이 학 박 사 학 위 논 문

Bandwidth Selection for Kernel Estimators of the Intensity and Pair Correlation Functions of Spatial Point Processes

공간점과정의 강도와 짝상관함수의 커널 추정치에서 띠너비 선택

2023년 2월

서울대학교 대학원

통계학과

정 양 하

Bandwidth Selection for Kernel Estimators of the Intensity and Pair Correlation Functions of Spatial Point Processes
공간점과정의 강도와 짝상관함수의 커널 추정치에서 띠너비 선택

지도교수 장 원 철

이 논문을 이학박사 학위논문으로 제출함
2022년 10월

서울대학교 대학원
통계학과
정 양 하

정양하의 이학박사 학위논문을 인준함
2022년 12월

위 원 장	김 용 대	(인)
부위원장	장 원 철	(인)
위 원	임 요 한	(인)
위 원	임 채 영	(인)
위 원	정 석 오	(인)

Bandwidth Selection for Kernel Estimators of the Intensity and Pair Correlation Functions of Spatial Point Processes

By

Yangha Chung

A Thesis
Submitted in fulfillment of the requirement
for the degree of
Doctor of Philosophy
in Statistics

Department of Statistics
College of Natural Sciences
Seoul National University
February, 2023

ABSTRACT

Bandwidth Selection for Kernel Estimators of the Intensity and Pair Correlation Functions of Spatial Point Processes

Yangha Chung
The Department of Statistics
The Graduate School
Seoul National University

In this thesis, we present an optimal bandwidth selection method for kernel estimator of the intensity function of the spatial point process and the LISA function of the pair correlation function. Particularly in estimating intensity function, we suggest a method to control smoothness with a small data size by using Bayesian bootstrap. We propose a bandwidth selection method that minimizes the mean integrated square error in the kernel estimation in the LISA function. We numerically compare our method with other existing methods and show that our method outperforms other methods in most cases. We also apply our method to two case studies.

Keywords: Kernel estimation, Bandwidth selection, Intensity function, Pair correlation function, Spatial point process

Student ID: 2015 − 20311

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this thesis, we study bandwidth selection for estimating intensity and local pair correlation functions of spatial data through kernel smoothing methods. Spatial data is quite similar with normal attribute data but has a large difference in that it has a fixed location. Geographic data is a typical example of spatial data and it includes geometric, topological or geographic properties and the dimension of spatial data is mainly two while three-dimensional properties can be used when exploring the data such as astronomical data. Since it has geographical characteristic itself, it has to be treated differently and analysing spatial data could make progress in its own field.

There are three typical spatial data types which are vector, raster and time-series data. We are interested in vector data, especially point data in this thesis. Vector data consists of discrete features on the surface of the earth as points, lines, and polygons and discrete objects like roads or locations of wildfires in a specific region are frequently represented as vector data. The point data

is most common and basic above all and the information for sites of events are mostly considerd as point data.

The behavior or realization of points is called a point pattern, and the word "point process" refers to the collection of points in such point pattern in spatial statistics. A point process is a type of stochastic process that is characterized by a set of points in space or time and there are many different types of point processes, and the specific type used can depend on applications and the nature of data. Some common types of point processes include homogeneous Poisson processes, inhomogeneous Poisson processes, Cox processes, and Gibbs point processes. Homogeneous Poisson process can be treated as a collection of completely random locations in a region and the expected number of location is constnatn. Inhomogeneous Poisson point process is a point process which the probability of finding a point in a given region depends on the location within the region. Clustered or thinned processes are originated from inhomogeneous Poisson point process and there is a generalization of inhomogeneous Poisson point process which is considered as doubly stochastic.

Furthermore, the asymptotic theory must be developed under the Poisson assumption in order to ensure the consistency of the kernel estimator of the density of event sites. However, the process in this study can also be used with non-Poisson point processes that we executed the simulation studies for both Poisson and non-Poisson point processes. Diggle (2003) stated that we cannot determine the difference between hetereogeneity and interaction in an observable point pattern without extra information,

such as variables or a parametric model. In order to test the Poisson assumption, it is usual practice in the study of spatial point patterns to first estimate the first-order intensity before predicting the second-order attributes. Therefore, the Poisson assumption is not very restrictive.

Normally, the first step to exploring the point process is to investigate the first order properties which describe an overall pattern and variability of the process and the accurate estimation of the first order intensity is quite essential. First-order properties are important because they provide information about the general characteristics of a spatial process, such as its average value and how much it varies across space. The second-order properties, also known as the spatial correlation or spatial dependence, describe the relationship between the values of the process at different locations. Second-order properties are important because they describe the spatial dependence of the process, which can have a significant impact on the analysis and modeling of the data. For example, if a spatial process has strong spatial dependence, this may indicate that the values of the process at nearby locations are highly correlated, and this could affect the accuracy of statistical estimates and predictions. On the other hand, if a spatial process has weak spatial dependence, this may indicate that the values of the process at different locations are less correlated and can be treated as independent observations. Understanding the first and second-order properties of a spatial process play an important role in accurately analyzing and modeling spatial data. When studying two properties, we usually deal with processes with some parame-

ters but nonparametric estimation is indispensable while studying the real data. So we chose kernel smoothing method and since our major concern of data is in two dimensional space, we considered bivariate kernel density estimation first and convert into spatial version.

A popular estimator for a probability density function in two dimensions is a bivariate kernel density estimator. Let $X_1, X_2, \cdots, X_N$ be the points in $\mathbb{R}^2$, then the bivariate kernel density estimator is

$$\hat{f}_{\mathbf{H}}(x) = \frac{1}{N}\sum_{i=1}^{N} K_{\mathbf{H}}(x - X_i) = \frac{1}{N|\mathbf{H}|^{-1/2}}\sum_{i=1}^{N} K\left(\mathbf{H}^{-1/2}(x - X_i)\right)$$

where $\mathbf{H}$ is a 2×2 symmetric and positive definite bandwidth matrix and K is an arbitrary bivariate kernel function. We can consider two types of the bandwidth matrices – diagonal and full. The class of symmetric and positive definite matrices is denoted by $\mathcal{F}$ and $\mathcal{D}$ a subset of $\mathcal{F}$ which is a set of diagonal matrices . Bandwidth matrices in each type can be expressed as below.

- $\mathbf{H} = \begin{pmatrix} h_1^2 & 0 \\ 0 & h_2^2 \end{pmatrix}$ if $\mathbf{H} \in \mathcal{D}$
- $\mathbf{H} = \begin{pmatrix} h_{11} & h_{21} \\ h_{21} & h_{22} \end{pmatrix}$ if $\mathbf{H} \in \mathcal{F}$

The rest of this thesis is organized as follows. Chapter 2 introduces an intensity version of kernel estimation and the bandwidth selection procedures with several numerical studies. LISA function of pair correlation function is a main theme in Chapter 3. We wrap up the thesis in Chapter 4 with concluding remarks and the proofs of the theorems and lemmas are in Appendix.

Chapter 2

Kernel Intensity Estimation

2.1 Overview

We introduce an motivating example of spatial data and how we utilize it.

Figure 2.1 is from Diggle (1990) and it shows the locations of lung and larynx cancer cases in Chorley, England. With this information, we would like to compare the population distribution of lung with that of larynx cancer cases. In spatial statistics, the distribution of locations or events – which involves absolute values rather than relative ones – is what we are most interested in. When analyzing spatial data, there is a concept of mean called intensity that is the average number of events per unit area. To be precise, "first-order intensity" is more precise expression, but the abbreviated term is in common use.

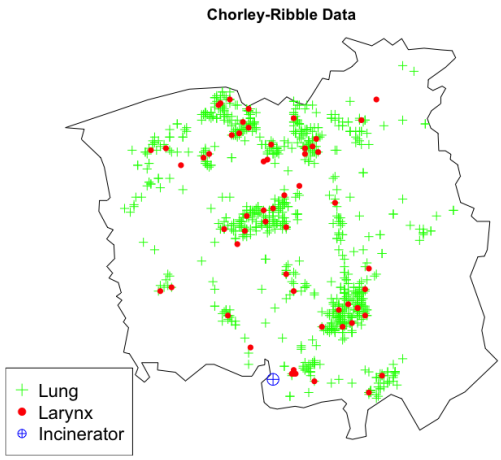


Figure 2.1: Locations of lung and larynx cancer cases and disused incinerator

Definition 1. *(Intensity)*

Let $W \subset \mathbb{R}^2$ and X be a spatial point process defined in W. $(X_1, \cdots, X_N)$ is a realization of X which is observed on a bounded region . If dx is an infinitesimal region with point $x \in \mathbb{R}^2$ as its center, the first-order intensity function of X (Diggle, 2003) is defined as

$$\lambda(x) = \lim_{|dx| \to 0} \frac{\mathsf{E}[N(dx)]}{|dx|}$$

where $|dx|$ and $N(dx)$ denote the area and the number of events of X in dx, respectively.

The intensity is comparable to the probability density function, but there are some distinctions between them. The intensity function is not a normalized property. As we can see from the definition of intensity, the value of intensity is related to "how many" events we expect to occur at a particular location. Another difference is

that the randomness of number of points. Recall we can count how many points in a point process. In general, the number of points in a bounded region acts as a random variable and it follows a Poisson distribution. Furthermore, when we consider choosing points from a density, they are chosen based on the likelihood so the locations themselves are independent. However, in spatial statistics, the locations of events are not necessarily independent.

Although intensity is somewhat different with density, the basic approach to analysis is quite similar, which is the kernel smoothing method. Kernel smoothing is a popular nonparametric method and it is mainly used in density estimation. Kernel density estimation and kernel intensity estimation are two closely related techniques used in statistics to estimate the underlying probability density function or intensity function of a random variable from a sample of observations. Both techniques involve using a kernel function to smooth the data and obtain an estimator of the probability density function or intensity function. The primary distinction between kernel density estimation and kernel intensity estimation is that the former is used to compute the probability distribution function of a continuous random variable, whereas the latter calculates the intensity function of a point process, a particular kind of random process that is defined by a collection of points in space or time. Diggle (1985) suggested the kernel intensity estimator of one-dimensional point process as follows

$$\hat{\lambda}_h(x) = \frac{1}{p_h(x)} \sum_{i=1}^{N} K_h(x - X_i)$$
$$= \frac{1}{p_h(x)h^2} \sum_{i=1}^{N} K\left(\frac{x - X_i}{h}\right)$$

where $K(\cdot)$ is a kernel function and $h > 0$ is the smoothing parameter, K_h denotes the smoothed kernel and $p_h(x) = \int_W \frac{1}{h^2} K\left(\frac{x-X_i}{h}\right) dy$ is the edge-correction term. There are many other kinds of edge-correction terms but the mentioned one is the most popular. A kernel function K should be real-valued, non-negative, integrable and it is preferable to define the function to meet two extra requirements for the majority of applications which are normalization – $\int_{-\infty}^{\infty} K(u)du = 1$ – and symmetry – $K(-u) = K(u)$ for all u.

All functions which satisfies the conditions can be used as a kernel function. There are several types of kernel function which are used popularly: Gaussian, Epanechnikov, uniform, triweight and so on.

Although it has acquired less attention, bandwidth selection in kernel estimation of spatial point processes is a crucial component of kernel estimation. The bandwidth selection for density estimation providing least-squares cross-validation (LSCV) (Silverman, 1986) is the same as that provided by the kernel intensity estimator, as demonstrated by Diggle and Marron (1988) for stationary Cox processes in $\mathbb{R}$, and Berman and Diggle (1989) introduced a data-driven method to choose this bandwidth. For inhomogeneous point processes, this equivalence has not yet been established. Brooks and Marron (1991) showed that the LSCV

bandwidth is asymptotically optimal in case of inhomogeneous point processes in $\mathbb{R}$. According to Taylor (1989), Cao (1993), and Cao et al. (1994), the approach to employ smooth bootstrap procedures can outperform cross-validation to determine the optimal bandwidth for the kernel estimators for density and hazard rate. In $\mathbb{R}$, Cowling et al. (1996) suggested a resampling procedure called smooth bootstrap to obtain the confidence region for kernel intensity estimator of inhomogeneous point processes and Loh and Jang (2010) used nonparametric bootstrap to determine the optimal bandwidth for estimating the two-point correlation function in the point process framework.

To discuss the estimator, the asymptotic framework is indispensable. The specific situation which needs the asymptotic framework will be introduced in Chapter 2. There are two types of asymptotic framework: increasing domain asymptotics and infill asymptotics. Increasing domain asymptotics is used when dealing with stochastic processes at an increasing number of sites such that any two sites are at least a fixed distance apart. Also, the observation region eventually becomes unbounded as $n \to \infty$ so it can be useful in epidemiology. The process over a regular lattice is another example for using increasing domain asymptotics. On the other hand, infill asymptotics – which is our concern in this thesis – is used when an observation region is necessarily bounded and more and more samples are taken from the given region. The minimum distance between the data-sites tends to 0 as $n \to \infty$, infill asymptotics often entails examining the behavior of estimators in the limit as the sample size approaches infinity. This frame-

work is suitable for geostatistical applications. The infill asymptotic framework, used by Diggle and Marron (1988), states that the expected number of events approaches to infinity within $\mathbb{R}^2$ or a bounded observation domain W, that is, $\int_{\mathbb{R}^2} \lambda(x)dx \to \infty$ or $\int_W \lambda(x)dx \to \infty$.

For kernel intensity estimation, we can take into consideration bandwidth matrices and define the kernel intensity estimator (Wand, 1992; Wand and Jones, 1994); Duong and Hazelton, 2003) as

$$\begin{aligned}\hat{\lambda}_{\mathbf{H}}(x) &= \frac{1}{p_{\mathbf{H}}(x)} \sum_{i=1}^{N} K_{\mathbf{H}}(x - X_i) \\ &= \frac{1}{p_{\mathbf{H}}(x)|\mathbf{H}|^{-1/2}} \sum_{i=1}^{N} K\left(\mathbf{H}^{-1/2}(x - X_i)\right)\end{aligned}$$

where $\mathbf{H}$ denotes the bandwidth matrix which should be symmetric and positive-definite and $|\mathbf{H}|$ is the determinant of $\mathbf{H}$. This kernel intensity estimator is asymptotically unbiased but not consistent, it acts as a major demerit for using as an estimator.

This problem can be addressed by considering the relationship between first-order intensity of a spatial point process and bivariate density . As abovementioned, we are dealing with Poisson point processes and N is the given number of events which follows $\mathsf{Poisson}(\int_W \lambda(x)dx) = \mathsf{Poisson}(m)$. Cucala (2006) introduced a concept of "density of event locations" as below.

$$f(x) = \frac{\lambda(x)}{\int_W \lambda(x)dx} = \frac{\lambda(x)}{m} = \lambda_0(x)$$

Then, the kernel estimator of $\lambda_0(x)$ can be represented as

$$\begin{aligned}\hat{\lambda}_{0,\mathbf{H}}(x) &= \frac{\hat{\lambda}_{\mathbf{H}}(x)}{N}\mathrm{I}(N \neq 0) \\ &= \frac{1}{N \cdot p_{\mathbf{H}}(x)|\mathbf{H}|^{-1/2}} \sum_{i=1}^{N} K\left(\mathbf{H}^{-1/2}(x - X_i)\right)\mathrm{I}(N \neq 0)\end{aligned} \tag{2.1}$$

where $\mathrm{I}(\cdot)$ stands for an indicator function. This indicator term is due to the randomness of the number of events in Poisson point processes.

The performance of the kernel estimator must then be measured, with the mean integrated squared error(MISE) being the most commonly used metric. It can be expressed as the sum of integrated squared bias term and integrated variance term as follows.

$$\begin{aligned}\mathrm{MISE}(\mathbf{H}) &= \mathbf{E}\left(\int_W \left(\hat{\lambda}_{0,\mathbf{H}}(x) - \lambda_0(x)\right)^2 dx\right) \\ &= \int_W B(x,\mathbf{H})^2 dx + \int_W \mathbf{Var}(x,\mathbf{H}) dx\end{aligned}$$

$\mathbf{E}$ represents the expectation over randomness in both location and number of events, and $B(x,\mathbf{H})$ and $\mathbf{Var}(x,\mathbf{H})$ denote the bias and variance of $\hat{\lambda}_{0,\mathbf{H}}(x)$, respectively. Here we need infill asymptotic framework to achieve the asymptotic version of MISE($\mathbf{H}$) which is denoted as AMISE($\mathbf{H}$). We find an optimal bandwidth matrix $\mathbf{H}$ by minimizing MISE($\mathbf{H}$) or AMISE($\mathbf{H}$). The relationship between MISE($\mathbf{H}$) and AMISE($\mathbf{H}$) is as below.

$$\mathrm{MISE}(\mathbf{H}) = \mathrm{AMISE}(\mathbf{H}) + o\left(A(m)|\mathbf{H}|^{-1/2} + tr(\mathbf{H})\right)$$

Here

$$\begin{aligned}\text{AMISE}(\mathbf{H}) &= \frac{1}{4}\mu_2(K)^2 \int_{\mathbb{R}^2} tr^2(\mathbf{H}D^2\lambda_0(x))dx + A(m)|\mathbf{H}|^{-1/2}R(K) \\ &= \frac{1}{4}\mu_2(K)^2 (vech^t\mathbf{H})\Psi_4(vech^t\mathbf{H}) + A(m)|\mathbf{H}|^{-1/2}R(K)\end{aligned} \tag{2.2}$$

where $R(K) = \int K^2(x)dx$, $\int_{\mathbb{R}^2} uu^T K(u)du = \mu_2(K)I_2$,$D^2\lambda_0(x)$ is a Hessian matrix of λ_0, *vech* is a half-vectorization of matrix and $A(m) = \mathbf{E}\left(\frac{1}{N}\mathrm{I}(N \neq 0)\right)$. Note that $A(m) = e^{-m}\sum_{k=1}^{\infty}\frac{m^k}{kk!} < e^{-m}\sum_{k=0}^{\infty}\frac{2m^k}{(k+1)!} = \frac{2}{m} \to 0$ as $m \to \infty$. Ψ_4 is the 3×3 matrix whose entities are the integrated density derivative functional. (See details in Appendix 1.) It is known that there is no such closed form for MISE($\mathbf{H}$) in general so we can use AMISE($\mathbf{H}$) in that case. In addition, we can only achieve the minimization of AMISE($\mathbf{H}$) by a numerical way. We must acquire an estimator of AMISE($\mathbf{H}$) to use the MISE as an error criterion to choose the proper bandwidth.

To attain fine results through this kernel estimating process, we have to set three assumptions on $\mathbf{H}, \lambda, K$.

1. $\mathbf{H}$ should be symmetric and positive-definite and all entries of $\mathbf{H} \to 0$ and $m^{-1}|\mathbf{H}|^{-1/2} \to 0$, as $m \to \infty$.

2. The partial derivatives of λ are up to order four and the 2nd and 4th partial derivatives are all bounded, square integrable and continuous.

3. The kernel function is a symmetric, square integrable, continuous density function such that $\int_{\mathbb{R}^2} uu^T K(u)du = \mu_2(K)I_2$ with $\mu_2(K) < \infty$

These are called as regularity conditions and it can be understood that these are an intensity version of regularity conditions for ker-

nel density estimation. We assume that the functions in this thesis satisfy the regularity conditions and note that the types of bootstrap kernel estimators in this thesis are all used for minimizing MISE and the optimal bandwidths from them is substituted into the equation 2.1 . Now we introduce an estimator of MISE which uses bootstrap procedure.

2.2 Smoothed bootstrap and Bayesian bootstrap

In general, "bootstrapping" means a nonparametric bootstrap with equal size. Nonparametric bootstrap techniques produce copies of a dataset and they enable inference and the creation of goodness-of-fit tests in a variety of statistical applications. there are many other types of bootstrapping such as smoothed bootstrap, Bayesian bootstrap, parametric bootstrap, block bootstrap and so on. The application of bootstrap techniques to spatial or temporal data including dependent observations has been quite successful when applied to independent data. For instance, Guan and Loh (2007) used bootstrap to fit models to stationary point patterns, and Lahiri et al. (1999) used it to derive inferences about the spatial cumulative distribution function. Marked-points bootstrap was employed by Loh and Stein (2004), Loh (2010), and Loh and Jang (2010) and they determined the bandwidth for the nonparametric estimator of the two-point correlation function. These studies introduced goodness-of-fit test for the K-function of homogeneous and inhomogeneous point patterns. K-function is related to second order

property of point processes and will be treated in Chapter 3. In this section, we introduce smoothed bootstrap and Bayesian bootstrap.

2.2.1 Smoothed bootstrap

In order to extend the smooth bootstrap approach for inhomogeneous Poisson point processes in $\mathbb{R}$ by Cowling et al. (1996), Fuentes-Santos et al. (2016) brought the smooth bootstrap procedure to $\mathbb{R}^2$ which estimates the MISE of the kernel estimator of the "density of event locations". Both two methods below explain the smoothed bootstrap sampling procedure equivalently. (Devroye and Györfi, 1985)

Method 1

1. Conditional on $(X_1, \cdots, X_N)$, let $N^* \sim \mathsf{Poisson}\left(\int_W \hat{\lambda}_{\mathbf{G}}(x)dx\right)$
2. Choose $(X_1^*, \cdots, X_{N^*}^*)$ by random sampling with replacement N^* times from the distribution with density $\hat{\lambda}_{\mathbf{G}}(x)/\int_W \hat{\lambda}_{\mathbf{G}}(x)dx$.

Method 2

1. Conditional on $(X_1, \cdots, X_N)$, let $N^* \sim \mathsf{Poisson}\left(\int_W \hat{\lambda}_{\mathbf{G}}(x)dx\right)$
2. The point pattern resampling is defined as follows:

   $X_i^* = Y_i^* + Z_i^*$, $i = 1, \cdots, N^*$ where Y_i^* is chosen by random sampling with replacement from $(X_1, \cdots, X_N)$ and Z_i^*s are independent and identically distributed with a bivariate density function $K(\cdot)$ with a pilot smoothing matrix $\mathbf{G}$

Refer to Cowling et al. (1996), the kernel function and bandwidth matrix used in the step of smoothing can be different with those used to estimate $\lambda(x)$. For the convenience of computation Gaussian kernel function is commonly used, that is, $Z^* \sim N(0, \mathbf{G})$.

