Intensity estimation for spatial point processes observed with noise

Lionel Cucala[1]

[1] Laboratoire de Statistiques et Probabilits, Universit Paul Sabatier, 31062 Toulouse cedex 9, France

Abstract: This article proposes new kernel estimators of the intensity function of spatial point processes taking into account position errors. The asymptotic properties of these estimators are derived. A simulation study compares their results to the results of the classical kernel estimator and shows that the edge-corrected deconvoluting kernel estimator is the most appropriate.

Keywords: Deconvolution; Kernel density estimation; Heterogeneous Poisson point process; Measurement error; Spatial point patterns.

1 Introduction

In this paper, we propose new kernel estimators of the intensity function of the spatial point processes which take into account the location errors by a deconvolution method. For simplicity, we develop them in the case of bidimensional point processes. Section 2 is an introduction to the perturbed point processes. We then define the new estimator and discuss its properties in Section 3. We present an asymptotic study in Section 4 and an adaptation of an existing bandwidth selection procedure to this specific problem in Section 5. The usefulness of the estimator is assessed by its application to simulated data in Section 6.

2 Perturbed point processes

Consider a Poisson point process $\mathbf{Y}$ in $\mathbb{R}^2$ with intensity function $\lambda_Y()$. We only observe the point pattern $Z = \{z_1, \cdots, z_N\}$ in the domain $D \subset \mathbb{R}^2$ according to the model:

$$z_i = y_i + \epsilon_i, \tag{1}$$

where $(y_i : i = 1, \cdots, N)$ are events issued from the process $\mathbf{Y}$ and $(\epsilon_i : i = 1, \cdots, n)$ are i.i.d. with known isometric density function $g(.)$ and represent the location errors. We will also assume that the errors ϵ_i are independant from the true locations y_i. Our goal is to estimate the intensity function $\lambda_Y(s)$ for every point $s \in D$.

3 The deconvoluting kernel intensity estimators

Denote $\lambda_Z(.)$ the intensity function of the perturbed process $\mathbf{Z}$.
Based on the observations Z, the edge-corrected kernel estimator for $\lambda_Z(.)$ is (Diggle, 1985):

$$\forall s \in \mathbb{R}^2, \hat{\lambda}_{Z,h}(s) = \begin{cases} \frac{\sum_{j=1}^{n} \frac{1}{h^2} K\left(\frac{s-z_j}{h}\right)}{\int_D \frac{1}{h^2} K\left(\frac{s-u}{h}\right)\nu(du)} & \text{if } \int_D \frac{1}{h^2} K\left(\frac{s-u}{h}\right)\nu(du) \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $K(.)$ is a kernel function and ν represents the Lebesgue measure.
A deconvoluting estimator for $\lambda_Y(s)$ inspired by Stefanski & Carroll (1990) 's density estimator is

$$\begin{aligned} \lambda^*_{Y,h}(s) &= \sum_{j=1}^{n} \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{is't} \Big\{ \int_{\mathbb{R}^2} e^{-it'z} \frac{1}{h^2} K\big(\frac{z-z_j}{h}\big)\nu(dz)/\mathcal{F}(g)(t) \Big\} \nu(dt) \\ &= \sum_{j=1}^{n} \frac{1}{h^2} K^*_h\Big(\frac{s-z_j}{h}\Big), \end{aligned}$$

where $K^*_h(t) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{it'y} \mathcal{F}(K)(y)/\mathcal{F}(g)(y/h)dy$.
A way of adapting the estimator $\lambda^*_{Y,h}$ to the limited domain context is to define

$$\lambda^{**}_{Y,h}(s) = \frac{\sum_{j=1}^{n} \frac{1}{h^2} K^*_h\left(\frac{s-z_j}{h}\right)}{\int_D \frac{1}{h^2} K^*_h\left(\frac{s-u}{h}\right)\nu(du)}.$$

4 Asymptotic study

It is shown that, in this framework, none of the intensity estimators is asymptotically unbiased.

5 The bandwidth selection procedure

We adapt the normal reference rule to the bidimensional case with noisy observations.

