A Method to Find the Robust Center for a Set of Demand Points, based on L_p-norms and Bootstrap

Didier Josselin[1,2] , *Philippe Michelon*[1] , *Serigne Gueye*[1], *Cyrille Genre-Grandpierre*[2]

[1] Laboratoire d'Informatique d'Avignon, Computer Science Department, Université d'Avignon, France

[2] UMR ESPACE 73000, CNRS, Department of Geography, Université d'Avignon, France

Abstract

We propose a method to find a new center, based on robustness assessment (bootstrap on data samples to infer variance). The facility optimal location problem is recalled. The objective function that includes the variance is explained. A robust center minimizes the variance among a large set of L_p-norms. Some validating results are presented in 1 dimension and an application illustrates the process in 2 dimensions. The robust center shows a location which resists to random and moderate changes of the demand. It is somehow sustainable in space, because it adapts to the shape of the (spatial) distribution of demand points.

Keywords: robust center, sustainability, 1-facility optimal location problem, L_p-norms, simulations, bootstrap.

1. Searching for an Accurate Center

Accuracy has several meanings. For instance, a first property of an accurate assessment is the *correctness*, the freedom for mistake or error. A second property is the *exactness:* the clue is exactly conforming to truth or to a standard. From these two definitions and in practice, finding the *degree of conformity* of a measure to a standard or a true value is also of high interest.

Some specific methods can be developed to assess spatial accuracy, *i.e.* in our case to locate an *accurate center* in space. This objective leads to a key question: what is the primary *truth* or *standard model* to be considered? Since a model is somehow a view of the world related to a (group of) subject(s), it can be defined as an objective to reach. That is basically the approach of the Operation Research, which consists in writing the objective function and in drawing the frame of constraints in which the solution should be found. In our study, a center may be a (collection of) optimal location(s). In optimization problems, an exact and unique solution can sometimes be fixed, or not. In those cases, it may require heuristic or stochastic processes to tend to a correct solution. Thence, statistics and resampling can be useful.

In this paper, we propose a method to compute the most robust center in space. Robustness is here set as the capacity of a center location to remain *stable in space and time* while keeping to serve the fixed objective on the long time. In spatial planning, this infers a property of (spatial) sustainability. According to a set of n demand points x_i (with their coordinates) and for a given L_p-norm (or Minkowski distance, called here D_{Lp}) minimization (cf. Equation (1)), it is indeed possible to approximate an optimal center (c^*) responding to this type of accuracy, where the *model* is the robust location (the center c) that serves the demands and the *degree of* *conformity* is computed by a spatial bootstrap. The value of the L_p-norm which aggregates the distances between the demand points and the center defines the objective of the metric chosen.

$$c^* = \arg\min_c \left(D_{L_p}\right) = \arg\min_c \left(\sum_{i=1}^{n} (x_i - c)^p\right)^{\frac{1}{p}} \quad (1)$$

2. Optimal Centers: Equity, Equality or Efficacy?

Most of the approaches to locate an optimal center are based on a cost assessment on a continuous space or along a road network (Hakimi, 1964, Nickel & Puerto, 2005). Since building a health center (figure 1), a waste collection center (figure 2) or a transport station (figure 3) has a non negligible cost, urban planners expect those centers to serve for a long time in the best way. For instance, these three types of centers may correspond to several objective functions and associated L_p-norms (*resp.* equity with $p=1$, equality with $p=2$ or efficacy with $p\rightarrow\infty$). But we do not know anything about their behavior regarding spatial situation sustainability.

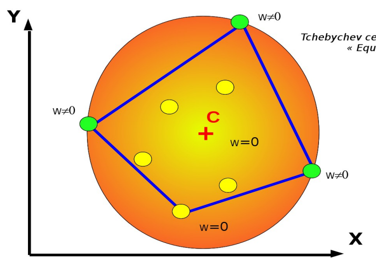

Figure 1: The *1*-center ($p\rightarrow\infty$) provides a quasi-boolean behaviour of the demand point influences: only the points on the Tchebychev circle have an influence (*equity*). This approach may be used to fix a health center.

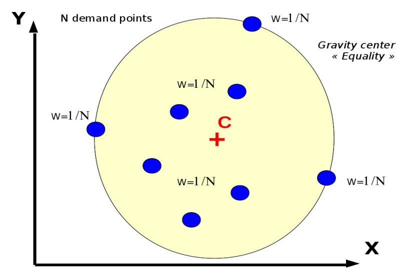

Figure 2: The gravity center ($p=2$) induces an equal value for all the point influences anywhere on space (*equality*). This approach may be used to fix a waste collection center.

