, 1)/(2*pi)))
```

`twosample.geodesic.torus.test` *Geodesic two-sample goodness-of-fit on the torus*

Description

Performs a two-sample goodness-of-fit test for measures supported on the torus, by testing the equality of their projected distributions on ng closed geodesics and combining the results into a global p-value.

Usage

```
twosample.geodesic.torus.test(
  sample_1,
  sample_2,
  n_geodesics = 1,
  NC_geodesic = 1,
  geodesic_list = NULL,
  sim_null = NULL,
  NR = 500,
  NC = 1,
  n = 30
)
```

Arguments

`sample_1`	n x 2 matrix containing n observations in the two-dimensional flat torus, parameterized as the periodic [0,1) x [0,1).
`sample_2`	n x 2 matrix containing n observations in the two-dimensional flat torus, parameterized as the periodic [0,1) x [0,1).
`n_geodesics`	The number of closed geodesics where samples must be projected.
`NC_geodesic`	The number of cores for computing parallelly the circular tests.
`geodesic_list`	If NULL, geodesics are chosen randomly with the function rgeodesic. Else, a list of vectors containing the director vectors of the chosen geodesics.
`sim_null`	The simulated null distribution of the circle test statistic. It NULL, the null distribution is simulated with the given parameters (very time consuming).
`NR`	The number of replicas if simulation is required.
`NC`	The number of cores if parallel simulation for sim_null is required.
`n`	The sample sizes of the simulated samples is simulation is required.

Value

The p-value for the two-sample test on the torus.

Examples

```
n <- 100 # Sample size

# Simulate the null distribution of the circle test statistic

sim_free_null <- sim.null.stat(500, NC = 2)

# Bivariate von Mises distributions
samp_1 <- BAMBI::rvmcos(n) / (2 * pi)
samp_2 <- BAMBI::rvmcos(n) / (2 * pi)

#4 geodesics are chosen randomly
twosample.geodesic.torus.test(samp_1, samp_2, n_geodesics = 3,
NC_geodesic = 2, sim_null = sim_free_null)

#4 geodesics are chosen a priori
glist <- list(c(1, 0), c(0, 1), c(1, 1), c(2, 3))
twosample.geodesic.torus.test(samp_1, samp_2, geodesic_list = glist,
NC_geodesic = 2, sim_null = sim_free_null)
```

`twosample.test.s1` *Two-sample goodness-of-fit on the circle*

Description

Performs a two-sample goodness-of-fit test for measures supported on the circle, based on a distribution-free Wasserstein statistic.

Usage

```
twosample.test.s1(x, y, sim_null = NULL, NR = 500, NC = 1, n = 30)
```

Arguments

`x`	numerical vector of sample in [0,1).
`y`	numerical vector of sample in [0,1).
`sim_null`	The simulated null distribution of the statistic. If NULL, the distribution is simulated with the given parameters (very time consuming).
`NR`	The number of replicas if simulation is required.
`NC`	The number of cores if parallel simulation is required.
`n`	The sample sizes of the simulated samples is simulation is required.

Value

- stat - The test statistic.
- pvalue - The test p-value.

References

[1] Delon, J., Salomon, J., Sobolevskii, A.: Fast transport optimization for Monge costs on the circle. SIAM J. Appl. Math. 70(7), 2239–2258 (2010).

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

Examples

```
n <- 50 # Sample size

# Simulate the statistic null distribution
NR <- 100
sim_free_null <- sim.null.stat(500, NC = 2)

x <- runif(n, 0, 1)
y <- runif(n, 0, 1)
twosample.test.s1(x, y, sim_free_null)

x <- as.numeric(circular::rvonmises(n, pi, 1)/(2*pi))
y <- as.numeric(circular::rvonmises(n, pi, 0)/(2*pi))
twosample.test.s1(x, y, sim_free_null)
```

---

twosample.ubound.torus.test

Asymptotic two-sample goodness-of-fit on the torus

Description

Performs a two-sample goodness-of-fit test for measures supported on the torus, using a p-value upper bound. The test is asymptotically consistent at level alpha.

Usage

```
twosample.ubound.torus.test(sample_1, sample_2)
```

Arguments

sample_1	n x 2 matrix containing n observations in the two-dimensional flat torus, parametrized as the periodic [0,1) x [0,1).
sample_2	n x 2 matrix containing n observations in the two-dimensional flat torus, parametrized as the periodic [0,1) x [0,1).

Value

The p-value for the two-sample test on the torus.

Examples

```
n <- 2000 # Sample size

set.seed(10)
 # Bivariate von Mises distribution
samp_1 <- BAMBI::rvmcos(n, kappa1 = 1, kappa2 = 1, mu1 = 0, mu2 = 0)/(2*pi)
samp_2 <- BAMBI::rvmcos(n, kappa1 = 1, kappa2 = 1, mu1 = 0, mu2 = 0)/(2*pi)
twosample.ubound.torus.test(samp_1, samp_2)

samp_1 <- BAMBI::rvmcos(n ,kappa1 = 0, kappa2 = 0, mu1 = 0.5, mu2 = 0.5)/(2*pi)
samp_2 <- BAMBI::rvmcos(n, kappa1 = 1, kappa2 = 1, mu1 = 0.5, mu2 = 0.5)/(2*pi)
twosample.ubound.torus.test(samp_1, samp_2)
```

Index