Used vs. Offered Densities of Human Settlement in Space: When the Statistical Population Matters

Fabien Leurent

Fabien Leurent, ENPC-CIRED

Used vs. Offered Densities of Human Settlement in Space: When the Statistical Population Matters

Abstract

To people living in areas, the denser is the area, the more numerous are the opportunities of inter-personal and social interaction, of employment and of amenities of all kinds. So far, the spatial density of human settlement has been studied according to places. The article's aim is to put density in the perspective of the people that experience it, by shifting from the statistical population of places to that of people. The user-centric, "Used density" is related to the place-based, "Offered density" by a consumption model – a specific instance of a well-established probabilistic model. The article provides the probabilistic framework to derive the statistical distribution of used density from that of offered density. The average used density is systematically larger than its offered counterpart: the ratio amounts to one plus the squared relative dispersion of offered density. The relation between the two statistical distributions can also be represented using a Lorenz curve and the associated Gini index. A case of France's population as of 2019 is studied to demonstrate the methodology.

Keywords

Spatial heterogeneity; Density metrics; Unit places; Consumption model; Lorenz curve; Gini index.

1/ Introduction

Background. The spatial density of human settlements stands out as a prominent feature of territories. Urban areas are endowed with high density of population and jobs: typical urban residential densities range from 1,000 inhabitants per square km to more than 100,000 in the densest parts of some Asian megacities e.g. Dhaka. Rural areas, on the other hand, exhibit sparse human settlements and low densities: typical values of rural density range from near zero to some hundreds inhabitants per square km. Thus, density is a relevant indicator of urbanization, together with the overall area population which is the primary indicator of city size in human geography. Spatial maps exhibiting zones colored according to their respective levels of density make a basic tool to understand the spatial structure of territories (e.g. Dijkstra & Poelman, 2014).

Problem setting. Despite its obvious relevance to depict the spatial conditions of human life in territories, the traditional density indicator pertains primarily to spatial units. The average density of a heterogeneous territory is calculated over the statistical population of spatial units. To recall the underlying statistical population, we shall refer to it as the Settled Density – or the Offered Density if we think of a spatial unit as a server of settlement. As for people that live in the territory, their respective spatial conditions would be represented better by considering the human population as the statistical population of primary interest, above that of spatial units. Let us call "Lived Density" the human-centric notion of spatial density – or Used Density if we think of people as users of settlement services. Having clearly distinguished the two statistical populations, it becomes obvious that the associated notions of average density are likely to differ.

Objective. This article provides a statistical model to analyze the Used Density and relate it to the Offered Density. We shall state the model using basic probabilistic concepts of mean value (or mathematical expectation), variance and probability distribution function (PDF, yet another notion of density). The relationship between UD and OD is analogous to those arising in various fields: for instance in traffic theory between the temporal and spatial distributions of vehicular speeds on a roadway (Wardop), or in queuing theory between inter-service times as opposed to waiting times experienced by the users. It is known in econometrics as a "consumption model": a well-known instance is the model underpinning the Gini index that measures the inequality in an income distribution.

Method. We shall apply the classical framework of consumption models to the issue of spatial density, Used versus Offered. The names are given in an economic perspective where spatial units offer settlement as a service to the people that use them.

Article structure. The rest of the article is organized in three parts: after stating the methodology, we will apply it to a case study of communal density in France as of 2019, before providing a discussion and conclusion.

2/ Methodology

2.1/ Territory, zoning system and local population

To analyze human spatial density in territorial system, let us model the territorial space as a set Z of zones z. Each zone z has its own ground area, A_z, and human population, P_z. Then its own density is simply

$$x_z := \mathrm{P}_z/\mathrm{A}_z, \quad \forall z \in \mathrm{Z} \tag{1}$$

Overall, the territory has total ground area of $\mathrm{A_Z} := \sum_{z\in \mathrm{Z}} \mathrm{A}_z$ and total population of $\mathrm{P_Z} := \sum_{z\in \mathrm{Z}} \mathrm{P}_z$. Its density averaged over space is

$$\bar{x}^{\mathrm{o}}_{\mathrm{Z}} := \mathrm{P_Z}/\mathrm{A_Z} \tag{2}$$

The "o" superscript reminds that the underlying statistical population is that of spatial units offering the settlement service.

