

# Package ‘dixon’

January 27, 2015

**Type** Package

**Title** Nearest Neighbour Contingency Table Analysis

**Version** 0.0-5

**Date** 2013-03-13

**Author** Marcelino de la Cruz Rot and Philip M. Dixon

**Maintainer** Marcelino de la Cruz Rot <marcelino.delacruz@upm.es>

**Depends** splancs, spatstat

**Description** Function to test spatial segregation and association based in contingency table analysis of nearest neighbour counts following Dixon (2002). Some fortran code has been included to the original dixon2002 function of the ecespa package to improve speed.

**License** GPL (>= 2)

**LazyLoad** yes

**LazyData** yes

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2013-03-13 21:10:26

## R topics documented:

dixon	2
p2colasr	4
swamp	5

<b>Index</b>	<b>7</b>
--------------	----------

---

dixon

*Dixon (2002) Nearest-neighbor contingency table analysis*


---

## Description

dixon is a wrapper to the functions of Dixon (2002) to test spatial segregation for several species by analyzing the counts of the nearest neighbour contingency table for a marked point pattern.

## Usage

```
dixon(datos, nsim = 99, fortran =TRUE)
```

## Arguments

datos	data.frame with three columns: x-coordinate, y-coordinate and sp-name. See <a href="#">swamp</a> .
nsim	number of simulations for the randomization approximation of the p-values.
fortran	should the fortran implementation be used?.

## Details

A measure of segregation describes the tendency of one species to be associated with itself or with other species. Dixon (2002) proposed a measure of the **segregation of species  $i$**  in a multispecies spatial pattern as:

$$S[i] = \log[(N[ii]/(N[i] - N[ii]))/((N[i] - 1)/(N - N[i]))]$$

where  $N[i]$  is the number of individuals of species  $i$ ,  $N[ii]$  is the frequency of species  $i$  as neighbor of species  $i$  and  $N$  is the total number of locations. Values of  $S[i]$  larger than 0 indicate that species  $i$  is segregated; the larger the value of  $S[i]$ , the more extreme the segregation. Values of  $S[i]$  less than 0 indicate that species  $i$  is found as neighbor of itself less than expected under random labelling. Values of  $S[i]$  close to 0 are consistent with random labelling of the neighbors of species  $i$ .

Dixon (2002) also proposed a **pairwise segregation index** for the off-diagonal elements of the contingency table:

$$S[ij] = \log[(N[ij]/(N[i] - N[ij]))/((N[i]/(N - N[j]) - 1)]$$

$S[ij]$  is larger than 0 when  $N[ij]$ , the frequency of neighbors of species  $j$  around points of species  $i$ , is larger than expected under random labelling and less than 0 when  $N[ij]$  is smaller than expected under random labelling.

As a **species/neighbor-specific test**, Dixon(2002) proposed the statistic

$$Z[ij] = (N[ij] - EN[ij])/sqrt(VarN[ij])$$

where  $j$  may be the same as  $i$  and  $EN[ij]$  is the expected count in the contingency table. It has an asymptotic normal distribution with mean 0 and variance 1; its asymptotic p-value can be obtained from the numerical evaluation of the cumulative normal distribution or by simulation, i.e. by conducting a randomization test (appropriate when the sample size is small).

An **overall test of random labelling** (i.e. a test that all counts in the  $k \times k$  nearest-neighbor contingency table are equal to their expected counts) is based on the quadratic form

$$C = (N - EN)'Sigma^{-1}(N - EN)$$

where  $N$  is the vector of all cell counts in the contingency table,  $Sigma$  is the variance-covariance matrix of those counts and  $Sigma^{-1}$  is a generalized inverse of  $Sigma$ . Under the null hypothesis of random labelling of points,  $C$  has a asymptotic Chi-square distribution with  $k(k - 1)$  degrees of freedom (if the sample sizes are small its distribution should be estimated using Monte-Carlo simulation). P-values are computed from the probability of observing equal or larger values of  $C$ . The overall statistic  $C$  can be partitioned into  $k$  **species-specific test** statistics  $C[i]$ . Each  $C[i]$  test if the frequencies of the neighbors of species  $i$  are similar to the expected frequencies if the points were randomly labelled. Because the  $C[i]$  are not independent Chi-square statistics, they do not sum to the overall  $C$ .

## Value

A list with the following components:

ON	Observed nearest neighbor counts in table format. From row sp to column sp.
EN	Expected nearest neighbor counts in table format.
Z	Z-score for testing whether the observed count equals the expected count.
S	Segregation measure.
pZas	P-values based on the asymptotic normal distribution of the Z statistic.
pZr	If nsim !=0, p-values of the Z-score based on the randomization distribution.
C	Overall test of random labelling.
Ci	Species-specific test of random labelling.
pCas	P-value of the overall test from the asymptotic chi-square distribution with the appropriate degrees of freedom.
pCias	P-values of the species-specific tests from the asymptotic chi-square distribution with the appropriate degrees of freedom.
pCr	If nsim !=0, p-value of the overall test from the randomization distribution.
pCir	If nsim !=0, p-values of the species-specific tests from the randomization distribution.
tablaZ	table with ON, EN, Z, S, pZas and pZr in pretty format, as in the table II of Dixon (2002).
tablaC	table with C, Ci, pCas, pCias, pCr and pCir in pretty format, as in the table IV of Dixon (2002).

