A unified framework for measuring industry location characteristics based on marked spatial point processes

Florent Bonneu and Christine Thomas-Agnan

June 26, 2011

We propose a unified framework for defining measures of industrial concentration based on micro-geographic data. These encompass the Duranton-Overman and the Marcon-Puech indices. We discuss the basic requirements for such measures introduded by Duranton and Overman (2005) and we propose five additional requirements. We describe several types of concentration depending on the second order characteristics of the marginal patterns of positions and of marks but also on their mutual dependence. We also discuss the null assumptions classically used for testing aggregation of a particular sector. The framework we propose is based on some second order characteristics of marked spatial point processes discussed in Illian et al. (2008). The general measure involves a cumulative and a non-cumulative version. This allows us to propose an alternative version of the Duranton-Overman index with a proper baseline as well as a cumulative version of this index.

Abstract

Abstract

Keywords : Keywords.

Mathematics Subject Classification (2000) : MSC.

1 Introduction

Krugman's theory of economic geography states that "instead of spreading out evenly around the world, production will tend to concentrate in a few countries, regions, or cities, which will become densely populated but also have higher levels of income." Empirical evidence brings out that jobs and industries are highly clustered in a limited number of regions.There are several forces that induce economics agglomeration. First of all, plants locate near to each other because of agglomeration spillovers (localization economies and urbanization economies) or local amenities. Returns to scale induce industries to concentrate their production in a small number of business units $\rightarrow$ and there is interdependence between firm's location choice (snowball effect mechanism). Note that imilar questions arise in other disciplines: for example in ecology when studying spatial concentration of biomass.

There are numerous motivations for studying the geographic concentration of economic sectors. Such a measure allows to understand the determinants of localization, compare different sectors with respect to agglomeration/dispersion and predict the evolutions of localization. A similar question is that of co-localization and interactions between sectors for which measures can be generally derived from the former. Another related issue is cluster detection but we do not include this problem in the present paper.

Until 2000, all studies about geographic concentration of economic activity use areal data for measuring spatial concentration. The localization of firms is not available and data consists only in counts aggregated on administrative zones. There is a large literature on this topic and many measures including the Herfindahl index, the locational Gini index (which is the Gini index of the localization ratio), the Ellison-Glaeser index, the Maurel-Sédillot's index and many others. However these measures depend upon the aggregation level (Modifiable Areal Unit Problem) and most importantly they do not take geography into account: a permutation of the sites does not affect the measure !

A new vein of this literature arises in 2002 considering the treatment of micro-geographic data. This type of data usually consists in the precise location of firms together with a size measure such as the number of employees. Duranton, G. and Overman, H.G. (2005) introduce a measure based on the distribution of inter-distances between firms. Marcon, E. and Puech, F. (2002, 2010) introduce another measure based on Ripley's K-function. Combes P-J., Meyer T., and Thisse J-F. (2008) survey this literature. Espa, Giuliani and Arbia (2010) use a model-based approach to assess concentration. Duranton et Overman (2002) list five properties that a good measure of industrial concentration should satisfy

1. DO1 The index must be comparable from one sector to the other (should not depend upon the number of firms in the sector)
2. DO2 The index must take into account the overall manufacturing geographical pattern
3. DO3 The index must take into account the structural differences of a particular sector / country
4. DO4 The index must be independent of the geographical scale of observation
5. DO5 The index must be assorted with a level of statistical significance

DO2 means that the benchmark should not be spatial homogeneity because obviously geographic and demographic factors influence industrial location. DO3 'means that one should take into account firm's sizes in the measure. DO4 is related to the MAUP. In this paper, we introduce five additional requirements namely

1. BTA1 The index must be an empirical measure associated to a well identified theoretical characteristic. This last point is not satisfied by the current candidates in the literature. This point may allow to satisfy DO5 without using Monte Carlo methods.
2. BTA2 The index must take into account spatial inhomogeneity of a particular sector (for example fishing)
3. BTA3 The index must take into account a possible inhomogeneity of the distribution of firm's sizes in space.
4. BTA4 The index must have a known and constant benchmark in the absence of concentration.
5. BTA5 For testing concentration, a null hypotheses must be correctly specified.

Note that the Duranton-Overman index as well as the Marcon-Puech index are both inspired from the marked point process theory. However none of them corresponds to a well identified statistical parameter. This weakness relates to the absence of clear definition of the theoretical meaning of spatial concentration only introduced through empirical measures. We intend to fill this gap and make progress in the understanding of spatial concentration. None of the cited measures takes satisfies BTA2 neither BTA3. With respect to BTA4, the Marcon-Puech index has a constant benchmark but not the Duranton-Overman's one.

In section 2, we present the mathematical tools of the spatial point process theory. In section 2.4, we introduce several types of spatial concentration.