2.2/ Spatial units as a statistical population

While zones are also often called spatial units, here we shall rather refer to them as "spatial entities" and keep the name of "spatial units" (or "land units") for elementary places of unit ground area, say a^{o}. Such land units are more convenient than zones to constitute the statistical population of places since, being identical in area, it is easier to compare them in other respects such as the human population.

The assignment of spatial units o (i.e. unit places) to any zone z is an idealization: thinking of the unit ground area as 1 square km or 1 hectare, we expect most zones to involve a non-integer number of land units. Let us nevertheless denote as "$o \in z$" the composition of zone z out of land units o. For every such land unit, denote as $\mathrm{p_o}$ its population: the associated human density is therefore

$$x_o := \mathrm{p_o}/\mathrm{a^o}, \quad \forall o \in O \tag{3}$$

Notionally, it holds that

$$\mathrm{A}_z := \sum_{o\in z} \mathrm{a^o}, \quad \forall z \in Z \tag{4a}$$

$$\mathrm{P}_z := \sum_{o\in z} \mathrm{p_o}, \quad \forall z \in Z \tag{4b}$$

We are now ready to study human population over the statistical population of land units in a standard probabilistic fashion. We shall consider density x as a Random Variable in this statistical population, with PDF $\mathrm{f_o}$ and CDF $\mathrm{F_o}$. From this stems the average human density over space:

$$\bar{x}_\mathrm{o} := \int x.\mathrm{f_o}(x)\, dx. \tag{5}$$

Denoting as $\overline{\mathrm{O}}$ the total number of land units, the discretized version is notionally equivalent to the continuous one:

$$\bar{x}_\mathrm{o} := \frac{\sum_{o\in \mathrm{O}} x_o}{\overline{\mathrm{O}}}$$

As $\overline{\mathrm{O}}.\mathrm{a^o} = \mathrm{A}_Z$, replacing x_o with $p_o/\mathrm{a^o}$ and $\mathrm{a^o}\overline{\mathrm{O}}$ with A_Z owing to (4b) aggregated over Z, it comes out that

$$\bar{x}_\mathrm{o} = \frac{\mathrm{P_Z}}{\mathrm{A_Z}}. \tag{6}$$

Thus $\bar{x}_\mathrm{o} = \bar{x}_\mathrm{Z}^\mathrm{o}$, as could be expected.

Higher order moments of density in the statistical population of land units are defined as follows. At order r,

$$\mathrm{E_o}[x^r] := \int x^r.\mathrm{f_o}(x)\, dx. \tag{7}$$

Assuming that density is homogenous among the land units composing any zone, then $\mathrm{E_o}[x^r] = \mathrm{E_Z}[x^r]$ wherein

$$\mathrm{E_Z}[x^r] := \sum_{z\in \mathrm{Z}} \frac{\mathrm{A}_z}{\mathrm{A_Z}} (\frac{\mathrm{P}_z}{\mathrm{A}_z})^r. \tag{8}$$

But in fact local density is likely to be heterogeneous among land units, even at the zone level. The intra-zone variance of human density is a metric for that heterogeneity. It is defined as

$$\mathrm{V_o^{(z)}}[x] := \mathrm{E_o}[x^2 | o \in z] - (\mathrm{E_o}[x | o \in z])^2$$

And satisfies that

$$\mathrm{V_o^{(z)}}[x] = \left\{ \sum_{o\in z} \frac{\mathrm{a^o}}{\mathrm{A}_z} x_o^2 \right\} - x_z^2. \tag{9}$$

Over the territory, the overall variance of human density can be recovered on the basis of the law of total variance: (its decomposition into intra-class variance and inter-class variance)

$$\mathrm{V}_{\mathrm{o}}^{(\mathrm{Z})}[x] = \left\{ \sum_{z \in \mathrm{Z}} \frac{\mathrm{A}_z}{\mathrm{A}_\mathrm{Z}} \mathrm{V}_{\mathrm{o}}^{(z)}[x] \right\} + \left\{ \sum_{z \in \mathrm{Z}} \frac{\mathrm{A}_z}{\mathrm{A}_\mathrm{Z}} (\bar{x}_z - \bar{x}_\mathrm{o})^2 \right\}. \tag{10}$$

The associated standard deviation and relative dispersion are of course

$$\mathrm{SD}_\mathrm{o}[x] := \sqrt{\mathrm{V}_\mathrm{o}^{(Z)}[x]}, \tag{11a}$$

$$\gamma_\mathrm{o}[x] := \mathrm{SD}_\mathrm{o}[x] \,/\, \bar{x}_\mathrm{o} \;\; . \tag{11b}$$

