Austrian Journal of Statistics
2026, Volume 55, 24–43.
http://www.ajs.or.at/
doi:10.17713/ajs.v55i2.2146

Boundary Performance of the Beta Kernel Intensity Estimators for Inhomogeneous Poisson Process

Marcel Sihintoé Badiane
Laboratoire de Recherche Appliquée en Sciences de la Nature
Faculté des Sciences
Université de Kindia
BP 212 Kindia (Guinée)

Papa Ngom
Laboratoire de Mathématiques Appliquées
Faculté des Sciences et Techniques
Université Cheikh Anta Diop
BP 5005 Dakar-Fann (Sénégal)

Saa Moussa Tenguiano
Laboratoire de Recherche Appliquée en Sciences de la Nature
Faculté des Sciences
Université de Kindia
BP 212 Kindia (Guinée)

Abstract

We consider the non-parametric estimation of the bivariate intensity function of non-homogeneous Poisson processes in a bounded data. To this end we use the class of kernel estimators with asymmetric beta kernel functions. The beta kernels are non-negative. They change their shape depending on the position on the semi-axis and possess good boundary properties for a wide class of intensity. The theoretical asymptotic properties of the bivariate intensity function like bias and variance are derived. We obtain the optimal bandwidth selection for both estimates as a minimum of the mean integrated squared error (MISE). Numerical studies indicate that the performance of our approach is better, comparing with other bandwidth selection techniques using bias, variance and integrated squared error as criterion. Some applications are made on real datasets.

Keywords: Intensity function, beta kernel, non-homogeneous Poisson process.

1. Introduction

Spatial point patterns arise in many scientific domains. Typical examples include the map of trees in a forest stand, the addresses of individuals infected with some disease, and the locations of cells in a tissue (Diggle 2013; Gelfand, Diggle, Guttorp, and Fuentes 2010; Illian, Penttinen, Stoyan, and Stoyan 2008). The analysis of such patterns usually involves estimating the intensity function, that is, the likelihood of finding a point as a function of location. Sometimes the scientific context suggests a parametric form for the intensity function, perhaps in terms of covariate information. More often, non-parametric estimation is called for. In both cases, it is important to estimate the intensity function in a reliable way. Indeed, this is more urgent than ever in light of the recent development of functional summary statistics that correct for spatial heterogeneity (Baddeley, Møller, and Waagepetersen 2000; Van Lieshout 2011).

Here we focus on non-parametric estimation. A few approaches have been suggested in the literature. For example, one may divide the observation window into quadrats and count the number of points that fall in each. An obvious drawback of this method is its strong dependence on the size and shape of the quadrats. A partial solution is to use spline (Ogata 1998) or kernel smoothing, where the shape of the kernel tends to have little impact; the size parameter, though, remains crucial. Therefore, data-dependent, adaptive methods have been proposed in which the quadrats are replaced by the cells in the pattern's Delaunay tessellation, as shown in a 2007 University of Groningen PhD thesis by W. E. Schaap and in Schaap and Van De Weygaert (2000), or the cells in the pattern's Voronoi tessellation (Bernardeau and van de Weygaert 1996; Ord 1978). The relative merits of these approaches were investigated by Barr and Schoenberg (2010) and van Lieshout (2012). Further details can be found in Diggle (2013).

The main drawback of the non-parametric intensity estimator is its lack of consistency. As pointed out by Guan and Loh (2007), if we consider a kernel $K(.)$ with finite support around the origin, only local information around each point is used to estimate the intensity. If the true intensity is continuous, local smoothing will provide an asymptotically unbiased estimator. But, as the number of events in any region is of order one, the variance of the estimate does not tend to zero. To overcome this problem, Diggle (1985) defined the density of event locations $\lambda_0(x) = \frac{\lambda(x)}{m}$ where $m = \int_W \lambda(x)dx$ is the expected number of events of X observed on W, and showed the consistency of its kernel estimator. In addition, Guan and Loh (2007) developed a consistent non-parametric intensity estimator assuming that the first-order intensity of a spatial point process is a continuous function of some observed covariates.

Kernel intensity estimation for spatial point processes has been addressed assuming a scalar bandwidth parameter (Diggle 2003; Cucala 2008; Guan 2008; Illian *et al.* 2008), which can be quite restrictive specially for anisotropic and highly inhomogeneous point processes. Following the philosophy of bivariate kernel intensity estimation, we can consider bandwidth matrices (Wand and Jones 1994; Duong and Hazelton 2003) and define the kernel intensity estimator as

$$\hat{\lambda}_h(x) = \frac{1}{p_h(x)} \sum_{i=1}^{N} K_{x,h}(X_i),$$

where $x = (x_1, x_2)^T$, $X_i = (X_{i1}, X_{i2})^T$, $\quad i = 1, ..., N$, the term

$$p_h(x) = \int_W \frac{1}{h^2} K\left(\frac{x-y}{h}\right) dy$$

is the edge-correction term and $h = (h_1, h_2)^T$ is the vector of bandwidths. Note that $K_{x,h}(.)$ is the bivariate continuous kernel.

Given that the number of events of an inhomogeneous Poisson point process, N, has distribution

$$Poisson\left(\int_W \lambda(x)dx\right) = Poisson(m),$$

we can establish the following relationship between a bivariate density and the first-order intensity of a spatial point process

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

Considering this relationship, Cucala (2006) defined the density of event locations as

$$\lambda_0(x) = \frac{\lambda(x)}{m}.$$

Let h be a vector of bandwidths, the kernel estimator of $\lambda_0(.)$ is given by

$$\hat{\lambda}_{0,h}(x) = \frac{\hat{\lambda}_h(x)}{N} \mathbb{1}_{[N \neq 0]}.$$

One of the aims of the work is to develop a bandwidth selector for the kernel estimator of the first-order intensity. Diggle and Marron (1988) pointed out that the edge-correction term in estimator given by Cucala can introduce a bias towards under-smoothing in bandwidth selection. To overcome this problem, these authors proposed to estimate and minimize the MSE in the interior of the observation domain in order to obtain the optimal bandwidth. However, the application of this idea implies the selection of a new parameter to define the region where the error measurement is computed.

When the support of some variables is bounded, for example, in the case of non negative data, the standard kernel estimator continues to give weight outside the supports. This causes a bias in the boundary region. The boundary bias problem of the standard kernel is well documented in the univariate case. An initial solution to the boundary problem is given by Schuster (1985), who proposes the reflection method. Müller (1991), Lejeune and Sarda (1992), Jones (1993), Jones and Foster (1996), and Cheng, Fan, and Marron (1997) suggest the use of adaptive and boundary kernels at the edges and a fixed standard kernel in the interior region. Marron and Ruppert (1994) investigate some transformations before using the standard kernels, and Cowling and Hall (1996) propose a pseudo data method. Recently, Chen (2000), Bouezmarni and Rolin (2003), and Bouezmarni and Scaillet (2005) study the gamma kernels for univariate non negative data. For data defined on the unit interval, Chen (1999) proposes to use a beta kernel.

Let X be a spatial point process defined in $\mathbf{R}^2$ and $X_1, ..., X_N$ a realization of X observed on a bounded region $W \in \mathbf{R}^2$. The bivariate continuous associated kernel estimator $\hat{\lambda}_{0,h}$ of λ_0 is defined as

$$\hat{\lambda}_{0,h}(x) = \frac{1}{p_h(x)N} \sum_{i=1}^{N} K_{x,h}(X_i) \mathbb{1}_{[N \neq 0]},$$

where $x = (x_1, x_2)^T \in W,\ X_i = (X_{i1}, X_{i2})^T,\ i = 1, ..., N,\ h = (h_1, h_2)^T$ is the vector of bandwidths and $p_h(x) = \int_W h^{-2} K\left(\frac{x-y}{h}\right) dy$ is the edge-correction term. Note that,

$$K_{x,h}(.) = \prod_{j=1}^{2} K_{x_j,h_j}$$

is the bivariate continuous associated kernel, parametrized by x and h, which verify the following conditions:

$$x \in \mathcal{S}_{x,h},\ \mathbf{E}(\mathcal{Z}_{x,h}) = x + a(x,h) \quad \text{and} \quad Cov(\mathcal{Z}_{x,h}) = B(x,h),$$

where $\mathcal{Z}_{x,h}$ denotes the random vector with probability mass function $K_{x,h}$ of support $\mathcal{S}_{x,h}$ and both $a(x,h) = (a_1(x,h), a_2(x,h))^T$, $B(x,h) = (b_{ij}(x,h))_{i,j=1,2}$ tend, respectively, to the null vector and the null matrix as h goes to 0_2.