2 The relevance of random spatial point patterns theory

2.1 The theoretical model

Spatial distribution of firms together with their sizes can be modeled using random spatial point patterns associated with a mark (the size, the sector). Spatial point processes (PP) are models for a random spatial configuration of a random number of points N (for us:location of firms for different industrial sectors). Two important aspects of their description are spatial inhomogeneity and spatial interaction. Spatial Inhomogeneity relates to the fact that some regions may have a mean number of points higher than others, for example when studying the spatial distribution of population, mountainous zones may be less populated. Spatial interaction relates to the dependence between points locations pairs. For example, the competition for food may generate repulsion between animals positions, whereas when looking at infectious disease cases, contagion generates attraction between spatial occurrences of a disease. One talks about a Marked PP when a random mark is associated to each position (for us: number of employees + sector).

Let $\mathcal{X}$ be a subset of $\mathbb{R}^2$, a configuration of n points of $\mathcal{X}$ is a non ordered set of n points $x = \{x_1, \cdots, x_n\}$. A PP model is a model for a random configuration with a random countable number N of points (possibly zero or infinity), repetitions being allowed. Two mathematical approaches exist for this theory: they are based on locally finite random sets of points of $\mathcal{X}$ or alternatively on random measures on $\mathcal{X}$ and we refer the reader to Moller and Waagepetersen (2004) or to Illian et al.(2008) for precise definitions and properties. A PP is **stationary** if its law is invariant under translations of the configurations. A PP is **isotropic** if its law is invariant under the rotations of the configurations. Figure 2.1 illustrates the notion of non stationarity on the left pane, the notion of anysotropy on the central panel and the absence of interaction on the right panel.

Let M be a space of "marks" and for each configuration X let m_X be a random variable with values in M. Then one says that (X, m_X) is a **marked** PP with **mark space** M. In practice, we consider the case M finite, or M subset of $\mathbb{R}^p$. Marks can be dependent or not from positions. The "Random labeling" hypotheses means that marks are independent from positions. Figures 2.1 and 2.1 show examples of realizations of such processes. On the right of Figure 2.1, the process is a Homogeneous Poisson process which is the model for homogeneity and absence of interaction between points. On the left of Figure 2.1 is an aggregated process which means that there is interaction and it is of the attraction kind. On the left of Figure 2.1 is a regular process which means that there is interaction and it is of the repulsion kind and on the right of Figure 2.1 is a marked PP with circles showing the mark through their radius.
The order 1 characteristic of a PP is given by the intensity. Let $N_X(B)$ be the number of points of PP X in B. The intensity measure is dcefined by

$$\Lambda(B) = \mathbb{E}(N_X(B)).$$

When Λ is absolutely continuous with respect to the Lebesgue measure, one can write

$$\Lambda(B) = \int_B \lambda(u)du,$$

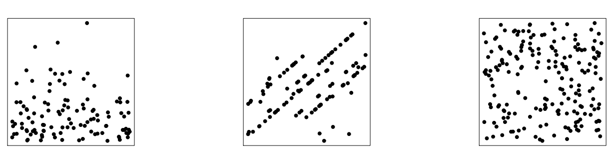

Figure 1: From left to right: non stationarity, anisotropy and absence of interaction

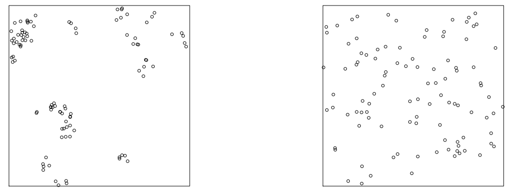

Figure 2: Aggregated PP (left) and Homogeneous Poisson PP (right)

where the **intensity function** can be interpreted as follows: $\lambda(u)du$ is the probability of occurrence of a point in the infinitesimal ball of center u and radius du, or the limit

$$\lambda(x) = \lim_{|dx| \longrightarrow 0} \frac{\mathbb{E}(N_X(dx))}{\mid dx \mid},$$

where $\mid dx \mid$ is the volume of dx.

There is a relationship between the density of points regarded as i.i.d. realizations of a measure and the intensity: the intensity is the product of the density by the total expected number of points.