Fuentes-Santos et al. (2016) suggested a consistent bootstrap kernel intensity estimator defined as follows. Let $(X_1^*, X_2^*, \cdots, X_{N^*}^*)$ be a bootstrap sample from the realization $(X_i)_{i=1}^N$ on $W \subset \mathbb{R}^2$. Then, Equation 2.1 with bootstrap would be expressed as

$$\begin{aligned}\hat{\lambda}^*_{0,\mathbf{H}}(x) &= \frac{\hat{\lambda}^*_{\mathbf{H}}(x)}{N^*}\mathrm{I}(N^* \neq 0) \\ &= \frac{1}{N^* \cdot p_{\mathbf{H}}(x)|\mathbf{H}|^{-1/2}} \sum_{i=1}^{N^*} K\left(\mathbf{H}^{-1/2}(x - X_i^*)\right)\mathrm{I}(N^* \neq 0)\end{aligned}$$

and when we let $\hat{\lambda}_{0,\mathbf{G}}(x)$ be the kernel estimator which is used to generate the bootstrap pattern, then

$$\begin{aligned}\mathrm{MISE}^*(\mathbf{H}) &= \mathbf{E}^*\left(\int_W \left(\hat{\lambda}^*_{0,\mathbf{H}}(x) - \hat{\lambda}_{0,\mathbf{G}}(x)\right)^2 dx\right) \\ &= \int_W B^*(x,\mathbf{H})^2 dx + \int_W \mathbf{Var}^*(x,\mathbf{H}) dx\end{aligned}$$

would be an expression of bootstrap MISE and we substitute $\arg\min_{\mathbf{H}} \mathrm{MISE}^*(\mathbf{H})$ to $\hat{\lambda}_{0,\mathbf{H}}(x)$. Some might be confused as to why we use $\hat{\lambda}_{0,\mathbf{H}}(x)$ rather than $\hat{\lambda}^*_{0,\mathbf{H}}(x)$ itself. We want to find an approximation for the true "density of event locations" function $\lambda_0(x)$ which we can never know. So we assume the hypothetical function $\hat{\lambda}_{0,\mathbf{G}}(x)$ by using an auxiliary bandwidth matrix $\mathbf{G}$ to act as a true function. If $\mathbf{H}$ makes a good approximation $\hat{\lambda}^*_{0,\mathbf{H}}(x)$ for the hypothetical function $\hat{\lambda}_{0,\mathbf{G}}(x)$ and also $\hat{\lambda}_{0,\mathbf{G}}(x)$ is assumed to be quite similar with $\lambda_0(x)$, then $\mathbf{H}$ would generate a nice approximation $\hat{\lambda}_{0,\mathbf{H}}(x)$ for $\lambda_0(x)$.

Fuentes-Santos et al. (2016) also proved consistency of bootstrap estimator and also suggested an ℓ-stage plug-in bandwidth algorithm. Bandwidth selector will be described in detail in Chapter 2.3.

2.2.2 Bayesian bootstrap

Bayesian bootstrap was first introduced by Rubin (1981). To state briefly, Bayesian bootstrap uses continuous weight instead of discrete weight as in nonparametric bootstrap. The Bayesian bootstrap is based on Bayesian statistics, which is a framework for modeling statistical uncertainty that allows for the incorporation of prior knowledge and the incorporation of uncertainty in statistical estimates.

To use the Bayesian bootstrap, one first needs to specify a prior distribution for the statistic or model parameter of interest. The prior distribution represents the researcher's initial beliefs about the value of the statistic or parameter. The weight vector follows Dirichlet distribution with hyperparameter α, and strong belief of given data makes α larger. The researcher can then use the Bayesian bootstrap to construct a set of simulated values for the statistic or parameter by sampling from the dataset and from the prior distribution. These simulated values can then be used to calculate the posterior distribution, which indicates the researcher's revised opinions about the value of the statistic or parameter after considering the data.

The Bayesian bootstrap has the advantage of being able to assess the uncertainty of a statistic or model in the situation of

small sample size or unknown fundamental distribution. When the assumptions of classic statistical approaches are not met, it can also be used to quantify the uncertainty of a statistic or model. We propose a method which uses both the smoothed bootstrap and Bayesian bootstrap.

2.2.3 Proposed method

Let $(X_i)_{i=1}^N$ be the realization of the inhomogeneous spatial Poisson point process X with intensity $\lambda(x)$ on a region $W \subset \mathbb{R}^2$ and $(X_1^*, X_2^*, \cdots, X_{N^*}^*)$ be a bootstrap sample based on a smoothing matrix $\mathbf{G}$. Then, the smoothed and Bayesian bootstrap estimator of "density of event locations" is represented as

$$\hat{\lambda}_{0,\mathbf{H}}^B(x) = \frac{1}{p_{\mathbf{H}}(x)} \sum_{i=1}^{N^*} W_i K_{\mathbf{H}}(x - X_i^*) \mathrm{I}(N^* \neq 0)$$

where $\mathbf{H}$ is the symmetric and positive definite bandwidth matrix and $W^B = (W_i)_{i=1}^{N^*} \sim \mathsf{Dir}(\alpha \cdot \mathbf{1}_{N^*})$ is the weight with $\mathbf{1}_{N^*}$ is an N^*-vector of ones which is independent of the data. $p_{\mathbf{H}}(\mathbf{x}) = \int_W K_{\mathbf{H}}(x - y)dy$ is an edge-correction term. As we mentioned above, the weight vector shows the belief and unless we do not have different belief within the given data points, we give same weight for each points in smoothed bootstrap pattern. Hereafter in this thesis, for brief expression, we let W is large enough in order to make $p_{\mathbf{H}}(x) \approx 1$ and leave the term out.

By the law of total expectation and from $\mathbf{E}(W_i) = \frac{1}{N}$ and

$N \sim \mathsf{Poi}(\int_W \lambda(x)dx) = \mathsf{Poi}(m)$,

$$\begin{aligned}
\mathbf{E}(W_i^2) &= \frac{2}{N(N+1)} \\
\mathbf{E}(W_i W_j) &= \mathbf{Cov}(W_i, W_j) + \mathbf{E}(W_i)\mathbf{E}(W_j) \\
&= -\frac{1}{N^2(N+1)} + \frac{1}{N^2} \\
&= \frac{1}{N(N+1)}
\end{aligned}$$

for $i \neq j$.

Proposition 2.2.1. *(Expectation)*

$$\mathbf{E}\ \hat{\lambda}^B_{0,\boldsymbol{H}}(x) = (1 - e^{-m})(K_{\boldsymbol{H}} * \hat{\lambda}_{0,\boldsymbol{G}})(x)$$

Proposition 2.2.2. *(Variance)*

$$\begin{aligned}
\mathbf{Var}\left(\hat{\lambda}^B_{0,\boldsymbol{H}}(x)\right) &= 2C_\alpha(\hat{m})(K^2_{\boldsymbol{H}} * \hat{\lambda}_{0,\boldsymbol{G}})(x) \\
&\quad - (2C_\alpha(\hat{m}) - e^{-\hat{m}} + e^{-2\hat{m}})(K_{\boldsymbol{H}} * \hat{\lambda}_{0,\boldsymbol{G}})^2(x)
\end{aligned}$$

Then,

$$\begin{aligned}
\mathrm{MISE}^B(\mathbf{H};\alpha) &= \mathbf{E}\left(\int_W \left(\hat{\lambda}_{0,\mathbf{H}}(x) - \hat{\lambda}_{0,\mathbf{G}}(x)\right)^2 dx\right) \\
&= \int_W B^B(x,\mathbf{H})^2 dx + \int_W \mathbf{Var}^B(x,\mathbf{H})dx \\
&= \mathrm{AMISE}(\mathbf{H};\alpha) + o_P\left(C_\alpha(\hat{m})|\mathbf{H}|^{-1/2} + tr^2(\mathbf{H})\right)
\end{aligned}$$

where

$$\mathrm{AMISE}(\mathbf{H}) = \frac{1}{4}\mu_2(K)^2 (vech\mathbf{H})^t \hat{\Psi}_{4,\mathbf{G}} (vech\mathbf{H}) + C_\alpha(\hat{m})|\mathbf{H}|^{-1/2} R(K)$$

Note that $C_\alpha(m)$ can be estimated by $C_\alpha(\hat{m}) = \frac{1}{\alpha N + 1}$

Theorem 2.2.3. *Let X be an inhomogeneous spatial Poisson point process and $\hat{\lambda}^B_{0,\boldsymbol{H}}(\boldsymbol{x})$ be the smoothed and Bayesian bootstrap MISE minimizer of the "density of event locations". Assuming the regularity conditions,*

$$MISE^B(\boldsymbol{H};\alpha) = AMISE^B(\boldsymbol{H};\alpha) + o_p\left(C_\alpha(\hat{m})|\boldsymbol{H}|^{-1/2} + tr^2(\boldsymbol{H})\right)$$

where

$$AMISE^B(\boldsymbol{H};\alpha) = \frac{1}{4}\mu_2(K)^2(vech\boldsymbol{H})^t\hat{\Psi}_{4,\boldsymbol{G}}(vech\boldsymbol{H}) + C_\alpha(\hat{m})|\boldsymbol{H}|^{-1/2}R(K)$$

$\to 0$ *when* $m \to \infty$.

Therefore, $\hat{\lambda}^B_{0,\boldsymbol{H}}(\boldsymbol{x})$ is a consistent MISE estimator of $\hat{\lambda}_{0,\boldsymbol{G}}(\boldsymbol{x})$. Moreover, $AMISE^B(\boldsymbol{H};\alpha)$ is a consistent estimator of $AMISE(\boldsymbol{H})$ if $\Psi_{4,\boldsymbol{G}}$ is a consistent estimator of Ψ_4.

Proof. The proof is in Appendix A.3. □

We then introduce how to achieve the optimal bandwidth matrix specifically.

2.3 Bandwidth selector

Refer to Silverman (1986), the most important part in kernel estimating process is selecting an appropriate bandwidth rather than selecting a nice kernel function. For this reason, Epanechnikov kernel, which is known to be optimal among the kernel functions, is not the most common kernel function. Gaussian kernel function is the most common since it makes various computations easier and we also use Gaussian kernel function in this thesis.

In the previous section, we introduced MISE which we have to minimize to obtain an optimal bandwidth. Recall equation 2.2 and we can notice that this equation needs a property about the true intensity function which is unknown. So we need to discover an equation that can replace MISE and compute in real. The types of bandwidth selectors are categorized by what objective equation we use or what constraints the bandwidth might have.

One of the simplest bandwidth is normal scale bandwidth or 'rule of thumb' selector which is

$$\hat{\mathbf{H}}_{NS} = \left(\frac{4}{d+2}\right)^{2/(d+4)} n^{-2/(d+4)} S$$

where d is the dimension of data and S is the sample variance.Since the normal density is one of the smoothest densities accessible, the normal scale selector produces bandwidths that result in over-smoothing for non-normal data (Chacón (2018)).

There are numerous additional selectors, but the two most commonly used are plug-in bandwidth selector and cross-validation bandwidth selector which we are interested in this thesis.

2.3.1 Plug-in bandwidth selector

Wand and Jones (1994), who extended univariate methods of Sheather and Jones (1991), were the first to provide plug-in bandwidth selectors for multivariate data for constrained matrices. The present method for unconstrained bandwidth matrices was developed by Duong and Hazelton (2003) and modified by Chacón and Duong (2010).

The plug-in bandwidth selector achieves optimal bandwidth matrix $\mathbf{H}$ by minimizing AMISE. We can only minimize AMISE

in a numerical way so in `R` program, we use `optim` function to find optimal bandwidth $\mathbf{H}$. `Hpi` may be utilized, and we can see that the function ultimately used `optim` by looking inside the code. Recall AMISE

$$\text{AMISE}(\mathbf{H}) = \frac{1}{4}\mu_2(K)^2 (vech\mathbf{H})^t \hat{\Psi}_{4,\mathbf{G}} (vech\mathbf{H}) + C_\alpha(\hat{m})|\mathbf{H}|^{-1/2} R(K) \tag{2.3}$$

and we can notice that we require an auxiliary bandwidth matrix $\mathbf{G}$ which is involved in computing density derivative functional matrix. We refer to $\mathbf{G}$ as a pilot bandwidth matrix.

Pilot bandwidth matrix

Many data-driven strategies for selecting a kernel density estimator's bandwidth rely on unknown constants related to auxiliary bandwidths that while estimating the functionals. Typically, these constants are either replaced by the equivalent constants for some reference distribution or estimated. (Park and Marron, 1992)

In order to substitute $\hat{\Psi}_{4,\mathbf{G}}$, we should take into account $\hat{\psi}_{r,\mathbf{G}} = \mathbf{E}\lambda_{0,\mathbf{G}}^{(r)}(\mathbf{x})$. Since the normal $\mathbf{G}$ is quite complicated, we suggest a scalar pilot bandwidth g. We let $\mathbf{G} = g^2 I_2$ and then now we need to know how to obtain an optimal g. The common method is to minimize SAMSE, the sum of the asymptotic mean squared errors, of $\hat{\Psi_{4,g}}$ and we call this g an SAMSE-optimal bandwidth.

$$\begin{aligned}\text{AMSE}\left(\hat{\psi_{r,g}}\right) &= 2n^{-2}g{-}2|r| - 2\psi_0 R(K^{(r)}) \\ &\quad + \left(n^{-1}g^{-|r|-2}K^{(r)}(0) + \frac{g^2}{2}\mu_2(K)\sum_{j=1}^{2}\psi_{r+2e_i} \right)^2\end{aligned}$$

Obtain SAMSE-optimal bandwidth by minimizing SAMSE($\hat{\Psi}_{4,\mathbf{G}}$) is

$$g_{SAMSE} = \left(\frac{24A_2}{(-4A_3 + \sqrt{16A_3^2 + 48A_2A_4})n} \right)^{1/8}.$$

Details are in Appendix A.4.

Back to the plug-in bandwidth selector, we introduce two-stage plug-in bandwidth selector algorithm. Aldershof (1991), Park and Marron (1992), Wand and Jones (1994), and Tenreiro (2003) all advocate the optimal stages of kernel functional estimation is two, theoretically and empirically.

Algorithm 1 Two-stage plug-in bandwidth selector (Chacón (2018)

Input: $\{X_1, \cdots, X_N\}$ Output: $\hat{\mathbf{H}}_{\mathrm{PI}}$

1: Compute sixth order SAMSE-optimal pilot bandwidth

$$\hat{\mathbf{G}}_6 := g_{6,\mathrm{SAMSE}}^2 I_2$$

2: Compute sixth order kernel functional estimate $\hat{\Psi}_6(\hat{\mathbf{G}}_6)$

(/* Stage 1 */)

3: Plug $\hat{\Psi}_6(\hat{\mathbf{G}}_6)$ into formula for pilot bandwidth $\hat{\mathbf{G}}_4$

$$\hat{\mathbf{G}}_4 := g_{4,\mathrm{SAMSE}}^2 I_2$$

4: Compute fourth order kernel functional estimate $\hat{\Psi}_4(\hat{\mathbf{G}}_4)$

(/* Stage 2 */)

5: $\hat{\mathbf{H}}_{\mathrm{PI}} :=$ minimizer of $\mathrm{AMISE}^B(\mathbf{H}; \hat{\mathbf{G}}_4)$

We can apply this algorithm to equation 2.3 to obtain optimal bandwidth. Meanwhile, if we consider only about the diagonal bandwidth matrix we can obtain it as a closed form. Rewrite AMISE as below.

$$\mathrm{AMISE}^B(\mathbf{H}; \alpha) = \frac{1}{4}\mu_2(K)^2 \left(h_1^4\psi_{40} + 2h_1^2h_2^2\psi_{22} + h_2^4\psi_{04}\right) + \frac{C_\alpha(\hat{m})}{h_1 h_2} R(K)$$

Let

$$\hat{\mathbf{H}}^B(\alpha) := \underset{\mathbf{H}\in\mathcal{D}}{\arg\min}\, \mathrm{AMISE}^B(\mathbf{H}; \alpha) = \begin{pmatrix} h_{1,\mathrm{AMISE}} & 0 \\ 0 & h_{2,\mathrm{AMISE}} \end{pmatrix},$$

then we can achieve optimal bandwidth matrix as a closed form for diagonal bandwidth matrix with plug-in procedure.

Proposition 2.3.1. *$h_{1,AMISE}$ and $h_{2,AMISE}$ can be expressed as*

$$h_{1,AMISE} = \left[\frac{\psi_{04}^{3/4} C_\alpha(m) R(K)}{\mu_2(K)^2 \psi_{40}^{3/4} (\psi_{40}^{1/2} \psi_{04}^{1/2} + \psi_{22})} \right]^{1/6}$$

$$h_{2,AMISE} = \left[\frac{\psi_{40}^{3/4} C_\alpha(m) R(K)}{\mu_2(K)^2 \psi_{04}^{3/4} (\psi_{40}^{1/2} \psi_{04}^{1/2} + \psi_{22})} \right]^{1/6}$$

and h's decrease as α increases.

The proof is in Appendix A.5.

2.3.2 Smoothed cross-validation bandwidth selector

Contrary to plug-in bandwidth selector, there exists another class to achieve bandwidth matrix $\mathbf{H}$ by pre-smoothing integrated squared bias part. The class of these method is called 'cross-validation' class. We can minimize both MISE and AMISE in this class. Unbiased cross-validation(UCV) method is what we minimize MISE and biased cross-validation(BCV) method is what we minimize AMISE. Smoothed cross-validation bandwidth is somewhat a hybrid of UCV and BCV. Though no estimator of the integrated density functional Ψ_4 is required in cross, the computation of double sums is substantially complicated.

The goal of this criterion is to estimate the accurate integrated squared bias(ISB) by replacing the true intensity λ_0 by a pilot kernel intensity estimator $\tilde{\lambda}_{0,\mathbf{G}}(z) = \frac{1}{n} \sum_{i=1}^{n} L_G(z - Z_i)$ and adding the dominant term of the integrated variance(IV) gives in

$$\begin{aligned} \mathrm{SCV}(\mathbf{H}; \mathbf{G}) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (K_{\mathbf{H}} * K_{\mathbf{H}} - 2K_{\mathbf{H}} + K_0) * L_{\mathbf{G}} * L_{\mathbf{G}}(Z_i - Z_j) \\ + \frac{1}{n} |\mathbf{H}|^{-1/2} R(K) \end{aligned}$$

as the SCV estimator of (A)MISE and $\hat{\mathbf{H}}_{\mathrm{SCV}} = \arg\min \mathrm{SCV}(\mathbf{H}; \mathbf{G})$. L can be different with K, but to ease the computation, we let $K = L = \Phi$. Then,

$$\mathrm{SCV}(\mathbf{H}; \mathbf{G}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\Phi_{2\mathbf{H}+2\mathbf{G}} - 2\Phi_{\mathbf{H}+2\mathbf{G}} + \Phi_{2\mathbf{G}})(X_i - X_j) + \frac{1}{N}(4\pi)^{-1}|\mathbf{H}|^{-1/2}$$

Applying our proposed methods to the upper equation brings

$$\mathrm{SCV}^B(\mathbf{H}; \mathbf{G}, \alpha) = \frac{1}{N^{*2}} \sum_{i=1}^{N^*} \sum_{j=1}^{N^*} (\Phi_{2\mathbf{H}+2\mathbf{G}} - 2\Phi_{\mathbf{H}+2\mathbf{G}} + \Phi_{2\mathbf{G}})(X_i^* - X_j^*) + \frac{1}{\alpha N + 1}(4\pi)^{-1}|\mathbf{H}|^{-1/2}$$

and we have to achieve optimal bandwidth by minimizing $\mathrm{SCV}^B(\mathbf{H}; \mathbf{G}, \alpha)$. We wrap up this subsection with introducing the two-stage SCV bandwidth selector algorithm

Algorithm 2 Two-stage SCV bandwidth selector (Chacón (2018))

Input: $\{X_1, \cdots, X_N\}$ Output: $\hat{\mathbf{H}}_{\mathrm{SCV}}$

1: Compute sixth order SAMSE-optimal pilot bandwidth
$\hat{\mathbf{G}}_6 := g_{6,\mathrm{SAMSE}}^2 I_2$
2: Compute sixth order kernel functional estimate $\hat{\Psi}_6(\hat{\mathbf{G}}_6)$
(/* Stage 1 */)
3: Plug $\hat{\Psi}_6(\hat{\mathbf{G}}_6)$ into formula for pilot bandwidth $\hat{\mathbf{G}}_4$
$\hat{\mathbf{G}}_4 := g_{4,\mathrm{SAMSE}}^2 I_2$
4: $\hat{\mathbf{H}}_{\mathrm{PI}} :=$ minimizer of $\mathrm{SCV}^B(\mathbf{H}; \hat{\mathbf{G}}_4)$ (/* Stage 2 */)

We developed `Hpi`, `Hpi.diag`, `Hscv` and `Hscv.diag` functions

in R package ks to Hpi.dir, Hpi.diag.dir, Hscv.dir and Hscv.diag.dir based on our proposed method.

2.4 Numerical studies

2.4.1 Simulation study

We set our intensity functions from Fuentes-Santos et al. (2016) to compare the result directly.

The first intensity function is

$$\lambda_1(x) = 3300 \cdot \exp(-3x_1)$$

and the second and third intensity functions are

$$\lambda(x) = 1000\ \phi(0.3 - 0.2x_2, \sigma)(x_1) + 25$$

where $\phi(\mu, \sigma)$ is the univariate normal density with mean μ and standard deviation σ. σ's are 0.1 and 0.02 for $\lambda_2(x)$ and $\lambda_3(x)$, respectively. λ_3 was used to evaluate the performance of the Voronoi estimator for the first order intensity (Barr and Schoenberg, 2010). The fourth is a (log-Gaussian) Cox process with the intensity function of

$$\lambda_4(x) = 2 \cdot \exp(6 + 4\ Y(x)).$$

where Y is a realization of a GRF(Gaussian Random Field) with mean zero and exponential model with covariance function $C(t) = \sigma^2 \exp(-t/\rho)$, $t > 0$, $\sigma = \rho = 0.1$. Cox point process is a generalization of Poisson point process and known as a doubly stochastic Poisson point process. The intensity function of Cox point process varies across the underlying mathematical space is itself a

stochastic process. The fifth and sixth intensity functions are

$$\lambda_5(x) = 600\left(\phi_{(x_2,0.1)}(x_1) + \phi_{(1-x_2,0.1)}(x_1) - \frac{1}{5}\phi_{2,(\mu,\Sigma)}(x)\right) + 30$$
$$\lambda_6(x) = 600\left(\phi_{(x_2,0.1)}(x_1) + \phi_{(1-x_2,0.2)}(x_1) - \frac{1}{20}\phi_{2,(\mu,\Sigma)}(x)\right) + 50$$

where $\phi_{2,(\mu,\Sigma)}(x)$ is the bivariate normal density with mean $\mu = (0.5, 0.5)$ and variance $\Sigma = 0.01I_2$.