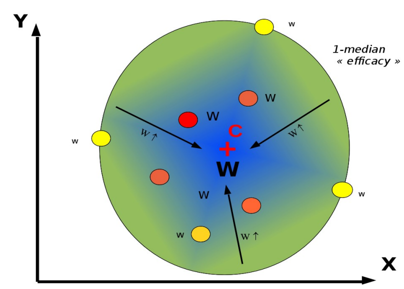

Figure 3: The *1*-median ($p=1$) favours the demand points close to the center (*efficacy* or *efficiency*). This approach may be used to fix a logistic transport station.

3. Toward a Robust Center

Thence, an interesting property of a center is its sustainability, *i.e.* its capacity to respond to the initial objective for a long term, despite many unexpected changes of the demand location in the future. This property is linked to statistical robustness (Hoaglin et al., 1984, Huber, 1981) and center sensitivity (Josselin & Ladiray, 2002, Josselin & Ciligot-Travain, 2013).

3.1. Using Bootstrap and Resampling

The objective function we propose aims at minimizing the center sensitivity, given a metric to compute the cost to access to the center. This sensitivity can be assessed using Monte-Carlo or bootstrap simulations (Thomas, 2002). The simulations can experiment random local change of the population, possibly in relation with the territorial planning which plans (or not) specific urban development. For both approaches, the objective is to minimize the variance *Var(c*)* of the center location (See Equation (2)), through all possible L_p-norms, according to stochastic change of demand location. A set of *m* bootstrapped spatial distributions (c_j*) of points is computed using a drawing with replacement. For each of them, the deviation from the initial center c_{init} is measured. Lower this variance, better the location of the center in terms of spatial sustainability. This means that, despite a random urban evolution based on the initial demand point structure, the robust center location (*r**) remains still relevant over time.

$$r^* = \arg\min_c \left(Var\left(c^*\right)\right) = \arg\min_c \left(\frac{1}{m}\sum_{j=1}^{m}\left(c^*_{init} - c^*_j\right)^2\right) \quad (2)$$

This approach can be useful in at least two cases:

- according to a fixed objective (equity, equality, efficiency in the access to the center), it becomes possible to measure how the center location responds to the initial objective in spite of possible or predictable urban changes;

- over all the possible metrics (L_p-norms) to be used for computing the distances from the demands to the center, it allows to find the most robust center over time, that is to say the center whose optimal location is the most stable for all the simulations of demand changes. For this second case, the solution we look for is an optimal couple Center/L_p-norm *i.e.* *{c*;p}* which minimizes *Var(c*)* (figure 4) and which depends on the spatial distribution of the demand.

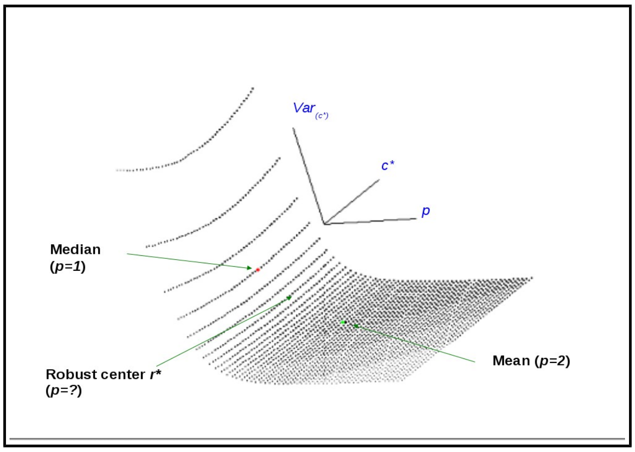

Figure 4: Among all the possible L_p-norms, the robust center *r** minimizes the variance *Var(c*)* for the set of centers c* of the bootstrapped spatial distributions

3.2. Results in 1 Dimension

To comprehend the relationship between the spatial configuration and the metric to maximize the robustness, we propose an algorithm based on a gradient descent which converges to the most robust center adapted to the probable location change of demand. We first show the algorithm efficiency in 1 dimension for finding the Median (minisum operator, $p=1$), the Mean ($p=2$) and the Mean of the extreme values (minimax operator, $p\rightarrow\infty$) of the distribution. Some results illustrate how the metric depends on the spatial distribution of the demand, when considering the sustainability (*i.e.* robustness) of the center. The figure depicts six different distributions. The table 1 shows the estimation of the central value of three of them (*H1*, *H2* and *H3*), once the value of p is fixed. The estimations are very close to the reference center values. The table 2 shows the couple $\{p;r^*\}$ that minimizes the variance (for *H2*, *H4*, *H5* and *H6)*. The most robust centers correspond to different L_p-norms according to the shape of the distribution (here in one dimension).