The order 2 characteristics of a PP can be specified by the order two factorial moment measure (the mean number of points pairs with a point in A and the other in B):

$$\Lambda^{(2)}(A \times B) = \mathbb{E}(\sum_{u,v \in X : u \neq v} 1(u \in A, v \in B))$$

When $\Lambda^{(2)}$ is absolutely continuous with respect to the Lebesgue measure, one can write

$$\Lambda^{(2)}(A \times B) = \int_A \int_B \lambda^{(2)}(u,v)dudv$$

where $\lambda^{(2)}(u,v)dudv$ can be interpreted as the probability of joint occurrence of a point in the infinitesimal ball of center u with radius du and of a point in the infinitesimal ball of center v and of radius dv, or the limit

$$\lambda^{(2)}(x,y) = \lim_{|dx| \longrightarrow 0, |dy| \longrightarrow 0} \frac{\mathbb{E}(N_X(dx)N_X(dy))}{\mid dx \mid \mid dy \mid}$$

Another way of characterizing the second order structure is through the pair correlation function which is defined by

$$g(x,y) = \frac{\lambda^{(2)}(x,y)}{\lambda(x)\lambda(y)}$$

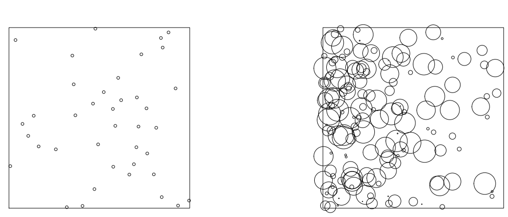

Figure 3: Regular PP (left) and Inhomogeneous Poisson PP (right)

with the convention $\frac{a}{0} = 0$ if $a \geq 0$.
A PP is said to be "second order reweighted stationarity" when g is translation invariant.
At last, a third way of characterizing the second order structure is through the Ripley's K function. If X is "second order reweighted stationary" and isotropic, the **Ripley's K function** is defined by

$$K(r) = \pi \int_0^r ug(u)du,$$

In this case, $\lambda K(r)$ is the mean number of points within radius r of the origin given that the origin belongs to the configuration.
The assumption of complete spatial randomness (CSR) is embodied by the Poisson homogeneous process (PPP) for which we have
$K(r) = \pi r^2$ and $g(r) \equiv 1$.

A scaled version of the K-function is sometimes used and called the L-function $L(r) = \sqrt{\frac{K(r)}{\pi}} - r$.

2.2 The estimators of theoretical characteristics

In the isotropic and homogeneous case, one can estimate the intensity by

$$\hat{\lambda}(x) = \frac{N}{\mid \mathcal{X} \mid}$$

and the Ripley's K-function by

$$\hat{K}(r) = \frac{\mid \mathcal{X} \mid}{N(N-1)} \sum_{i \neq j} w_{i,j} 1(\| x_i - x_j \| \leq r)$$

where $w_{i,j}$ is a boundary correction factor.

In the isotropic inhomogeneous case, one can estimate the intensity by

$$\hat{\lambda}(x) = \sum_{\xi \in \mathcal{X}} \kappa((x - \xi)/h)/h$$

and the Ripley's K-function by

$$\hat{K}_{inhom} = \frac{1}{\mid \mathcal{X} \mid} \sum_{i \neq j} w_{i,j,r} \frac{1(\| x_i - x_j \| \leq r)}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)}$$

where $w_{i,j,r}$ is a boundary correction factor. The following is an estimator of the pair correlation function

$$\hat{g}(r) = \frac{1}{2\pi r} \sum_{i=1}^{n} \sum_{j \neq i} w_{i,j,r} \frac{h^{-1}\kappa\left(\frac{r - \|x_i - x_j\|}{h}\right)}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)}.$$

Figure 2.2 shows three processes, from left to right a regular, a Homogeneous Poisson, and an aggregated process. A circle centered on a configuration point illustrates the fact that the K-function counts mean the number of points within a given radius of a point in the configuration. The last panel shows the corresponding L-functions and we see that the regular process has an L-function below 1, the aggregated process above 1 and the Homogeneous Poisson is very near 1.

2.3 Use of the K-function to test complete spatial randomness

Even though we said that CSR was not a good benchmark for studying spatial concentration of industry, we will here illustrate how the Ripley's K-function is used to test for CSR. The reason for showing this is that this methodology will inspire the technique we introduce later on in the paper. Figure 2.3 shows a realization of a inhomogeneous Poisson process on the left panel. The central panel shows the ordinary K-function and the right panel the inhomogeneous K-function: both are displayed together with an empirical envelope obtained by Monte Carlo simulations of a process with the same intensity under the CSR assumption. The central K-curve is out of the envelope whereas the right K-curve is inside the envelope: the reason is that the estimation of the K-function in the central panel does not take into account inhomogeneity whereas this is done using the inhomogeneous K estimator on the right panel. This allows to conclude that the fact that the curve is outside the envelope in the central panel is not due to interaction but rather due to inhomogeneity. A parallel can be done with a time series situation when the unaccounted presence of a trend may reveal a wrong serial correlation.