Although the consequences of the boundary problem in multivariate dimensions are much more severe, because the boundary region increases with dimension, solutions to the problem are not well investigated. Müller and Stadtmüller (1999) propose boundary kernels for multivariate data defined on arbitrary support by selecting the kernels that minimize a variational problem. In fact, they extend the minimum variance selection principle kernel used to select the optimal kernel in the interior region, as in Epanechnikov (1969) and Granovsky and Müller

(1991). Although this estimator has interesting properties, it remains complicated in practice. In addition, it requires an additional bandwidth parameter and a weighting function.

This paper proposes a non parametric product kernel intensity estimator for bivariate bounded data. Estimation is based on a beta kernel when the support is a compact set. Based on these kernels, the intensity estimators are robust to the boundary problem. The method is easy in conception and implementation. We provide the asymptotic properties of these estimators and show that the optimal rate of convergence of the mean integrated squared error is obtained. For the bivariate intensity function, we show that the estimator. We examine the finite sample performance in several simulations. As for any non parametric kernel estimator, the performance is sensitive to the choice of the bandwidth parameters. We suggest the application of the least squares cross-validation method to select these parameters. We prove the consistency of this method for the proposed estimators and investigate its performance in the simulations.

The rest of the paper is organized as follows. We derive the bivariate beta kernel intensity estimator for bivariate bounded data in Section 2. Section 3 introduce the bivariate beta kernel intensity estimator modified for bivariate bounded data. We introduce the asymptotic property in Section 4. In Section 5, we apply the method to simulated data and to three species of trees in the Lansing data. Section 6 concludes. The proofs of the theorems are presented in the Appendix.

2. Beta kernel intensity estimators

Let X be a point process defined on a bounded region $W \subset \mathbf{R}^2$, where W is assumed to have positive area. Let $\{X_1, X_2, ..., X_N\}$ be a realisation of the process where N is the random variable counting the number of events. The first-order intensity, from now on referred as intensity, is defined following Diggle (2013) as:

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

where $|dx|$ denotes the area of an infinitesimal region containing the point $x \in \mathbf{R}^2$. Let $\{(X_{i1}, X_{i2}),\ i = 1, ..., N\}$ be a inhomogeneous Poisson process from with an unknown intensity function λ_0 which has a compact support. We assume that the compact support W is known and, without loss of generality, is $W = [0,1]^2$. In this work, we assume the intensity function λ_0 satisfies:

- $W = [0,1]^2$ (compact support),
- λ_0 twice continuously differentiable.

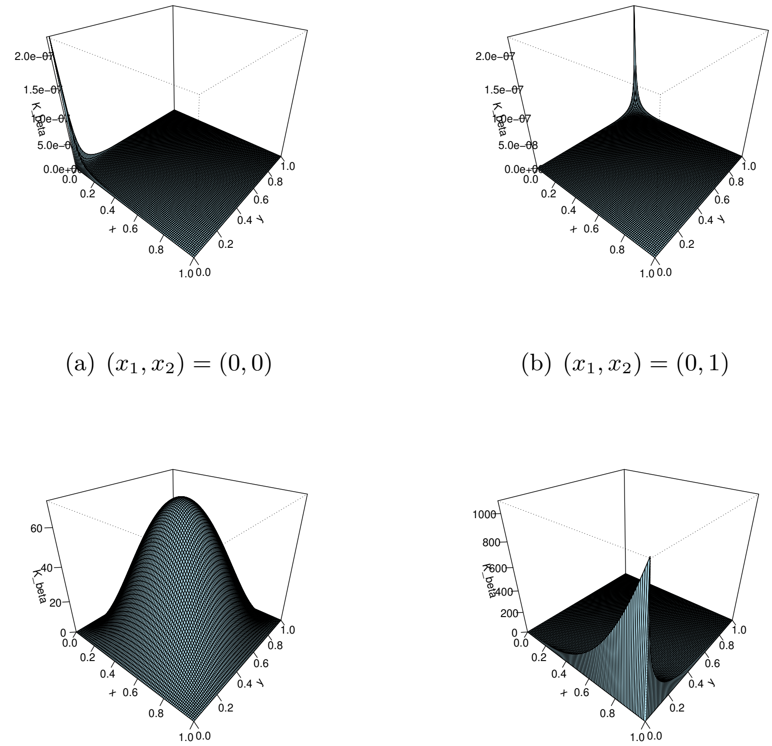
and λ_0 has continuous second derivative. Let $K_{p,q}$ be the density function of a $Beta(p,q)$ random variable. We consider two beta kernel estimators for λ_0. The $Beta(p,q)$ kernel estimator uses $K_{\frac{y}{h}+1,\frac{1-y}{h}+1}(t)$ as the kernel at $y \in [0,1]$ (Chen 1999), where h is a smoothing parameter satisfying the condition that $h \to 0$ as $m \to \infty$, and is defined as, for $y \in [0,1]$,

$$\hat{\lambda}_{0,h}(x) = \frac{1}{N} \sum_{i=1}^{N} \prod_{j=1}^{2} K_{\frac{x_j}{h_j}+1,\frac{1-x_j}{h_j}+1}(X_{ij}) \mathbb{1}_{[N\neq 0]}, \tag{1}$$

where

$$K_{\frac{x_j}{h_j}+1,\frac{1-x_j}{h_j}+1}(t) = \frac{t^{\frac{x_j}{h_j}}(1-t)^{\frac{1-x_j}{h_j}}}{B\left(\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1\right)}$$

and B is the beta function (Chen 1999).



(a) $(x_1, x_2) = (0, 0)$ (b) $(x_1, x_2) = (0, 1)$

(c) $(x_1, x_2) = (0.5, 0.5)$ (d) $(x_1, x_2) = (0.8, 0.1)$

Figure 1: Shape of beta kernels at (x_1, x_2). The bandwidth parameters are $h_1 = h_2 = 0.2$.

The bias of the estimator is given:

$$\text{Bias}(\hat{\lambda}_{0,h}(x)) = -e^{-m}\lambda_0(x) + (1 - e^{-m})\sum_{j=1}^{2} h_j u_j(x) + o\left((1 - e^{-m})\sum_{j=1}^{2} h_j\right),$$

where

$$u_j(x) = (1 - 2x_j)\lambda_0^j(x) + \frac{x_j}{2} h_j (1 - x_j)\lambda_0^{jj}(x), \quad \lambda_0^j(x) = \frac{\partial \lambda_0(x)}{\partial x_j}, \quad \lambda_0^{jj}(x) = \frac{\partial^2 \lambda_0(x)}{\partial x_j^2}$$

and the variance is given by:

$$\text{Var}(\hat{\lambda}_{0,h}(x)) = A(m)\prod_{j=1}^{2}(A_{h_j}(x_j))\lambda_0(x) + o\left(A(m)\sum_{j=1}^{2} h_j^2\right), \tag{2}$$

where

$$A(m) = \mathbf{E}\left[\frac{1}{N}\mathbb{1}_{N \neq 0}\right],$$

$$A_{h_j}(x_j) = \begin{cases} \frac{h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}}, & \text{if} \quad \frac{x_j}{h_j} \quad \text{or} \quad \frac{1-x_j}{h_j} \to \infty \\ \frac{h_j^{-1}\Gamma(2c_j+1)}{2^{2c_j+1}\Gamma^2(c_j+1)}, & \text{if} \quad \frac{x_j}{h_j} \to c_j \quad \text{or} \quad \frac{1-x_j}{h_j} \to c_j. \end{cases}$$

Hence, the MSE of the beta kernel is:

$$MSE(\hat{\lambda}_{0,h}(x)) = \left(-e^{-m}\lambda_0(x) + (1-e^{-m})\sum_{j=1}^{2} h_j u_j(x)\right)^2 + A(m)\prod_{j=1}^{2}\left(A_{h_j}(x_j)\right)\lambda_0(x).$$

The mean integrated square error ($MISE$) is determined as:

$$MISE(\hat{\lambda}_{0,h}(x)) = \int_W \left(-e^{-m}\lambda_0(x) + (1-e^{-m})\sum_{j=1}^{2} h_j u_j(x)\right)^2 dx + \int_W A(m)\prod_{j=1}^{2}\left(A_{h_j}(x_j)\right)\lambda_0(x))dx.$$

In view of (2), the variance of $\hat{\lambda}_{0,h}(x)$ takes this form

$$\mathrm{Var}(\hat{\lambda}_{0,h}(x)) = \begin{cases} A(m)\prod_{j=1}^{2}\frac{h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}}(\lambda_0(x)+o(1)), & \text{if } \frac{x_j}{h_j} \text{ or } \frac{1-x_j}{h_j}\to\infty \\ A(m)\prod_{j=1}^{2}\frac{h_j^{-1}\Gamma(2c_j+1)}{2^{2c_j+1}\Gamma^2(c_j+1)}(\lambda_0(x)+o(1)), & \text{if } \frac{x_j}{h_j}\to c_j \text{ or } \frac{1-x_j}{h_j}\to c_j, \end{cases}$$

where c_j is a positive constant.

