Kernel Smoothing in Spatialkernel

This web page tries to illustrate the kernel smoothing methods implemented and other useful functions for spatial point patterns analysis in the R package Spatialkernel. Two kinds of kernel smoothing methods are implemented in the R package Spatialkernel, kernel regression estimate of the *type-specific probabilities* in a multivariate Poisson point process and kernel density estimate of the intensity function of an inhomogeneous Poisson point process with *edge-correction* algorithm implemented against an arbitrary polygon area. An example also gives a brief application of the functionality in the package.

If you cannot view this web page properly, please download a pdf version of this page from Section Downloads.

Spatial Multivariate Poisson Point Process

The multivariate point process we discussed here is a inhomogeneous Poisson process that generates points in two-dimensional space with each point marked with one of several categorical types. The data are a partial realisation of the Poisson point process with categorical component processes being stochastically independent. Data of this kind are usually denoted as (x_i, m_i), where x_i are the spatial locations and m_i are the marked categorical types, within a study area of A. Without losing generalization, we suppose $m_i = 1, 2, \dots ,$ and $m_i = k$ means that the point at spatial location x_i is of kth category of types.

Spatial segregation exists if particular types of points predominate in particular regions in the study area. Spatial segregation effects can be describe in terms of the component intensity functions $\lambda_k(x)$. If $\lambda_j(x) / \lambda_k(x) = \rho_{ij}$ is a constant, then no spatial segregation exists, different categorical type of data points are randomly intermingled.

For the purpose of spatial segregation analysis, we do not have to estimate the component intensity functions. We introduce the *type-specific probabilities*, $p_k(x) = \lambda_k(x) / \sum_j \lambda_j(x)$, the conditional probability that we know a point is at location x, that point is of kth categorical type with probability $p_k(x)$. Therefore, the null hypothesis of no segregation can be described as $p_k(x) = p_k$, a constant.

Kernel Regression Estimation

A kernel regression estimator is adapted to estimate the type-specific probabilities,

$$p_k(x) = \sum_j w_h(x - x_j)\, I(m_j = k) / \sum_j w_h(x - x_j),$$

where $w_h(x) = w(x / h) / h^2$, $w(x)$ is a standard kernel function, I is the indicator function. Note that we use $p_k(x)$ for both variables and their estimators. Without causing confusion, we will use the same notations for both variables and their estimators thereafter.

Bandwidth Selection

We proposed to select a bandwidth for the kernel regression by maximizing the cross-validated log-likelihood function based on the *leave-one-out* type-specific probability estimator at data points,

$$p_k^{(i)}(x_i) = \sum_{j\neq i} w_h(x_i - x_j)\, I(m_j = k) / \sum_{j\neq i} w_h(x_i - x_j).$$

Spatial Segregation

Simulations for the Monte Carlo spatial segregation test are sampled by randomly re-labelling of the categorical marks whilst preserving the observed number of points of each categorical type. Pointwise segregation test also being carried out to mark the areas where the estimated type-specific probabilities are significantly greater or smaller than the spatial average. The test statistics chosen is a measurement of the total deviance of estimated type-specific probabilities from their typewise mean values,

$$T = \sum_k \sum_j (p_k(x_j) - p_k^{(.)})^2,$$

where $p_k^{(.)}$ is the mean of $p_k(x_j)$ over those j where $m_j = k$.

Temporal Changes

Spatial segregation analysis can be generalized to multivariate spatial-temporal Poisson point process where each point is marked with a time group (time-period) sequence number. For the spatial-temporal point process, the data are denoted as (x_i, m_i, t_i), where x_i are the spatial locations, m_i are the marked categorical types, and t_i are the time-periods. The spatial locations are within a study area of A.

Within each time-period, we can estimate the type-specific probabilities using kernel regression methods with a common bandwidth selected by the cross-validated log-likelihood functions pooled over time-periods. The kernel regression estimator of thetype-specific probabilities within each time-period is

$$p_k(x, t) = \sum_j w_h(x - x_j)\, I(m_j = k)\, I(t_j = t) / (\sum_j w_h(x - x_j)\, I(t_j = t)).$$

The null hypothesis of the temporal changes over time-periods is that the spatial patterns of the type-specific probability surfaces of each categorical type will not change over time-periods, that is, $p_k(x, t) = p_k(x)$, which is constant with respect to the time-periods t. The test statistics adopted is

$$P = \sum_t \sum_k \sum_j (p_k(x_j, t) - p_k^{(.)}(x_j))^2,$$

where $p_k^{(.)}(x_j)$ is the mean of $p_k(x_j, t)$ over time-period t. The simulations are sampled from the type-specific probability surfaces, $p_k^{(.)}(x_j)$, which are the approximate *true* type-specific probability surfaces under the null hypothesis of no temporal changes over time-periods, preserving the number of points in each time-period.