2.4 Characteristics of a marked PP

Let (X, M) be a marked PP, homogeneous for positions, and let $f(m_1, m_2)$ be a weighting function, we define a weighted version of $\alpha^{(2)}$ by

$$\alpha_f^{(2)}(A \times B) = \mathbb{E}\left[\sum_{u,v \in X: u \neq v} f(m_1, m_2) \mathbb{1}_A(u) \mathbb{1}_B(v)\right].$$

When $\alpha^{(2)}$ is absolutely continuous wrt the Lebesgue measure, one can write

$$\alpha_f^{(2)}(A \times B) = \int_A \int_B \rho_f^{(2)}(u, v) du dv$$

then $\rho_f^{(2)}$ is called second order product density of X for weighting scheme f.

3 The different faces of spatial concentration

In this section, we want to discuss the definition of spatial concentration and distinguish between several types. Figure 3 shows two examples of spatially concentrated marked processes. In the left panel, the marks are constant, the point process is an inhomogeneous Poisson PP and the concentration aspect of the configuration is due to the inhomogeneity of positions and not to interaction. In the right panel,

Figure **??** shows three examples of spatially concentrated marked processes. In the left panel, the marks are constant and the concentration aspect of the configuration is due to aggregation between the points. In the central panel, the marks are called constructed marks: they are equal

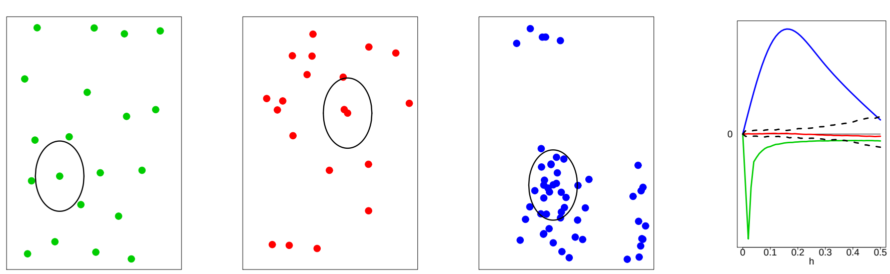

Figure 4: From left to right: Regular PP, PPP, Aggregated PP and L-functions

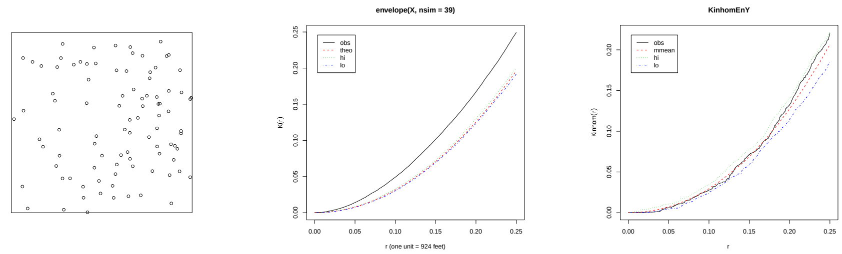

Figure 5: Use of K to test CSR

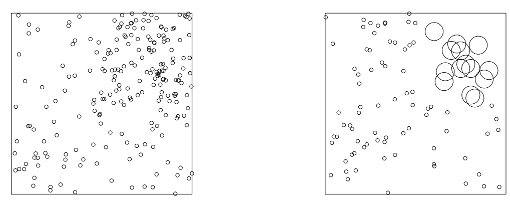

Figure 6: The different faces of concentration: order 1 concentration

to the number of neighbors at $dist \leq 0.1$ and the concentration is due to the correlation between marks and point positions. In the right panel, the marks are also constructed marks but now they are equal to the distance between each point and its nearest neighbor, the concentration is therefore also due to the correlation between marks and point positions.

4 Indices based on inter-points distances

4.1 The Duranton-Overman and the Marcon-Puech indices

The Duranton-Overman index (Duranton and Overman, 2005) and the Marcon-Puech index (Marcon and Puech, 2002, 2010) are both based on inter-point distances. The Duranton-Overman index is a non cumulative index defined by

$$i_{DO}(r) = \frac{\sum_i \sum_{j \neq i} h^{-1} w\left(\frac{r - \|x_i - x_j\|}{h}\right) m_i m_j}{\sum_i \sum_{j \neq i} m_i m_j}$$

When the mark is a count, which is the case for the number of employees, it can be compared to the Parzen-Rosenblatt density estimator associated to a replicated point process of positions (number of replications equal to the mark) considering that the points positions are i.i.d.