3. Beta kernel modified for the intensity function estimator

In order to reduce the bias of $\hat{\lambda}_{0,h}(x)$, another beta kernel estimator is

$$\tilde{\lambda}_{0,h}(x) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{2} K_{\frac{x_j}{h_j},\frac{1-x_j}{h_j}}(X_{ij})\mathbb{1}_{[N\neq 0]}, & \text{if } x_j\in[2h_j, 1-2h_j] \\ \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{2} K_{\rho(x_j),\frac{1-x_j}{h_j}}(X_{ij})\mathbb{1}_{[N\neq 0]}, & \text{if } x_j\in[0,2h_j) \\ \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{2} K_{\frac{x_j}{h_j},\rho(1-x_j)}(X_{ij})\mathbb{1}_{[N\neq 0]}, & \text{if } x_j\in(1-2h_j, 1], \end{cases} \tag{3}$$

where the edge-correction term $\rho_{h_j}(x_j) = 2h_j{}^2 + 2.5 - \sqrt{4h_j^4 + 6h_j^2 + 2.25 - x_j^2 - \frac{x_j}{h_j}}$. Note that for each fixed h_j $\rho_{x_j}(h_j)$ is a monotonic increasing function of x_j between 0 and $2h_j$. In particular, it has $y = x_j/h_j$ as its tangent line at $x_j = 2h_j$ and $\rho_0(h_j) = 1$, where $\overline{h} = (h_1, h_2)$ is the vector of the bandwidth parameters such that $h_j \to 0$ as $m\to\infty$. The bias of $\tilde{\lambda}_{0,h}(x)$ is equal to

$$\begin{aligned}\mathrm{Bias}(\tilde{\lambda}_{0,h}(x)) \approx & -e^{-m}\lambda_0(x) + (1-e^{-m})\sum_{j=1}^{2}\Big[\frac{1}{2}h_j x_j(1-x_j)\lambda_0^{(jj)}(x)\mathbb{1}_{[2h_j,\,1-2h_j]}(x_j) \\ & + \xi(x_j)h_j\lambda_0^{(j)}(x)\mathbb{1}_{[0,\,2h_j)}(x_j) - \xi(1-x_j)h_j\lambda_0^{(j)}(x)\mathbb{1}_{(1-2h_j,\,1]}(x_j)\Big]\end{aligned} \tag{4}$$

with $\xi(x_j) = (1-x_j)(\rho(x_j) - \frac{x_j}{h_j})/(1 + h_j\rho(x_j) - x_j)$ and the variance of $\tilde{\lambda}_{0,h_j}(x_j)$ is equal to

$$\mathrm{Var}(\tilde{\lambda}_{0,h_j}(x_j)) = A(m)\left(\prod_{j=1}^{2} A^*_{h_j}(x_j)\right)\lambda_0(x) + o(A(m)\sum_{j=1}^{2} h_j),$$

$$A^*_{h_j}(x_j) = \frac{B\left(\frac{2x_j}{h_j}, \frac{2(1-x_j)}{h_j}\right)}{B^2\left(\frac{x_j}{h_j}, \frac{(1-x_j)}{h_j}\right)} \mathbb{1}_{[2h_j, 1-2h_j]}(x_j) + \frac{B\left(2\rho(x_j), \frac{2(1-x_j)}{h_j}\right)}{B^2\left(\rho(x_j), \frac{(1-x_j)}{h_j}\right)} \mathbb{1}_{[0, 2h_j)}(x_j)$$
$$+ \frac{B\left(\frac{2x_j}{h_j}, 2\rho(1-x_j)\right)}{B^2\left(\frac{x_j}{h_j}, \rho(1-x_j)\right)} \mathbb{1}_{(1-2h_j, 1]}(x_j).$$

It is obvious from the bias expansion of $\tilde{\lambda}_{0,h}(x)$ that its bias is free of λ_0^j in the interval $[2h_j, 1-2h_j]$. Combining the above bias and variance expansions, we have the following asymptotic MSE for $\tilde{\lambda}_{0,h}(x)$.

$$MSE(\tilde{\lambda}_{0,h}(x)) \approx \Bigg(-e^{-m}\lambda_0(x) + (1-e^{-m}) \sum_{j=1}^{2} \Big[\frac{1}{2} h_j x_j (1-x_j) \lambda_0^{(jj)}(x) \mathbb{1}_{[2h_j, 1-2h_j]}(x_j)$$
$$+ \xi(x_j) h_j \lambda_0^{(j)}(x) \mathbb{1}_{[0, 2h_j)}(x_j) - \xi(1-x_j) h_j \lambda_0^{(j)}(x) \mathbb{1}_{(1-2h_j, 1]}(x_j) \Big] \Bigg)^2 \tag{5}$$
$$+ A(m) \left(\prod_{j=1}^{2} A^*_{h_j}(x_j) \right) \lambda_0(x).$$

It is claimed in Chen that $A^*_{h_j}(x_j)$ has similar asymptotic expansion to that of $A_{h_j}(x)$, and hence $\mathrm{Var}(\tilde{\lambda}_{0,h}(x)) \approx \mathrm{Var}(\hat{\lambda}_{0,h}(x))$. Based on this, the following simplified asymptotic $MISE$ of $\tilde{\lambda}_{0,h}(x)$ is given by:

$$MISE(\tilde{\lambda}_{0,h}) \approx \int_{[0,1]^2} \Bigg[\Bigg(-e^{-m}\lambda_0(x) + (1-e^{-m}) \sum_{j=1}^{2} \Big(\frac{1}{2} h_j x_j (1-x_j) \lambda_0^{(jj)}(x) \mathbb{1}_{[2h_j, 1-2h_j]}(x_j)$$
$$+ \xi(x_j) h_j \lambda_0^{(j)}(x) \mathbb{1}_{[0, 2h_j)}(x_j) - \xi(1-x_j) h_j \lambda_0^{(j)}(x) \mathbb{1}_{(1-2h_j, 1]}(x_j) \Big) \Bigg)^2$$
$$+ A(m) \left(\prod_{j=1}^{2} A^*_{h_j}(x_j) \right) \lambda_0(x) \Bigg] dx. \tag{6}$$

4. The asymptotic properties of the estimators

In this section we obtain the asymptotic properties of the estimator (3). To this end we derive the bias and the variance in the following lemmas. All proofs are relegated to the Appendix and hold assuming that all the components of the bandwidth vector are different. Regarding the practical use we give a simpler formulation of the lemmas for equal bandwidths $h_1 = h_2 = h$. In the next lemmas devoted to the bias and the variance we assume that $\lambda_0(x)$ is a twice continuously differentiable function. The next lemma states the bias of the non-parametric density estimative.

Lemma 1 *The bias of the estimator is given:*

$$Bias(\hat{\lambda}_{0,h}(x)) = -e^{-m}\lambda_0(x) + (1-e^{-m}) \sum_{j=1}^{2} h_j u_j(x) o \left((1-e^{-m}) \sum_{j=1}^{2} h_j \right)$$
$$Bias(\tilde{\lambda}_{0,h}(x)) = -e^{-m}\lambda_0(x) + (1-e^{-m}) \sum_{j=1}^{2} h_j v_j(x) o \left((1-e^{-m}) \sum_{i=1}^{2} h_j \right),$$

$$\text{where} \quad v_j(x) = \frac{x_j}{2}(1-x_j)\frac{\partial^2 \lambda_0(x))}{\partial x_j^2}.$$

Remark 1 *If $h_1 = h_2 = h$ and $h \to 0$, $mh \to 0$ as $m \to \infty$, then the bias of the intensity function estimates are equal*

$$Bias(\hat{\lambda}_{0,h}(x)) = -e^{-m}\lambda_0(x) + (1-e^{-m})h\sum_{j=1}^{2} u_j(x) + o\left((1-e^{-m})h\right)$$

$$Bias(\tilde{\lambda}_{0,h}(x)) = -e^{-m}\lambda_0(x) + (1-e^{-m})h\sum_{j=1}^{2} v_j(x) + o((1-e^{-m})h).$$

Lemma 2 *The variance are given by*

$$\begin{aligned} Var(\hat{\lambda}_{0,h}(x)) &= A(m)\left(\prod_{j=1}^{2} A_{h_j}(x_j)\right)\lambda_0(x) + o\left(A(m)\sum_{j=1}^{2} h_j\right) \quad \text{and} \\ Var(\tilde{\lambda}_{0,h}(x)) &= A(m)\left(\prod_{j=1}^{2} A^{*}_{h_j}(x_j)\right)\lambda_0(x) + o\left(A(m)\sum_{j=1}^{2} h_j\right). \end{aligned}$$

Remark 2 *If $h_1 = h_2 = h$ and $h \to 0$, $mh \to 0$ as $m \to \infty$ then the variance expansion of the intensity function estimate is equal to*

$$\begin{aligned} &Var(\hat{\lambda}_{0,h}(x)) \approx Var(\tilde{\lambda}_{0,h}(x)) \\ &\qquad = \frac{A(m)}{h}\left(\prod_{j=1}^{2}\frac{(x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}}\right)\lambda_0(x) + o\left(A(m)\prod_{i=1}^{2} h^{-1/2}\right). \end{aligned}$$

The Assumptions (**H**) on the bandwidth parameters are $h_j \longrightarrow 0$ and $m^{-1}\prod_{j=1}^{2} h_j^{-1/2} \longrightarrow 0$, as $m \longrightarrow 0$. The mean integrated square error ($MISE$) is determined as

$$MISE(\hat{\lambda}_{0,h}(x)) = \int_W \left(\lambda_0(x) - \hat{\lambda}_{0,h}\right)^2 dx = \int_W \text{Bias}(\hat{\lambda}_{0,h}(x)^2 dx + \int_W \text{Var}(\hat{\lambda}_{0,h}(x))dx.$$

The following result states the variance of the non-parametric estimator.

