Package 'selectspm'

January 5, 2023

Type Package

Title Select Point Pattern Models Based on Minimum Contrast, AIC and Goodness of Fit

Version 0.6

Date 2022-12-26

Author Marcelino de la Cruz

Depends spatstat (>= 2.0-0), ecespa

Imports spatstat.geom, spatstat.explore, spatstat.model, spatstat.random

Maintainer Marcelino de la Cruz <marcelino.delacruz@urjc.es>

Description
Fit and selects point pattern models based on minimum contrast, AIC and and goodness of fit.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2023-01-05 16:00:02 UTC

R topics documented:

LF.gof *Loosmore and Ford Goodness of Fit Test*

Description

Performs the Loosmore and Ford (2006) test or the Maximum Absolute Deviation test for a spatial point pattern.

Usage

```
LF.gof(X, rmin=NULL, rmax=NULL, na.rm=TRUE)
```

Arguments

X	An object resulting from the function envelope, i.e., with an attribute "*simfuns*" (obtained using the argument savefuns=TRUE in envelope) ,which is an object of class "fv" containing the summary functions computed for each of the simulated patterns.
rmin	Minimum value of the function argument r over which the maximum absolute deviation, or the integral, will be computed for the test.
rmax	Maximum value of the function argument r over which the maximum absolute deviation, or the integral, will be computed for the test.
na.rm	Should NA's be removed to compute the integral?

Details

These function perform a tests for goodness-of-fit of a point pattern dataset to a point process model, based on Monte Carlo simulation from the model. The simulations should have been previously computed with the function envelope, applied with the argument savefuns=TRUE in order to save all the simulated functions, required for the computation of the test.

The test, popularized in the ecological field by Loosmore and Ford (2006) is also described in Diggle (2003, page 14), and according to Baddeley and Turner (2005) also in Diggle (1986) and Cressie (1991, page 667, equation (8.5.42)). If the arguments rmin and rmax are set to NULL, the integral of the GoF statistics will be computed over the complete range of r values.

Value

A list with the following components:

u The GoF statistic, i.e., the value of the integral over the range of *r*'s

p The p-value of the test

na.count.by.r Number of NA values for each r. It helps to evaluate the reliability of the computed u's, specially for small *r*'s

Author(s)

Marcelino de la Cruz <marcelino.delacruz@urjc.es>

References

Cressie, N.A.C. (1991) *Statistics for spatial data*. John Wiley and Sons, 1991.

Diggle, P. J. (1986). Displaced amacrine cells in the retina of a rabbit : analysis of a bivariate spatial point pattern. *J. Neuroscience Methods* 18, 115-125.

Diggle, P.J. (2003) *Statistical analysis of spatial point patterns*, Second edition. Arnold.

Loosmore, N.B. and Ford, E.D. (2006) Statistical inference using the G or K point pattern spatial statistics. *Ecology* 87, 1925-1931.

See Also

`dclf.test` for an alternative implementation of the test in **spatstat**.

`select.model.gof` *Fit and Select Point Pattern Models Based on Minimum Contrast and Goodness of Fit*

Description

Fits Poisson, Poisson cluster, several inhomogeneous Poisson and several inhomogeneous Poisson cluster processes to a spatial point pattern and select the best fitting based on goodness of fit.

Usage

```
select.model.gof(pp, sigmas, r, nlarge = 10000, q = 1/4, p = 2, correction = "trans",
                 sigma2=NULL, rho=NULL, lower=NULL, upper=NULL, parscale=c(1,1),
                   dimyx=c(128,128), nsim=99, seed=1, correct.lambda=10)
  ## S3 method for class 'selectedmodgof'
plot(x,...)
  ## S3 method for class 'selectedmodgof'
print(x,...)
  ## S3 method for class 'selectedmodgof'
envelope(Y,fun=NULL,nrank=1,nsim=99,dimyx=c(128,128),...)
  ## S3 method for class 'selectedmodgof'
simulate(object, nsim=99, seed=1, dimyx=c(128,128),...)
```

Arguments

`pp`	Unmarked point pattern with the ppp format of spatstat
`sigmas`	Vector with the sigma values (standard deviations of the Gaussian kernel) for estimating different intensity surfaces with density.ppp.
`r`	Vector of values for the argument r at which the (in)homogeneous K function should be evaluated. First element should be 0.
`nlarge`	A number of points. In spatstat, if the number of points exceeds nlarge, then only the border correction will be computed (by default) for K(r). If you have a large number of points (n) and you want your border correction to be applied, set nlarge > n.