Marcon and Puech note that i_{DO} does not account for order 1 inhomogeneity. They propose to perform this correction by using the union of all the available sectors. No correction is then possible if only one sector is available. The Marcon-Puech index is a cumulative index defined by

$$I_{MP}(r) = \sum_{i=1}^{N_s} \frac{\sum_{j=1, j \neq i}^{N_s} m_j \mathbb{1}(\|x_{i,s} - x_{j,s}\| \leq r)}{\sum_{j=1, j \neq i}^{N} m_j \mathbb{1}(\|x_{i,s} - x_j\| \leq r)} / \sum_{i=1}^{N_s} \frac{\sum_{j=1, j \neq i}^{N_s} m_j}{\sum_{j=1, j \neq i}^{N} m_j} \quad \forall r > 0,$$

$I_{MP}(r) > 1$ indicates that there are proportionally more employees close to plants of sector s within a radius r than in the whole area. Note that $I_{MP}(r)$ can be written $J_{MP}(r)/J_{MP}(\infty)$ where

$$J_{MP}(r) = \sum_{i=1}^{N_s} \frac{\sum_{j=1, j \neq i}^{N_s} m_j \mathbb{1}(\|x_{i,s} - x_{j,s}\| \leq r)}{\sum_{j=1, j \neq i}^{N} m_j \mathbb{1}(\|x_{i,s} - x_j\| \leq r)}.$$

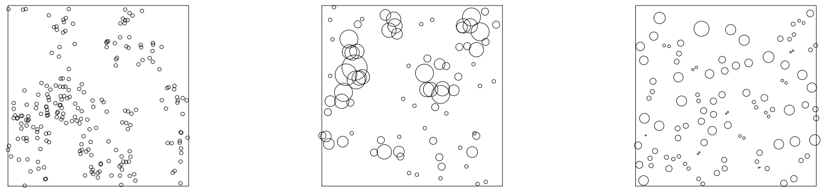

Figure 7: The different faces of concentration: order 2 concentration

$J_{MP}(r)$ is the average proportion of employees of sector s within a given radius r.

4.2 The weaknesses of the Duranton-Overman and the Marcon-Puech indices

Both in the Duranton-Overman and the Marcon-Puech framework, the simulations are done conditionally upon the positions and marks (sector + number of employees) are randomly reassigned to the observed positions. This simulation framework is not compatible with BTA3. There are a number of other drawbacks namely

1. there are no theoretical characteristics clearly associated to these indices (cf BTA1)
2. the possible dependence between marks and positions is not incorporated in the index formula (cf BTA3)
3. DO does not take into account inhomogeneity of location intensity (cf BTA2)
4. no clear benchmark for DO (cf BTA4)
5. no edge correction (implies bias for large r)
6. underlying assumption that all sectors are issued from the same type of process ("overall manufacturing")(cf simulations under H_0)

5 Another step towards a unified theory

In this section, we present theoretical characteristics of spatial marked point processes which will allow us to cast the previous two approaches in a same mould and to correct their weaknesses.

5.1 The theoretical characteristics

In the non stationary case, for any weight function k, we introduce an order one characteristic called the weighted intensity measure α_k

$$\alpha_k(D) = \mathbb{E} \sum_{u \in X} k(m) \mathbb{1}_D(u).$$

For $k(m) = m$, $\alpha_k(D)$ is the expected number of employees in D whereas $\Lambda(D)$ was the expected number of firms in D. If $\alpha_k(D) = \int_D \lambda_k(u) du$ then λ_k is the weighted intensity function for weighting function k.
For order, let k and q be two weighting functions, and for a multiplicative scheme $f(m_1, m_2) = k(m_1)q(m_2)$ we introduce the weighted measure $\beta_f^{(2)}$, corresponding to the unweighted $\alpha^{(2)}$,

$$\beta_f^{(2)}(A \times B) = \mathbb{E} \left[\sum_{u,v \in X : u \neq v} \frac{f(m_1, m_2)}{\lambda_k(u)\lambda_q(v)} \mathbb{1}_A(u) \mathbb{1}_B(v) \right]$$

with $\lambda_k(x) > 0$ and $\lambda_q(x) > 0$ ps for all $x \in A$. If $\beta_f^{(2)}(A \times B) = \int_A \int_B g_f(u,v)dudv$ then g_f is the weighted pair correlation function for weighting function f.
Note that these characteristics are introduced and studied in the homogeneous case by Schlater (2001) and Illian et al. (2008).