We compared the outcomes for each intensity function to those obtained using other methods that were already in use. As we formerly mentioned, the performance of the bandwidth selectors can be compared by MISE of the kernel estimator of $\hat{\lambda}_{0,\mathbf{G}}(x)$ obtained with the respective optimal bandwidths:

$$\int_W \left(\hat{\lambda}_{0,\mathbf{H}}(x) - \hat{\lambda}_{0,\mathbf{G}}(x)\right)^2 dx$$

For the scalar bandwidth selector, we compared with Diggle's criterion (Diggle and Marron (1988)) and pseudo-likelihood cross-validation (PLCV) (Loader (1999), Baddeley and Turner (2005)). For the diagonal bandwidth selector, we compared with Scott's rule of thumbs (Scott (1992)), LSCV (Brooks and Marron (1991)), and full-matrix least-squares cross-validation were contrasted with full-matrix plug-in (Flscv). Although we lack prior knowledge on the efficiency of the bivariate kernel density estimator for intensity estimation in spatial point processes, full matrix cross-validation was carried out in accordance with a proposal in Duong and Hazelton (2005). Edge-correction is not taken into account by full matrix cross-validation or Scott's rule of thumb. Also, we compared two other bandwidths; one is computed in empirical way and the other is an oracle bandwidth. We can notice that the boxplots of ISEs

for oracle bandwidths are located quite below other boxplots for all cases which means that the results were properly computed.

The simulation study was run with the `spatstat` (Baddeley & Turner, 2005) and `ks` (Duong, 2013) packages of `R`. We set the range of α as $\{1, 2, 3, 4, 5\}$ after some trials. Note that there occurs errors when α gets larger or the true intensity function behaves almost linear. It is because the leading minor of order 1 is not positive definite during the Cholesky decomposition step. In practice, in case of $\lambda_3(x)$, there were several times of error even if α was not that big such as 1 or 2. Despite the fact that the errors did not occur in the other cases shown, if we change α as 10, 50, 100, there occurred errors from time to time even in the case of $\lambda_1(x)$ and $\lambda_2(x)$.

The images for intensity functions are introduced in Figure 2.2. We used 100 bootstrap samples and introduce ISE boxplots for each method below. The mean of optimal bandwidths and ISEs are in Appendix B.1.

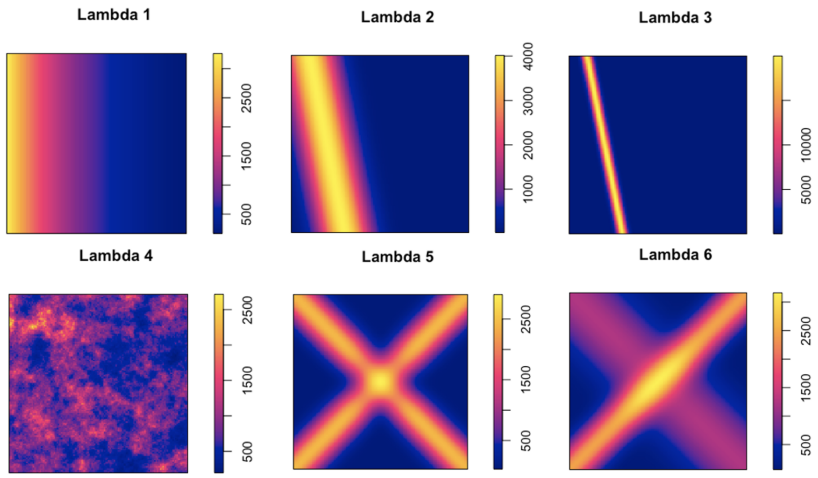


Figure 2.2: Images of intensity functions for simulations

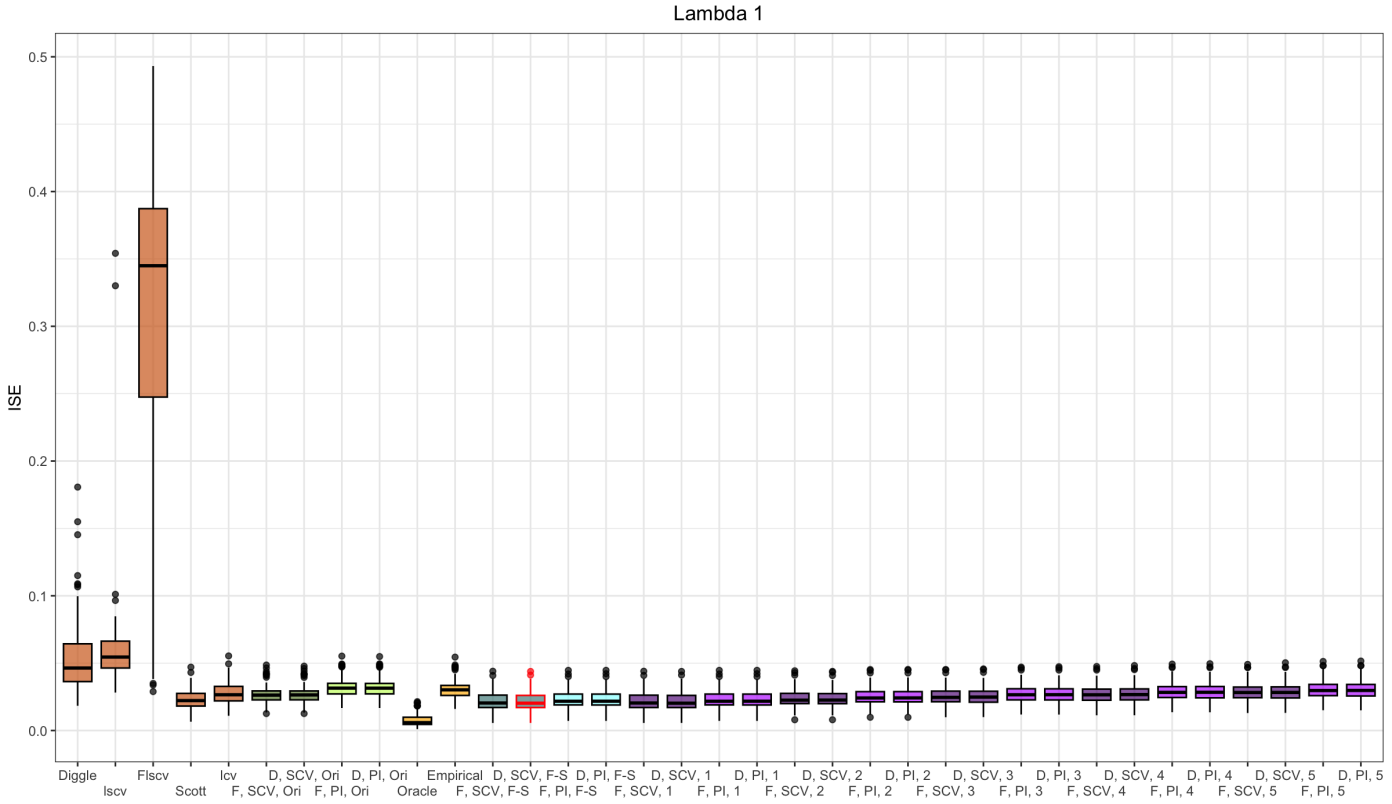


Figure 2.3: ISE boxplots for various bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_1(x)$; original

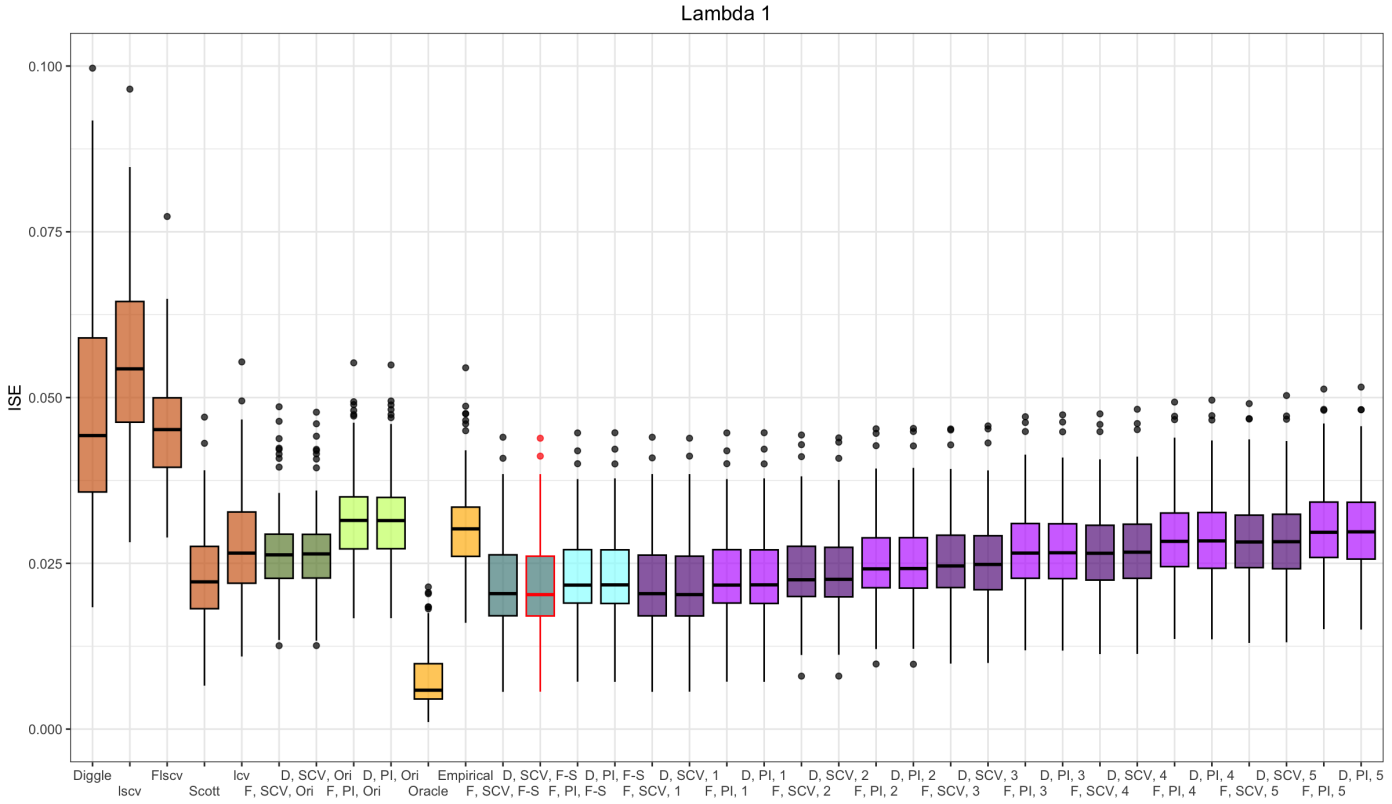


Figure 2.4: ISE boxplots for various bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_1(x)$; detailed, boxplot for Flscv is made of truncated data

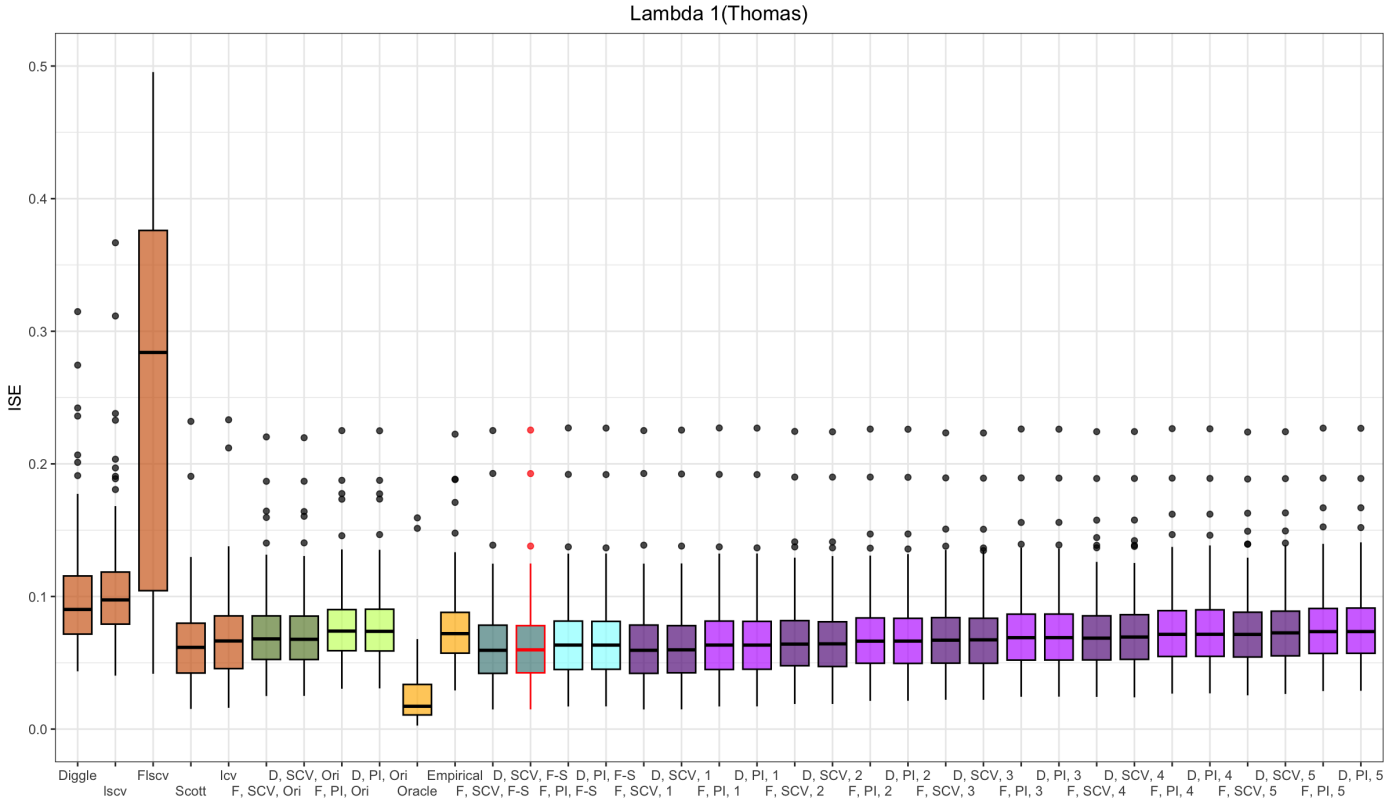


Figure 2.5: ISE boxplots for various bandwidth selectors for Thomas cluster point processes with intensity $\lambda_1(x)$

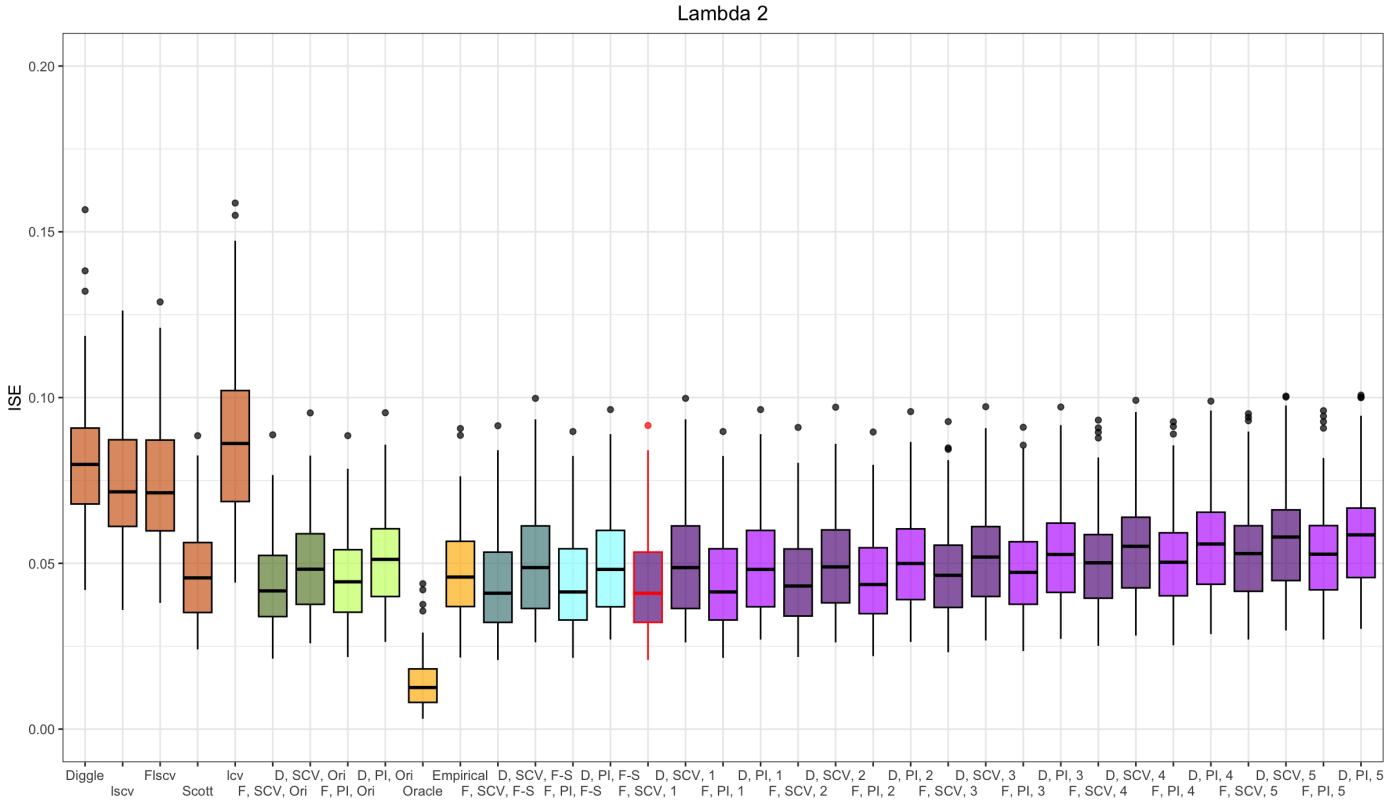


Figure 2.6: ISE boxplots for various bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_2(x)$

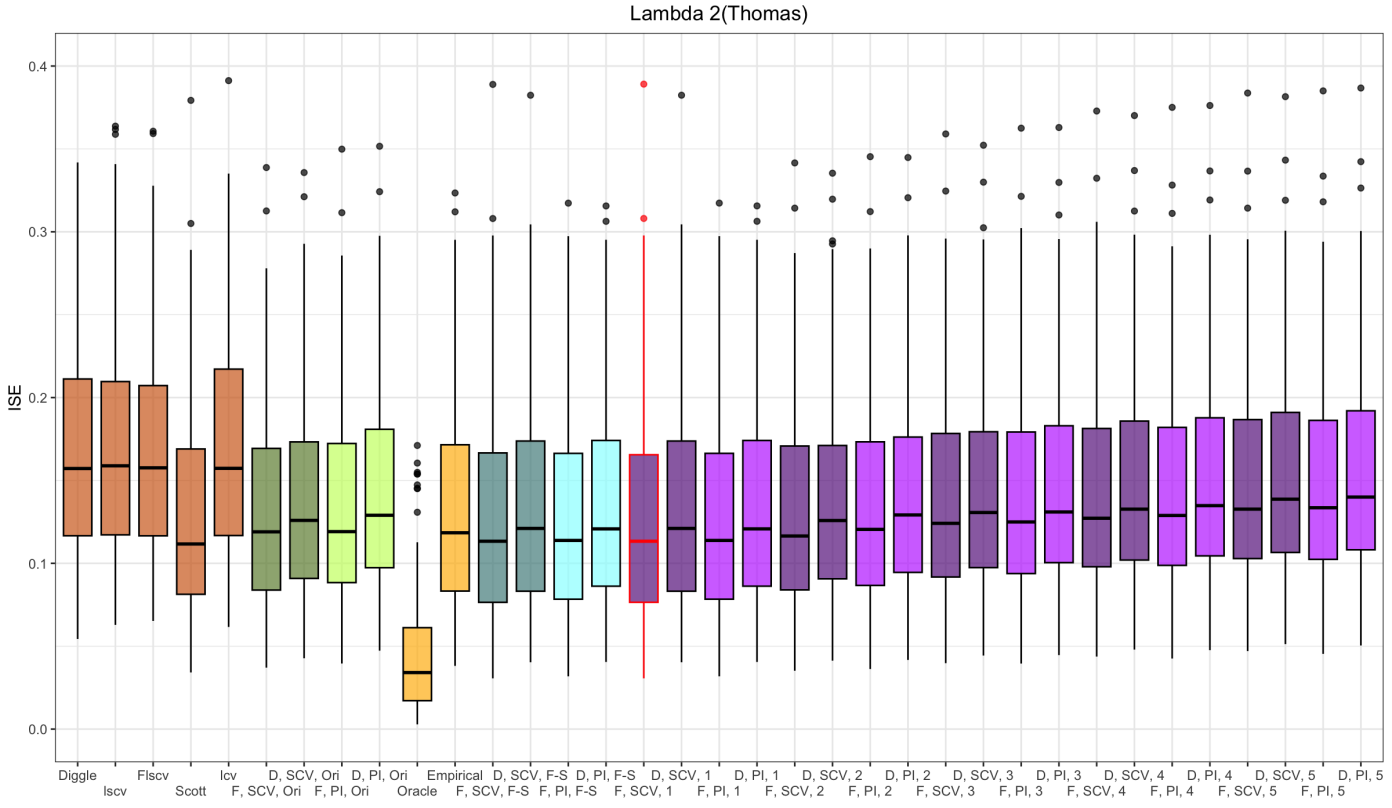


Figure 2.7: ISE boxplots for various bandwidth selectors for Thomas cluster point processes with intensity $\lambda_2(x)$