q	q parameter of the dtheta (i.e., minimum contrast) function.
p	p parameter of the dtheta (i.e., minimum contrast) function.
correction	Any selection of the options "border", "bord.modif", "isotropic", "Ripley", "translate", "translation", "none" or "best". It specifies the edge correction(s) to be applied.
sigma2	Starting value in the optimization for the squared standard deviation of the Gaussian dispersion around parent points in the in(homogeneous) Poisson cluster process.
rho	Starting value in the optimization for the intensity of parent points in the in(homogeneous) Poisson cluster process.
lower	Vector of length two with the lowest allowed values for sigma2 and rho.
upper	Vector of length two with the largest allowed values for sigma2 and rho.
parscale	Initial values for sigma2 and rho for one of the optimization approaches .
dimyx	Pixel array dimensions to estimate the intensity of the point pattern when simulating from inhomogeneous models.
nsim	Number of simulated point patterns to be generated for the estimation of the GoF of each fitted model, or when using simulate.selectedmodgof or envelope.selectedmodgof.
seed	A single value to set the random generator.
correct.lambda	Fraction of the lowest positive intensity value that will be employed to replace intensity estimates == 0.
...	Additional arguments passed to the plot, print, envelope and simulate methods.
x	An object of class "selectedmodgof", i.e., the result of using function select.model.gof().
Y	An object of class "selectedmodgof", i.e., the result of using function select.model.gof().
object	An object of class "selectedmodgof", i.e., the result of using function select.model.gof().
fun	Function to compute the desired summary statistic for the point pattern.
nrank	Integer. Rank of the envelope value amongst the nsim simulated values. A rank of 1 means that the minimum and maximum simulated values will be used.

Details

select.model.gof is a wrap to fit and select different point processes using standard tools in **spatstat** and in **ecespa**. The basic framework consists in choosing among the provided *sigma* values the bandwith which produces the best fitting of an inhomogeneous Poisson and the one which produces the best fitting of an inhomogeneous Poisson cluster process. The goodness of fit (GoF) of these models are compared to the GoF a homogeneous Poisson process and tho the GoF of a homogeneous Poisson cluster process, and the model with the best fit is returned as the final result. To avoid optimization problems (i.e., obtaining non-realistic parameters for the Poisson cluster models), arguments *lower* and *upper* alllow restricting the range of values that these parameters can attain. If these ranges are not set by the user, *select.model.gof* will select by default the most extreme and sensible values (e.g., from just one cluster to as many clusters as points in the pattern). As the experience shows that different optimization algorithms provide different results, the basic framework is repeated using both the "L-BFGS-B" and the "Nelder-Mead" algorithms of optim. In adition to the

initial parameters provided with the argument *parscale*, an aditional otimization using as *parscale* the maximum and minimum values of the *lower* and *upper* parameters range is also tried with the "L-BFGS-B" algorithm.

The goodness of fit is based on the K function (estimated by Kest or Kinhom) and is evaluated following the approach of Loosmore and Ford (2006) implemented in LF.gof. This implies computing the sum of squared differences between the observed K function and the mean of the K functions of *nsim* simulations from the fitted models (the *u* statistic), which means lots of computations and makes the process time comsuming.

Whereas the model selection approach based on AIC implemented in select.model2 is more appropriated for inference purposes, the approach followed by *select.model.gof* in general is able to select models whose simulations resemble more closely the original pattern.

Value

select.model.gof returns an object of class "*selectedmodgof*",i.e., a list with components:

`gof.u`	vector with the *u* statistic of each model.
`best.gof`	the minimum *u* value.
`best.model`	The best of the fitted models.
`models`	vector with the names of the fitted models.
`gof`	A list the results of the LF.gof test applied to the selected model, i.e., the *u* value and the associated *p*-value of the GoF test.
`envelopes`	An envelope object with envelopes computed form the best model.
`pp`	The analyzed point pattern, with the ppp format of spatstat.
`best.sigma`	The value of the selected bandwith.

envelope.selectedmod returns a an object of class "fv", see fv.object, which can be printed and plotted directly. Essentially a data frame containing columns

`r`	the vector of values of the argument r at which the summary function fun has been estimated.
`obs`	values of the summary function for the data point pattern.
`lo`	lower envelope of simulations.
`hi`	upper envelope of simulations.
`mmean`	estimated theoretical value of the summary function, computed by averaging simulated values.

simulate.selectedmod returns a list of *nsim* point patterns (with the ppp format of spatstat, simulated according to the best fitted model.

Author(s)

Marcelino de la Cruz

References

Chacon-Labella, J., De la Cruz, M. and Escudero, A. Beyond the classical nurse species effect: diversity assembly in a Mediterranean semiarid dwarf shrubland. *Journal of Vegetaation Science*. Accepted for publication, July 2015.