Theorem 1 *Suppose that $\lambda_0(x)$ is twice differentiable. Let $\hat{\lambda}_{0,h}(x)$ and $\tilde{\lambda}_{0,h}(x)$ be the non parametric estimators with the beta kernel. Under assumption **H**.*

$$\begin{aligned} AMISE(\hat{\lambda}_{0,h}(x)) &= (1-e^{-m})^2\int_W\left(\sum_{j=1}^{2} h_j u_j(x)\right)^2 dx \\ &+ A(m)\int_W \prod_{j=1}^{2} h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}\lambda_0(x)dx, \\ AMISE(\tilde{\lambda}_{0,h}(x)) &= (1-e^{-m})^2\frac{h^2}{4}\int_W\left(\sum_{j=1}^{2} x_j(1-x_j)\lambda_0(x)\right)^2 dx \\ &+ A(m)h^{-1}\int_W \prod_{j=1}^{2}(x_j(1-x_j))^{-\frac{1}{2}}\lambda_0(x)dx. \end{aligned}$$

The optimal bandwidths that minimize the asymptotic mean integrated squared error are $h_j = c_j m^{-1/3}$, for some positive constants c_1, c_2.

Theorem 1 proves that the rate of convergence of the bias of the non-parametric estimator is uniform, hence it is free of boundary bias. The rate of convergence of the mean integrated squared error becomes slower when the dimension of the random variable increases. This is known as the curse of dimensionality. We see that in the boundary region, the variance of the product kernel estimator is larger in comparison with the variance in the interior region. However, the increase of the variance is compensated by a smaller bias in this region. Away from the boundaries, we have the opposite effect, that is, a lower variance and a slightly higher bias. Fortunately, the second derivative of the intensity function is negligible away from zero.

Remark 3 *In conditions of Lemmas 1,2, the mean squared error (MSE) is given by*

$$\begin{aligned} AMISE(\hat{\lambda}_{0,h}(x)) &= (1-e^{-m})^2 h^2 \int_W \left(\sum_{j=1}^{2} u_j(x)\right)^2 dx \\ &+ A(m)h^{-1}\int_W \prod_{j=1}^{2}(x_j(1-x_j))^{-1/2}\lambda_0(x)dx, \\ AMISE(\tilde{\lambda}_{0,h}(x)) &= (1-e^{-m})^2 \frac{h^2}{4}\int_W \left(\sum_{j=1}^{2} v_j(x)\right)^2 dx \\ &+ A(m)h^{-1}\int_W \prod_{j=1}^{2}(x_j(1-x_j))^{-1/2}\lambda_0(x)dx. \end{aligned}$$

Using the results of the latter we can obtain the upper bound of the MISE and find the expression of the optimal bandwidths $\hat{h}$ and $\tilde{h}$ as the minimum of the latter. Hence, the following theorem holds.

Theorem 2 *Under assumption **H**, the optimal bandwidths are given by:*

$$\begin{aligned} \hat{h}_1 &= \left[\frac{\int U(x)dx}{4aK^{-\frac{3}{2}}[K\int u_1^2(x)dx + \int u_1(x)u_2(x)dx]}\right]^{\frac{1}{3}}, \\ \hat{h}_2 &= \left[\frac{\int U(x)dx}{4aK^{\frac{3}{2}}[K\int u_1^2(x)dx + \int u_1(x)u_2(x)dx]}\right]^{\frac{1}{3}}, \\ \tilde{h}_1 &= \left[\frac{\int U(x)dx}{4aC^{-\frac{3}{2}}[C\int v_1^2(x)dx + \int v_1(x)v_2(x)dx]}\right]^{\frac{1}{3}}, \\ \tilde{h}_2 &= \left[\frac{\int U(x)dx}{4aC^{\frac{3}{2}}[C\int v_1^2(x)dx + \int v_1(x)v_2(x)dx]}\right]^{\frac{1}{3}}, \end{aligned}$$

where

$$u_j(x) = (1-2x_j)\lambda_0^j(x) + x_j(1-x_j)\lambda_0^{jj}, \quad v_j(x) = x_j(1-x_j)\lambda_0^{jj}, \quad a = (1-e^{-m})^2,$$

$$U(x) = A(m)\prod_{j=1}^{2}\frac{(x_j(1-x_j))^{-1/2}}{2\sqrt{\pi}}, \quad K = \frac{\int u_2(x)dx}{\int u_1(x)dx}, \quad C = \frac{\int v_2(x)dx}{\int v_1(x)dx}.$$

Remark 4 *In conditions of Lemmas 1,2 the optimal bandwidths that provide a minimum of the MISE are given by*

$$\hat{h} = \left[\frac{2A(m) \int_W U(x)\lambda_0(x)dx}{(1-e^{-m})^2 \int_W \left(\sum_{j=1}^{2} u_j(x) \right)^2 dx} \right]^{\frac{1}{3}}, \quad \tilde{h} = \left[\frac{A(m) \int_W U(x)\lambda_0(x)dx}{(1-e^{-m})^2 \int_W \left(\sum_{j=1}^{2} v_j(x) \right)^2 dx} \right]^{\frac{1}{3}}.$$

5. Simulation

In this section, we make applications on simulated data and on real data from Lansing data to evaluate the performance of this beta kernel intensity estimator.

5.1. Simulation studies

In this section we study the finite sample properties for the non-parametric intensity estimator for bivariate data with bounded supports. We compare the bias, variance and the mean integrated squared error distribution of the non-parametric product kernel estimator using the following kernels: beta, modified beta and Gaussian. We consider the following six data generating processes:

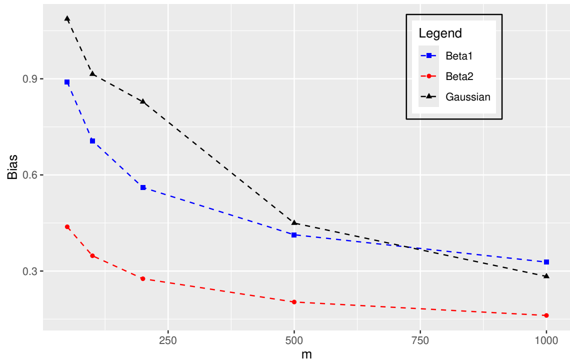
1. $\lambda_1(x) = b\exp(-2x_1 - x_2)$, $x = (x_1, x_2) \in [0,1]^2$, using several values of b ($b = 183$ or 365 or 730 or 1825 or 3650) to obtain $m = \int_W \lambda(x)dx \approx 20$ or 50 or 100 or 200 or 400 or 500 respectively.

2. $\lambda_2(x) = b\Phi_{(0.3-0.2x_2,0.02)}(x_1)+25$, $x = (x_1, x_2) \in [0,1]^2$, where Φ is the univariate normal density with mean $\mu = 0.3 - 0.2x_2$ and standard deviation $\sigma = 0.1$ and $\sigma = 0.02$, using several values of b ($b = 50$ or 100 or 200 or 500 or 1000) to obtain $m = \int_W \lambda(x)dx \approx 50$ or 100 or 200 or 500 or 1000 respectively. This model was used by Barr and Schoenberg (2010) to analyze the performance of the Voronoi estimator of the first order intensity.