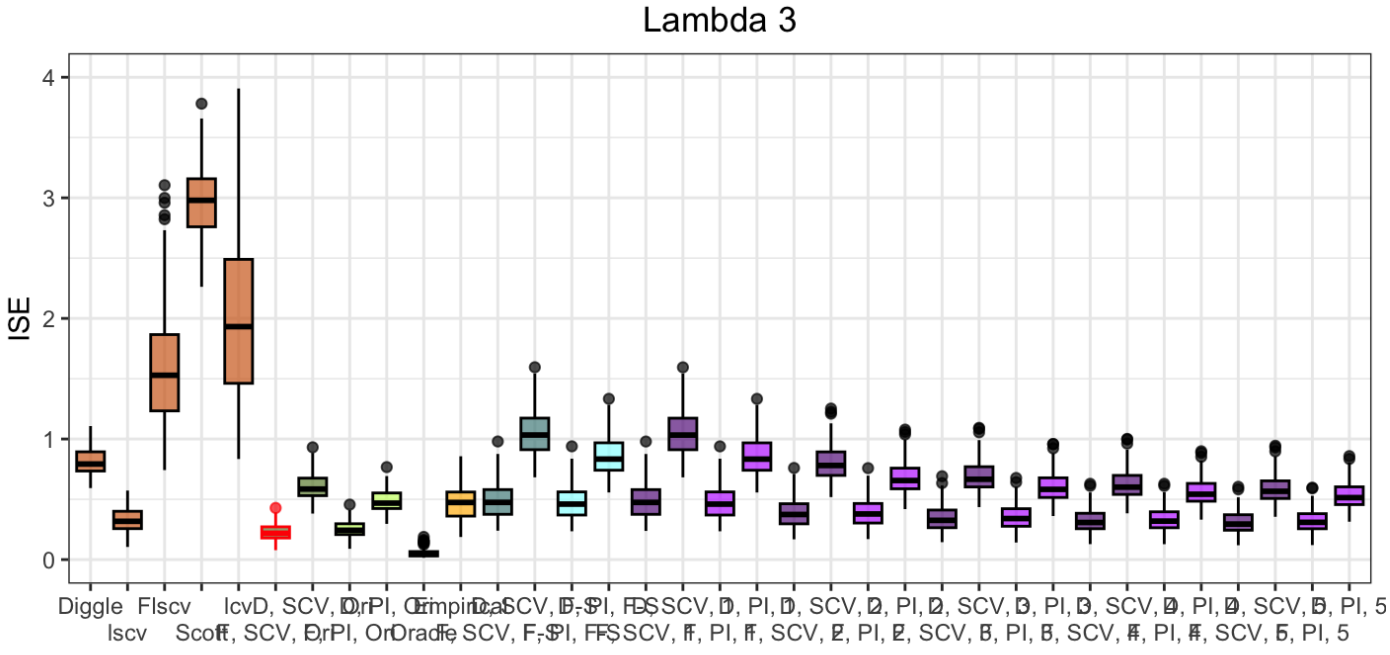


Figure 2.8: ISE boxplots for various bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_3(x)$

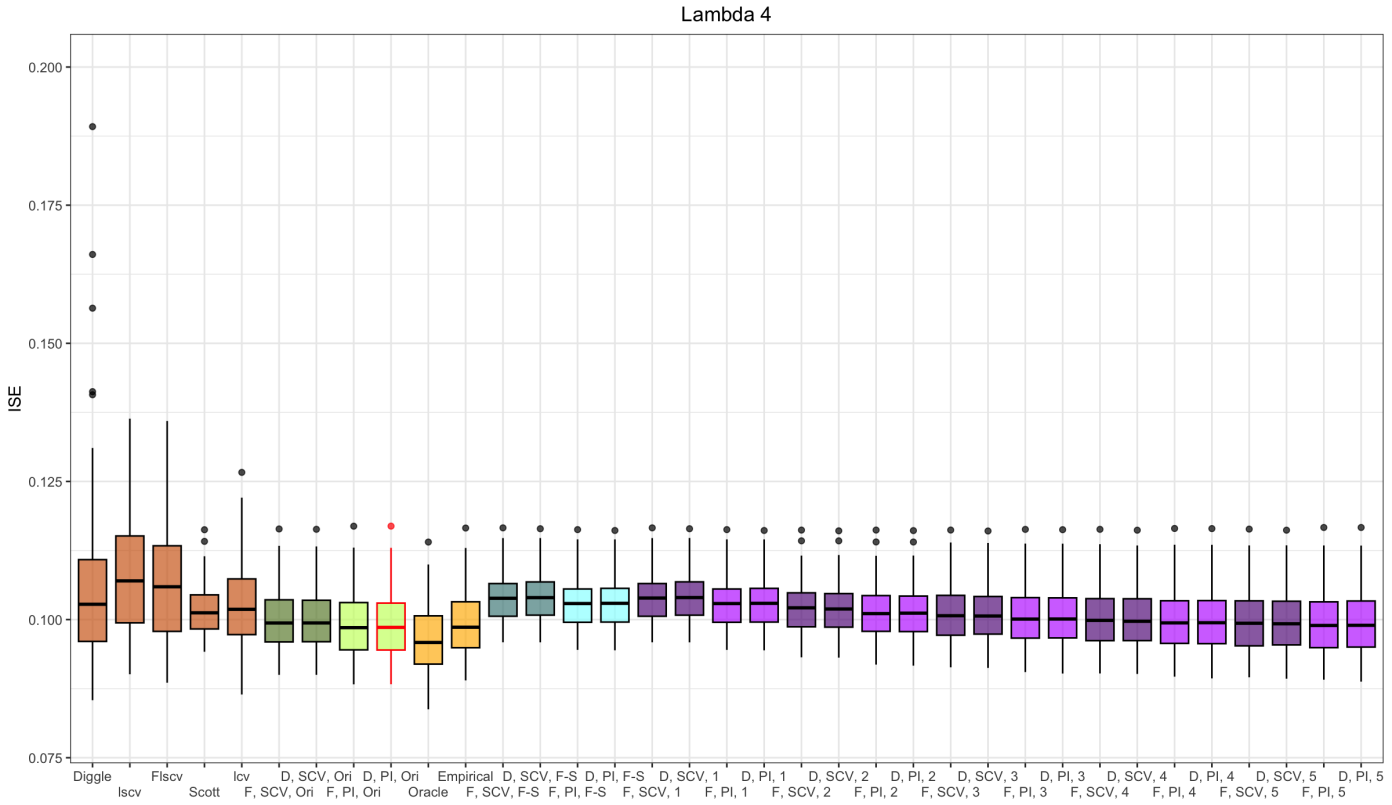


Figure 2.9: ISE boxplots for various bandwidth selectors for (log-Gaussian) Cox point processes with intensity $\lambda_4(x)$

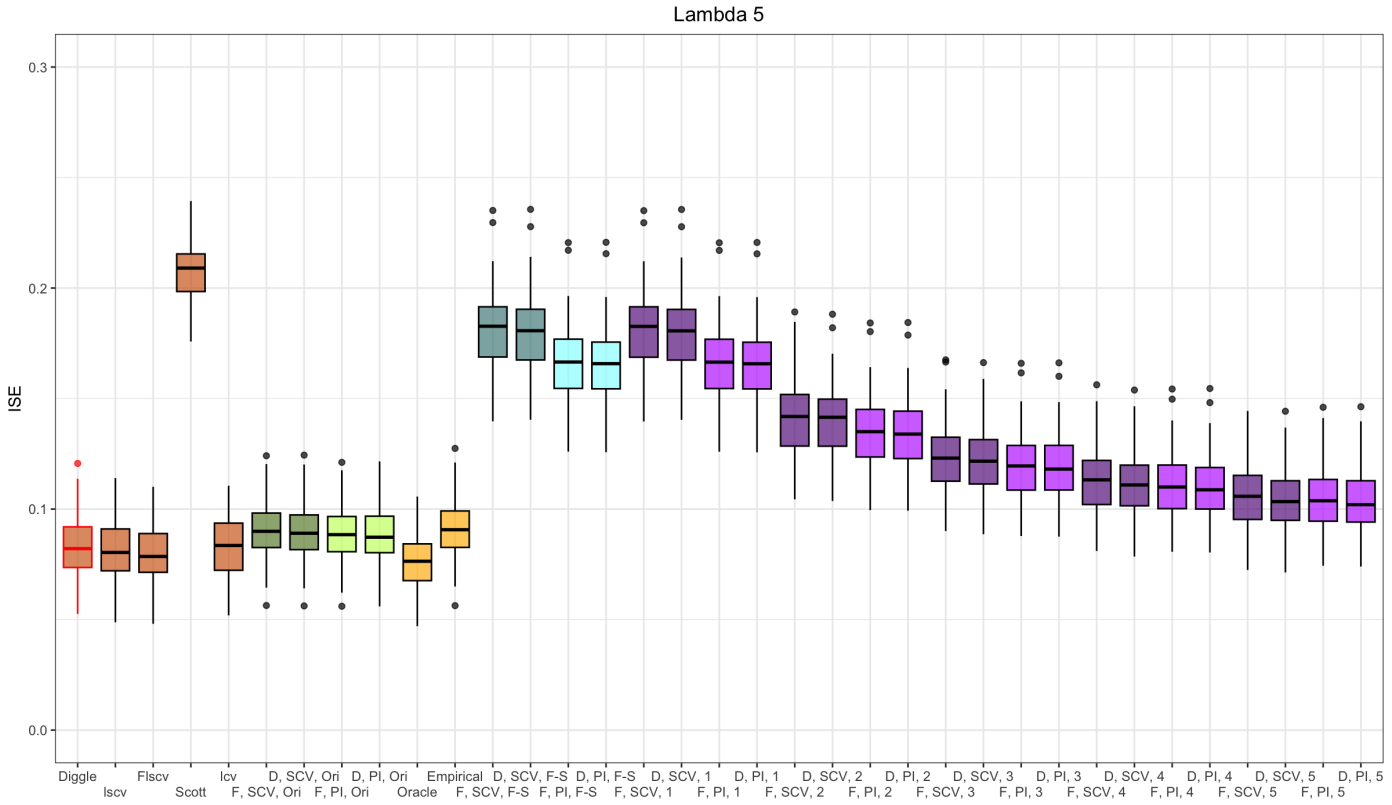


Figure 2.10: ISE boxplots for various bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_5(x)$

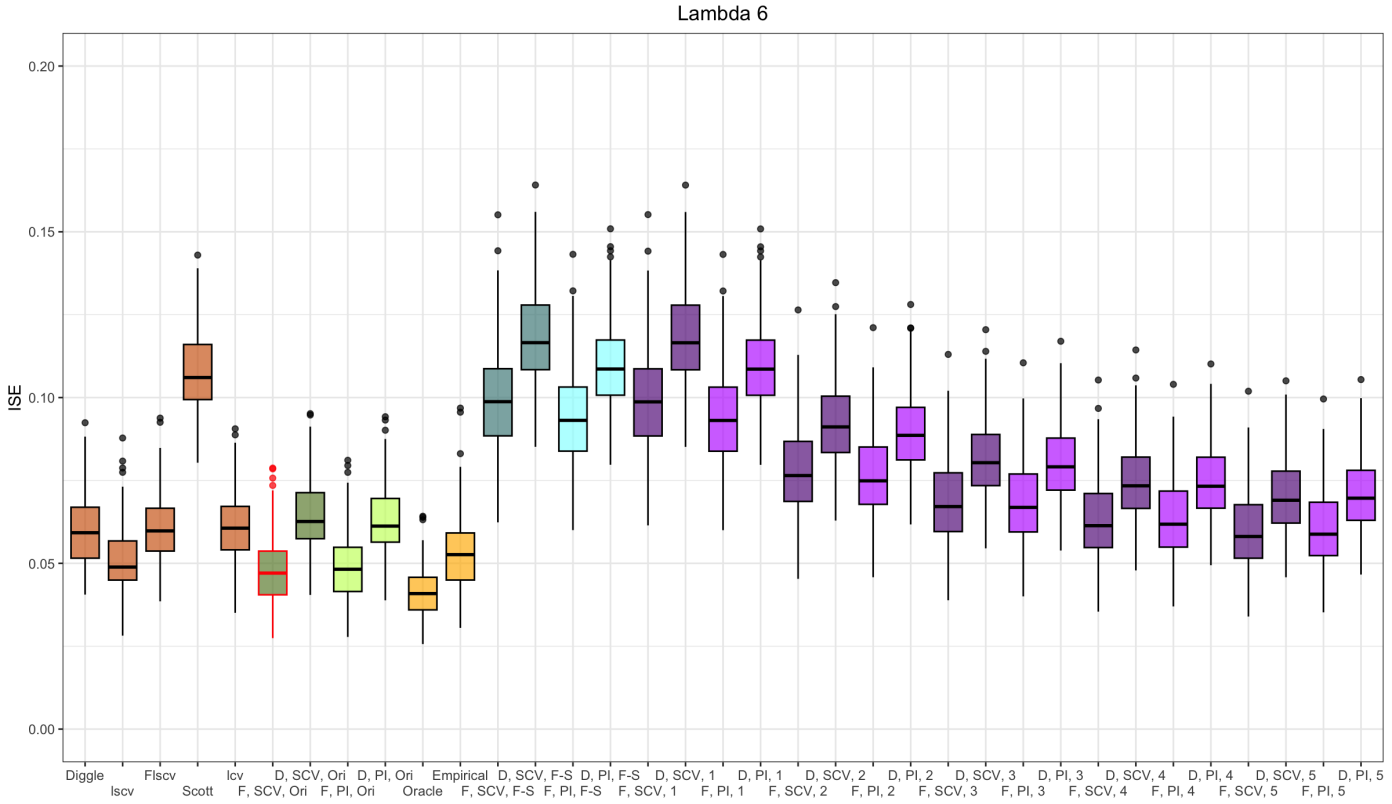


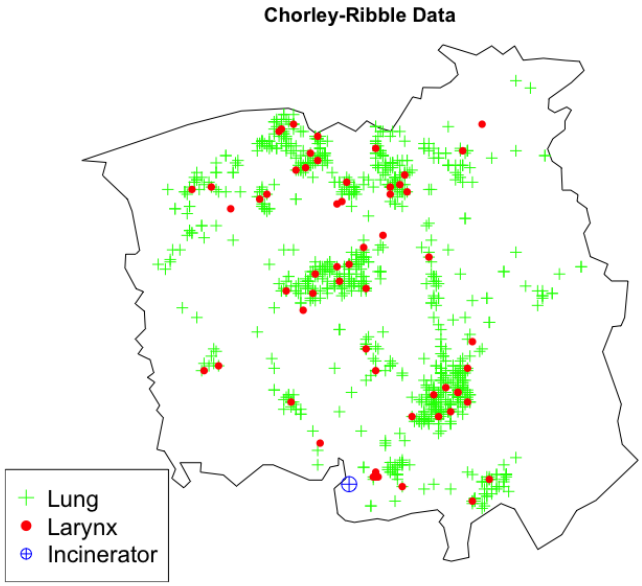
Figure 2.11: ISE boxplots for various bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_6(x)$

2.4.2 Case study

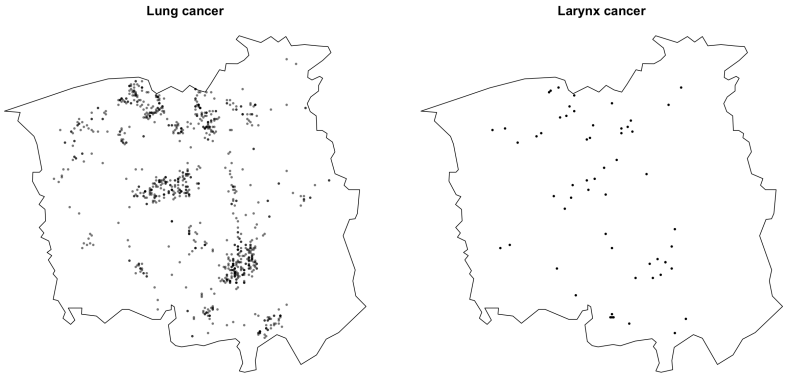
Chorley-Ribble cancer data were first presented by Diggle in 1990 and the data provide the precise address of cases of larynx cancer and lung cancer. They were recorded between 1974 and 1983 in the region of Chorley and South Ribble Health Authority of Lancashire, England. The locations are presented in Figure 2.12. Additionally, the location of an abandoned industrial incinerator location is also revealed with blue mark in Diggle (1990) presented and examined the data initially and they were further examined by Diggle and Rowlingson (1994) and Baddeley et al. (2005). The objective is to determine whether there is any proof that the area around the now-defunct industrial incinerator has a higher incidence of laryngeal cancer. The spatially variable density of the sensitive population is proxied by the lung cancer cases. The information is shown as a marked point pattern, with the marks indicating whether each point is a case of lung cancer or laryngeal cancer and the points showing the precise location of each person's home address. The resolution is 0.1 kilometers, and the coordinates are given in kilometers.

We want to demonstrate the density of locations of each cases but the number of two cases are quite different. However, since the number of cases for larynx(58 cases) is far smaller compared to that for lung(978 cases), it is hard to compare two densities at the same level and we can adjust the level of smoothness by using α instead of collecting a high amount of data to get a more detailed density.

As we can see in Figure 2.13, kernel estimation results of two



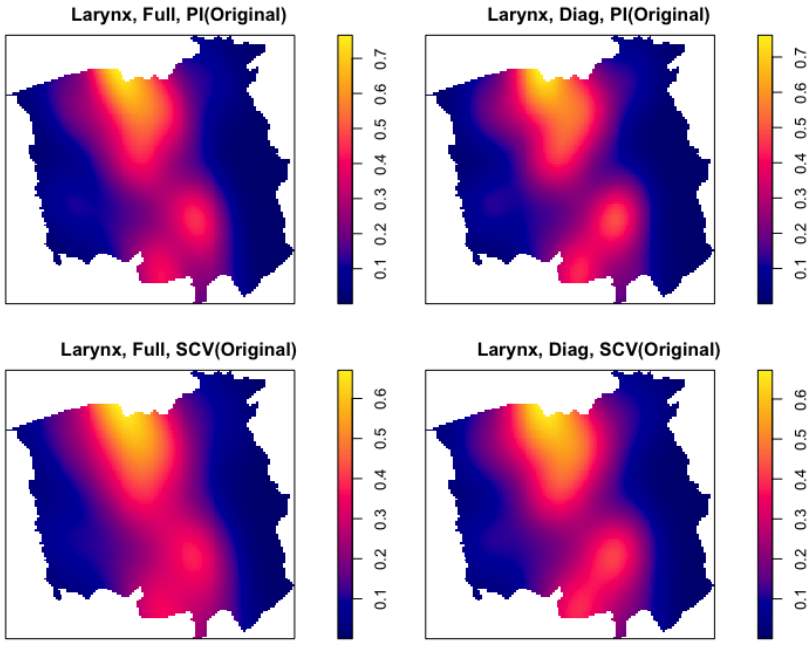
(a) Chorley-Ribble cancer data



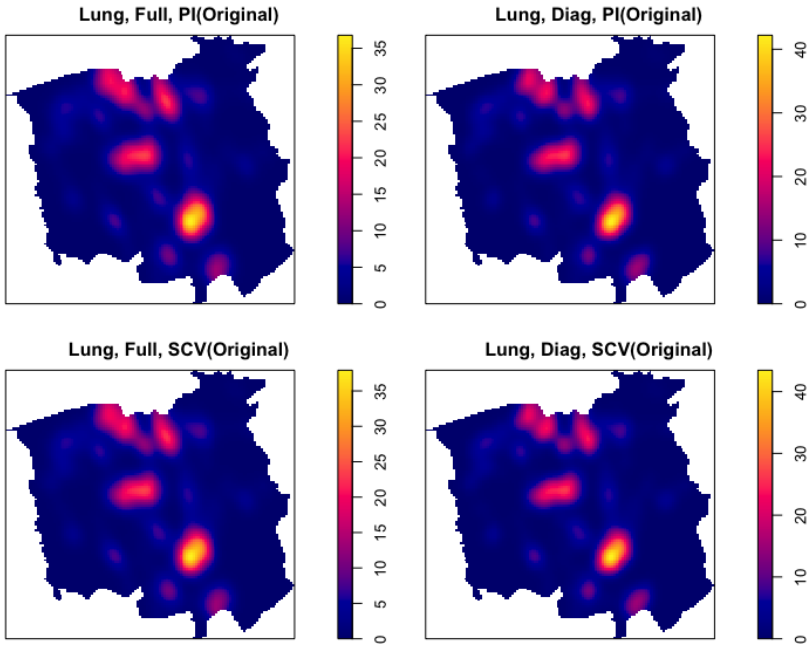
(b) Locations of each cancer cases

Figure 2.12: Chorley-Ribble cancer data

cases show the difference of smoothness. Here we use hyperparameter α and we can notice that varying α affects larger in larynx cancer cases than lung cancer cases. In fact, there is no big difference among the lung cancer cases' result in the rough. We recommend $\alpha = 3$ in this case which seperates upper and lower subregions but quite smooth.

