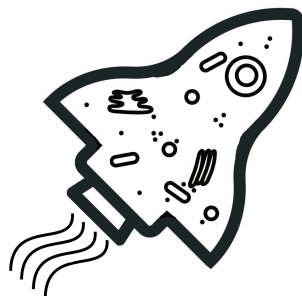


SpaceR

Spatial Cell Resampling
for quantitative analysis
of cytometry data



Package authors: **Giorgio Bertolazzi** and **Michele Tumminello**

Reference: Giorgio Bertolazzi¹², Michele Tumminello¹, Gaia Morello², Beatrice Belmonte², Claudio Tripodo²³, *Resampling approaches for the quantitative analysis of spatially distributed cells* (2023)

¹ Department of Economics, Business and Statistics, University of Palermo, Palermo, Italy

² Tumor Immunology Unit, Department of Health Sciences, University of Palermo, Palermo, Italy

³ Histopathology Unit, Institute of Molecular Oncology Foundation (IFOM) ETS - The AIRC Institute of Molecular Oncology, Milan, Italy

Introduction

The **SpaCeR** functions allow running novel resampling methods for testing spatial randomness and evaluating relationships among different cell populations. Our methods obviate the need for domain area estimation and provide non-biased clustering measurements.

Our approach can be applied to any point pattern analysis in which the study of sub-populations is limited by an inaccurate estimation of the domain area. The methodology is reported in Bertolazzi et al., *Resampling approaches for the quantitative analysis of spatially distributed cells* (2023).

1 Spatial randomness test

Description

Spatial randomness test to evaluate the presence of cell clustering behavior of cell populations.

Usage

```
spatial_randomness_test(data,X, Y, markers, B=5000, seed ,
  scatterplot=TRUE, pch=16, colors , color_neg="blue",
  name_plot="scatterplot")
```

Arguments

data	dataframe containing the cell spatial coordinates and marker positivities (1 = positive, 0 = negative). Negative cells must be included in the dataset
X	vector containing the names of variables that indicate the spatial coordinates of cells along the X-axis. The average values will be considered in the case of two X-axis variables
Y	vector containing the names of variables that indicate the spatial coordinates of cells along the Y-axis. The average values will be considered in the case of two Y-axis variables
markers	vector containing the names of variables that indicate marker positivities
B	number of random sample extractions
seed	seed of the resampling

scatterplot logicals. If TRUE, the scatterplot of the cell spatial distributions will be plotted

colors vector of colours for the scatterplot representation. The colour order should be the same as the marker vector

color_neg colour of the negative cells for the scatterplot representation

name_plot name of the output PDF file

Examples

```
# data loading
Data = data.frame(read_xlsx("Supplementary_data.xlsx", sheet=1))

# rename the variable names
names(Data)[which(names(Data)=="ACE2.Positive.classification")]="ACE2"

#loading and running the function to test the spatial randomness
source("function_spatial_randomness_test.R")

res = spatial_randomness_test(data=Data,
                              X = c("XMin", "XMax"),
                              Y= c("YMin", "YMax"),
                              markers= c("ACE2"),
                              B=5000, seed=123,
                              scatterplot=TRUE,
                              colors=c("brown") )

# Spatial randomness testing...
# 20% Done
# 40% Done
# 60% Done
# 80% Done
# 100% Done

#          n average_NN rand_average_NN      R    pval
# ACE2 203      47.34      61.35 0.772 <2e-04 ***
# ---
# alternative hypothesis: spatial clustering
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output

```
res
#   marker   n average_NN rand_average_NN      R    pval
# 1   ACE2 203    47.34      61.38 0.771 <2e-04
```

marker	cell population under exam
n	number of cells which compose the cell population
average_NN	average nearest neighbour distance inside the cell population
rand_average_NN	average nearest neighbour distance inside the random populations of size n
R	empirical R index. $R < 1$ indicates aggregation (i.e., clustering) $R > 1$ indicates segregation $R \approx 1$ indicates spatial randomness
pval	spatial randomness p -value. Null hypothesis: spatial randomness. Alternative hypothesis: cell aggregation

2 Spatial dependence test

Description

Spatial dependence test to evaluate the presence of aggregation and segregation effects between two cell populations.

Usage

```
spatial_randomness_test(data,X, Y, markers,
                        alternative = "two.sided",
                        B=5000, seed,
                        scatterplot=TRUE, pch=16,
                        colors, color_neg="blue",
                        name_plot="scatterplot")
```

Arguments

data	dataframe containing the cell spatial coordinates and marker positivities (1 = positive, 0 = negative). Negative cells must be included in the dataset
X	vector containing the names of variables that indicate the spatial coordinates of cells along the X-axis. The average values will be considered in the case of two X-axis variables
Y	vector containing the names of variables that indicate the spatial coordinates of cells along the Y-axis. The average values will be considered in the case of two Y-axis variables
markers	vector containing the names of variables that indicate marker positivities. The vector's first element refers to the cell population whose coordinates will be fixed during the resampling. The distances are calculated between each cell of the first population and the nearest neighbour cells of the i -th population (where i is the position of the marker name in the input vector, $i \geq 2$)
alternative	alternative hypothesis. Options: two.tailed, aggregation, segregation
B	number of random sample extractions
seed	seed of the resampling
scatterplot	logicals. If TRUE, the scatterplot of the cell spatial distributions will be plotted
colors	vector of colours for the scatterplot representation. The colour order should be the same as the marker vector.
color_neg	colour of the negative cells for the scatterplot representation
name_plot	name of the output PDF file

Examples