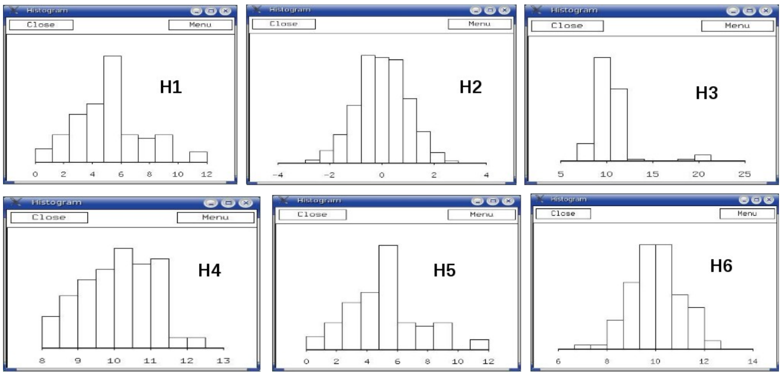

Figure 5: Six different distributions of data in one dimension

Table 1: Comparison between the center of a distribution (in one dimension) and its robust estimation using bootstrap (m=500), given a value of p and a distribution H. They are very close, proving the algorithm efficiency.

Distribution	p_{fixed}	c^*_{init}	r^*_{robust}
H1	1	5	5
H1	2	5.14	5.14
H1	∞	6	6
H2	1	-0.16	-1.21
H2	2	0.22	0.23
H2	∞	0.1	0.09
H3	1	10.21	10.19
H3	2	13.1	12.8
H3	∞	15.3	16

Table 2: Couples $\{p;r^*\}$ which minimize the variance $Var(c^*)$ are processed. The value fund for p differs according to the shape of the distribution.

Distribution	p_{found}	r^*_{robust}	*Comment*
H2	1.97	-0.3	Close to Gaussian distribution and mean
H4	3.4	10.9	Tends to an infinite norm (minimax)
H5	1	5	Fits a median
H6	2.95	9.97	Between a minimax and a mean

3.3. Application on the Geographical Space

The approach can be extended to 2D spaces to locate different centers associated to particular spatial distributions.When the spatial distribution is Gaussian, the most robust center is the *gravity center* with $p=2$ (that minimizes the sum of squared distances). When the spatial distribution is asymmetric with a strong core and a few outliers, the most robust center is likely the *1-median* ($p=1$, that minimizes the sum of distances). When we deal with two separate groups of demand points, generally the most robust center becomes the *1-center* ($p \rightarrow \infty$, that minimizes the maximum distance to the center). The figure 6 provides a set of optimal centers, including the robust center in a continuous space in two dimensions.

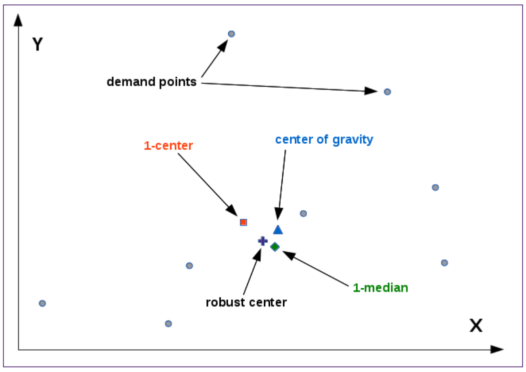

Figure 6: An example of robust location: close but different from the other centers

Conclusion

Once the method reliability is validated by experiments, we generalize the process to any distribution whose spatial law is unknown and show how our adaptive method can be used to find the 'robust' center. For any spatial distribution and according to a certain way to introduce noise in the demand location (random, constrained by urban planning), there exists a 'robust' center which corresponds to a specific value p of the L_p-norm. This center may be somehow considered as a 'sustainability' location, since it keeps its properties for serving the population over time. Complementary experiments must be done to assess the robust center properties and to verify which statistical distribution it fits. More mathematical proofs are expected in a near future to fix the robust center formalism.

References

Hakimi S (1964), Optimum Locations of Switching Center and the absolute Center and Medians of a Graph. *Operations Research* 12 pp. 450-459.

Hoaglin D, Mosteller F, Tukey JW. (1983), *Understanding Robust and Exploratory Data Analysis*, Wiley Series in probability and mathematical statistics, New-York.

Huber P. (1981), *Robust Statistics*, Wiley, New York.

Josselin D. & Ciligot-Travain M. (2013), Revisiting the center optimal location based on robustness, sensitivity and influence analysis, *Environment & Planning B*, Vol 40 (5), pp. 923-941.

Josselin D. & Ladiray D. (2002), Combining L1 and L2 norms, for a more robust spatial analysis : the “meadian attitude”. *Cybergeo: European Journal of Geography*, 222, 2002, http://cybergeo.revues.org/3458.

Nickel S & Puerto J. (2005), *Location theory*. A unified approach, Springer.

Thomas I. (2002), *Transportation Networks and the Optimal Location of Human Activities, a numerical geography approach*, Transports economics, management and policy Edward Elgar Northampton Massachusetts.