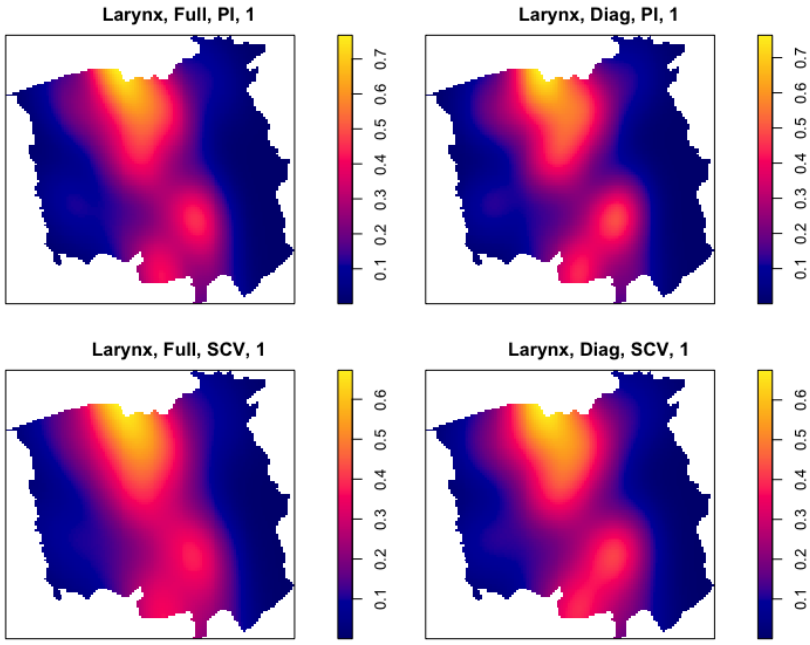


(a) Larynx cancer

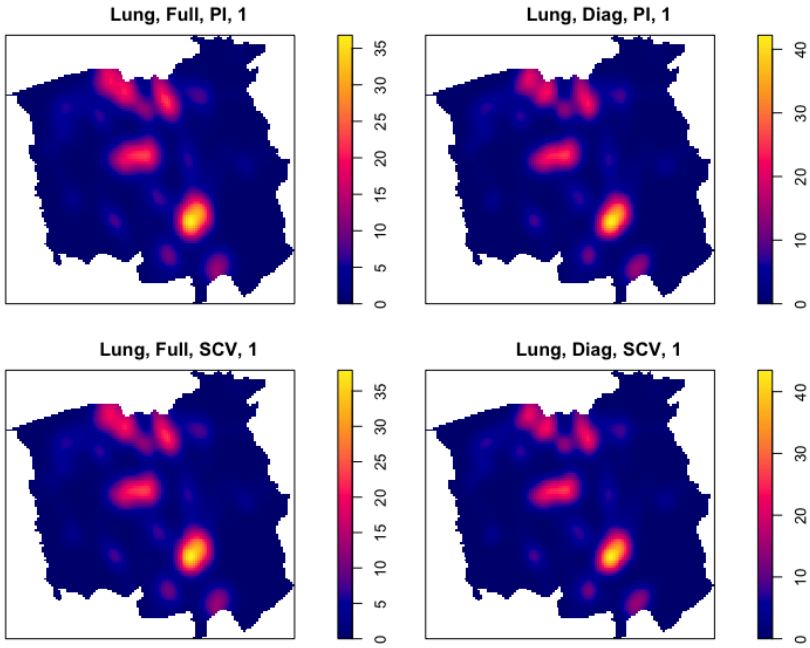


(b) Lung cancer

Figure 2.13: Kernel intensity estimation for Chorley-Ribble cancer data, original method

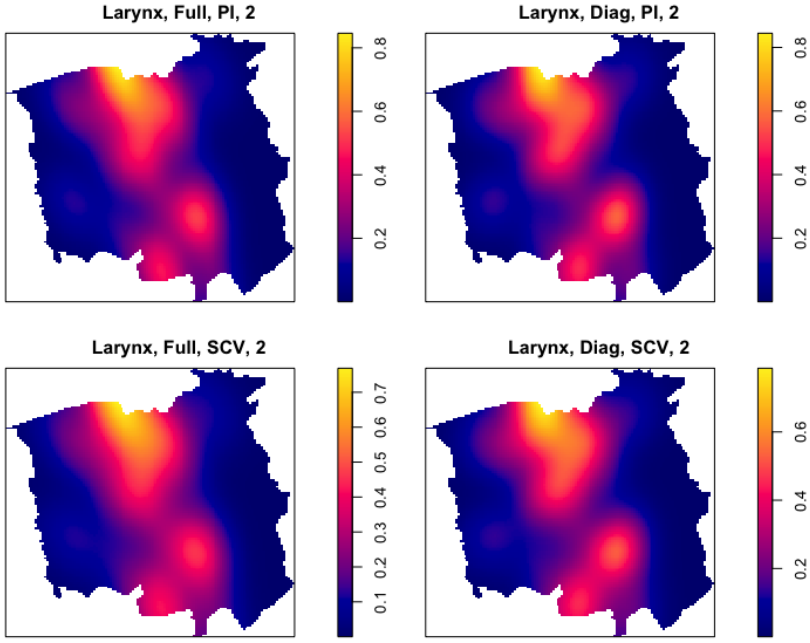


(a) Larynx cancer

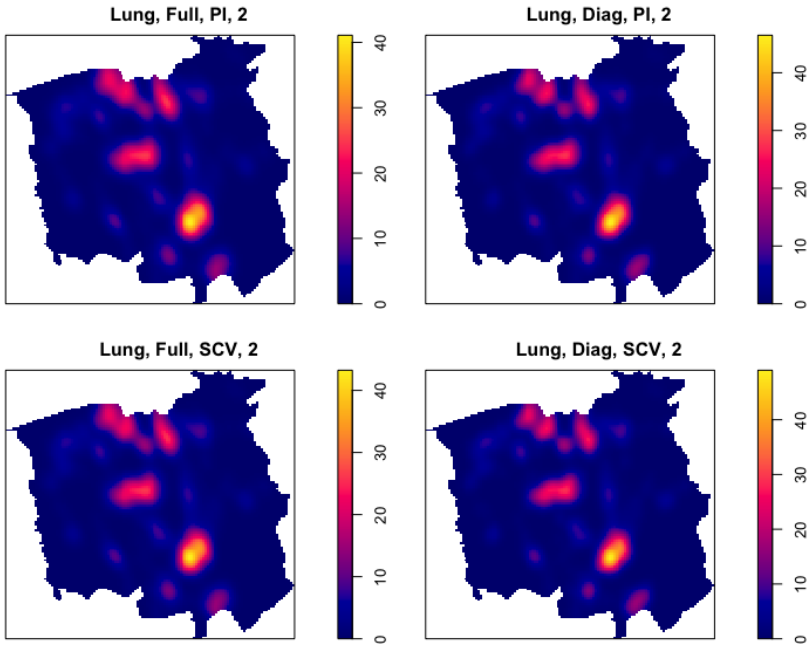


(b) Lung cancer

Figure 2.14: Kernel intensity estimation for Chorley-Ribble cancer data, $\alpha = 1$

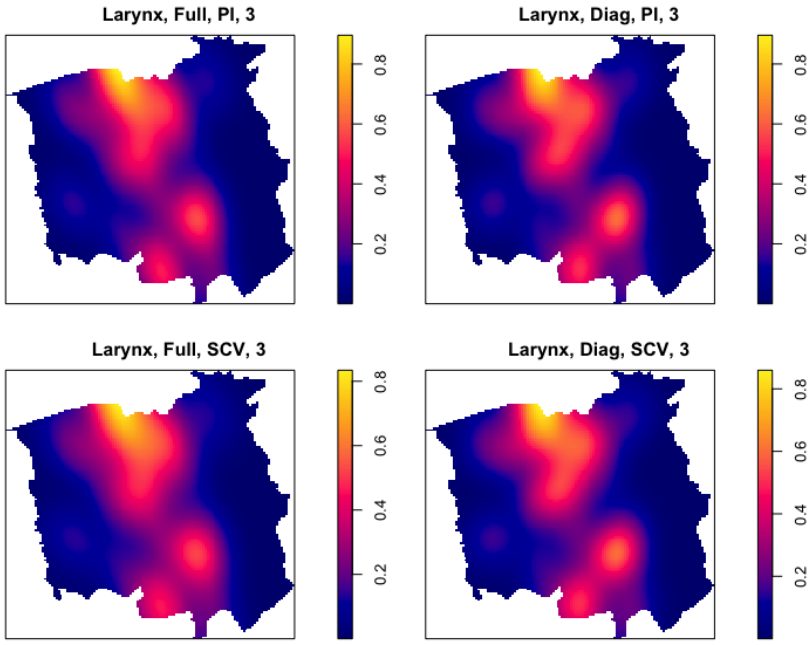


(a) Larynx cancer

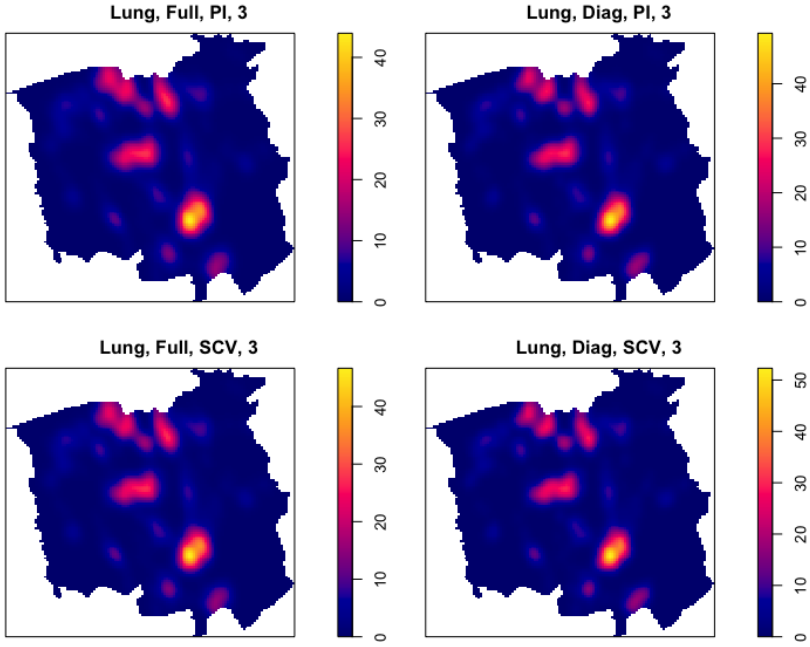


(b) Lung cancer

Figure 2.15: Kernel intensity estimation for Chorley-Ribble cancer data, $\alpha = 2$

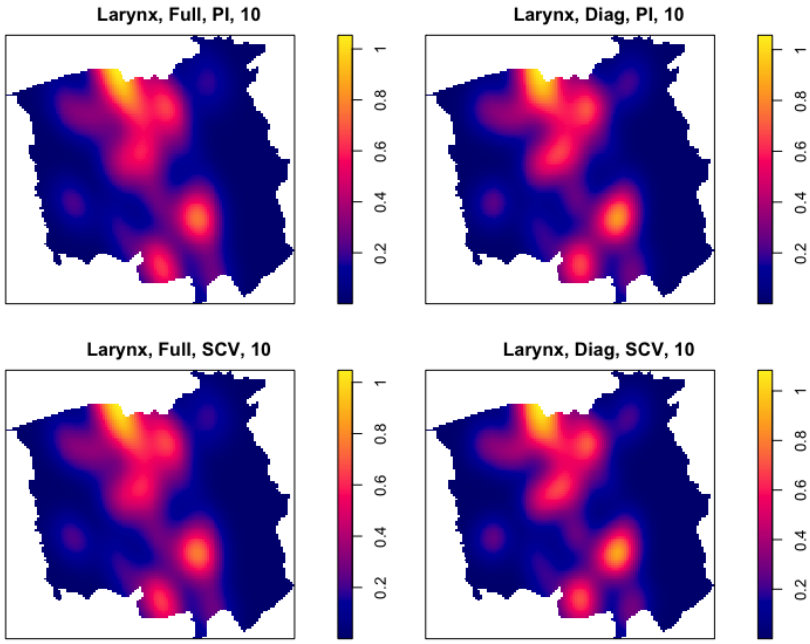


(a) Larynx cancer

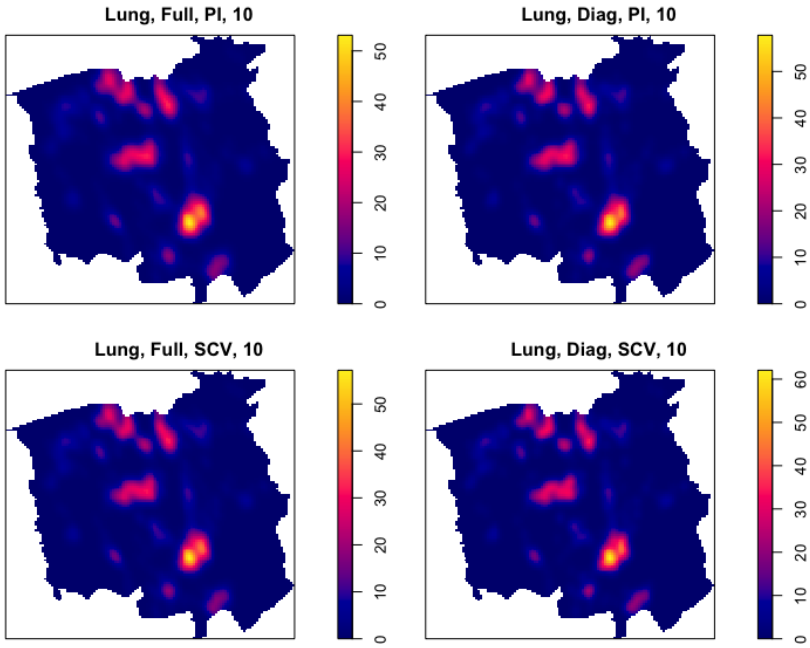


(b) Lung cancer

Figure 2.16: Kernel intensity estimation for Chorley-Ribble cancer data, $\alpha = 3$

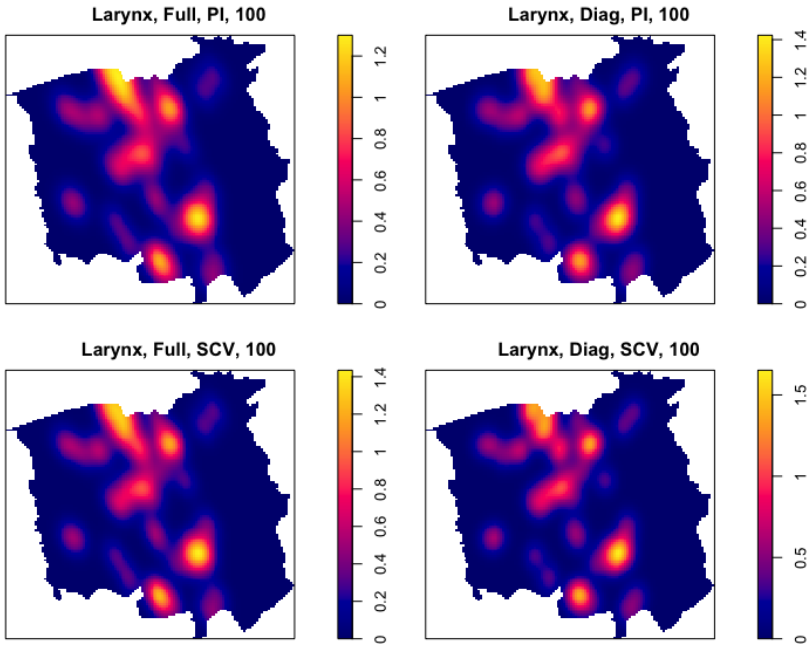


(a) Larynx cancer

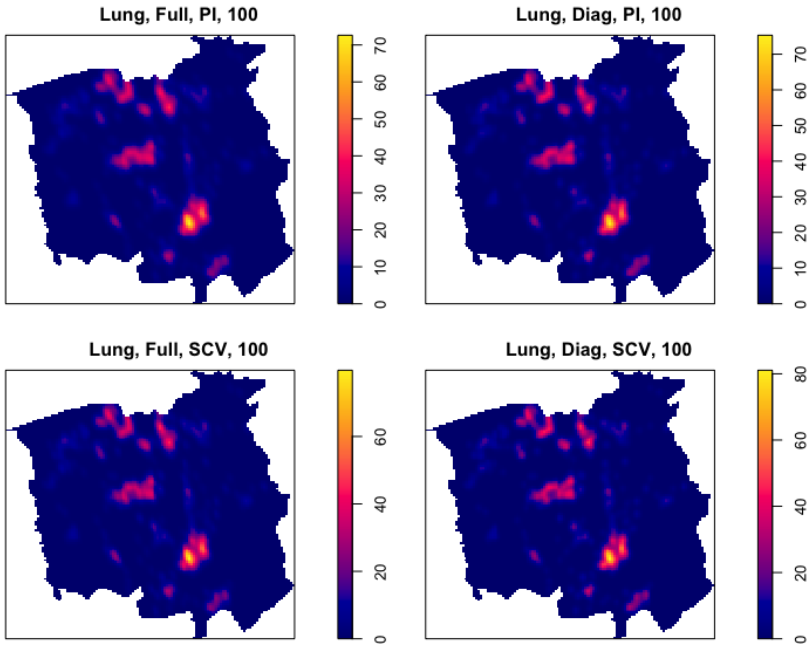


(b) Lung cancer

Figure 2.17: Kernel intensity estimation for Chorley-Ribble cancer data, $\alpha = 10$



(a) Larynx cancer



(b) Lung cancer

Figure 2.18: Kernel intensity estimation for Chorley-Ribble cancer data, $\alpha = 100$

Chapter 3

Kernel Estimation of Pair Correlation LISA Function

3.1 Overview

In the analysis of spatial association, it is quite unrealistic to assume stationarity or structural stability over with a large number of spatial observations. Random points in a bounded region frequently appear as data in a variety of scientific situations. Often we want to find clusters based on their second order properties within random spatial point patterns. In this chapter, we use a statistical approach to characterize the spatial distribution of points and provide a sound basis for constructing a stochastic model of their locations.

Ripley (1977) as follows

Definition 2. *(Ripley's K function)*

$$\begin{aligned} K(t) &= \lambda^{-2}\mu^2[\{(x,y) : x \in I, 0 < d(x,y) < t\}] \\ &= \lambda^{-1}\mu(b(0,t)\ 0), t > 0 \end{aligned}$$

where λ is an intensity, $b(0,t)$ is a ball of radius t centered at zero, $d(x,y)$ is a distance between x and y.

proposed a second-order property function K-function is a cumulative function and we can investigate the differential of function that is another interpretable function. It is called the product density function.

Definition 3. *(Product density function)*

$$\rho(t) = \frac{\lambda^2 K'(t)}{2\pi t}, \quad t > 0$$

When seeking to describe relationships in point patterns, it is necessary to consider the relevant features, referred to as second-order measures. Let du and dv has the same definition with in Chapter 2 and the areas are $|du|$ and $|dv|$, respectively. $\lambda^{(2)}(u,v)dudv$ can be regarded as the joint probability that both infinitesimal regions contain random points of X. In case of complete spatial randomness, $\lambda^{(2)}(u,v) = \lambda(u)\lambda(v)$ for $u, v \in \mathbb{R}^2$. Define the pair correlation function as the normalized product density, as follows:

$$g(u,v) = \frac{\lambda^{(2)}(u,v)}{\lambda(u)\lambda(v)}, \quad u, v \in \mathbb{R}^2,$$

where $a/0 = 0$ for $a \geq 0$. The relationship between intensity and the normalized product density is utilized to assess the degree of repulsion or aggregation in a point pattern: $\lambda^{(2)} > 1$ for aggregation and $\lambda^{(2)} < 1$ for repulsion.

The pair correlation function can be used to distinguish between many spatial point structures, such as a Poisson point process with $g(r) = 1$. Higher pair correlation function values for small distances indicate an abundance of short inter-event distances. If small inter-event distances are uncommon, this suggests the presence of an inhibitory structure, and places tend to disperse from one another.

Dasgupta and Raftery (1998) explored the application of model-based clustering and various distributions to identify linear minefields, which is treated as features, amidst clutter. Byers and Raftery (1998) proposed using k-th nearest neighbor distances to identify features from noise. Mateu et al. (2007, 2010) considered this distinguishment using product density LISA functions. To tackle the classification problem, they create a mixture modeling strategy.

As we mentioned at the beginning of this chapter, the stationarity assumption can be too strong and unrealistic for real data. That is why we would like to assume the Second Order Intensity Reweighted Stationarity throughout this chapter.

Definition 4. *(**S**econd-**O**rder **I**ntensity **R**eweighted **S**tationary, SOIRS)*

The point process Y is "second-order intensity-reweighted stationary" if the random measure is second-order stationary. Equivalently $M(A, B) = M(A + x, B + x)$ for all $x \in \mathbb{R}^2$ where $A + x$ denotes the translation of A by the vector x.

It was first introduced by Baddeley(2000) andThis assumption asserts that the pair correlation $g(\cdot, \cdot)$ is solely dependent on $r = ||u-v||$, where the Euclidean norm $||\cdot||$ is indicated. We can simply

notice that second order stationary point process is also a second order intensity reweighted stationary.

3.2 Pair correlation LISA function and clustering

Anselin (1995) proposed evaluating the local contributions of a global estimator of second-order properties and applied it to a clustering issue. These components are referred to as local indicators of spatial association (LISA). It was first developed in geostatistics and played a role as local Moran's I. Moran's I is a measure of spatial autocorrelation characterized by correlations between signals in close proximity in space and widely used in areas such as health care or housing values. By analyzing how each individual point behaves in relation to its neighbors, LISA functions were mainly used to look at the local structure of a point pattern.

An edge-corrected estimator of pair correlation function is

$$\hat{g}_h(r) = \frac{1}{2\pi r|W|} \sum_{i=1}^{n} \sum_{j \neq i} \frac{K_h(||u_i - u_j|| - r)}{\lambda(u_i)\lambda(u_j)} \cdot \frac{|\partial b(u_i, ||u_i - u_j||)|}{|\partial b(u_i, ||u_i - u_j||) \cap W|}, \quad r > h > 0$$

The $|\partial b(u_i, ||u_i - u_j||)|$ / $|\partial b(u_i, ||u_i - u_j||) \cap W|$ is an edge-corrected term which makes a correction for the intensity near the edge of the bounded region. It carries out an equivalent role with the boundary-correction term in the density estimation.