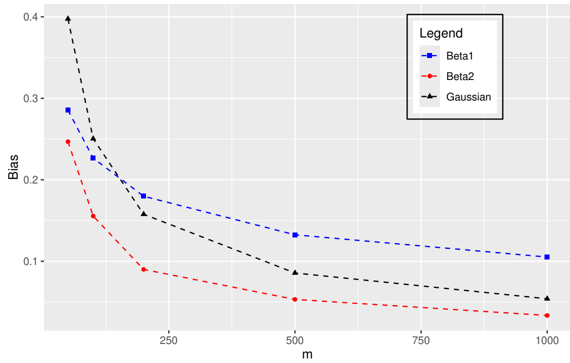
3. $\lambda_2(x) = b(\cos(\sqrt{(x_1-3)^2 + (x_2-3)^2}))^2$, $x = (x_1, x_2) \in [0,1]^2$, using several values of b ($b = 127$ or 330 or 254 or 634 or 1267) to obtain $m = \int_W \lambda(x)dx \approx 50$ or 100 or 200 or 500 or 1000 respectively.

4. $\lambda_4(x) = b(x_1^2 + x_2^3)$, $x = (x_1, x_2) \in [0,1]^2$, using several values of b ($b = 86$ or 172 or 344 or 860 or 1720) to obtain $m = \int_W \lambda(x)dx \approx 50$ or 100 or 200 or 500 or 1000 respectively.

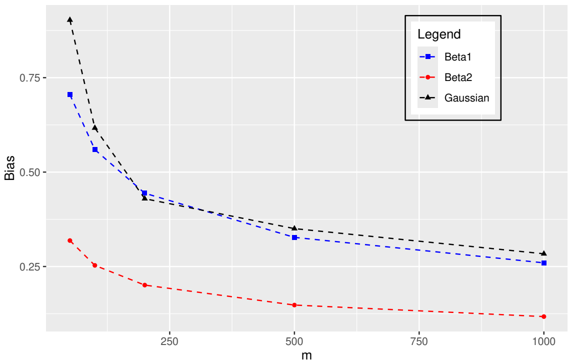
In simulations, we consider the sample sizes 50, 100, 200, 500 and 1000 and perform 100 replications for each model. In each replication the bandwidth is chosen such that the integrated squared error is minimized. We report the integrated bias, the integrated variance and the mean integrated squared error in Figure 2, 3, 4 and Table 1, 2. As a general remark we observe that the integrated bias, integrated variance decrease for all models, as expected when the sample size increases.
In Figure 2, we notice that the bias of the estimators is decreasing for all the models and all values of m. We notice that the modified beta nucleus has a smallest bias than the beta kernel and the Gaussian kernel. But also the bias of the beta kernel intensity estimator is smaller than the Gaussian kernel estimator one small m values and larger for large m values.



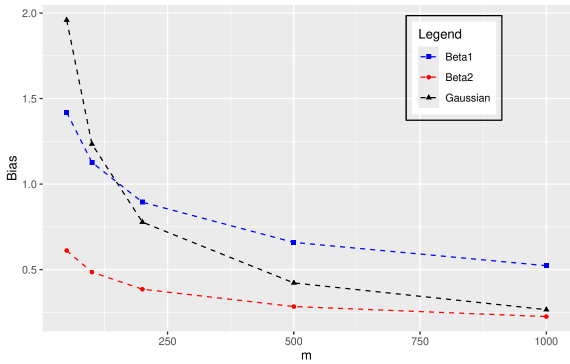
(a) λ_1



(b) λ_2

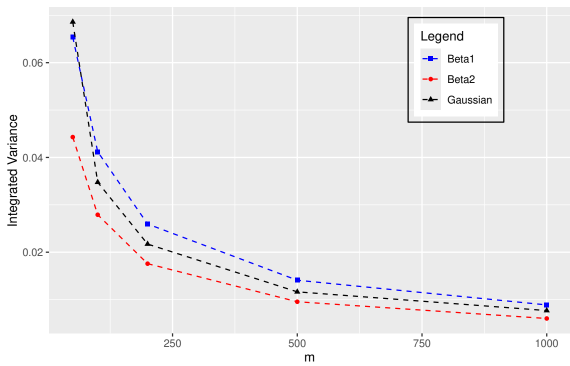


(c) λ_3

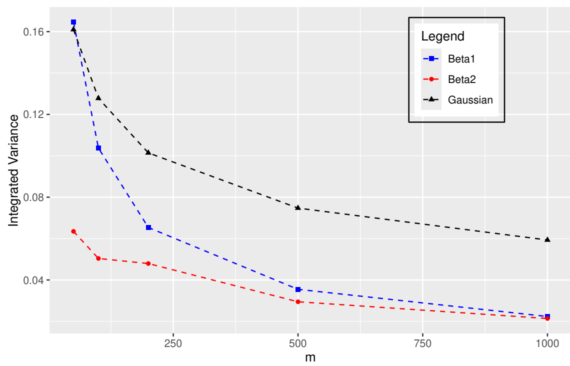


(d) λ_4

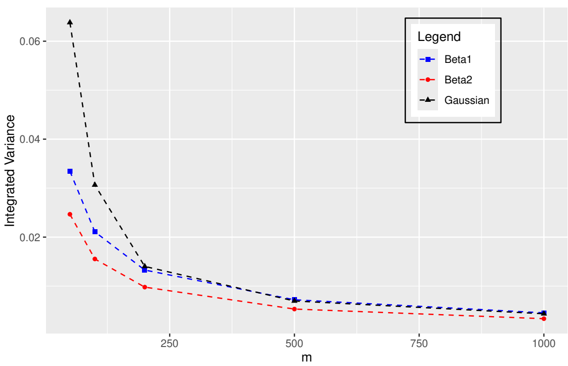
Figure 2: Integrated bias for the intensities functions



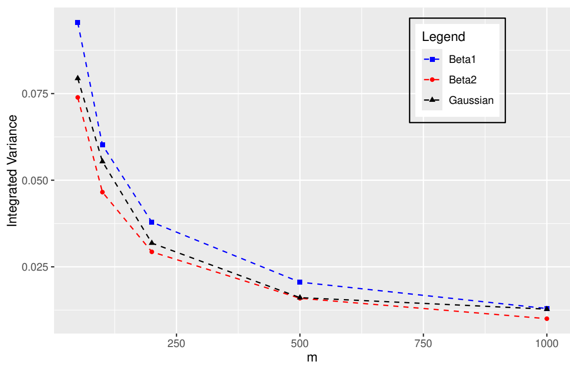
(a) λ_1



(b) λ_2



(c) λ_3



(d) λ_4

Figure 3: Integrated variance for the intensity functions

Table 1: Integrated bias and variance for intensity functions $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$ and $\lambda_4(x)$ respectively

m	Beta1	Beta2	Gaussian
		$\lambda_1(x)$	
50	(0.890175, 0.065398)	(0.438080, 0.044296)	(1.086967, 0.068601)
100	(0.706532, 0.041198)	(0.347705, 0.027904)	(0.914707, 0.034763)
200	(0.560775, 0.025953)	(0.275973, 0.017579)	(0.828213, 0.021718)
500	(0.413182, 0.014089)	(0.203339, 0.009543)	(0.449623, 0.011633)
1000	(0.327943, 0.008875)	(0.161390, 0.006011)	(0.283245, 0.007685)
		$\lambda_2(x)$	
50	(0.285453, 0.164689)	(0.246807, 0.063492)	(0.397468, 0.160992)
100	(0.226564, 0.103747)	(0.155479, 0.050393)	(0.250389, 0.127779)
200	(0.179824, 0.065356)	(0.089997, 0.047945)	(0.157735, 0.101418)
500	(0.132495, 0.035481)	(0.053173, 0.029470)	(0.085631, 0.074726)
1000	(0.105162, 0.022351)	(0.033496, 0.021390)	(0.053944, 0.059310)
		$\lambda_3(x)$	
50	(0.704867, 0.033462)	(0.318732, 0.024667)	(0.902631, 0.063814)
100	(0.559453, 0.021080)	(0.252978, 0.015539)	(0.616783, 0.030649)
200	(0.444038, 0.013279)	(0.200789, 0.009789)	(0.429521, 0.014020)
500	(0.327170, 0.007209)	(0.147942, 0.005314)	(0.350333, 0.006961)
1000	(0.259675, 0.004541)	(0.117422, 0.003347)	(0.283692, 0.004350)
		$\lambda_4(x)$	
50	(1.419621, 0.095508)	(0.611610, 0.073895)	(1.959342, 0.079393)
100	(1.126754, 0.060166)	(0.485435, 0.046551)	(1.234308, 0.055387)
200	(0.894305, 0.037902)	(0.385290, 0.029325)	(0.777565, 0.031857)
500	(0.658929, 0.020576,)	(0.283884, 0.015920)	(0.422127, 0.016104)
1000	(0.522992, 0.012962)	(0.225319, 0.010029)	(0.265923, 0.012782)

In Figure 3, we notice that the bias of the estimators is decreasing for all the models and all values of m. We notice that the modified beta kernel has a smallest bias than the beta kernel and the Gaussian kernel. But also the bias of the beta kernel intensity estimator is smaller than the Gaussian kernel estimator for small m values and larger for large m values. In terms of integrated bias, the beta and the modified beta kernel estimator is outperform the beta and Gaussian kernel estimators, and the beta kernel is better than the Gaussian estimator for large m values and more than for small m values. For example, for $m = 500$, the bias for the modified beta, beta and Gaussian kernel is 0.203339, 0.413182 and 0.449623 respectively for λ_1. In terms of variance, the estimators have also almost the same performance. For example, for $m = 500$,the integrated variance for the modified beta, beta and Gaussian kernel is 0.005314, 0.007209 and 0.006961, respectively for λ_3.