5.2 The Bonneu-Thomas-Agnan index: non cumulative version

For all $r > 0$, we introduce the non-cumulative Bonneu-Thomas-Agnan index by

$$i_{BT}(r) = \hat{g}_f(r) = \frac{1}{2\pi r} \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \frac{h^{-1} w\left(\frac{r - \|x_i - x_j\|}{h}\right) k(m_i) q(m_j)}{|A \cap (A - x_i + x_j)| \hat{\lambda}_k(x_i) \hat{\lambda}_q(x_j)}$$

with

$$\hat{\lambda}_k(x) = \hat{\lambda}(x) \mathbb{E}[k(\hat{M})|X]$$

Our index is an estimator of the theoretical g_f characteristic. It is important to note that this index can be calculated under the assumption of homogeneity of the intensity of positions as well as under the assumption of inhomogeneity using one of the two estimators of the intensity and this leads to two versions of our index called BThom and BTinhom thereafter.
For the estimations, for a given sector, we estimate :
1) The intensity of positions λ is estimated locally by a non parametric kernel method or by an non parametric iterative and adaptative method based on Voronoï cells.
2) The expectation of the mark conditionally on the position is estimated by a non-parametric kernel method or by an non parametric iterative and adaptative method based on Voronoï cells.
The strategy we propose is to use for null hypotheses a Poisson point process for positions with marks depending only on their own position. It is the only one under which it is easy so simulate realizations. For the simulations under H_0, we first generate a realization of a Poisson PP with the same intensity as in the estimation step. Then for each point of the realization, we estimate the conditional cumulative distribution function of the mark conditionally on the position by a non-parametric kernel method. We then simulate a mark realization from this empirical cumulative distribution function.

5.3 The Bonneu-Thomas-Agnan index: cumulative version

For a given multiplicative weighting scheme, a corresponding cumulative version of this index is given by the following estimator of the weighted K-function

$$I_{BT}(r) = \hat{K}_f(r) = \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \frac{k(m_i) q(m_j) \mathbb{1}(\|x_i - x_j\| \leq r)}{|A \cap (A - x_i + x_j)| \hat{\lambda}_k(x_i) \hat{\lambda}_q(x_j)} \quad \text{pour tout } r > 0.$$

5.4 Consequences for the Duranton-Overman index

We establish a link between the Duranton-Overman index and a theoretical characteristic g_f (the weighted pair correlation function) for the following choice of weighting scheme $k(m) = m$ and $q(m) = m$

$$i_{DO}(r) = \frac{2\pi r}{|A|} \hat{g}_f(r).$$

Hence we derive a natural normalization of this index with a clear benchmark: under H_0 we have $g_f \equiv 1$
We can also propose a cumulative version of this index

$$I_{DO}(r) = \frac{\sum \sum_{j\neq i} m_i m_j \mathbb{1}(\|x_i - x_j\| \leq r)}{\sum \sum_{j\neq i} m_i m_j} = \frac{\hat{K}_f(r)}{|A|}$$

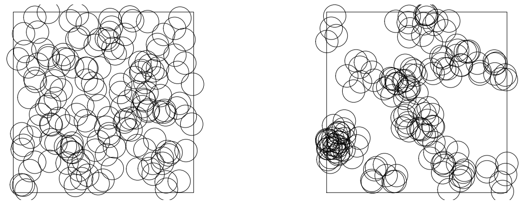

Figure 8: Scenario 1: one realization

$$\hat{K}_f(r) = |A| \frac{\sum\sum_{j\neq i} m_i m_j \mathbb{1}(\|x_i - x_j\| \leq r)}{\sum\sum_{j\neq i} m_i m_j}$$

5.5 Consequences for the Marcon-Puech index

Comparing

$$J_{MP}(r) = \sum_{i=1}^{N_s} \frac{\sum_{j=1,j\neq i}^{N_s} m_j \mathbb{1}(\|x_{i,s} - x_{j,s}\| \leq r)}{\sum_{j=1,j\neq i}^{N} m_j \mathbb{1}(\|x_{i,s} - x_j\| \leq r)}$$

and

$$I_{BT}(r) = \hat{K}_f(r) = \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} \frac{k(m_i) q(m_j) \mathbb{1}(\|x_i - x_j\| \leq r)}{|A \cap (A - x_i + x_j)| \hat{\lambda}_k(x_i) \hat{\lambda}_q(x_j)}.$$

for $k(m) = m$ and $q(m) = 1$, we understand that the correction for inhomogeneity of the location intensity of sector s is missing in the MP index.

5.6 Testing strategy

6 Simulation results

The framework of our simulations is the following. We simulate two sectors, non necessarily of the same type (random, aggregated, regular).

We compare

- the normalized DO index (non cumulative version)
- the cumulative MP index
- the indices BThom and BTinhom (non cumulative versions)

They all have a benchmark of 1 under H_0.

6.1 Results for scenario 1

Scenario 1 is composed of two sectors :the first one is homogeneous Poisson with constant marks and the second is an aggregated process with constant marks. Figure 6.1 presents one realization of these two point processes.
On Figures 6.1 and 6.1, we observe that the indices DO and MP detect a concentration of sector 2. The indices BThom and BTinhom correctly detect that the origin of concentration of sector 2 comes from the second order.