Definition 5. *(Pair correlation **L**ocal **I**ndicator of **S**patial **A**ssociation*

- LISA function)

$$\hat{g}_h^{(i)}(r) = \frac{n-1}{r|W|} \sum_{j \neq i} \frac{K_h(||u_i - u_j|| - r)}{\lambda(u_i)\lambda(u_j)} \frac{||u_i - u_j||)}{|\partial b(u_i, ||u_i - u_j||) \cap W|}$$

for $r > h > 0$

The global estimator is proportional to the sum of the individual pair correlation LISA function. This can be easily shown as below.

$$\begin{aligned}
\frac{1}{n-1} \sum_{i=1}^{n} \hat{g}_h^{(i)}(r) &= \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{n-1}{r|W|} \sum_{j \neq i} \frac{K_h(||u_i - u_j|| - r)}{\lambda(u_i)\lambda(u_j)} \frac{||u_i - u_j||)}{|\partial b(u_i, ||u_i - u_j||) \cap W|} \right) \\
&= \frac{1}{2\pi r|W|} \sum_{i=1}^{n} \sum_{j \neq i} \frac{K_h(||u_i - u_j|| - r)}{\lambda(u_i)\lambda(u_j)} \frac{|\partial b(u_i, ||u_i - u_j||)|}{|\partial b(u_i, ||u_i - u_j||) \cap W|} \\
&= \hat{g}_h(r)
\end{aligned}$$

To acquire various aspects of the above estimator, we employ Palm distributions, which are crucial in studying the conditional probability in the presence of a fixed event (Chiu et al. 2013). The Palm expectation is typically expressed by $\mathbf{E}_!(\cdot)$ and can be thought as the expectation conditional on a specific event. Refer to González (2021), the expected value and variance with respect to reduced Palm process are expressed as

$$\mathbf{E}_! \left(\hat{g}_h^{(i)}(r) \right) = \frac{\Lambda(W) + 1}{\lambda(u_i)|W|}$$

$$\begin{aligned}
\mathbf{Var}_! \left(\hat{g}_h^{(i)}(r) \right) = &\frac{\Lambda(W) + \Lambda^{-1}(W) + 3}{(\lambda(u_i)|W|)^2 2\pi r^2} \int_W \frac{2\pi(||u_i - s||^2 K_h^2(||u_i - s|| - r)}{\lambda(s)|\partial b(u_i, ||u_i - s||) \cap W|} ds \\
&+ \frac{3\Lambda(W) + 3}{(\lambda(u_i)|W|)^2}
\end{aligned}$$

where $\Lambda(W) = \int_W \lambda(u) du$.

3.2.1 Bandwidth selection

As we estimate the LISA function via kernel function, selection of optimal bandwidth is crucial. One option for determining the bandwidth for the pair correlation function kernel estimation is to use cross-validation, which involves selecting a smoothing bandwidth through composite likelihood (Guan, 2007). We suggest an empirical method to minimize. The algorithm is as below.

1. For a given h, obtain the non-parametric estimate $\hat{g}_h^{(i)}(r)$.
2. Use block bootstrapping to resample the point data and obtain the bootstrap estimates $\hat{g}_h^{(i),*}(r)$. 100 bootstrap samples were used in this thesis.
3. Compute $M(h) = \sum_{j=1}^{n} \left\{ \left(g_h^{(i)}(r_j) - \hat{g}_h^{(i),*}(r_j) \right)^2 + Var_*(\hat{g}_h^{(i)}(r_j)) \right\}$ for MISE where r_j's are in interest

Since we use Epanechnikov kernel function, it is critical to establish an acceptable range to prevent the kernel values from being 0. To rewrite,LISA functions are comprised of a set of real-valued functions $\{g_h^{(i)}\}_{i=1}^n$ defined on a compact interval $T = [\epsilon, r_{\max}]$, where the maximum value $r_{\max}$ must be specified beforehand as less than a quarter of min(width, height) by Diggle's rule.

The properties of these functions can be understood from a functional perspective as compositional functional observations. For any r, we define the following transformation

$$g_i^*(r) = \hat{g}_h^{(i)}(r) + \text{const}$$

where the constant value is arbitrary but it is known that the value around 1 tends to make a better classification performance. Indi-

vidual curves can be interpreted as long as they are not isolated from the rest. The points with the most congested neighborhoods are related with the higher transformed LISA functions. We can make that related information easier to understand by standardizing it.

$$\varrho_i(r) = \frac{g_i^*(r)}{\sum_i g_i^*(r)}, \quad i = 1, \cdots, n \quad \text{and} \quad r \in T$$

Using the logarithm of the right hand side does not matter since it only changes the y-axis value not the shape of lines. In this example, each component $\varrho_i(r)$ represents a relative weight that has been transformed linearly from the pair correlation. These components are non-negative functions defined on a compact domain T and have logarithms that are square-integrable. We use the Euclidean metric in L_2. Let ϱ_i and ϱ_j be the transformed LISA functions and we used the distance below:

$$d^2(\varrho_i, \varrho_j) = \int_T (\varrho_i(r) - \varrho_j(r))^2 \, dr$$

With this distance, all of the typical multivariate statistics methods that rely on dissimilarity measures such as multidimensional scaling can be used. Now we consider the clustering method.

3.2.2 Clustering method

In the context of identifying clusters of LISA functions, we explore functional classification methods. Most of these methods employ one of three strategies:

- Dimension reduction prior to clustering

- Nonparametric methods utilizing specific metrics between curves

- Model-based techniques

Jacques and Preda (2014) offer a comprehensive and instructive analysis of various tactics. Here, we compare and contrast supervised and unsupervised learning systems as solutions to the classification problem.

One of popular clustering methods is a multidimensional scaling (MDS) method (Leeuw and Mair, 2009). The inputs are pairwise distances between functions and the outputs correspond to a coordinate configuration in $\mathbb{R}^2$. It is important to note that the points in the new space are the same as the original points in the point pattern, but they are viewed in an MDS configuration space based on the proximity of the LISA functions. We also incorporate Kruskal's standardised stress as part of the approach to assess the fit quality, with lower values being desirable (Kruskal, 1964). To further process the data, we suggest utilizing a support vector machine (SVM, Steinwart and Christmann, 2008). The data is transformed into a higher-dimensional space through a kernel, which allows for the determination of the hyperplane with the greatest margin that can separate the data linearly.

We utilize a bagged clustering algorithm(BCA) as an alternate unsupervised technique. A hierarchical clustering algorithm is used to merge the resulting cluster centers in order to find clusters. The algorithm begins with as a number of groups as there are dissimilarities in the data and progresses to a single cluster. A cluster is able to combined with another in each phase. BCA

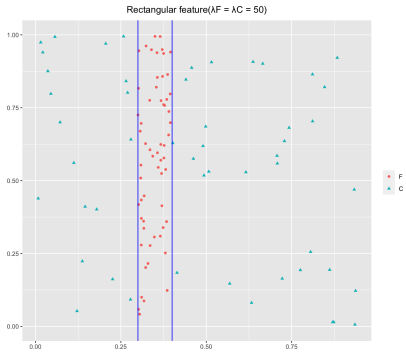
is nice method to preserve stability. Stability refers to the ability to maintain meaningful clusters when the dataset is altered in a trivial manner. It is widely acknowledged that stability is heavily influenced by the data. (Hennig, 2007). To assess stability, resampling techniques such as bagging can be employed. In this case, new training sets are created through bootstrap sampling and incorporated into the cluster analysis framework (Dudoit and Fridlyand, 2003).

3.3 Numerical studies

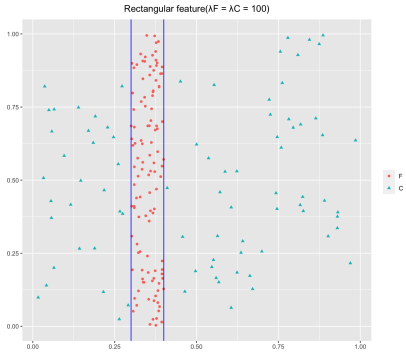
3.3.1 Simulation study

We simulated a rectangular feature which is superimposed with clutter outside. We set various intensities of the features (λ_F) and the clutter (λ_C). For each case, we computed the average of misclassification rates of 100 simulations.

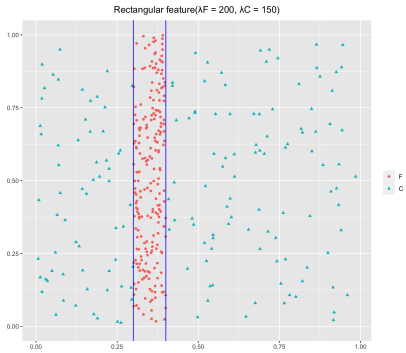
Both MDS + SVM and BCA approaches perform well and are comparable in terms of mean values. In general, the mean values for MDS + SVM are lower, suggesting better performance. This outcome is predicted, however, because BCA requires only the distance matrix, when MDS + SVM requires both training sets and the distance matrix. Figure 3.1 and 3.2 are the point patterns and the plots for transformed LISA functions, respectively. We also introduce the results for average and standard errors of 100 simulations.



(a) $\lambda_F = \lambda_C = 50$

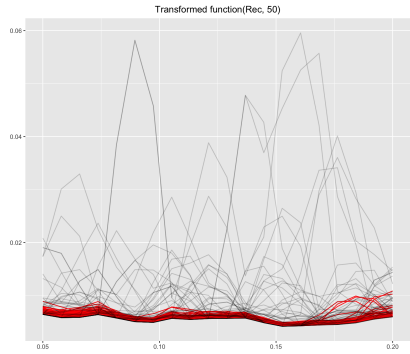


(b) $\lambda_F = \lambda_C = 100$

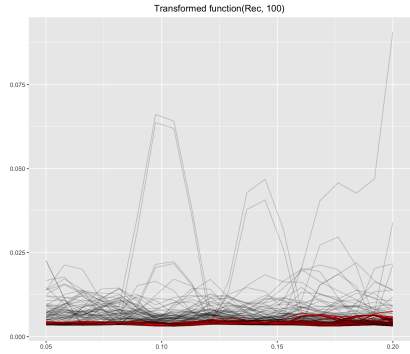


(c) $\lambda_F = 200, \lambda_C = 150$

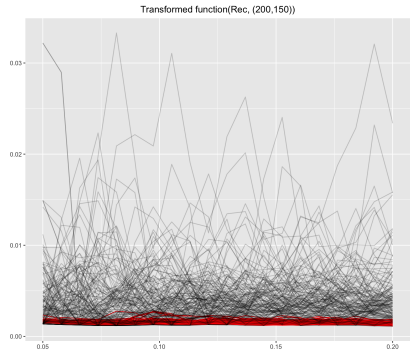
Figure 3.1: Point patterns with rectangular figure, various intensities



(a) $\lambda_F = \lambda_C = 50$



(b) $\lambda_F = \lambda_C = 100$



(c) $\lambda_F = 200, \lambda_C = 150$

Figure 3.2: Transformed LISA functions of point patterns with rectangular figure, various intensities

		MDS+SVM		BCA	
λ_F	λ_C	LCV	Empirical	LCV	Empirical
50	50	0.078(0.0028)	0.067(0.0022)	0.094(0.0076)	0.091(0.0089)
100	100	0.041(0.0016)	0.028(0.0032)	0.049(0.0039)	0.043(0.0028)
200	150	0.021(0.0009)	0.019(0.0011)	0.037(0.0011)	0.029(0.0012)

3.3.2 Case study

We chose an earthquake data called Bucaramanga nest where Bucaramanga is the capital and largest city of the department of Santander, Colombia. Refer to Prieto et al.(2012), they introduced three major intermediate-depth earthquake nests which are Vrancea, Hindu Kush and Bucaramanga. The detailed epicenter locations are in Figure 3.3.

The analysis of earthquakes on a macro-scale is useful in determining earthquake clusters and gaining insights into the geological dynamics of regions that may be prone to earthquakes. In this study, we use a dataset based on the epicenters of earthquakes in the Santander region of Colombia. Santander is well-known for having the world's densest concentration of intermediate-depth earthquakes at a location known as The Bucaramanga Nest (Prieto et al. 2012). Because of the zone's complicated tectonic topography, this area is characterized by a high seismic hazard. Zarifi et al. (2007) suggested the definition of a seismic nest as a region exhibiting an unusual concentration of seismic activity compared to its surroundings.Although this region is called as a term of 'nest', the spatial delimitation is not clear. For this reason, we would like to give our suggestion for the delimitation of Bucaramanga

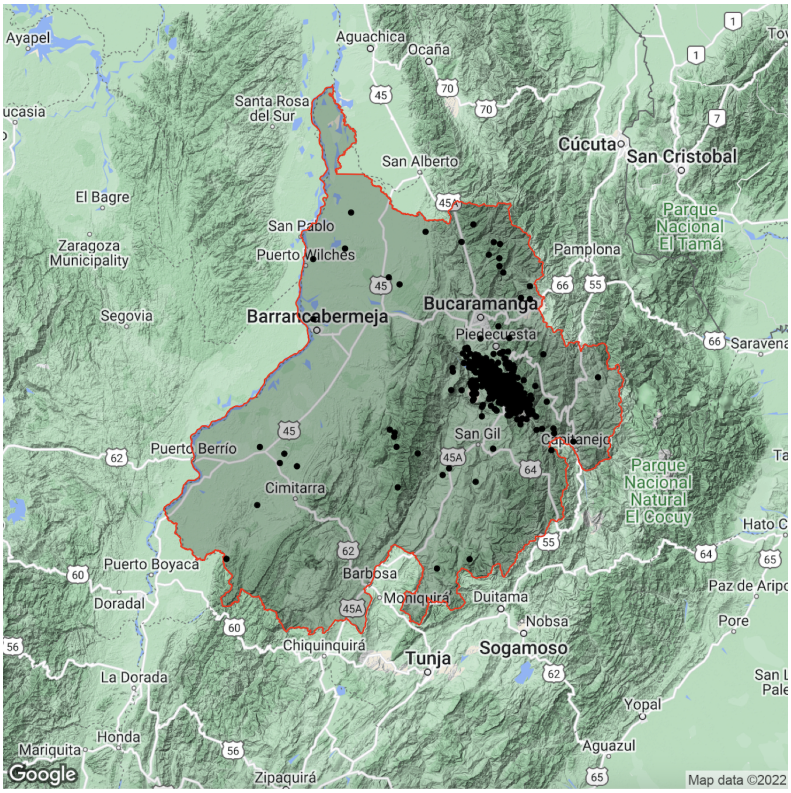


Figure 3.3: Santander earthquake epicenters

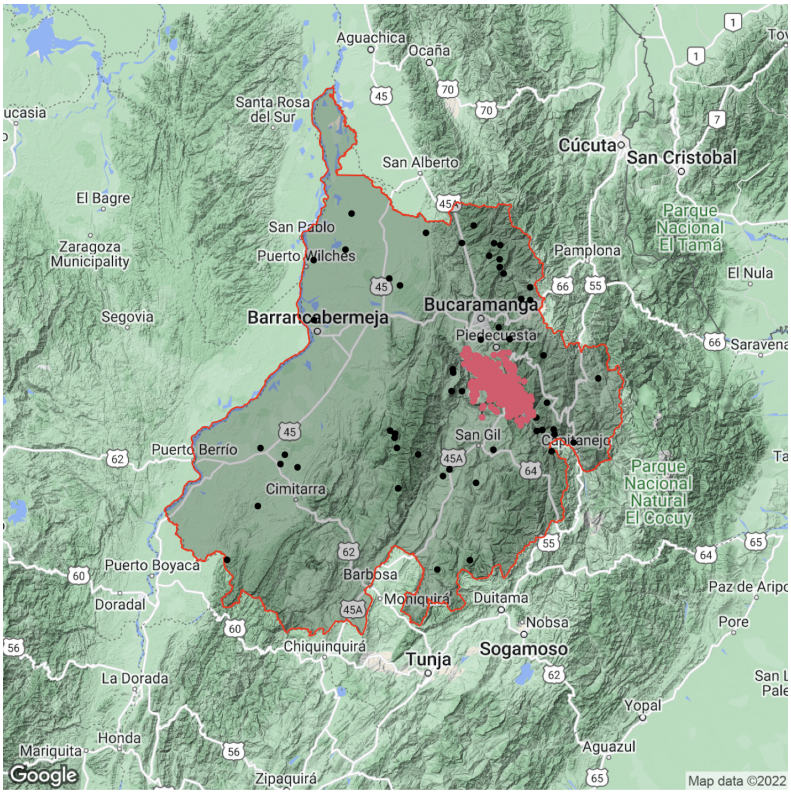


Figure 3.4: Santander earthquake epicenters with cluster

nest. We analyzed earthquake events in Santander-Colombia that took place between January 1, 2012 and October 31, 2022 and had magnitudes greater than 4.0 on the Richter local magnitude scale (M_L).

Chapter 4

Conclusions

In this thesis, we proposed the bandwidth selection procedure for estimating intensity estimation and LISA function. We used double bootstrap – smooth bootstrap and Bayesian bootstrap – for estimating intensity and empirical (spatial) bootstrap for estimating LISA function.

When applying smoothed bootstrap, we have to generate bootstrap sample from $\hat{\lambda}_{0,\mathbf{G}}(x)$ where $\mathbf{G}$ denotes a pilot smoothing matrix which determines the random noise applied to the events.

There is a limitation for selecting optimal bandwidth of kernel estimator of pair correlation LISA function. Just as in the case study, if the points are extremely clustered, guessing range of r might be intractable. Moreover, to apply an empirical method, we have to set the range discretely. As a result, it may be possible to omit potential optimal bandwidth.

To apply Bayesian bootstrap, we use a hyperparameter α. Hyperparameter optimization is not clearly solved normally and we used grid search to find α. Through our research, $\alpha \leq 5$ is an ap-

propriate range to apply grid search. Since the values of entries in bandwidth matrix is getting smaller as α gets larger, quite large α might cause undersmoothing problem. We leave a better optimal choice of α for future work.

Appendix A

Proofs

A.1 Proof of Proposition 2.2.1

Proof. By the law of total expectation,

$$
\begin{aligned}
\mathbf{E}\,\hat{\lambda}_{0,\mathbf{H}}(\mathbf{x}) &= \mathbf{E}\left(\frac{1}{p_{\mathbf{H}}(\mathbf{x})}\sum_{i=1}^{N} W_i K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_i) I(N\neq 0)\right) \\
&= \frac{1}{p_{\mathbf{H}}(\mathbf{x})}\,\mathbf{E}\left(\mathbf{E}\left(\sum_{i=1}^{N} W_i K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_i)\;\middle|\; N>0\right)\right) \\
&= \frac{1}{p_{\mathbf{H}}(\mathbf{x})}\sum_{k=1}^{\infty}\mathbf{E}\left(\sum_{i=1}^{N} W_i K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_i)\;\middle|\; N=k\right)\mathbf{P}(N=k) \\
&= \frac{1}{p_{\mathbf{H}}(\mathbf{x})}\sum_{k=1}^{\infty}\sum_{i=1}^{k}\mathbf{E}(W_i)\mathbf{E}(K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}))\,\mathbf{P}(N=k) \\
&= \sum_{k=1}^{\infty}(K_{\mathbf{H}} * \lambda_0)(\mathbf{x})\,\mathbf{P}(N=k) \\
&= (1-e^{-m})(K_{\mathbf{H}} * \lambda_0)(\mathbf{x})
\end{aligned}
$$

□

A.2 Proof of Proposition 2.2.2

Proof.

$$\begin{aligned}
\mathbf{E}\left(\left(\hat{\lambda}_{0,\mathbf{H}}(\mathbf{x})\right)^2\right) &= \mathbf{E}\left(\left(\sum_{i=1}^{N} W_i K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_i) I(N\neq 0)\right)^2\right) \\
&= \left(\mathbf{E}\left(\sum_{i=1}^{N} W_i^2 K_{\mathbf{H}}^2(\mathbf{x}-\mathbf{X}_i) \,\middle|\, N>0\right)\right. \\
&\qquad \left. + \mathbf{E}\left(\sum_{i\neq j}^{N} W_i W_j K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_i) K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_j) \,\middle|\, N>0\right)\right) \\
&= \left(\sum_{k=1}^{\infty}\left(\sum_{i=1}^{k} \mathbf{E}(W_i^2)\mathbf{E}(K_{\mathbf{H}}^2(\mathbf{x}-\mathbf{X}_i))\right)\mathbf{P}(N=k)\right. \\
&\qquad \left. + \sum_{k=1}^{\infty}\left(\sum_{i\neq j}^{k} \mathbf{E}(W_i W_j)\mathbf{E}(K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_i))\mathbf{E}(K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_j))\right)\mathbf{P}(\right. \\
&= \sum_{k=1}^{\infty} \frac{e^{-\hat{m}}\hat{m}^k}{k!}\left(\sum_{i=1}^{k} \frac{2}{k(k+1)}\mathbf{E}(K_{\mathbf{H}}^2(\mathbf{x}-\mathbf{X}_i))\right. \\
&\qquad \left. + \sum_{i\neq j}^{k} \frac{1}{k(k+1)}\mathbf{E}(K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_i))\mathbf{E}(K_{\mathbf{H}}(\mathbf{x}-\mathbf{X}_j))\right) \\
&= \sum_{k=1}^{\infty} \frac{e^{-\hat{m}}\hat{m}^k}{k!}\left(\frac{2}{k+1}(K_{\mathbf{H}}^2 * \lambda_0)(\mathbf{x}) + \frac{k-1}{k+1}(K_{\mathbf{H}} * \lambda_0)^2(\mathbf{x})\right) \\
&= 2(K_{\mathbf{H}}^2 * \lambda_0)(\mathbf{x})\ \mathbf{E}\left(\frac{1}{N+1} I(N\neq 0)\right) \\
&\qquad + (K_{\mathbf{H}} * \lambda_0)^2(\mathbf{x})\ \mathbf{E}\left(\frac{N-1}{N+1} I(N\neq 0)\right) \\
&= \left(2C_\alpha(\hat{m})(K_{\mathbf{H}}^2 * \lambda_0)(\mathbf{x}) + (1-e^{-m}-2C_\alpha(\hat{m}))(K_{\mathbf{H}} * \lambda_0)^2(\mathbf{x})\right)
\end{aligned}$$

where $C_\alpha(\hat{m}) := \mathbf{E}\left(\frac{1}{\alpha N+1} I(N\neq 0)\right) \to 0$ as $\hat{m}\to\infty$. Thus,

$$\begin{aligned}\mathbf{Var}\left(\hat{\lambda}_{0,\mathbf{H}}(\mathbf{x})\right) &= \mathbf{E}\left(\left(\hat{\lambda}_{0,\mathbf{H}}(\mathbf{x})\right)^2\right) - \left(\mathbf{E}\,\hat{\lambda}_{0,\mathbf{H}}(\mathbf{x})\right)^2 \\ &= \left(2C_\alpha(\hat{m})(K_{\mathbf{H}}^2 * \lambda_0)(\mathbf{x}) - (2C_\alpha(\hat{m}) - e^{-\hat{m}} + e^{-2\hat{m}})(K_{\mathbf{H}} * \lambda_0)^2(\mathbf{x})\right)\end{aligned}$$

□

A.3 Proof of Theorem 2.2.3

Proof. Let $\hat{\lambda}_{0,\mathbf{G}}(x)$ be the kernel estimator of $\lambda_0(x)$ and $\hat{\lambda}^B_{0,\mathbf{H}}(x)$ its smoothed + Bayesian bootstrap counterpart. As N^* follows $\mathsf{Poisson}(\int_W \hat{\lambda}_{\mathbf{G}}(x)dx) = \mathsf{Poisson}(\hat{m})$, then we obtain

$$\begin{aligned}p_{\mathbf{H}}(x)\mathbf{E}^B\left(\hat{\lambda}^B_{0,\mathbf{H}}(x)\right) &= (1-e^{-\hat{m}})|\mathbf{H}|^{-1/2}\int_W K(\mathbf{H}^{-1/2}(x-y))\hat{\lambda}_{0,\mathbf{G}}(y)dy \\ &= (1-e^{-\hat{m}})\int_{B_{x,\mathbf{H}}} K(\mathbf{H}^{-1/2}(x-y))\hat{\lambda}_{0,\mathbf{G}}(y)dy\end{aligned}$$

where $B_{x,\mathbf{H}} = \{\mathbf{H}^{-1}(x-y); y \in W\}$ and $p_{\mathbf{H}}(x) = \int_W K_{\mathbf{H}}(x-y)dy = \int_{B_{x,\mathbf{H}}} K(u)du$.