2.3/ Human density as lived by the statistical population of inhabitants

Any individual u inhabits a zone z_u and inside it a spatial unit $\mathrm{o}(u)$. Let us define the "individual density" or "user-centric density" as the density in the spatial unit inhabited by the user:

$$x_u := x_{\mathrm{o}(u)}, \tag{12}$$

Among the population of users, of size $\overline{\mathrm{U}} = \mathrm{P}_Z$, the user-centric density is a random variable. Its PDF denoted $\mathrm{f_u}$ is related to the PDF $\mathrm{f_o}$ of local density in the following way:

$$\mathrm{f_u}(x) \propto x.\mathrm{f_o}(x). \tag{13}$$

The reason is that the spatial units of which the density belongs to $[x, x + \delta x[$, in proportion $\mathrm{f_o}(x).\delta x$ in their distribution, do contain $x.\mathrm{a^o}$ users each: hence their total number of users amounts to $\mathrm{a^o}.x.\mathrm{f_o}(x).\delta x.\overline{\mathrm{O}}$. These users are those with user-centric density in $[x, x + \delta x[$ and those users only: thus their number is also $\mathrm{f_u}(x).\delta x.\overline{\mathrm{U}}$. On combining both formulas, as $\overline{\mathrm{O}}.\mathrm{a^o} = \mathrm{A}_Z$ and $\overline{\mathrm{U}} = \mathrm{P}_Z$, we recover that

$$\mathrm{f_u}(x) = \frac{\mathrm{A_Z}}{\mathrm{P_Z}} x.\mathrm{f_o}(x),$$

Which implies (13) with proportionality coefficient $\frac{\mathrm{A_Z}}{\mathrm{P_Z}} = 1/\bar{x}_\mathrm{o}$. To sum up,

$$\mathrm{f_u}(x) = \frac{1}{\bar{x}_\mathrm{o}} x.\mathrm{f_o}(x), \tag{14}$$

From this stems a relation between the moments: at order r,

$$\mathrm{E_u}[x^\mathrm{r}] = \frac{1}{\bar{x}_\mathrm{o}} \mathrm{E_o}[x^{\mathrm{r}+1}]. \tag{15}$$

At order $\mathrm{r} = 1$, the average density as experienced by the users satisfies that

$$\bar{x}_\mathrm{u} = \frac{1}{\bar{x}_\mathrm{o}} \mathrm{E_o}[x^2] = \bar{x}_\mathrm{o}(1 + \gamma_\mathrm{o}^2). \tag{16}$$

If the density is homogenous in the territory, then $\gamma_\mathrm{o} = 0$ and the average densities according to either statistical population are equal. But the larger the heterogeneity as measured by the relative dispersion, the higher the ratio $\bar{x}_\mathrm{u}/\bar{x}_\mathrm{o}$ of used to offered average densities.

Figure 1 illustrates the used versus offered PDF. Its assumptions are the following: that x_{o} is distributed log-normal with parameters $m_{\mathrm{o}} = 3.24$ (mean of $\ln(x_{\mathrm{o}})$) and $s_{\mathrm{o}} = 1.76$ (standard deviation of $\ln(x_{\mathrm{o}})$), hence $\bar{x}_{\mathrm{o}} = 120$ and $\gamma_{\mathrm{o}} = 4.6$. The related x_{u} is log-normal, too, with parameters $m_{\mathrm{u}} = 6.34$ and $s_{\mathrm{u}} = s_{\mathrm{o}}$. In that particular instance, the ratio of average densities, $\bar{x}_{\mathrm{u}}/\bar{x}_{\mathrm{o}}$, amounts to 22 – indeed a very large value.