Loosmore, N.B. and Ford, E.D. (2006) Statistical inference using the G or K point pattern spatial statistics. *Ecology* 87, 1925-1931

Examples

```
## Not run:
# Get the data
data(teucrium)

# Define the sequence of r's at which estimate K(r)
  r<-seq(0,1.5, by=0.01)

# Define different standard deviations for the Gaussian kernel
# to estimate different intensity surfaces
 sigmas <- seq(0.5, 3.5, by=0.25)



# Fit 28 models (1 Poisson, 1 Poisson cluster, 13 inhomogeneous Poisson
#  and 13 inhomogeneous Poisson cluster)  to  teucrium and select the better ones

teucrium.model <- select.model.gof(teucrium, sigmas=sigmas, r=r)

teucrium.model

# Show the empirical K function,
#  and the envelopes based on this model

 plot( teucrium.model, sqrt(./pi)-r~r, legend=F, ylab="L(r)", las=1)


# Compute and plot envelopes for the pcf function according to the best fitted model.

teucrium.env <- envelope(teucrium.model, fun=pcf,  nsim=19)
plot(teucrium.env, legend=F)

# simulate 10 point patterns according to the best fitted model

teucrium.simu <- simulate(teucrium.model, nsim=10)

teucrium.simu



## End(Not run)
```

select.model2 *Fit and Select Point Pattern Models Based on Minimum Contrast and AIC*

Description

Fits Poisson, Poisson cluster, several inhomogeneous Poisson and several inhomogeneous Poisson cluster processes to a spatial point pattern and select the best fitting based on AIC.

Usage

```
select.model2(pp, sigmas, r, nlarge = 10000, q = 1/4, p = 2, correction = "trans")
ipc.estK2(mippp, lambda = NULL, correction = "iso", r = NULL, sigma2 = NULL,
             rho = NULL, q = 1/4, p = 2, nlarge = NULL, ...)
aic.function(r, dtheta, npar)
  ## S3 method for class 'selectedmod'
envelope(Y,fun=NULL,nrank=1,nsim=99,dimyx=c(128,128),...)
  ## S3 method for class 'selectedmod'
simulate(object,nsim=99, seed=1,dimyx=c(128,128),...)
  ## S3 method for class 'selectedmod'
plot(x,...)
  ## S3 method for class 'selectedmod'
print(x,...)
```

Arguments

pp	Unmarked point pattern with the ppp format of spatstat
sigmas	Vector with the sigma values (standard deviations of the Gaussian kernel) for estimating different intensity surfaces with density.ppp.
r	Vector of values for the argument r at which the (in)homogeneous K function should be evaluated. First element should be 0.
nlarge	A number of points. In spatstat, if the number of points exceeds nlarge, then only the border correction will be computed (by default) for K(r). If you have a large number of points (n) and you want your border correction to be applied, set nlarge > n.
q	q parameter of the dtheta (i.e., minimum contrast) function.
p	p parameter of the dtheta (i.e., minimum contrast) function.
correction	Any selection of the options "border", "bord.modif", "isotropic", "Ripley", "translate", "translation", "none" or "best". It specifies the edge correction(s) to be applied.
mippp	Unmarked point pattern with the ppp format of spatstat
lambda	Optional. Values of the estimated intensity function. Either a vector giving the intensity values at the points of the pattern mippp, a pixel image (object of class "im") giving the intensity values at all locations, a fitted point process model

	(object of class "ppm" or "kppm") or a function(x,y) which can be evaluated to give the intensity value at any location.
sigma2	Starting value in the optimization for the squared standard deviation of the Gaussian dispersion around parent points in the in(homogeneous) Poisson cluster process.
rho	Starting value in the optimization for the intensity of parent points in the in(homogeneous) Poisson cluster process.
...	Additional arguments passed to the optim function (or to the plot, print, envelope and simulate methods).
dtheta	Minimum contrast discrepancy, i.e., sum of squared diferences between the normalized empirical and theoretical K functions.
npar	Number of parameters fitted in the model.
x	An object of class "selectedmod", i.e., the result of using function select.model2().
Y	An object of class "selectedmod", i.e., the result of using function select.model2().
object	An object of class "selectedmod", i.e., the result of using function select.model2().
fun	Function to compute the desired summary statistic for the point pattern.
nsim	Number of simulated point patterns to be generated with simulate.selectedmod or when computing the envelopes with envelope.selectedmod.
seed	A single value to set the random generator.
nrank	Integer. Rank of the envelope value amongst the nsim simulated values. A rank of 1 means that the minimum and maximum simulated values will be used.
dimyx	Pixel array dimensions to estimate the intensity of the point pattern when simulating from inhomogeneous models.

Details

select.model2 is a wrap to fit and select different point processes using standard tools in of **spatstat** and of **ecespa**. ipc.estK2 fits (in)homogeneous models as the function ipc.estK of **ecespa** but, in addition, it allows seting the argument nlarge and passing options to optim. AIC calculation (actually AICc) is made by aic.function. See more details in Jara et al. (2015).