A second order Taylor expansion yields the following expression for the bias of $\hat{\lambda}^B_{0,\mathbf{H}}(x)$

$$\begin{aligned}B^B(x,\mathbf{H}) = e^{-\hat{m}}\hat{\lambda}_{0,\mathbf{G}}(x) &- \frac{1-e^{-\hat{m}}}{p_{\mathbf{H}}(x)}\int_{B_{x,\mathbf{H}}} u^t\mathbf{H}^{1/2}D\hat{\lambda}_{0,\mathbf{G}}(x)K(u)du \\ &+ \frac{1-e^{-\hat{m}}}{p_{\mathbf{H}}(x)}\left(\frac{1}{2}\int_{B_{x,\mathbf{H}}} u^t\mathbf{H}^{1/2}D\hat{\lambda}_{0,\mathbf{G}}(x)K(u)du + o(tr(\mathbf{H}))\right)\end{aligned}$$

$\hat{\lambda}^B_{0,\mathbf{H}}(x)$ is asymptotically unbiased provided that, by regularity conditions, K is symmetric and $\int_{\mathbb{R}^2} uu^tK(u)du = \mu_2(K)I_2$, with

$\mu_2(K) < \infty$, thus

$$B^B(x, \mathbf{H}) = \mathbf{E}^B\left(\hat{\lambda}^B_{0,\mathbf{H}}(x)\right) - \hat{\lambda}_{0,\mathbf{G}}(x)$$
$$= \frac{1 - e^{-\hat{m}}}{2}\mu_2(K)^2 tr(\mathbf{H}D^2\hat{\lambda}_{0,\mathbf{G}}(x)) + o_p((1 - e^{-\hat{m}})tr(\mathbf{H}))$$

has order $O_p(tr(\mathbf{H}))$. The variance of $\hat{\lambda}^B_{0,\mathbf{H}}(x)$ is given by

$$\mathbf{Var}\left(\hat{\lambda}_{0,\mathbf{H}}(\mathbf{x})\right) = 2C_\alpha(\hat{m})\int_W K^2_{\mathbf{H}}(x - y)\hat{\lambda}_{0,\mathbf{G}}(y)dy$$
$$- (2C_\alpha(\hat{m}) - e^{-\hat{m}} + e^{-2\hat{m}})\left(\int_W K_{\mathbf{H}}(x - y)\hat{\lambda}_{0,\mathbf{G}}(y)dy\right)^2$$

By Taylor expansion, the first term in the right hand side of the previous expression is given by

$$\int_W K^2_{\mathbf{H}}(x - y)\hat{\lambda}_{0,\mathbf{G}}(y)dy = |\mathbf{H}|^{-1/2}\int_{B_{x,\mathbf{H}}} K(u)^2\hat{\lambda}_{0,\mathbf{G}}(x - u^t\mathbf{H}^{1/2})du$$
$$= |\mathbf{H}|^{-1/2}\hat{\lambda}_{0,\mathbf{G}}(x)\int_{B_{x,\mathbf{H}}} K(u)^2du + o_p(|\mathbf{H}|^{-1/2})$$

Given that $\hat{\lambda}_{0,\mathbf{G}}(x)$, as a consistent estimator of $\lambda_0(x)$, is bounded and K is a bivariate density function, the integral in the second term in the right hand side of $\mathbf{Var}^B(x, \mathbf{H})$ has order 1 and

$$\mathbf{Var}^B(x, \mathbf{H}) = \frac{C_\alpha(\hat{m})|\mathbf{H}|^{-1/2}}{p_{\mathbf{H}}(x)}\hat{\lambda}_{0,\mathbf{G}}(x)\int_{B_{x,\mathbf{H}}} K(u)^2du$$
$$+ o_p(C_\alpha(\hat{m})|\mathbf{H}|^{-1/2})$$

tends to 0 as, by $m^{-1}|\mathbf{H}|^{-1/2} \to 0$ when $m \to \infty$. When $W = \mathbb{R}^2$, the variance is

$$\mathbf{Var}^B(x, \mathbf{H}) = C_\alpha(\hat{m})|\mathbf{H}|^{-1/2}\hat{\lambda}_{0,\mathbf{G}}(x)R(K) + o_p(C_\alpha(\hat{m})|\mathbf{H}|^{-1/2})$$

Therefore, if we replace $B^B(x, \mathbf{H})$ and $\mathbf{Var}^B(x, \mathbf{H})$ in the equation for computing MISE and let $\hat{m} \to \infty$, we obtain expression

$$AMISE^B(\mathbf{H}; \alpha) = \frac{1}{4}\mu_2(K)^2(vech\mathbf{H})^t\hat{\Psi}_{4,\mathbf{G}}(vech\mathbf{H}) + C_\alpha(\hat{m})|\mathbf{H}|^{-1/2}R(K).$$

□

A.4 Proof of Proposition 2.3.1

Proof. Assume that $\mathbf{H} \in \mathcal{D}$ and let $\mathbf{H} = diag(h_1^2, h_2^2)$, then AMISE is

$$\text{AMISE}(\mathbf{H}; \alpha) = \frac{1}{4}\mu_2(K)^2(h_1^4\psi_{40} \ + \ 2h_1^2h_2^2\psi_{22} \ + \ h_2^4\psi_{04}) + \frac{C_\alpha(m)}{h_1h_2}R(K)$$

Denote $\mathbf{H}_{\text{AMISE}} = \arg\min_{\mathbf{H}\in\mathcal{D}} \text{AMISE}(\mathbf{H}; \alpha) = diag(h_{1,\text{AMISE}}^2, h_{2,\text{AMISE}}^2)$, then the gradient of AMISE would be zero to minimize AMISE:

$$\mu_2(K)^2(h_1^3\psi_{40} + h_1h_2^2\psi_{22}) - \frac{C_\alpha(m)}{h_1^2h_2} = 0$$
$$\mu_2(K)^2(h_1^2h_2\psi_{22} + h_2^3\psi_{04}) - \frac{C_\alpha(m)}{h_1h_2^2} = 0$$

We can obtain $h_{2,\text{AMISE}} = h_{1,\text{AMISE}}(\psi_{40}/\psi_{04})^{1/4}$ and

$$\mu_2(K)^2\left(h_1^6\frac{\psi_{40}^{5/4}\psi_{04}^{1/2}}{\psi_{04}^{3/4}} + h_1^6\frac{\psi_{40}^{3/4}}{\psi_{04}^{3/4}}\psi_{22}\right) - \frac{C_\alpha(m)R(K)}{h_1h_2^2} = 0$$

solving the equation and yields

$$h_{1,\text{AMISE}} = \left[\frac{\psi_{04}^{3/4}C_\alpha(m)R(K)}{\mu_2(K)^2\psi_{40}^{3/4}(\psi_{40}^{1/2}\psi_{04}^{1/2} + \psi_{22})}\right]^{1/6}$$
$$h_{2,\text{AMISE}} = \left[\frac{\psi_{40}^{3/4}C_\alpha(m)R(K)}{\mu_2(K)^2\psi_{04}^{3/4}(\psi_{40}^{1/2}\psi_{04}^{1/2} + \psi_{22})}\right]^{1/6}$$

Since $C_\alpha(m)$ can be estimated by $\frac{1}{\alpha n+1}$, we can notice that $h_{1,\text{AMISE}}$ and $h_{2,\text{AMISE}}$ gets smaller as α gets larger.

□

Appendix B

Supplementary information

B.1 Mean of optimal bandwidths and ISEs in the simulation study

The tables below are the results of mean of optimal bandwidths and MISEs for each intensity functions. The least MISE is expressed in bold.

		h			MISE
Diggle		4.519E-02			5.476E-02
PLCV		9.668E-02			2.764E-02
		h_1^2	-	h_2^2	MISE
Scott		7.420E-02	-	9.074E-02	2.301E-02
LSCV(Diag)		1.270E-04	-	1.535E-02	6.185E-02
Original	Plug-in	2.215E-03	-	4.903E-03	3.161E-02
	SCV	2.527E-03	-	6.731E-03	2.676E-02
Smoothed bootstrap	Plug-in	4.547E-03	-	8.017E-03	2.314E-02
	SCV	5.008E-03	-	9.978E-03	**2.175E-02**
Proposed method, $\alpha = 1$	Plug-in	4.544E-03	-	8.015E-03	2.314E-02
	SCV	5.006E-03	-	9.972E-03	2.176E-02
Proposed method, $\alpha = 2$	Plug-in	3.607E-03	-	6.362E-03	2.543E-02
	SCV	3.861E-03	-	7.467E-03	2.385E-02
Proposed method, $\alpha = 3$	Plug-in	3.151E-03	-	5.558E-03	2.738E-02
	SCV	3.320E-03	-	6.374E-03	2.575E-02
Proposed method, $\alpha = 4$	Plug-in	2.863E-03	-	5.050E-03	2.904E-02
	SCV	2.986E-03	-	5.685E-03	2.748E-02
Proposed method, $\alpha = 5$	Plug-in	2.658E-03	-	4.688E-03	3.050E-02
	SCV	2.750E-03	-	5.216E-03	2.898E-02
		h_{11}	h_{21}	h_{22}	MISE
LSCV(Full)		7.016E-04	-3.989E-05	4.882E-03	2.995E-01
Original	Plug-in	2.218E-03	-1.506E-05	4.910E-03	2.676E-02
	SCV	2.540E-03	-1.711E-05	6.695E-03	2.684E-02
Smoothed bootstrap	Plug-in	4.556E-03	-5.493E-05	8.039E-03	2.315E-02
	SCV	5.048E-03	-6.396E-05	9.858E-03	2.188E-02
Proposed method, $\alpha = 1$	Plug-in	4.554E-03	-5.490E-05	8.036E-03	2.315E-02
	SCV	5.045E-03	-6.341E-05	9.855E-03	2.188E-02
Proposed method, $\alpha = 2$	Plug-in	3.615E-03	-4.371E-05	6.380E-03	2.543E-02
	SCV	3.876E-03	-4.703E-05	7.427E-03	2.391E-02
Proposed method, $\alpha = 3$	Plug-in	3.158E-03	-3.810E-05	5.574E-03	2.737E-02
	SCV	3.324E-03	-4.821E-05	6.433E-03	2.567E-02
Proposed method, $\alpha = 4$	Plug-in	2.870E-03	-3.461E-05	5.064E-03	2.903E-02
	SCV	2.970E-03	-3.424E-05	5.803E-03	2.735E-02
Proposed method, $\alpha = 5$	Plug-in	2.664E-03	-3.204E-05	4.701E-03	3.048E-02
	SCV	2.738E-03	-2.455E-05	5.230E-03	2.889E-02

Table B1: Results for various bandwidth selectors from 100 simulations, inhomogeneous Poisson point processes with intensity $\lambda_1(x)$

		h			MISE
Diggle		4.854E-02			1.019E-01
PLCV		9.047E-02			7.086E-02
		h_1^2	-	h_2^2	MISE
Scott		7.486E-02	-	9.124E-02	6.560E-02
LSCV(Diag)		2.731E-04	-	1.430E-02	1.085E-01
Original	Plug-in	2.261E-03	-	4.868E-03	7.935E-02
	SCV	2.581E-03	-	6.565E-03	7.320E-02
Smoothed bootstrap	Plug-in	4.677E-03	-	7.973E-03	6.711E-02
	SCV	5.158E-03	-	9.763E-03	**6.454E-02**
Proposed method, $\alpha = 1$	Plug-in	4.676E-03	-	7.970E-03	6.711E-02
	SCV	5.155E-03	-	9.764E-03	6.454E-02
Proposed method, $\alpha = 2$	Plug-in	3.712E-03	-	6.327E-03	7.103E-02
	SCV	3.972E-03	-	7.356E-03	6.868E-02
Proposed method, $\alpha = 3$	Plug-in	3.243E-03	-	5.527E-03	7.385E-02
	SCV	3.412E-03	-	6.260E-03	7.170E-02
Proposed method, $\alpha = 4$	Plug-in	2.946E-03	-	5.022E-03	7.610E-02
	SCV	3.069E-03	-	5.583E-03	7.419E-02
Proposed method, $\alpha = 5$	Plug-in	2.735E-03	-	4.662E-03	7.799E-02
	SCV	2.826E-03	-	5.109E-03	7.626E-02
		h_{11}	h_{21}	h_{22}	MISE
LSCV(Full)		7.372E-04	-1.371E-06	5.064E-03	2.681E-01
Original	Plug-in	2.269E-03	-1.600E-05	4.884E-03	7.937E-02
	SCV	2.594E-03	-9.237E-06	6.555E-03	7.333E-02
Smoothed bootstrap	Plug-in	4.698E-03	-1.014E-05	8.012E-03	6.712E-02
	SCV	5.208E-03	1.463E-05	9.723E-03	6.472E-02
Proposed method, $\alpha = 1$	Plug-in	4.696E-03	-1.014E-05	8.009E-03	6.713E-02
	SCV	5.203E-03	1.384E-05	9.723E-03	6.472E-02
Proposed method, $\alpha = 2$	Plug-in	3.728E-03	-8.144E-06	6.358E-03	7.102E-02
	SCV	3.994E-03	1.657E-06	7.357E-03	6.877E-02
Proposed method, $\alpha = 3$	Plug-in	3.257E-03	-7.044E-06	5.555E-03	7.383E-02
	SCV	3.417E-03	2.860E-06	6.326E-03	7.166E-02
Proposed method, $\alpha = 4$	Plug-in	2.959E-03	-6.390E-06	5.047E-03	7.607E-02
	SCV	3.073E-03	-2.039E-06	5.722E-03	7.396E-02
Proposed method, $\alpha = 5$	Plug-in	2.747E-03	-5.881E-06	4.686E-03	7.795E-02
	SCV	2.810E-03	-1.000E-05	5.292E-03	7.587E-02

Table B2: Results for various bandwidth selectors from 100 simulations, Thomas cluster point processes with intensity $\lambda_1(x)$

		h			MISE
Diggle		5.068E-02			8.076E-02
PLCV		5.755E-02			8.748E-02
		h_1^2	-	h_2^2	MISE
Scott		3.870E-02	-	9.060E-02	4.668E-02
LSCV(Diag)		1.452E-03	-	2.847E-03	7.395E-02
Original	Plug-in	1.105E-03	-	4.741E-03	5.127E-02
	SCV	1.228E-03	-	5.408E-03	4.929E-02
Smoothed bootstrap	Plug-in	1.567E-03	-	6.093E-03	4.971E-02
	SCV	1.730E-03	-	6.972E-03	5.032E-02
Proposed method, $\alpha = 1$	Plug-in	1.567E-03	-	6.091E-03	4.971E-02
	SCV	1.730E-03	-	6.969E-03	5.032E-02
Proposed method, $\alpha = 2$	Plug-in	1.244E-03	-	4.835E-03	5.071E-02
	SCV	1.336E-03	-	5.231E-03	4.999E-02
Proposed method, $\alpha = 3$	Plug-in	1.087E-03	-	4.224E-03	5.306E-02
	SCV	1.152E-03	-	4.445E-03	5.210E-02
Proposed method, $\alpha = 4$	Plug-in	9.874E-04	-	3.838E-03	5.549E-02
	SCV	1.035E-03	-	3.984E-03	5.449E-02
Proposed method, $\alpha = 5$	Plug-in	9.166E-04	-	3.563E-03	5.780E-02
	SCV	9.536E-04	-	3.636E-03	5.699E-02
		h_{11}	h_{21}	h_{22}	MISE
LSCV(Full)		1.533E-03	-3.368E-04	1.993E-03	8.460E-02
Original	Plug-in	1.292E-03	-1.022E-03	5.824E-03	4.535E-02
	SCV	1.384E-03	-1.136E-03	6.404E-03	4.371E-02
Smoothed bootstrap	Plug-in	1.770E-03	-1.141E-03	7.067E-03	4.394E-02
	SCV	1.908E-03	-1.294E-03	7.868E-03	4.363E-02
Proposed method, $\alpha = 1$	Plug-in	1.770E-03	-1.140E-03	7.065E-03	4.394E-02
	SCV	1.905E-03	-1.290E-03	7.863E-03	**4.363E-02**
Proposed method, $\alpha = 2$	Plug-in	1.405E-03	-9.052E-04	5.608E-03	4.563E-02
	SCV	1.461E-03	-9.621E-04	5.905E-03	4.497E-02
Proposed method, $\alpha = 3$	Plug-in	1.227E-03	-7.908E-04	4.899E-03	4.816E-02
	SCV	1.259E-03	-8.151E-04	5.025E-03	4.775E-02
Proposed method, $\alpha = 4$	Plug-in	1.115E-03	-7.185E-04	4.451E-03	5.062E-02
	SCV	1.129E-03	-7.301E-04	4.495E-03	5.044E-02
Proposed method, $\alpha = 5$	Plug-in	1.035E-03	-6.670E-04	4.132E-03	5.292E-02
	SCV	1.044E-03	-6.664E-04	4.127E-03	5.294E-02

Table B3: Results for various bandwidth selectors from 100 simulations, inhomogeneous Poisson point processes with intensity $\lambda_2(x)$

		h			MISE
Diggle		4.820E-02			1.716E-01
PLCV		5.687E-02			1.774E-01
		h_1^2	-	h_2^2	MISE
Scott		3.948E-02	-	9.063E-02	1.331E-01
LSCV(Diag)		1.375E-03	-	2.054E-03	1.908E-01
Original	Plug-in	1.104E-03	-	4.508E-03	1.482E-01
	SCV	1.220E-03	-	5.081E-03	1.442E-01
Smoothed bootstrap	Plug-in	1.582E-03	-	5.918E-03	1.407E-01
	SCV	1.741E-03	-	6.767E-03	1.387E-01
Proposed method, $\alpha = 1$	Plug-in	1.581E-03	-	5.916E-03	1.407E-01
	SCV	1.741E-03	-	6.764E-03	1.387E-01
Proposed method, $\alpha = 2$	Plug-in	1.255E-03	-	4.697E-03	1.463E-01
	SCV	1.348E-03	-	5.090E-03	1.440E-01
Proposed method, $\alpha = 3$	Plug-in	1.097E-03	-	4.103E-03	1.511E-01
	SCV	1.157E-03	-	4.344E-03	1.490E-01
Proposed method, $\alpha = 4$	Plug-in	9.965E-04	-	3.728E-03	1.552E-01
	SCV	1.040E-03	-	3.877E-03	1.534E-01
Proposed method, $\alpha = 5$	Plug-in	9.250E-04	-	3.461E-03	1.588E-01
	SCV	9.591E-04	-	3.554E-03	1.573E-01
		h_{11}	h_{21}	h_{22}	MISE
LSCV(Full)		1.500E-03	-2.695E-04	2.055E-03	2.210E-01
Original	Plug-in	1.275E-03	-9.178E-04	5.461E-03	1.398E-01
	SCV	1.359E-03	-1.003E-03	5.916E-03	1.372E-01
Smoothed boostrap	Plug-in	1.776E-03	-1.082E-03	6.799E-03	1.332E-01
	SCV	1.904E-03	-1.208E-03	7.533E-03	1.311E-01
Proposed method, $\alpha = 1$	Plug-in	1.776E-03	-1.082E-03	6.797E-03	1.332E-01
	SCV	1.901E-03	-1.207E-03	7.534E-03	**1.311E-01**
Proposed method, $\alpha = 2$	Plug-in	1.410E-03	-8.587E-04	5.396E-03	1.395E-01
	SCV	1.461E-03	-9.050E-04	5.683E-03	1.380E-01
Proposed method, $\alpha = 3$	Plug-in	1.231E-03	-7.502E-04	4.714E-03	1.445E-01
	SCV	1.260E-03	-7.707E-04	4.825E-03	1.437E-01
Proposed method, $\alpha = 4$	Plug-in	1.119E-03	-6.816E-04	4.283E-03	1.487E-01
	SCV	1.137E-03	-6.868E-04	4.329E-03	1.482E-01
Proposed method, $\alpha = 5$	Plug-in	1.039E-03	-6.327E-04	3.976E-03	1.523E-01
	SCV	1.049E-03	-6.347E-04	3.968E-03	1.523E-01