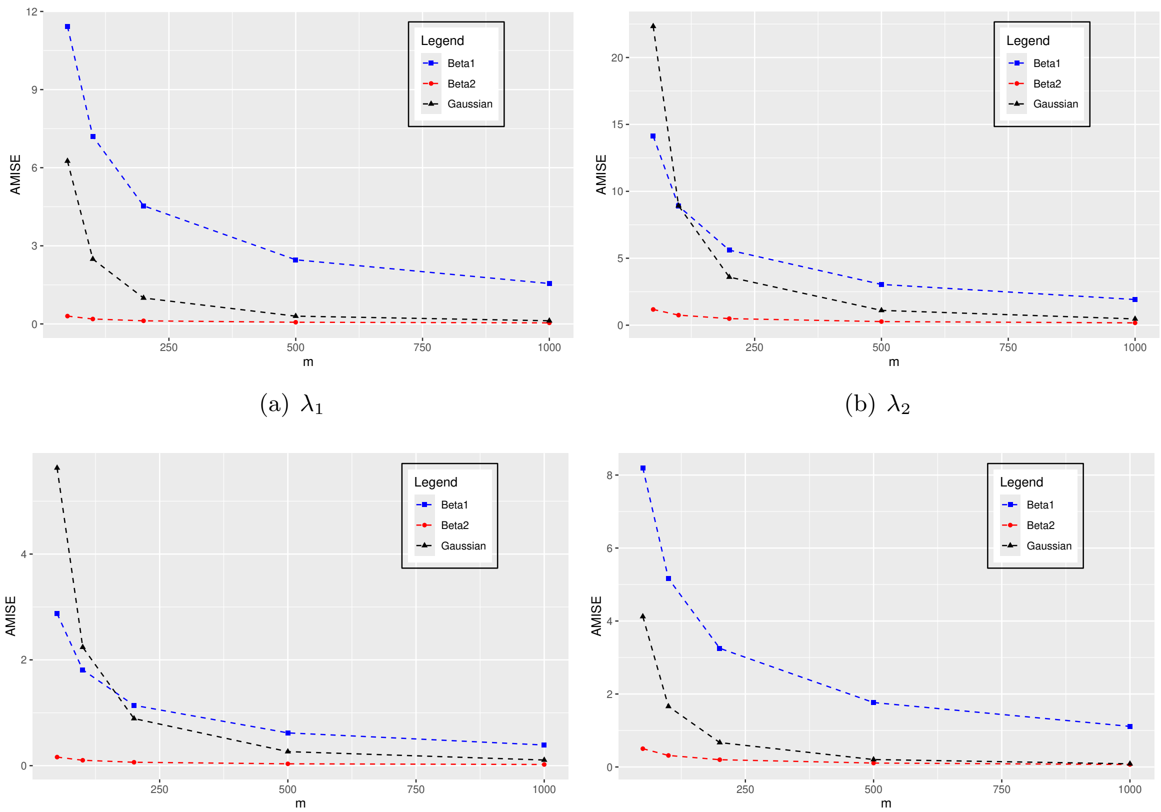
5.2. Application to Lansing data.

We here use the dataset from an investigation of a 924 ft $\times$ 924 ft (19.6 acre) plot in Lansing Woods, Clinton County, Michigan USA by D.J. Gerrard. The data give the locations of 2251 trees and their botanical classification (into hickories, maples, red oaks, white oaks, black oaks and miscellaneous trees). The original plot size (924×924 feet) has been rescaled to the unit square. We consider three particular species of trees: the hickories (703 trees), maples (514 trees), and red oaks (346 trees) species.

Table 2: Integrated square bias and AMISE for intensity functions $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$ and $\lambda_4(x)$ respectively

m	Beta1	Beta2	Gaussian
		$\lambda_1(x)$	
50	(11.368660, 11.43058)	(0.256379, 0.300675)	(6.187647, 6.256248)
100	(7.161809, 7.203007)	(0.161509, 0.189413)	(2.455569, 2.490332)
200	(4.511657, 4.537610)	(0.101744, 0.119323)	(0.974493, 0.996211)
500	(2.449304, 2.463393)	(0.055235, 0.064778)	(0.287205, 0.298838)
1000	(1.542965, 1.551840)	(0.034796, 0.040807)	(0.113977, 0.121662)
		$\lambda_2(x)$	
50	(13.96110, 14.125790)	(1.110869, 1.174361)	(22.17056, 22.33155)
100	(8.794944, 8.898691)	(0.699803, 0.750196)	(8.798394,8.926173)
200	(5.540467, 5.605823)	(0.440848, 0.488793)	(3.491645, 3.593063)
500	(3.007828, 3.043309)	(0.239329, 0.268799)	(1.029066, 1.103792)
1000	(1.894813, 1.917164)	(0.150768, 0.172158)	(0.408385, 0.467695)
		$\lambda_3(x)$	
50	(2.842427, 2.875889)	(0.137356, 0.162023)	(5.564476, 5.62829)
100	(1.790617, 1.811697)	(0.086529, 0.102068)	(2.208264, 2.238913)
200	(1.128018, 1.141297)	(0.054509, 0.064298)	(0.876350, 0.890370)
500	(0.612382, 0.619591)	(0.029592, 0.034906)	(0.258280, 0.265241)
1000	(0.385776, 0.390317)	(0.018642, 0.021989)	(0.102498, 0.106848)
		$\lambda_4(x)$	
50	(8.099679, 8.195187)	(0.427505, 0.501400)	(4.043768, 4.123161)
100	(5.102478, 5.162644)	(0.269311, 0.315862)	(1.604770, 1.660157)
200	(3.214360, 3.252262)	(0.169655, 0.198980)	(0.636853, 0.668710)
500	(1.745023, 1.765599)	(0.092103, 0.108023)	(0.187695, 0.203799)
1000	(1.099296, 1.112258)	(0.058021,0.068050)	(0.074486, 0.087268)

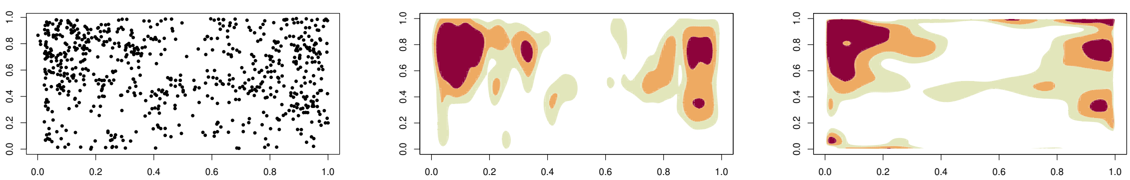
We obtained intensity estimates for these three species of trees using the Gaussian kernel and beta kernel show our results for the Hickory, Maple, and red oak species, respectively. Figures 5, 6 and 7 display the locations plot for the trees and the contour for the non-parametric intensity estimates. We obtained intensity estimates for these three species using the kernel intensity estimators, the beta kernel and the Gaussian kernel. Figures 5, 6 and 7 show our results for the Hickory, Maple, and Read oak species, respectively. Figures 5, 6, and 7 show the representations of (a) an inhomogeneous Poisson process, (b) contour for the Gaussian kernel estimator, and (c) contour for the modified beta kernel estimator for three tree species: Hickory, Maple, and Red Oak. These figures confirm the performance of the beta kernel compared to the Gaussian kernel and support the data presented in Tables 1 and 2.