```
# data loading
Data = data.frame(read_xlsx("Supplementary_data.xlsx", sheet=2))

## rename the variable names
```

```

names(Data)[which(names(Data)=="AID.Positive.Classification")] ="AID"
names(Data)[which(names(Data)=="CD3.Positive.Classification")] ="CD3"

# loading the function for segregation/aggregation analysis
source("function_spatial_dependence_test.R")

# running the test to evaluate the segregation behavior
res = spatial_dependence_test(data=Data,
                              X=c("XMin", "XMax"),
                              Y= c("YMin", "YMax"),
                              markers= c("AID", "CD3"),
                              alternative = "segregation",
                              colors = c("red", "green"),
                              B=5000, seed=1234)

# Cell resampling...
# 20% Done
# 40% Done
# 60% Done
# 80% Done
# 100% Done
# ---
# Average distances of nearest neighbour cells to AID (n = 199):
#      n average_NN rand_average_NN      R  pval
# CD3 158      265.84      100.48 2.646 <2e-04 ***
# ---
# alternative hypothesis: segregation
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Output

```

res
# markers  n average_NN rand_average_NN      R  pval
# 1      CD3 158      265.84      100.16 2.654 <2e-04

```

marker cell population whose segregation/aggregation is studied with respect to the first population

n number of cells of the population under exam

average_NN average distance between each cell of the first cell population and the nearest neighbour cells of the population under exam

rand_average_NN	average distance between each cell of the first cell population and the nearest neighbour cells of random populations of size n
R	<p>empirical R_{ab} index.</p> <p>$R < 1$ indicates aggregation effect between the two populations</p> <p>$R > 1$ indicates segregation effect between the two populations</p> <p>$R \approx 1$ indicates the absence of spatial dependence between the two populations</p>
pval	spatial dependence p -value. Null hypothesis: spatial independence. Alternative hypothesis: cell aggregation and/or segregation

3 Distribution equality test

Description

Spatial dependence test to evaluate if two cell populations are differentially distributed over the spatial domain.

Usage

```
distribution_equality_test(data, X, Y, M1, M2,
                           colM1="green", colM2="red",
                           rm.double=FALSE,
                           B=5000, seed )
```

Arguments

data	dataframe containing the cell spatial coordinates and marker positivities (1 = positive, 0 = negative)
X	vector containing the names of variables that indicate the spatial coordinates of cells along the X-axis. The average values will be considered in the case of two X-axis variables
Y	vector containing the names of variables that indicate the spatial coordinates of cells along the Y-axis. The average values will be considered in the case of two Y-axis variables

M1	name of the variable which contains the positivity to the first marker
M2	name of the variable which contains the positivity to the second marker
colM1	colour of the first marker for the scatterplot representation
colM2	colour of the second marker for the scatterplot representation
rm.double	logicals. If TRUE, the double positive cells will be removed
B	number of random sample extractions
seed	seed of the resampling

Examples

```
# data loading
Data = data.frame(read_xlsx("Supplementary_data.xlsx", sheet=4))

# rename the variable names
names(Data)[which(names(Data)=="AID.Positive.Classification")] ="AID"
names(Data)[which(names(Data)=="CD3.Positive.Classification")] ="CD3"

# loading the function for distribution equality testing
source("function_distribution_equality_test.R")

# distribution equality test
res = distribution_equality_test(data=Data,
                                X= c("XMin", "XMax"),
                                Y= c("YMin", "YMax"),
                                M1="CD3", M2="AID",
                                colM1="green", colM2="red",
                                rm.double=TRUE, B=5000)

# warning: 2 double positive cells have been removed
# Distribution equality testing...
# 20% Done
# 40% Done
# 60% Done
# 80% Done
# 100% Done
```



```
# test.statistic = 1679.62 , p-value < 2e-04
# alternative hypothesis: CD3 and AID cells have different
# spatial distributions
# dissimilarity index = 12.07
```

Output

```
res
# $n_M1
# [1] 211

# $n_M2
# [1] 292

# $M1_proportion
# [1] 0.419

# $Chi_statistic
# [1] 1679.624

# $pvalue
# [1] 0

# $index
# [1] 12.06879
```

n_M1	number of cells in the first population
n_M2	number of cells in the second population
M1_proportion	proportion of cells of the first population
Chi_statistic	empirical value of the statistical test
pvalue	distribution equality test p -value. Null hypothesis: distribution equality. Alternative hypothesis: different spatial distributions
index	dissimilarity index I_C <div> $I_C \neq 1$ indicates distribution differences $I_C \approx 1$ indicates equal distribution </div>