6 A simulation study

The choice of the kernel is of secondary importance for the quality of our estimator. Here, for practical purpose, we choose a bidimensional kernel whose Fourier transform has compact support. The chosen kernel is a product kernel $K(x, y) = K_0(x)K_0(y)$, where

$$K_0(t) = \frac{48}{\pi} \frac{t^3 \cos(t) - 6t^2 \sin(t) + 15 \sin(t) - 15t \cos(t)}{t^7}.$$

TABLE 1. Gaussian error, σ=0.02

	ISE	ISE^*	ISE^{**}
1st quartile ($*10^3$)	1.0600	1.6745	0.9038
median ($*10^3$)	1.3939	1.9613	1.0279
3rd quartile ($*10^3$)	1.5899	2.2432	1.3158

TABLE 2. Gaussian error, σ=0.05

	ISE	ISE^*	ISE^{**}
1st quartile ($*10^3$)	0.8185	1.4153	0.6655
median ($*10^3$)	1.2474	1.7199	0.9298
3rd quartile ($*10^3$)	1.5281	1.8908	1.2138

An inhomogeneous Poisson process is simulated in $[0,1]^2$ enlarged by a guard area with intensity

$$\lambda_Y(s) = C\big[1 + 0.7\cos\big(2\pi(||s|| - 0.5)\big)\big],$$

where C is a constant chosen such that the expected number of events in $[0,1]^2$ is 100. This is done by an acceptation-rejection method (Gentle, 2002).
The location errors $\{\epsilon_i,\ i = 1, \cdots, n\}$ are then simulated and added to the simulated locations:

$$z_i = y_i + \epsilon_i.$$

Only the observations z_i in $[0,1]^2$ will be used to estimate the intensity.
From the simulated sample, we compute the estimates $\hat{\lambda}_{Z,h_{opt}}$, λ^*_{Y,h^*} and λ^{**}_{Y,h^*}, where h_{opt} is the bandwidth obtained by the classical cross-validation procedure (Silverman, 1986) and h^* is the bandwidth obtained via the procedure described in section 5.
Denote $ISE = \int_{[0,1]^2} \big(\hat{\lambda}_{Z,h_{opt}} - \lambda_Y(s)\big)^2 \nu(ds)$,
$ISE^* = \int_{[0,1]^2} \big(\lambda^*_{Y,h^*}(s) - \lambda_Y(s)\big)^2 \nu(ds)$, $ISE^{**} = \int_{[0,1]^2} \big(\lambda^{**}_{Y,h^*}(s) - \lambda_Y(s)\big)^2 \nu(ds)$.
This procedure is repeated m times and we compute the empirical quartiles of ISE, ISE^* and ISE^{**}. Tables 1, 2 and 3 give the results when ϵ follows a Gaussian distribution with mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and variance matrix $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$, and the number m of realizations is equal to 10.
In each case, the estimator λ^{**}_{Y,h^*} gives the best results. The results of the estimator λ^*_{Y,h^*} are not better, or even worse, than the ones obtained by the classical Diggle estimator $\hat{\lambda}_{Z,h_{opt}}$, suggesting that deconvolution and

TABLE 3. Gaussian error, σ=0.1

	ISE	ISE^*	ISE^{**}
1st quartile ($*10^3$)	0.7669	1.2194	0.7223
median ($*10^3$)	0.8854	1.4123	0.8733
3rd quartile ($*10^3$)	1.4305	1.6451	1.2544

edge-correction should both be considered when dealing with perturbed locations in a bounded domain.

Acknowledgments: This research project is within the scope of my PhD thesis work under the direction of Christine Thomas-Agnan. I gratefully thank her for her constant help and support. It was initiated by my visit to Ohio State University and my rewarding discussions with Noel Cressie.

References

Diggle, P.J. (1985).
A Kernel Method for Smoothing Point Process Data.
Applied Statistics, **34**, 138-147.

Gentle, J.E. (2002).
Elements of computational statistics.
Springer-Verlag.

Silverman, B.W. (1986).
Density Estimation for Statistics and Data Analysis.
Chapman and Hall.

Stefanski, L. and Carroll, R.J. (1990).
Deconvoluting kernel density estimators.
Statistics, **21**, 169-184.