(a) λ_1 (b) λ_2

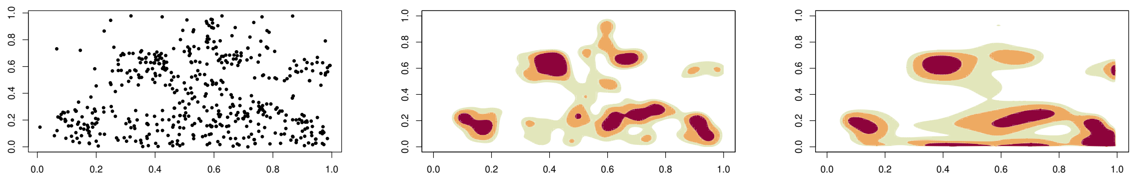
(c) λ_3 (d) λ_4

Figure 4: AMISE for the intensity functions



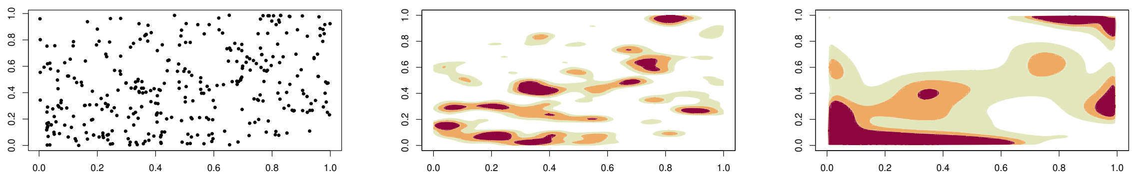
(a) (b) (c)

Figure 5: (a) inhomogeneous Poisson process, (b) contour for Gaussian kernel estimator, (c) contour for beta kernel estimator with Hickory trees



(a) (b) (c)

Figure 6: (a) inhomogeneous Poisson process, (b) contour for Gaussian kernel estimator, (c) contour for beta kernel estimator with Maple trees



(a) (b) (c)

Figure 7: (a) inhomogeneous Poisson process, (b) contour for Gaussian kernel estimator, (c) contour for beta kernel estimator with red oak trees

6. Conclusion

This paper proposes a non-parametric estimator for bivariate intensity functions of bounded data using beta kernel. The estimator is based on the beta kernel when the support is a compact set. By using boundary kernels, the intensity function estimators are robust to the boundary problem. We provide the asymptotic properties of the estimator and show that the optimal rate of convergence of the mean integrated squared error is obtained. We examine the finite sample performance in several simulations. In fact, we find that the estimators we propose perform almost as well as the Gaussian estimator when there are no boundary problems.

In light of our contributions, some perspectives emerge, leading to future research directions:

- Application to more complex models (spatial, multidimensional processes).
- Development of adaptive approach for parameter optimization.

References

Baddeley AJ, Møller J, Waagepetersen R (2000). "Non- and Semi-parametric Estimation of Interaction in Inhomogeneous Point Patterns." *Statistica Neerlandica*, **54**(3), 329–350. doi:10.1111/1467-9574.00144.

Barr CD, Schoenberg FP (2010). "On the Voronoi Estimator for the Intensity of an Inhomogeneous Planar Poisson Process." *Biometrika*, **97**(4), 977–984. URL https://www.jstor.org/stable/29777151.

Bernardeau F, van de Weygaert R (1996). "A New Method for Accurate Estimation of Velocity Field Statistics." *Monthly Notices of the Royal Astronomical Society*, **279**(2), 693–711. doi:10.1093/mnras/279.2.693.

Bouezmarni T, Rolin JM (2003). "Consistency of the Beta Kernel Density Function Estimator." *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pp. 89–98. doi:10.2307/3315905.

Bouezmarni T, Scaillet O (2005). "Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data." *Econometric Theory*, **21**(2), 390–412. doi:10.1017/S026646660505021X.

Chen SX (1999). "Beta Kernel Estimators for Density Functions." *Computational Statistics & Data Analysis*, **31**(2), 131–145. doi:10.1016/S0167-9473(99)00010-9.

Chen SX (2000). "Probability Density Function Estimation Using Gamma Kernels." *Annals of the Institute of Statistical Mathematics*, **52**, 471–480. doi:10.1023/A:1004165218295.

Cheng MY, Fan J, Marron JS (1997). "On Automatic Boundary Corrections." *The Annals of Statistics*, **25**(4), 1691–1708. URL https://www.jstor.org/stable/2959068.

Cowling A, Hall P (1996). "On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation." *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **58**(3), 551–563. doi:10.1111/j.2517-6161.1996.tb02100.x.

Cucala L (2006). *Espacements Bidimensionnels et Données Entachées d'Erreurs dans l'Analyse des Processus Ponctuels Spatiaux*. Ph.D. thesis, Université des Sciences Sociales-Toulouse I. URL https://theses.hal.science/tel-00135890v1.

Cucala L (2008). "Intensity Estimation for Spatial Point Processes Observed with Noise." *Scandinavian Journal of Statistics*, **35**(2), 322–334. doi:10.1111/j.1467-9469.2007.00583.x.

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. doi:10.2307/2347366.

Diggle P, Marron JS (1988). "Equivalence of Smoothing Parameter Selectors in Density and Intensity Estimation." *Journal of the American Statistical Association*, **83**(403), 793–800. doi:10.2307/2289308.

Diggle PJ (2003). *Statistical Analysis of Spatial Point Patterns*. Hodder Arnold Publication, Oxford.

Diggle PJ (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. CRC press. doi:10.1201/b15326.

Duong T, Hazelton M (2003). "Plug-in Bandwidth Matrices for Bivariate Kernel Density Estimation." *Journal of Nonparametric Statistics*, **15**(1), 17–30. doi:10.1080/10485250306039.

Epanechnikov VA (1969). "Non-parametric Estimation of a Multivariate Probability Density." *Theory of Probability & Its Applications*, **14**(1), 153–158. doi:10.1137/1114019.

Gelfand AE, Diggle P, Guttorp P, Fuentes M (2010). *Handbook of Spatial Statistics*. CRC press. doi:10.1201/9781420072884.

Granovsky BL, Müller HG (1991). "Optimizing Kernel Methods: A Unifying Variational Principle." *International Statistical Review/Revue Internationale de Statistique*, pp. 373–388. doi:10.2307/1403693.

Guan Y (2008). "On Consistent Nonparametric Intensity Estimation for Inhomogeneous Spatial Point Processes." *Journal of the American Statistical Association*, **103**(483), 1238–1247. doi:10.1198/016214508000000526.

Guan Y, Loh JM (2007). "A Thinned Block Bootstrap Variance Estimation Procedure for Inhomogeneous Spatial Point Patterns." *Journal of the American Statistical Association*, **102**(480), 1377–1386. doi:10.1198/016214507000000879.

Illian J, Penttinen A, Stoyan H, Stoyan D (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons. doi:10.1002/9780470725160.

Johnson NL, Kotz S, Balakrishnan N (1995). *Continuous Univariate Distributions, Volume 2*, volume 289. John Wiley & Sons.

Jones MC (1993). "Simple Boundary Correction for Kernel Density Estimation." *Statistics and Computing*, **3**, 135–146. doi:10.1007/BF00147776.

Jones MC, Foster PJ (1996). "A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation." *Statistica Sinica*, pp. 1005–1013. URL https://www.jstor.org/stable/24306056.

Lejeune M, Sarda P (1992). "Smooth Estimators of Distribution and Density Functions." *Computational Statistics & Data Analysis*, **14**(4), 457–471. doi:10.1016/0167-9473(92)90061-J.

Marron JS, Ruppert D (1994). "Transformations to Reduce Boundary Bias in Kernel Density Estimation." *Journal of the Royal Statistical Society: Series B (Methodological)*, **56**(4), 653–671. doi:10.1111/j.2517-6161.1994.tb02006.x.

Müller HG (1991). "Smooth Optimum Kernel Estimators Near Endpoints." *Biometrika*, **78**(3), 521–530. doi:10.2307/2337021.

Müller HG, Stadtmüller U (1999). "Multivariate Boundary Kernels and a Continuous Least Squares Principle." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**(2), 439–458. doi:10.1111/1467-9868.00186.

Ogata Y (1998). "Space-Time Point-process Models for Earthquake Occurrences." *Annals of the Institute of Statistical Mathematics*, **50**, 379–402. doi:10.1023/A:1003403601725.

Ord J (1978). "How Many Trees in a Forest." *Mathematical Scientist*, **3**, 23–33.

Schaap WE, Van De Weygaert R (2000). "Continuous Fields and Discrete Samples: Reconstruction through Delaunay Tessellations." *arXiv preprint*. URL https://arxiv.org/abs/astro-ph/0011007.

Schuster EF (1985). "Incorporating Support Constraints into Nonparametric Estimators of Densities." *Communications in Statistics-Theory and Methods*, **14**(5), 1123–1136. doi:10.1080/03610928508828965.

van Lieshout MCNM (2012). "On Estimation of the Intensity Function of a Point Process." *Methodology and Computing in Applied Probability*, **14**, 567–578. doi:10.1007/s11009-011-9244-9.

Van Lieshout MNM (2011). "A J–function for Inhomogeneous Point Processes." *Statistica Neerlandica*, **65**(2), 183–201. doi:10.1111/j.1467-9574.2011.00482.x.

Wand MP, Jones MC (1994). "Multivariate Plug-in Bandwidth Selection." *Computational Statistics*, **9**(2), 97–116.