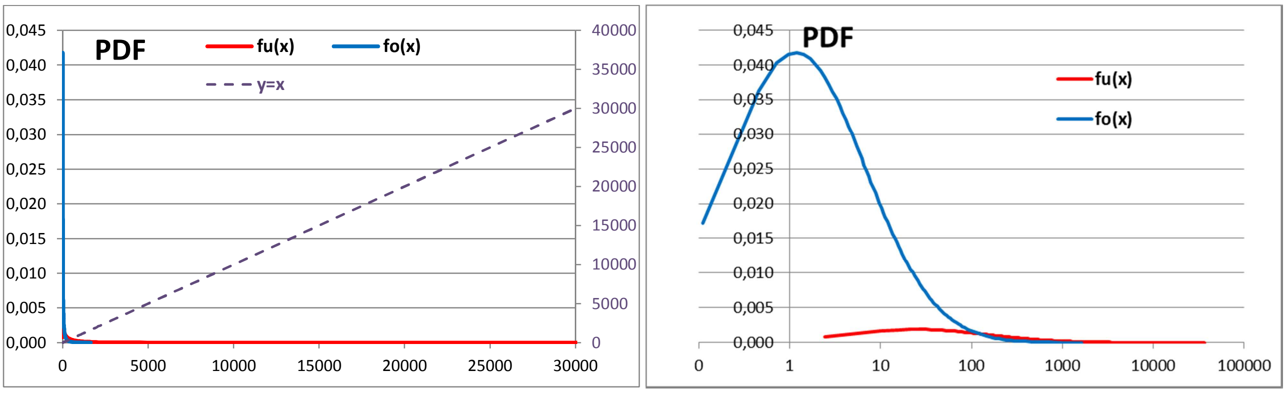

Fig.1. Used & Offered PDFs of Human Density: (a) standard scales, (b) abscissas in log-scale.

2.4/ Lorenz curve and Gini index

In Gini's analysis of income inequalities, the sum of all individual incomes in a group of people is decomposed according to specific sub-groups of people. A typical sub-group gathers the people that earn each less than a given level of income. Then, to the proportion $\mathrm{F_p}(x)$ of people whose income is less than x is associated the proportion $\mathrm{F_I}(x)$ of total income that stems from the aggregation of their individual incomes. The relation linking $\mathrm{F_p}(x)$ and $\mathrm{F_I}(x)$ is stated as the Lorenz curve. The basic illustration consists in a diagram of income proportion vs. people proportion: a proportion α along the horizontal axis induces a quantile $x_{\mathrm{p}}^{(\alpha)} := \mathrm{F_p}^{(-1)}(\alpha)$ of people ranked in increasing order of income: thus $\alpha = \mathrm{F_p}(x_{\mathrm{p}}^{(\alpha)})$. In turn, the specific value $x_{\mathrm{p}}^{(\alpha)}$ induces a share $\mathrm{F_I}(x_{\mathrm{p}}^{(\alpha)})$ of total income. Precisely, the Lorenz function is defined as $\mathrm{L}\colon \alpha \mapsto \mathrm{L}(\alpha) := \mathrm{F_I} \circ \mathrm{F_p}^{(-1)}(\alpha)$.

The derivative $\dot{\mathrm{L}}$ of L satisfies that

$$\dot{\mathrm{L}}(\alpha) = \frac{\mathrm{f_I}\left(x_{\mathrm{p}}^{(\alpha)}\right)}{\mathrm{f_p}\left(x_{\mathrm{p}}^{(\alpha)}\right)} = \frac{1}{\bar{x}_{\mathrm{p}}} x_{\mathrm{p}}^{(\alpha)}$$

It is non-negative and increasing with α since $\mathrm{F_p}^{(-1)}$ is increasing: this makes L an increasing and convex function.

In the diagram, the graph of function L lies below the straight line from point (0,0) to point (1,1). The area between the straight line and the function graph, divided by the area below the straight line i.e. ½, is known as the Gini index. Its value belongs to [0,1]. Between different income distributions, the larger the heterogeneity, the larger the Gini index: it is a metric of inequality.

This line of reasoning applies to our distributions of density: to the $\mathrm{F_o}(x)$ share of space with density less than x corresponds the $\mathrm{F_u}(x)$ share of people each experiencing individual density less than x.

Here the Lorenz function is $\mathrm{L} := \mathrm{F_u} \circ \mathrm{F_o^{(-1)}}$. The resulting Gini index, $G := 2\int_0^1 (\alpha - \mathrm{L}(\alpha))d\alpha$, constitutes another metric of density heterogeneity, along with γ_o and γ_u.

In the Appendix, a log-normal instance is addressed to give insight in the consumption model and illustrate the properties of relative dispersions and the Gini index.

Figure 2 depicts the Lorenz curve of density, assuming the same used and offered distributions of human density as in figure 1.This particular instance exhibits a Gini index of 0.79 – indeed a very high value for that kind of index.