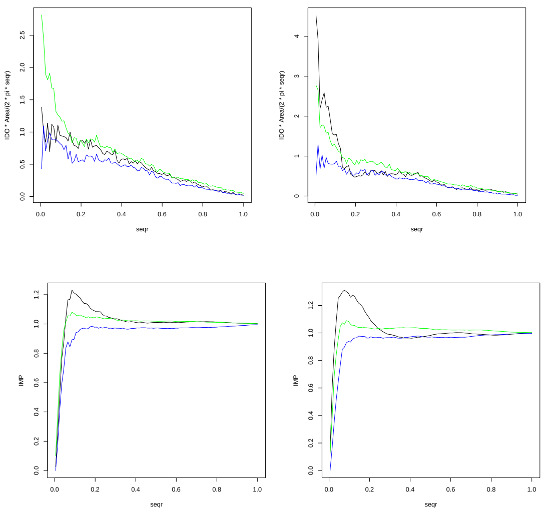

Figure 9: Scenario 1: results for DO (first row) and MP (second row)

6.2 Results for scenario 2

Scenario 2 is composed of two sectors :the first one is homogeneous Poisson with random marks independent from positions.The second one is also homogeneous Poisson with marks depending upon the positions. Figure 6.2 presents one realization of these two point processes.

On Figure 6.2, we see that the index DO detects concentration for sector 2 and MP does not detect anything.

On Figure 6.2, we see that the index BThom detects concentration and BTinhom does not, hence the origin of the concentration of sector 2 comes from the first order.

6.3 Results for scenario 3

Scenario 3 is composed of two sectors :the first one is homogeneous Poisson with constant marks and the second one is not a Poisson process, it is described in Baddeley, Moller and Waagepetersen (2000) and is such that $g_f = 1$. Figure 6.3 presents one realization of these two point processes.

On Figure 6.3, we see that the indices DO and MP do not detect any concentration for sector 2.

On Figure 6.3, we see that the indices BThom and BTinhom do not detect any concentration for sector 2 and this is of course a limitation of this approach. However if a process satisfies $g_f = 1$ for a particular weighing scheme, it is unlikely that it satisfies the same for another choice of weighing scheme so this could be a way out by testing for two different choices.

7 Conclusion

The BT index satisfies the ten objectives DO1 to DO5 and BT1 to BT5. It is clear that an extension can be easily derived to study co-localization. As the MP and the DO index, the BT index depends upon the distance r: this can be viewed as an advantage because it is a richer tool or as a disadvantage because it is more complex to handle. Some more work should be done to study the influence of the choice of weighting scheme.

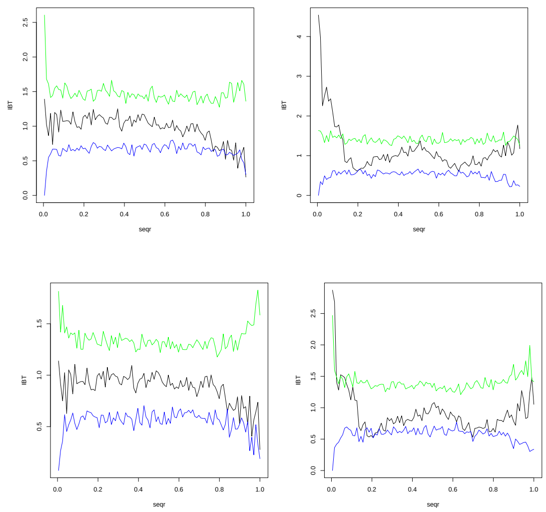

Figure 10: Scenario 1: results for BThom (first row) and BTinhom (second row)

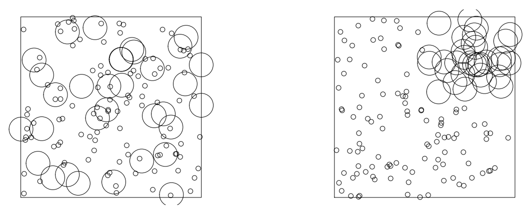

Figure 11: Scenario 2: one realization

References

[1] Duranton, G. and Overman, H.G. (2005) Testing for localization using micro-geographic data. *Review of Economic Studies* **72** 1077-1106.

[2] Illian, J., Illian P, Stoyan H. and Stoyan D. (2008) Statistical analysis and modelling of spatial point patterns, Wiley, Statistics in practice.

[3] Marcon, E. and Puech, F. (2002) A new method to evaluate spatial economic activity and its application to two french areas, preprint.

[4] Marcon, E. and Puech, F. (2010) Measures of the geographic concentration of industries : improving distance-based methods. *Journal of Economic Geography* **10(5)** 745-762.