A. Proof for Lemma 1

Recall that N is not a constant but a random variable, so we have a double stochastic scenario, on one side the randomness provided by N and on the other the randomness of the point process. To deal with this we use the conditional mean and we consider also some tools related to real number series. Consider first the mean,

$$\begin{aligned}\mathbf{E}[\hat{\lambda}_{0,h}(x)] &= \mathbf{E}[\mathbf{E}[\hat{\lambda}_{0,b}|N = n > 0]] = \sum_{n=1}^{\infty} \mathbf{E}[\hat{\lambda}_{0,h}(x)|N = n]\mathbf{P}(N = n) \\ &= \sum_{n=1}^{\infty} \mathbf{E}[\hat{\lambda}_{0,h}(x)|N = n]\frac{e^{-m}m^n}{n!} = (1 - e^{-m})\mathbf{E}[\hat{\lambda}_{0,h}(x)|N = n],\end{aligned}$$

where we have used that

$$\mathbf{E}[\hat{\lambda}_{0,h}(x)|N = n] = \int \lambda_0(x) \prod_{j=1}^{2} K^*_{x_j,h_j}(t)dt = \mathbf{E}_\eta(\lambda_0(Y)), \tag{7}$$

where the Y_j are i.i.d beta distributed $(B(\frac{x_j}{h_j} + 1, \frac{1-x_j}{h_j} + 1))$. Let μ_j and σ_j^2 be the mean and the variance of Y_j respectively. From Johnson, Kotz, and Balakrishnan (1995) it may be shown that there exists a constant M such that

$$\begin{aligned}\mu_{Y_j} &= x_j + h_j(1 - 2x_j) + \Delta_1(x_j) \\ \sigma^2_{Y_j} &= b_j x_j(1 - x_j) + \Delta_2(x_j),\end{aligned}$$

where $\Delta_j(x_j) \leq M h_j^2$ for $j = 1$ and 2. Hence, the remainder terms in the above expansions for μ_{Y_j} and $\sigma^2_{Y_j}$ are uniformly $O(h^2)$.

The product kernel $\prod_{j=1}^{2} K^*_{x_j,h_j}(t)$ is used in Equation (7) as the intensity function. To derive the asymptotic properties, we perform a second-order Taylor expansion of $\lambda_0(y)$ around the point μ_j

$$\begin{aligned}\lambda_0(y) = \lambda_0(\mu) &+ \sum_{j=1}^{2}(y_j - \mu_j)\frac{\partial\lambda_0(y)}{\partial x_j} + \sum_{j=1}^{2}(y_j - \mu_j)^2\frac{\partial^2\lambda_0(y)}{\partial x_j^2} \\ &+ \sum_{j\neq l}(y_j - \mu_j)(y_l - \mu_l)\frac{\partial\lambda_0(y)}{\partial x_j \partial x_l} + o\left(\sum_{j=1}^{2} h_j\right).\end{aligned}$$

Taking the expectation from both sides of the latter equation we can write

$$\begin{aligned}\mathbf{E}(\hat{\lambda}_{0,h}(x)) &= (1 - e^{-m})\lambda_0(x) + (1 - e^{-m})\sum_{j=1}^{2} h_j(1 - 2x_j)\lambda_0^j(x)) \\ &+ \frac{(1 - e^{-m})}{2}\sum_{j=1}^{2} x_j h_j(1 - x_j)\lambda_0^{jj}(x) + o((1 - e^{-m})\sum_{j=1}^{2} h_j^2).\end{aligned}$$

Hence, the bias of the bivariate intensity function estimate is given by

$$\text{Bias}(\hat{\lambda}_{0,h}(x)) = -e^{-m}\lambda_0(x) + (1 - e^{-m})\sum_{j=1}^{2} u_j(x) + o\left((1 - e^{-m})\sum_{j=1}^{2} h^2\right), \tag{8}$$

where

$$u_j(x) = h_j\left((1 - 2x_j)\frac{\partial\lambda_0(x)}{\partial x_j} + \frac{x_j}{2}(1 - x_j)\frac{\partial^2\lambda_0(x)}{\partial x_j^2}\right).$$

When $h_1 = h_2 = h$ holds, hence

$$\int \left(\mathbf{E}(\hat{\lambda}_{0,h}(x) - \lambda_0(x)\right)^2 dx = \int \left(-e^{-m}\lambda_0(x) + (1-e^{-m})\sum_{j=1}^{2} u_j(x)\right)^2 dx + o\left((1-e^{-m})\sum_{j=1}^{2} h^2\right).$$

B. Proof for Lemma 2

The variance of the approximation, by definition, thus, if

$$A(m) = \mathbf{E}\left(\frac{1}{N}\mathbb{1}_{N\neq 0}\right) \quad \text{and} \quad g(X_{ij}) = \prod_{j=1}^{2} K^*_{x_j,h_j}(X_{ij})$$

we have thus

$$\mathbf{E}(g(x)^2) = \int g(x)^2\lambda_0(x)dx = \mathbf{E}\left(\prod_{j=1}^{2}(K^*_{x_j,h_j})^2(X_j)\right) = \prod_{j=1}^{2} A_{h_j}(x_j)\mathbf{E}(\lambda_0(Z)),$$

where γ_{x_j} is a $Beta\left(\frac{2x_j}{h_j}+1, \frac{2(1-x_j)}{h_j}+1\right)$ random variable.

$$A_{h_j}(x_j) = \frac{B\left(\frac{2x_j}{h_j}+1, \frac{2(1-x_j)}{h_j}+1\right)}{B\left(\frac{x_j}{h_j}+1, \frac{(1-x_j)}{h_j}+1\right)}.$$

For Chen (1999),

$$A_{h_j}(x_j) = \begin{cases} \frac{h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}}, & \text{if } \frac{x_j}{h_j} \text{ and } \frac{1-x_j}{h_j} \to \infty \\ \frac{h_j^{-1}\Gamma(2k_j+1)}{2^{2k_j+1}\Gamma^2(k_j+1)}, & \text{if } \frac{x_j}{h_j} \to k_j \text{ or } \frac{1-x_j}{h_j} \to k_j. \end{cases}$$

$$\mathrm{Var}(\hat{\lambda}_{0,h}(x)) = A(m)\prod_{j=1}^{2} A_{h_j}(x_j)\mathbf{E}(\lambda_0(Z(x)), \tag{9}$$

where $Z = (Z_1, Z_2)$ and the random variables Z_j are independent and beta distributed.
From a Taylor expansion,

$$\mathbf{E}(\lambda_0(Z)) = \lambda_0(x) + O(h).$$

Therefore

$$\frac{1}{A(m)}\mathrm{Var}(\hat{\lambda}_{0,h}(x)) = \lambda_0(x)\left(\prod_{j\in I}\frac{h^{-\frac{1}{2}}(x(1-x))^{-\frac{1}{2}}}{2\sqrt{\pi}}\right)\left(\prod_{j\in I^c}\frac{\Gamma(2k_j+1)}{2^{2k_j+1}\Gamma^2(k_j+1)}h_j^{-1}\right),$$

where $I = \{j, \frac{x_j}{h_j} \longrightarrow \infty\}$.

The second product disappears in the integrated variance. Let $\delta_j = h_j^{1-\epsilon}$, $0 < \epsilon < 1/2$ and $\delta = (\delta_1, \delta_2)$, then

$$\begin{aligned}\int \mathrm{Var}(\hat{\lambda}_{0,h}(x)) &= A(m)\int \mathrm{Var}(\hat{\lambda}_{0,h}(x))dx + \int \mathrm{Var}(\hat{\lambda}_{0,h}(x)) \\ &= O\left(\prod_{j\in I} h_j^{1/2-\epsilon}\prod_{j\in I^c} h_j^{-\epsilon}\right) + \int\left(\prod_{j\in I}\frac{h^{-\frac{1}{2}}(x(1-x))^{-\frac{1}{2}}}{2\sqrt{\pi}}\right)\lambda_0(x)dx \\ &= o\left(\prod_{j=1}^{2} h_j^{-1/2}\right) + \frac{1}{2\sqrt{\pi}}\int\left(\prod_{j=1}^{2} h^{-\frac{1}{2}}(x(1-x))^{-\frac{1}{2}}\right)\lambda_0(x)dx\end{aligned}$$

$$\int \mathrm{Var}(\hat{\lambda}_{0,h}(x)) = \frac{1}{2\sqrt{\pi}}\int\left(\prod_{j=1}^{2} h^{-\frac{1}{2}}(x(1-x))^{-\frac{1}{2}}\right)\lambda_0(x)dx + o\left(\prod_{j=1}^{2} h_j^{-1/2}\right).$$

By combining Equation (8) and Equation (9), we obtain the mean integrated squared error of the non-parametric estimator with the beta kernel intensity.

Affiliation:

Marcel Sihintoé Badiane
Applied Research Laboratory in Natural Sciences (LARASCINA)
Kindia University (Guinea)
E-mail: m.badiane238@zig.univ.sn

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 55 *Submitted:* 2025-02-13
2026 *Accepted:* 2025-08-22