The Lorenz curve depicts the relation between space and people. It relates a proportion of people, on the vertical axis, to the proportion of land that accommodates them, on the horizontal axis. For instance, in Figure 2 it appears that about 20% of people are accommodated in 80% of space. The relation pertains to the spatial density of human settlement: both the spatial units (horizontal axis) and the individuals (vertical axis) are ranked in increasing order of density x. As proportion $\mathrm{F_o}(x)$ of space accommodates proportion $\mathrm{F_u}(x)$ of people, all of them at density lower than x, conversely the residual $1 - \mathrm{F_o}(x)$ share of space accommodates the residual $1 - \mathrm{F_u}(x)$ share of people, all of them at density greater than x. Thus the point $(\mathrm{F_o}(x), \mathrm{F_u}(x))$ splits the diagram in two parts: lower density space and people on the left hand side, higher density space and people on the right hand side. Between the two parts, there is a striking contrast of average density: $\frac{\mathrm{F_u}(x)}{\mathrm{F_o}(x)}$ versus $\frac{1-\mathrm{F_u}(x)}{1-\mathrm{F_o}(x)}$: continuing our instance, $\frac{20\%}{80\%}$ as opposed to $\frac{80\%}{20\%}$ means that the average density in the higher part is about 16 times that in the lower part.

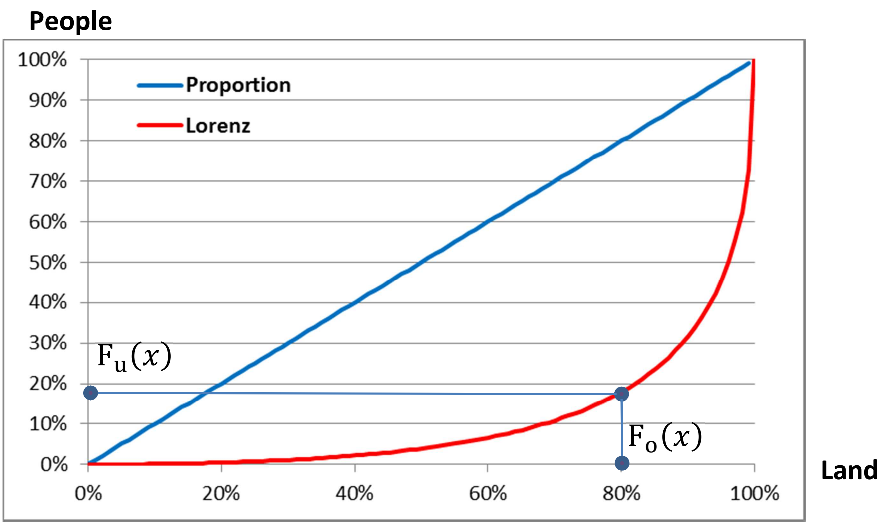

Fig.2. Lorenz function of human density.

3/ A case study of France as of 2019

3.1/ The territory under study

Metropolitan France comprises about 34,750 municipalities called “communes”. We take them as zones in the country. The country area of about 543 thousand square km yields an average commune area of 15.6 km².

As of 2019, the French metropolitan population amounts to about 65 M inhabitants (Insee, 2021b). Thus the average commune population is 1,800 people only and the overall spatial density of population is about 120 persons per km².

Figure 3 exhibits the map of French communes colored according to density level. It shows that most of the country area has low population density.