Table B4: Results for various bandwidth selectors from 100 simulations, Thomas cluster point processes with intensity $\lambda_2(x)$

		h			MISE
Diggle		1.458E-02			8.143E-01
PLCV		2.946E-02			2.042E+00
		h_1^2	-	h_2^2	MISE
Scott		3.135E-02	-	1.027E-01	2.959E+00
LSCV(Diag)		4.623E-14	-	1.308E-01	2.078E+00
Original	Plug-in	1.133E-04	-	1.741E-03	4.850E-01
	SCV	1.478E-04	-	2.568E-03	5.990E-01
Smoothed bootstrap	Plug-in	2.472E-04	-	2.778E-03	8.535E-01
	SCV	2.871E-04	-	3.714E-03	1.051E+00
Proposed method, $\alpha = 1$	Plug-in	2.199E-04	-	2.462E-03	8.532E-01
	SCV	2.555E-04	-	3.296E-03	1.051E+00
Proposed method, $\alpha = 2$	Plug-in	1.745E-04	-	1.954E-03	6.781E-01
	SCV	1.982E-04	-	2.502E-03	7.967E-01
Proposed method, $\alpha = 3$	Plug-in	1.525E-04	-	1.707E-03	6.023E-01
	SCV	1.713E-04	-	2.138E-03	6.872E-01
Proposed method, $\alpha = 4$	Plug-in	1.385E-04	-	1.551E-03	5.595E-01
	SCV	1.544E-04	-	1.917E-03	6.244E-01
Proposed method, $\alpha = 5$	Plug-in	1.286E-04	-	1.440E-03	5.323E-01
	SCV	1.428E-04	-	1.760E-03	5.842E-01
		h_{11}	h_{21}	h_{22}	MISE
LSCV(Full)		2.259E-04	-6.502E-04	3.292E-03	1.619E+00
Original	Plug-in	2.413E-04	-7.262E-04	3.643E-03	2.535E-01
	SCV	3.368E-04	-1.174E-03	5.898E-03	**2.283E-01**
Smoothed bootstrap	Plug-in	4.153E-04	-9.487E-04	4.920E-03	4.834E-01
	SCV	4.891E-04	-1.269E-03	6.553E-03	4.984E-02
Proposed method, $\alpha = 1$	Plug-in	3.673E-04	-8.333E-04	4.329E-03	4.832E-01
	SCV	4.334E-04	-1.120E-03	5.788E-03	4.984E-01
Proposed method, $\alpha = 2$	Plug-in	2.915E-04	-6.615E-04	3.436E-03	3.873E-01
	SCV	3.324E-04	-8.456E-04	4.389E-03	3.880E-01
Proposed method, $\alpha = 3$	Plug-in	2.547E-04	-5.779E-04	3.002E-03	3.497E-01
	SCV	2.832E-04	-7.098E-04	3.687E-03	3.448E-01
Proposed method, $\alpha = 4$	Plug-in	2.314E-04	-5.251E-04	2.728E-03	3.302E-01
	SCV	2.547E-04	-6.317E-04	3.280E-03	3.242E-01
Proposed method, $\alpha = 5$	Plug-in	2.148E-04	-4.874E-04	2.532E-03	3.190E-01
	SCV	2.359E-04	-5.906E-04	3.075E-03	3.095E-01

Table B5: Results for various bandwidth selectors from 100 simulations, inhomogeneous Poisson point processes with intensity $\lambda_3(x)$

		h			MISE
Diggle		5.303E-02			1.058E-01
PLCV		9.508E-02			1.027E-01
		h_1^2	-	h_2^2	MISE
Scott		9.501E-02	-	9.560E-02	1.017E-01
LSCV(Diag)		2.932E-03	-	2.509E-03	1.080E-01
Original	Plug-in	5.978E-03	-	5.643E-03	**9.902E-02**
	SCV	7.223E-03	-	6.594E-03	9.977E-02
Smoothed bootstrap	Plug-in	1.087E-02	-	1.069E-02	1.029E-01
	SCV	1.261E-02	-	1.249E-02	1.041E-01
Proposed method, $\alpha = 1$	Plug-in	1.087E-02	-	1.069E-02	1.029E-01
	SCV	1.260E-02	-	1.250E-02	1.041E-01
Proposed method, $\alpha = 2$	Plug-in	8.627E-03	-	8.484E-03	1.012E-01
	SCV	9.561E-03	-	9.405E-03	1.020E-01
Proposed method, $\alpha = 3$	Plug-in	7.537E-03	-	7.412E-03	1.003E-01
	SCV	8.138E-03	-	8.018E-03	1.009E-01
Proposed method, $\alpha = 4$	Plug-in	6.848E-03	-	6.734E-03	9.980E-02
	SCV	7.285E-03	-	7.168E-03	1.002E-01
Proposed method, $\alpha = 5$	Plug-in	6.357E-03	-	6.252E-03	9.941E-02
	SCV	6.703E-03	-	6.561E-03	9.967E-02
		h_{11}	h_{21}	h_{22}	MISE
LSCV(Full)		3.049E-03	8.535E-05	2.448E-03	1.176E-01
Original	Plug-in	6.004E-03	-1.613E-04	5.671E-03	9.904E-02
	SCV	7.186E-03	-1.484E-04	6.631E-03	9.978E-02
Smoothed bootstrap	Plug-in	1.094E-02	-3.217E-04	1.077E-02	1.029E-01
	SCV	1.261E-02	-3.854E-04	1.253E-02	1.041E-01
Proposed method, $\alpha = 1$	Plug-in	1.094E-02	-3.216E-04	1.076E-02	1.029E-01
	SCV	1.260E-02	-3.878E-04	1.253E-02	1.041E-01
Proposed method, $\alpha = 2$	Plug-in	8.685E-03	-2.556E-04	8.546E-03	1.012E-01
	SCV	9.539E-03	-2.822E-04	9.440E-03	1.019E-01
Proposed method, $\alpha = 3$	Plug-in	7.587E-03	-2.234E-04	7.466E-03	1.004E-01
	SCV	8.163E-03	-2.462E-04	8.034E-03	1.008E-01
Proposed method, $\alpha = 4$	Plug-in	6.894E-03	-2.027E-04	6.784E-03	9.981E-02
	SCV	7.301E-03	-2.208E-04	7.189E-03	1.001E-01
Proposed method, $\alpha = 5$	Plug-in	6.400E-03	-1.883E-04	6.297E-03	9.943E-02
	SCV	6.698E-03	-1.874E-04	6.625E-03	9.968E-02

Table B6: Results for various bandwidth selectors from 100 simulations, log-Gaussian Cox point processes with intensity $\lambda_4(x)$

		h			MISE
Diggle		4.336E-02			**8.286E-02**
PLCV		4.199E-02			8.312E-02
		h_1^2	-	h_2^2	MISE
Scott		9.010E-02	-	9.006E-02	2.081E-01
LSCV(Diag)		1.648E-03	-	2.055E-03	8.995E-02
Original	Plug-in	2.886E-03	-	2.887E-03	8.823E-02
	SCV	2.978E-03	-	2.980E-03	8.985E-02
Smoothed bootstrap	Plug-in	6.322E-03	-	6.261E-03	1.654E-01
	SCV	6.965E-03	-	6.876E-03	1.804E-01
Proposed method, $\alpha = 1$	Plug-in	6.320E-03	-	6.258E-03	1.654E-01
	SCV	6.963E-03	-	6.873E-03	1.803E-01
Proposed method, $\alpha = 2$	Plug-in	5.017E-03	-	4.968E-03	1.341E-01
	SCV	5.283E-03	-	5.216E-03	1.403E-01
Proposed method, $\alpha = 3$	Plug-in	4.383E-03	-	4.340E-03	1.192E-01
	SCV	4.490E-03	-	4.453E-03	1.218E-01
Proposed method, $\alpha = 4$	Plug-in	3.982E-03	-	3.944E-03	1.100E-01
	SCV	4.039E-03	-	3.982E-03	1.111E-01
Proposed method, $\alpha = 5$	Plug-in	3.697E-03	-	3.661E-03	1.038E-01
	SCV	3.713E-03	-	3.665E-03	1.040E-01
		h_{11}	h_{21}	h_{22}	MISE
LSCV(Full)		1.579E-03	8.383E-06	1.655E-03	1.372E-01
Original	Plug-in	2.896E-03	3.504E-05	2.897E-03	8.870E-02
	SCV	2.990E-03	4.151E-05	2.997E-03	9.047E-02
Smoothed bootstrap	Plug-in	6.372E-03	6.954E-05	6.308E-03	1.665E-01
	SCV	7.017E-03	8.200E-05	6.946E-03	1.816E-01
Proposed method, $\alpha = 1$	Plug-in	6.370E-03	6.952E-05	6.306E-03	1.665E-01
	SCV	7.019E-03	8.706E-05	6.949E-03	1.816E-01
Proposed method, $\alpha = 2$	Plug-in	5.056E-03	5.516E-05	5.006E-03	1.315E-01
	SCV	5.339E-03	6.457E-05	5.270E-03	1.418E-01
Proposed method, $\alpha = 3$	Plug-in	4.417E-03	4.829E-05	4.373E-03	1.202E-01
	SCV	4.536E-03	5.843E-05	4.528E-03	1.235E-01
Proposed method, $\alpha = 4$	Plug-in	4.014E-03	4.368E-05	3.974E-03	1.110E-01
	SCV	4.085E-03	4.803E-03	4.026E-03	1.125E-01
Proposed method, $\alpha = 5$	Plug-in	3.726E-03	4.056E-05	3.689E-03	1.047E-01
	SCV	3.751E-03	4.630E-05	3.710E-03	1.054E-01

Table B7: Results for various bandwidth selectors from 100 simulations, inhomogeneous Poisson point processes with intensity $\lambda_5(x)$

		h			MISE
Diggle		5.160E-02			6.021E-02
PLCV		4.882E-02			6.100E-02
		h_1^2	-	h_2^2	MISE
Scott		8.508E-02	-	8.517E-02	1.079E-01
LSCV(Diag)		1.983E-03	-	1.950E-03	5.079E-02
Original	Plug-in	3.572E-03	-	3.572E-03	6.342E-02
	SCV	3.712E-03	-	3.716E-03	6.474E-02
Smoothed bootstrap	Plug-in	7.413E-03	-	7.368E-03	1.098E-01
	SCV	8.054E-03	-	7.985E-03	1.180E-01
Proposed method, $\alpha = 1$	Plug-in	7.411E-03	-	7.366E-03	1.097E-01
	SCV	8.049E-03	-	7.988E-03	1.180E-01
Proposed method, $\alpha = 2$	Plug-in	5.883E-03	-	5.847E-03	8.984E-02
	SCV	6.106E-03	-	6.068E-03	9.273E-02
Proposed method, $\alpha = 3$	Plug-in	5.140E-03	-	5.108E-03	8.048E-02
	SCV	5.239E-03	-	5.194E-03	8.165E-02
Proposed method, $\alpha = 4$	Plug-in	4.670E-03	-	4.641E-03	7.484E-02
	SCV	4.688E-03	-	4.650E-03	7.504E-02
Proposed method, $\alpha = 5$	Plug-in	4.335E-03	-	4.309E-03	7.102E-02
	SCV	4.325E-03	-	4.254E-03	7.071E-02
		h_{11}	h_{21}	h_{22}	MISE
LSCV(Full)		2.662E-03	1.259E-03	2.592E-03	6.062E-02
Original	Plug-in	4.136E-03	1.666E-03	4.137E-03	4.947E-02
	SCV	4.513E-03	2.134E-03	4.491E-03	**4.834E-02**
Smoothed bootstrap	Plug-in	7.745E-03	1.724E-03	7.699E-03	9.389E-02
	SCV	8.471E-03	2.116E-03	8.416E-03	9.938E-02
Proposed method, $\alpha = 1$	Plug-in	7.742E-03	1.724E-03	7.696E-03	9.386E-02
	SCV	8.462E-03	2.115E-03	8.414E-03	9.932E-02
Proposed method, $\alpha = 2$	Plug-in	6.146E-03	1.369E-03	6.110E-03	7.625E-02
	SCV	6.476E-03	1.599E-03	6.413E-03	7.772E-02
Proposed method, $\alpha = 3$	Plug-in	5.369E-03	1.195E-03	5.338E-03	6.828E-02
	SCV	5.538E-03	1.361E-03	5.491E-03	6.836E-02
Proposed method, $\alpha = 4$	Plug-in	4.879E-03	1.086E-03	4.849E-03	6.360E-02
	SCV	4.985E-03	1.209E-03	4.916E-03	6.318E-02
Proposed method, $\alpha = 5$	Plug-in	4.529E-03	1.008E-03	4.502E-03	6.051E-02
	SCV	4.562E-03	1.096E-03	4.522E-03	5.984E-02

Table B8: Results for various bandwidth selectors from 100 simulations, inhomogeneous Poisson point processes with intensity $\lambda_6(x)$

Bibliography

Aldershof, B. (1991) *Estimation of integrated squared density derivatives.* Ph.D. thesis, University of North Carolina, Chapel Hill.

Anselin, L. (1995). Local Indicators of Spatial Associations - LISA. *Geographical Analysis* **27(2)**, 93-115.

Baddeley, A., Møller, J. and Waagepetersen, R. (2000). Non- and Semi-Parametric Estimation of Interaction in Inhomogeneous Point Patterns. *Statistica Neerlandica* **54(3)**, 329-350.

Baddeley, A., Turner, R., Møller, J. and Hazelton, M. (2005) Residual Analysis for Spatial Point Processes. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **67**, 617–666.

Baddeley, A. Rubak, E. and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R.* London: Chapman and Hall/CRC Press.

Barr, C. D. and Schoenberg, F. P. (2010). On the Voronoi Estimator for the Intensity of an Inhomogeneous Planar Poisson Process *Biometrika* **97(4)**, 977-984.

Berman, M. and Diggle, P. (1989). Estimating Weighted Integrals of the Second-order Intensity of a Spatial Point Process. *Journal of Royal Statistical Society. Series B (Statistical Methodology)* **51(1)**, 81-92.

Brooks, M. M. and Marron, J. S. (1991). Asymptotic Optimality of the Least-squares Cross-validation Bandwidth for Kernel Estimates of Intensity Functions. *Stochastic Process and their Applications* **38(1)**, 157–165.

Cao, R. (1993). Bootstrapping the Mean Integrated Squared Error. *Journal of Multivariate Analysis* **45(1)**, 137–160.

Cao, R., Cuevas, A. and González Manteiga, W. (1994). A Comparative Study of Several Smoothing Methods in Density Estimation. *Computational Statistics and Data Analysis* **17(2)**, 153–176.

Chacón, J. E. and Duong, T. (2010). Multivariate Plug-in Bandwidth Selection with Unconstrained Pilot Bandwidth Matrices. *Test* **19(2)**, 375-398.

Chacón, J. E. and Duong, T. (2018). *Multivariate Kernel Smoothing and Its Applications*. Chapman & Hall/CRC, London.

Cowling, A., Hall, P. and Phillips, M. J. (1996). Bootstrap Confidence Regions for the Intensity of a Poisson Point Process. *Journal of American Statistical Association* **91(436)**, 1516–1524.

Cressie, N. and Collins, L. (2001a) Analysis of Spatial Point Patterns Using Bundles of Product Density LISA Functions.

Journal of Agricultural, Biological, and Environmental Statistics **6(1)**, 118-135.

Cressie, N. and Collins, L. (2001b) Patterns in Spatial Point Locations: Local Indicators of Spatial Association in a Minefield with Clutter. *Naval Research Logistics* **48**, 333-347.

Cucala, L. (2006). *Espacement bimensionnels et données entachées d'erreurs dans l'analyse des processus ponctuels spatiaux*, Ph.D. thesis, Université des Sciences Sociales - Toulouse I, 2006.

Dasgupta., A. and Raftery, A.E. (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of American Statistical Association* **93(441)**, 294–302.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: the ℓ_1 View, Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics,* Wiley, New York.

Diggle, P. (1985). A Kernel Method for Smoothing Point Process Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **34(2)**, 138-147.

Diggle, P. and Marron, J. S. (1988). Equivalence of Smoothing Parameter Selectors in Density and Intensity Estimation.*Journal of American Statistical Association* **83**, 793–800.

Diggle, P. (1990). A Point Process Modelling Approach to Raised Incidence of a Rare Phenomenon in the Vicinity of a Prespecified Point *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **153(3)**, 349-362.

Diggle, P. and Rowlingson, B. S. (1994). A Conditional Approach to Point Process Modelling of Elevated Risk. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* **157(3)**, 433-440.

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*, (2nd ed.)., A Hodder Arnold Publication, Oxford, UK.

Dudoit, S. and Fridlyand, J. (2003). Bagging to improve the accuracy of a clustering procedure. *Bioinformatics* **19(9)**, 1090–1099.

Duong, T. and Hazelton, M. L. (2003). Plug-in bandwidth matrices for bivariate kernel density estimation. *Journal of Nonparametric Statistics* **15**, 17-30.

Duong, T. and Hazelton, M. L. (2005). Cross-validation Bandwidth Matrices for Multivariate Kernel Density Estimation. *Scandinavian Journal of Statistics* **32**, 485-506.

Duong, T (2022). *ks: Kernel Smoothing*. R package version 1.13.5.

Faraway, J. J. and Jhun, M. (1990). Bootstrap Choice of Bandwidth for Density Estimation. *Journal of the American Statistical Association* **85(412)**, 1119-1122.

Fiksel, T. (1988). Edge Corrected Density Estimators for Point Processes. *A Journal of Theoretical and Applied Statistics* **19(1)**, 67-75.

Fuentes-Santos, I., González-Matniega, W., Mateu, J. (2016). Consistent Smooth Bootstrap Kernel Intensity Estimation for Inho-

mogeneous Spatial Poisson Point Processes. *Scandinavian Journal of Statistics* **43(2)**, 416-435.

González, J., Rodríguez-Cortés, F., Romano, E. and Mateu, J. (2021). Classification of Events Using Local Pair Correlation Function for Spatial Point Patterns. *Journal of Agricultural, Biological, and Environmental Statistics* **26(4)**, 538-559.

Guan, Y. and Loh, J. M. (2007). A Thinned Block Bootstrap Variance Estimation Procedure for Inhomogeneous Spatial Point Patterns. *Journal of American Statistical Association* **102(480)**, 1377–1386.

Hall, P., Marron, J. S. and Park, B. U. (1992). Smoothed Cross-validation. *Probability Theory and Related Fields* **92**, 1-20.

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis* **52(1)**, 258–271.

Jacques, J., Preda, C. (2014). Functional data clustering: a survey. *Advances in Data Analysis and Classification* **8(3)**, 231–255.

Kruskal, J.B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29(2)**, 115–129.

Lahiri, S. N., Kaiser, M. S., Cressie, N. and Hsu, N. J. (1999). Prediction of spatial cumulative distribution functions using subsampling.*Journal of American Statistical Association* **94(445)**, 86–97.

Loader, C. R. (1999). Classical or Plug-in? *The Annals of Statistics* **27(2)**, 415-438.

Loh, J. M. and Stein, M. L. (2004). Bootstrapping a spatial point process. *Statistical Sinica* **14(1)**, 69-101.

Loh, J. M. and Jang, W. (2010). Estimating a cosmological mass bias parameter with bootstrap bandwidth selection. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **59**, 761-779.

Loh, J. M. (2010). Bootstrapping an inhomogeneous point process. *Journal of Statistical Planning and Inference* **140**, 734-749.

Mateu, J., Lorenzo, G., Porcu, E. (2007) Detecting features in spatial point processes with clutter via local indicators of spatial association. *Journal of Computational and Graphical Statistics* **16(4)**, 968–990.

Park, B. U. and Marron, J. S. (1992). On the use of pilot estimators in bandwidth selection. *Journal of Nonparametric Statistics* **1(3)**, 231-240.

Prieto, G. A., Beroza, G. C., Barrett, S. A., López, G. A. and Florez, M. (2012). Earthquake nests as natural laboratories for the study of intermediate-depth earthquake mechanics. *Tectonophysics* **570-571**, 42-56.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability* **13**, 255-266.

Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **39**, 172-212.

Rosenblatt, M. (1956). Remarks on some estimates of a density function. *The Annals of Mathematical Statistics* **27(3)**, 832-837.

Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of Statistics* **9(1)**, 130-134.

Scott, Rb. (1992) Practice Valuation - Thumb Rules and Common-Sense. *Journal of Accountancy* **174(6)**, 74-77.

Sheather, S. J. and Jones, M. C. (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation . *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* ,**53(3)**, 683-690.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall/CRC, London.

Steinwart, I. and Christmann, A. (2008) *Support vector machines.* Springer, New York.

Taylor, C. C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika* **76(4)**, 705-712.

Tenreiro, Carlos. (2003). On the Asymptotic Normality of Multistage Integrated Density Derivatives Kernel Estimators. *Statistics and Probability Letters* **643**, 311-322.

Wand, M. P. (1992). Error analysis for general multtvariate kernel estimators. *Journal of Nonparametic. Statistics* **2(1)**, 1–15.

Wand, M. and Jones, M. (1994). Multivariate plug-in bandwidth selection. *Computational Statistics* **9**, 97–116.

van Lieshout, M. N. M. (2020). Infill asymptotics and bandwidth selection for kernel estimators of spatial intensity functions. *Methodology and Computing in Applied Probability* **22**, 995-1008.

Zarifi, Z., Havskov, J., Hanyga, A. (2007) An insight into the Bucaramanga nest. *Tectonophysics* **443(1)**, 93–105.

국문초록

본 학위논문에서는 공간점과정의 강도함수와 짝상관함수의 LISA 함수를 추정하는 데 있어 커널 추정 사용 시 최적의 띠너비 선택 방법을 제시한다. 특히, 강도함수의 추정에서 베이지안 붓스트랩을 사용하여 작은 자료 크기를 통한 커널 추정에도 평활도를 조절할 수 있는 방법을 제시하며, 짝상관함수의 LISA 함수에서의 커널 추정에서도 평균적분제곱오차를 최소로 만드는 띠너비를 구하는 법에 대해 제안한다. 마지막으로, 우리의 방법을 기존의 방법들과 수치적으로 비교하고 대부분의 경우 제안하는 방법이 기존의 방법들을 능가한다는 것을 보여주며 이를 두 가지 사례 연구에 적용한다.

주 요 어 : 커널 추정, 띠너비 선택, 강도함수, 짝상관함수, 공간점과정

학　　번 : 2015 – 20311