Value

select.model2 returns an object of class "selectedmod",i.e., a list with components:

dtheta vector with the minimized discrepancy values for each fitted model.

best.dtheta the minimum of the minimized discrepancy values.

best.model The best of the fitted models.

models vector with the names of the fitted models

HPPs A list with the intensity objects employed to fit each inhomogeneous model

sigmas vector with the sigma values (standard deviations of the Gaussian kernel) employed to estimate the different intensity surfaces.

aics vector with the AIC values of each model.

Kas data.frame with the empirical K functios employed to fit each model.

pp analyzed point pattern, with the ppp format of spatstat.

`ipc.estK` gives an object of class '`ecespa.minconfit`', basically a list with the following components:

sigma2 Parameter $sigma^2$.

rho Parameter rho.

d.theta Minimized value of the contrast criterion $D(theta)$.

Kobs Values of the observed K-function.

Kfit Values of the fitted K-function.

r Sequence of distances at which `Kobs` and `Kfit` have been estimated.

data Original point pattern.

lambda Original intensity object.

dataname Name of the original point pattern.

lambdaname Name of the original intensity object.

q q exponent of the contrast criterion.

p p exponent of the contrast criterion.

`aic.function` returns a one-row data.frame, with the following items:

n number of observations employed to compute AIC; i.e., number of r values where K(r) was estimated.

K number of parameters of the model.

RSS "Resiidual Sum of Squares". It is the value of the discrepacy function Dtheta.

LL Loglikelihood.

AIC AIC value.

AICc Small sample AIC value.

`envelope.selectedmod` returns a an object of class "fv", see fv.object, which can be printed and plotted directly. Essentially a data frame containing columns

r the vector of values of the argument r at which the summary function fun has been estimated.

obs values of the summary function for the data point pattern.

lo lower envelope of simulations.

hi upper envelope of simulations.

mmean estimated theoretical value of the summary function, computed by averaging simulated values.

`simulate.selectedmod` returns a list of *nsim* point patterns (with the ppp format of spatstat, simulated according to the best fitted model.

Author(s)

Marcelino de la Cruz

References

Jara, A., De la Cruz, M., Espinosa, C.I., Mendez, M. & Escudero, A. (2015). Does spatial heterogeneity blur the signature of dispersal syndromes on spatial patterns of woody species? A test in a tropical dry forest. *Oikos*. http://dx.doi.org/10.1111/oik.02098.

Examples

```
## Not run:
# Get the data
data(lansing)

# Split the multivariate pp in their individual components
lansing.sp<-split(lansing)

# Define the sequence of r's at which estimate K(r)
r<-  seq(0,0.25,le=101)

# Define different standard deviations for the Gaussian kernel
# to estimate different intensity surfaces
sigmas<- seq(0.1,1,by=0.05)

# Note that lansing is defined in a (0,1) x (0,1) window and this affects
# the election of r and sigma values

# Fit 40 models (1 Poisson, 1 Poisson cluster, 19 inhomogeneous Poisson
#  and 19 inhomogeneous Poisson cluster)  to maple and select the better ones
maple.model <- select.model2(lansing.sp$maple, sigmas=sigmas, r=r)
# show the AICc value and the fitted parameters for the best model in each class
maple.model

# Draw the empirical and theoretical models to visually asses the fitting.
# P = Poisson; HPP= heterogeneous (i.e. inhomogeneous) Poisson;
# PC = Poisson cluster; HPC=heterogeneous (i.e. inhomogeneous) Poisson cluster
plot(maple.model)

# Compute and plot envelopes for the K function according to the best fitted model.
maple.env <- envelope(maple.model, nsim=19)
plot(maple.env, sqrt(./pi)-r~r, legend=F)

# simulate 10 point patterns according to the best fitted model

maple.simu <- simulate(maple.model, nsim=10)

maple.simu

# FIt and select models to all species
lansing.models<-lapply(lansing.sp, function(x) select.model2(x, sigmas=sigmas, r=r))
lapply(lansing.models, function(x) x)

## End(Not run)
```

`teucrium` *Teucrium capitatum point pattern*

Description

Locations of *Teucrium capitatum* plants in a dwarf-shrub community in Central Spain. They are part of a more extensive dataset collected and analysed by Chacon-Labella et al. (2015). The coordinates of the plants are given in meters.

Usage

```
data(teucrium)
```

Format

An object of class "ppp" representing the point pattern of tree locations. See ppp.object for details of the format of a ppp object.

References

Chacon-Labella, J., De la Cruz, M. and Escudero, A. Beyond the classical nurse species effect: diversity assembly in a Mediterranean semiarid dwarf shrubland. *Journal of Vegetaation Science*. Accepted for publication, July 2015.

Examples

```
data(teucrium)
plot(teucrium)
```

Index