Intensity Estimation of Inhomogeneous Poisson Point Process

The intensity function of the inhomogeneous spatial Poisson point process can be estimated by the kernel density estimator,

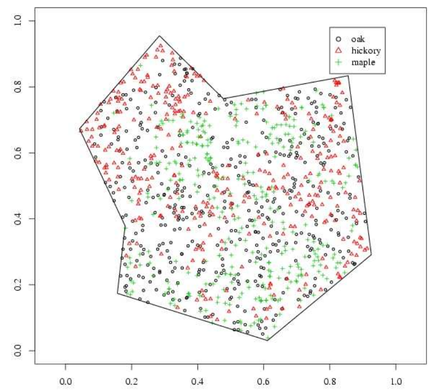
$$\lambda(x) = \sum_j w_h(x - x_j) / a_h(x),$$

where $a_h(x) = \int_A w_h(u - x)\, du$ is the edge-correction adjustment factor, proposed by Berman and Diggle (1989).

An example: The Lansing Woods Tree Data

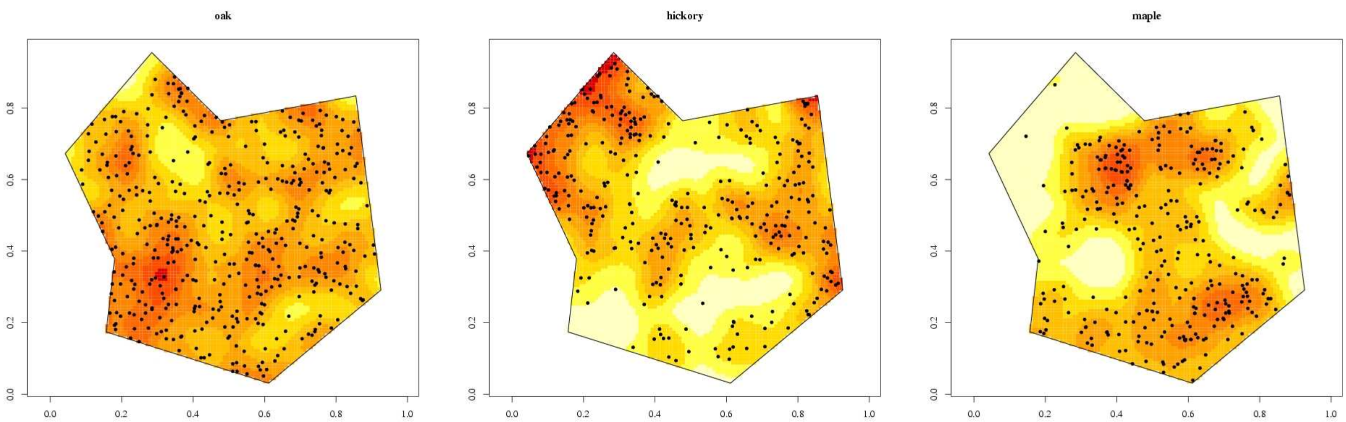
We use the Lansing Woods tree data to illustrate the basic usages of the functionality in the R package spatialkernel. We present the spatial distribution of the three different kind of trees in an arbitrary polygonal area, the estimated type-specific probabilities and the estimation of spatial segregation. We also present the estimated intensity for overall threes with edge-correction applied against the polygon boundary. Both the kernel regression estimation of the type-specific probabilities and kernel density estimation of the overall intensity use *Gaussian* kernel.

Spatial distribution of the Lansing Woods trees in an arbitrary polygonal area



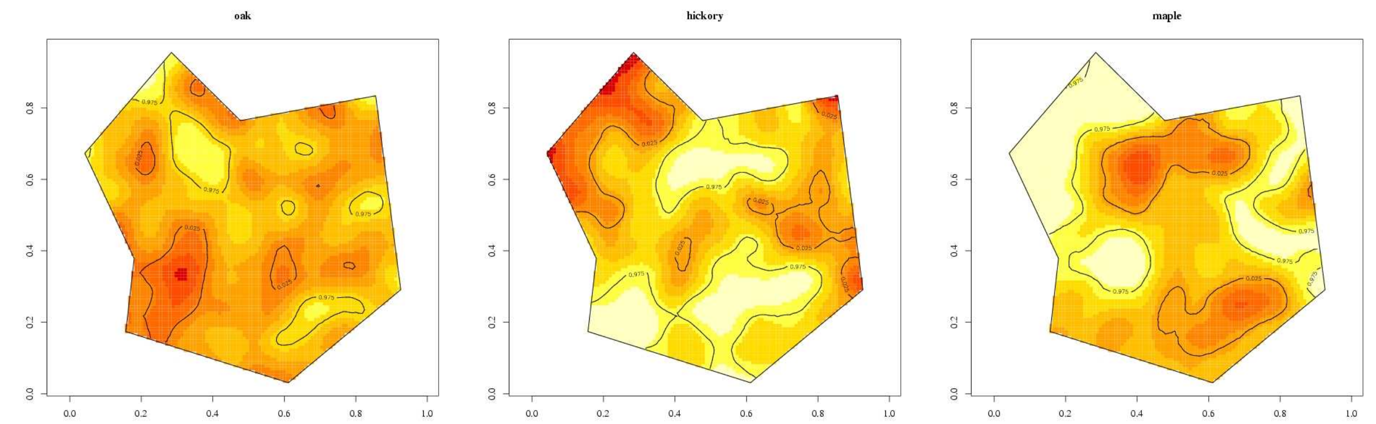
The Lansing Woods tree data we used consist of oak, hickory and maple trees, confined within a arbitrary polygonal area.