**Warning**

The  $S[i]$  and  $S[ij]$  statistics assume that the spatial nearest-neighbor process is stationary, at least to second order, i.e., have the same sign in every part of the entire plot. A biologically heterogeneous process will violate this assumption.

**Note**

This function is an improvement of function `dixon2002` of the package **ecespa**. It includes also a small typo correction of the original code.

**Author(s)**

Philip M. Dixon <http://www.public.iastate.edu/~pdixon/>. Marcelino de la Cruz <marcelino.delacruz@upm.es> wrote the wrapper code and the fortran implementation for this package.

**References**

Dixon, P.M. 2002. Nearest-neighbor contingency table analysis of spatial segregation for several species. *Ecoscience*, **9** (2): 142-151.

De la Cruz, M. 2008. Métodos para analizar datos puntuales. In: *Introducción al Análisis Espacial de Datos en Ecología y Ciencias Ambientales: Métodos y Aplicaciones* (eds. Maestre, F. T., Escudero, A. and Bonet, A.), pp 76-127. Asociación Española de Ecología Terrestre, Universidad Rey Juan Carlos y Caja de Ahorros del Mediterráneo, Madrid. ISBN: 978-84-9849-308-5.

**See Also**

`K012` in the package **ecespa** for another segregation test, based in the differences of univariate and bivariate  $K$ -functions.

**Examples**

```
## Not run:  
  
data(swamp)  
  
dixon(swamp,nsim=99)  
  
## End(Not run)
```

---

p2colasr

*P-value for a discrete distribution on small sample data*

---

**Description**

Computes the p-value for a two-sided hypothesis test following Dixon's (2002:145) description of the method of Agresti & Min (2001).

**Usage**

```
p2colasr(Z, nsim = length(Z))
```

**Arguments**

**Z** vector with the observed Z-score in the first position and all the simulated values behind.

**nsim** Number of simulated values.

**Value**

P-value of the two-sided hypothesis test

**Note**

This function is usually not to be called by the user. It is internally used by [dixon](#).

**Author(s)**

Marcelino de la Cruz Rot. <marcelino.delacruz@upm.es>

**References**

- Agresti, A. & Min, Y. 2001. On small-sample confidence intervals for parameters in discrete distributions. *Biometrics*, **57**: 963-971.
- Dixon, P.M. 2002. Nearest-neighbor contingency table analysis of spatial segregation for several species. *Ecoscience*, **9**(2): 142-151.

**Description**

Locations and botanical classification of trees in a plot in the Savannah River. Locations are given in metres, rounded to the nearest 0.1 metre. The data come from a 1-ha (200 m x 50 m) plot in the Savannah River Site, South Carolina, USA. The 734 mapped stems included 156 Carolina ash (*Fraxinus caroliniana*), 215 Water tupelo (*Nyssa aquatica*), 205 Swamp tupelo (*Nyssa sylvatica*), 98 Bald cypress (*Taxodium distichum*) and 60 stems of 8 additional species. Although the plots were set up by Bill Good and their spatial patterns described in Good and Whipple(1982), the plots have been maintained and resampled by Rebecca Sharitz and her colleagues of the Savannah River Ecology Laboratory. There are slightly different versions of the Good plot data. Every time the plots are resampled, some errors are corrected. This is mostly a concern for the biologists. The different versions are very similar; they are all very good examples of a marked spatial point pattern.

**Usage**

```
data(swamp)
```

**Format**

A data frame with 734 observations on the following 3 variables.

x Cartesian x-coordinate of tree

y Cartesian y-coordinate of tree

sp a factor with levels indicating the species of each tree:

FX	Carolina ash ( <i>Fraxinus caroliniana</i> )
NS	Swamp tupelo ( <i>Nyssa sylvatica</i> )
NX	Water tupelo ( <i>Nyssa aquatica</i> )
TD	Bald cypress ( <i>Taxodium distichum</i> )
OT	Other species

**Source**

Philip Dixon's personal web page <http://www.public.iastate.edu/~pdixon/>

**References**

Dixon, P.M. 2002. Nearest-neighbor contingency table analysis of spatial segregation for several species. *Ecoscience*, **9**(2): 142-151.

Good, B. J. & Whipple, S.A. 1982. Tree spatial patterns: South Carolina bottomland and swamp forest. *Bulletin of the Torrey Botanical Club*, **109**: 529-536.

Jones et al. 1994. Tree population dynamics in seven South Carolina mixed-species forests. *Bulletin of the Torrey Botanical Club*, **121**:360-368.

**Examples**

```
data(swamp)
plot(swamp$x,swamp$y, col=as.numeric(swamp$sp),pch=19,
      xlab="",ylab="",main="Swamp forest")
```

# Index

\*Topic **datasets**

swamp, [5](#)

\*Topic **distribution**

p2colasr, [4](#)

\*Topic **math**

dixon, [2](#)

\*Topic **spatial**

dixon, [2](#)

dixon, [2](#), [5](#)

dixon2002, [4](#)

K012, [4](#)

p2colasr, [4](#)

swamp, [2](#), [5](#)