[5] Moller, J. et Waagepetersen, R.P. (2004) *Statistical inference and simulation for spatial point processes.* vol. 100. Chapman & HallCRC.

[6] Schlather, M. (2001) On the second-order characteristics of marked point processes. *Bernoulli* **7**(1) 99-117.

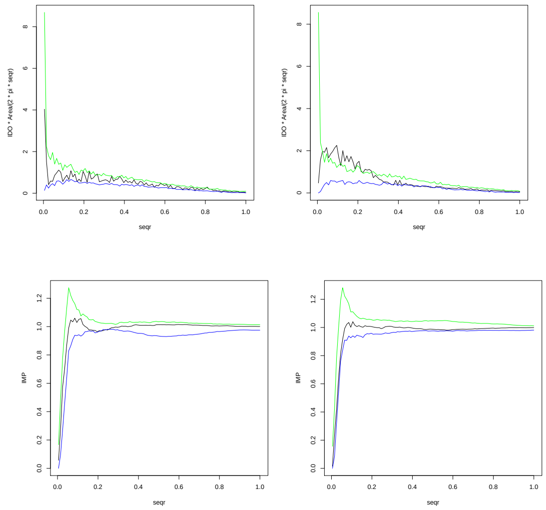

Figure 12: Scenario 2: results for DO (first row) and MP (second row)

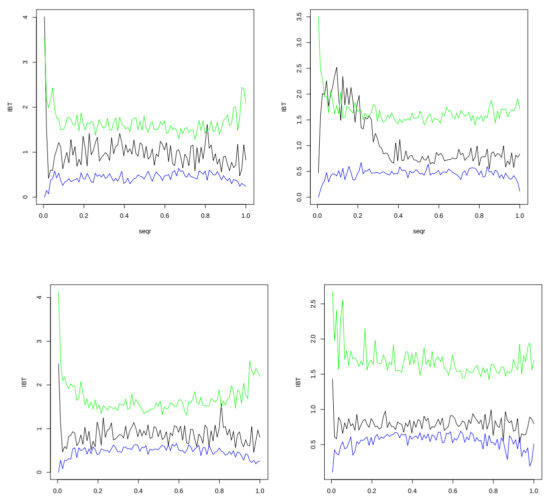

Figure 13: Scenario 2: results for BThom (first row) and BTinhom (second row)

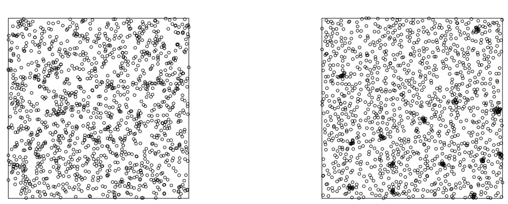

Figure 14: Scenario 3: one realization

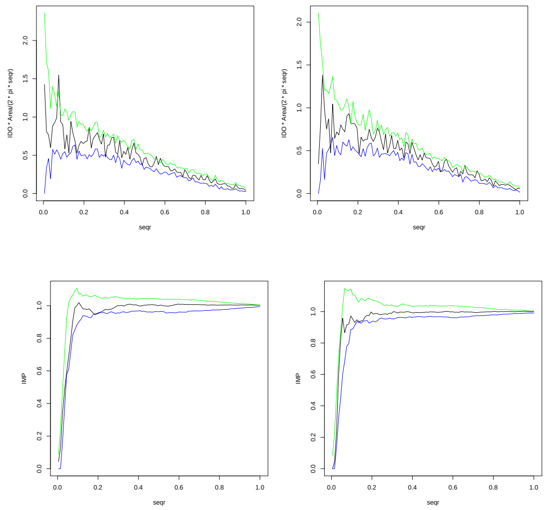

Figure 15: Scenario 3: results for DO (first row) and MP (second row)

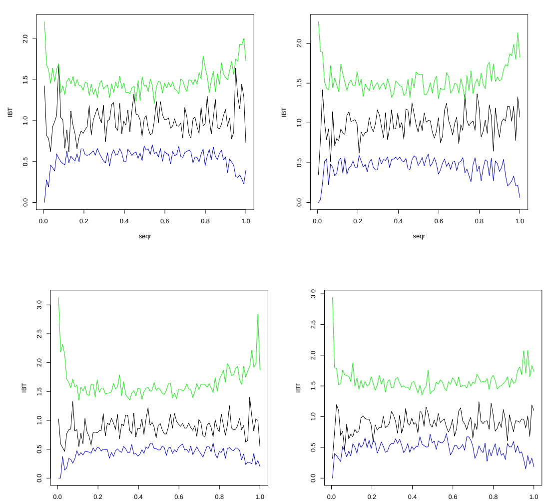

Figure 16: Scenario 3: results for BThom (first row) and BTinhom (second row)