Estimated type-specific probabilities



The estimated type-specific probability surfaces show obvious spatial segregation. Each estimated type-specific probability surface shows spatial variations over the polygonal area. This is confirmed in the spatial segregation test below.

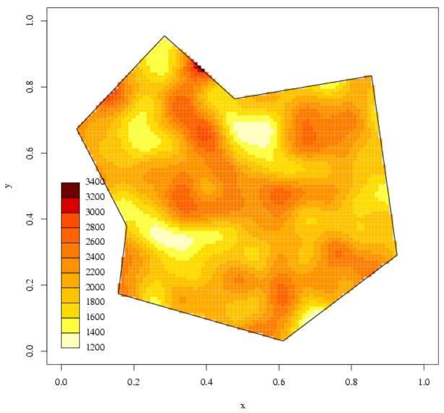
Contour lines of pointwise spatial segregation test

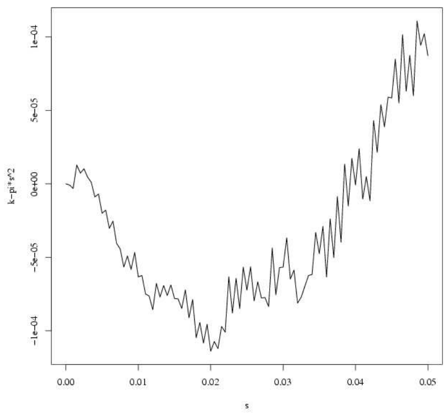


The Monte Carlo spatial segregation test gives a p-value of 0.001 in 999 simulations, which clearly reject the null hypothesis of no segregation. The contour lines show the areas where the estimated type-specific probabilities are significant great than the average (labelled with 0.025) and areas where the estimated type-specific probabilities are significant smaller than the average (labelled with 0.0975).

Estimated intensity and inhomogeneous K function

We use the kernel density method to estimate the univariate point process of overall trees and then estimate the inhomogeneous K function.





The edge-correction method proposed by Berman and Diggle (1989) is applied. The kernel density estimate of the overall intensity uses the same bandwidth selected by the cross-validated log-likelihood function in the type-specific probabilities. The estimate of the inhomogeneous K function use the estimated overall intensity.
Caution should be taken when estimate the inhomogeneous K function and the intensity using the same data. See Diggle, P.J. *et al* (2006) for a cautious note.

References

- M. Berman and P.J. Diggle (1989) Estimating weighted integrals of the second-order intensity of a spatial point process, *J. R. Stat. Soc, B*, **51**, 81–92.
- Kelsall, J.E. and Diggle, P.J. (1998) Spatial variation in risk: a nonparametric binary regression approach, *Applied Statistics*, **47**, 559–573.
- Baddeley, A.J., Møller, J. and Waagepetersen, R. (2000) Non and semi-parametric estimation of interaction in inhomogeneous point patterns, *Statistica Neerlandica,* **54***,* 3, 329–350*.*
- P. Zheng, P.A. Durr and P.J. Diggle (2004) Edge-Correction for Spatial Kernel Smoothing — When Is It Necessary? *Proceedings of the GisVet Conference 2004*, University of Guelph, Ontario, Canada, June 2004.
- Diggle, P.J., Zheng, P. and Durr, P.A. (2005) Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK. *J. R. Stat. Soc. C*, **54**, 3, 645–658.
- Diggle, P.J., V. Gómez-Rubio, P.E. Brown, A.G. Chetwynd and S. Gooding (2006) *Biometrics*, submitted.

Downloads

If your browser supports math poorly, here is a pdf version of this page. If you want to try the package, please download the latest version in source file or binary for Windows.

Thanks for reading this page.