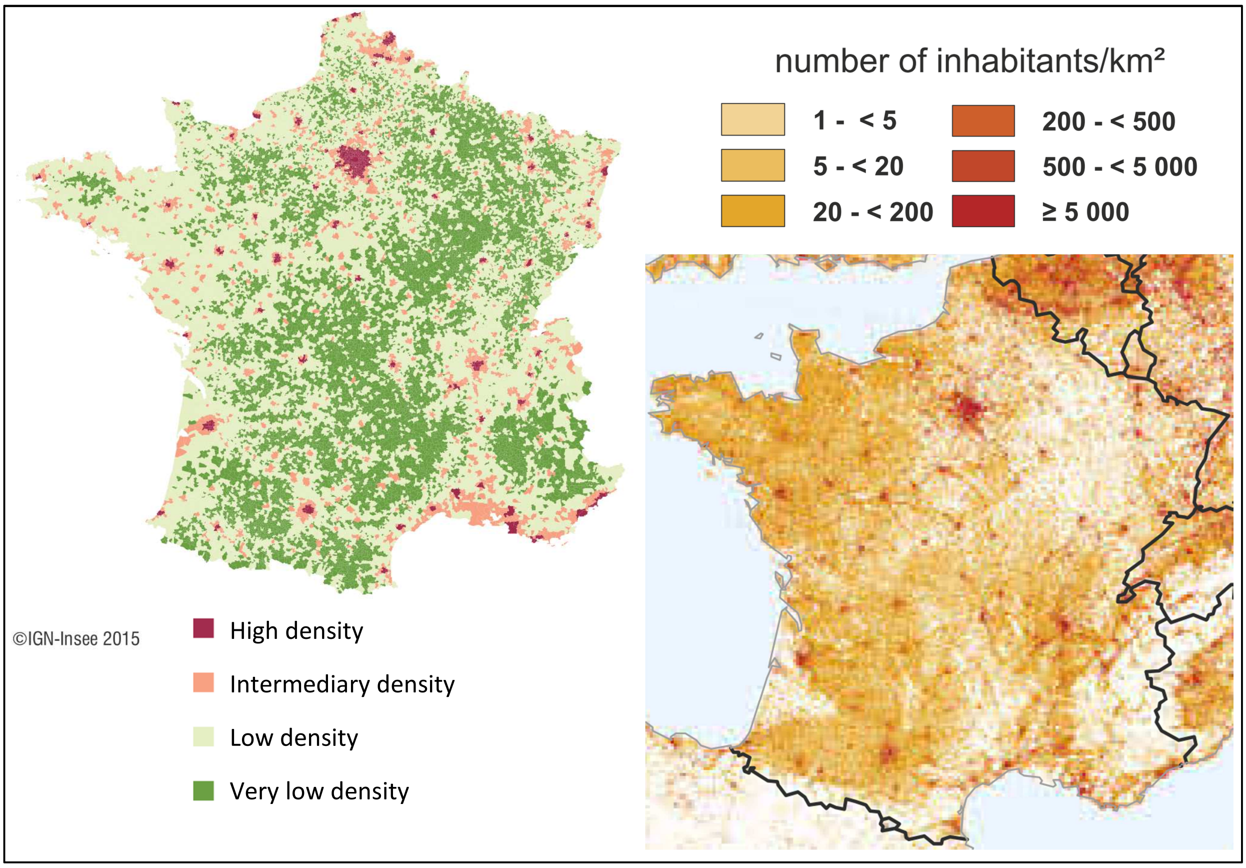

***Fig. 3**. **Human density of French communes as of 2015** (from Aliaga, 2015) **& 2018** (Eurostat, 2019)*

3.2/ Used density vs. Offered density

To obtain the statistical distribution of used density, we made the following assumption: that each commune’s population is distributed evenly in its area. Of course this is only an approximation as large communes (meaning communes of large area) are likely to exhibit significant intra-communal heterogeneity of human settlement.

Based on this assumption, we ranked the communes in order of increasing average density. According to the ranking we calculated two cumulated variables: first the land area, second the population. By dividing the cumulated area up to commune z by the total country area, the F_{O} CDF is obtained at point x_z. Similarly, by dividing the cumulated population up to commune z by the total country area, the $\mathrm{F_u}$ CDF is obtained at point x_z.

The next step is to draw the diagram of $\mathrm{F_u}$ vs. $\mathrm{F_o}$, i.e. the Lorenz curve (figure 5). The Gini index is easy to calculate, by accumulating $2(\mathrm{F_o}(x_z) - \mathrm{F_u}(x_z)).(\mathrm{F_o}(x_z) - \mathrm{F_o}(x_{z-1}))$ over communes z. The outcome is 0.76, again a very large value.

Also easy to calculate are the mean value, variance, standard deviation and relative dispersion of the density variable either offered or used. The results are, in persons per square kilometer:

- for x_o: $\mathrm{E_o}[x] = 120$ and $\mathrm{SD_o}[x] = 548$, yielding $\gamma_o = 4.58$ (dimensionless).
- for x_u: $\mathrm{E_u}[x] = 2{,}628$ and $\mathrm{SD_u}[x] = 4{,}691$, yielding γ_u =1.79 (dimensionless).

Figure 6 depicts the empirical CDFs $\mathrm{F_o}$ and $\mathrm{F_u}$, along with log-normal approximations that mimic the mean and variance of each distribution. To obtain PDFs (figure 7), we discretized the CDFs and derived the respective PDFs as the average value between two successive points.

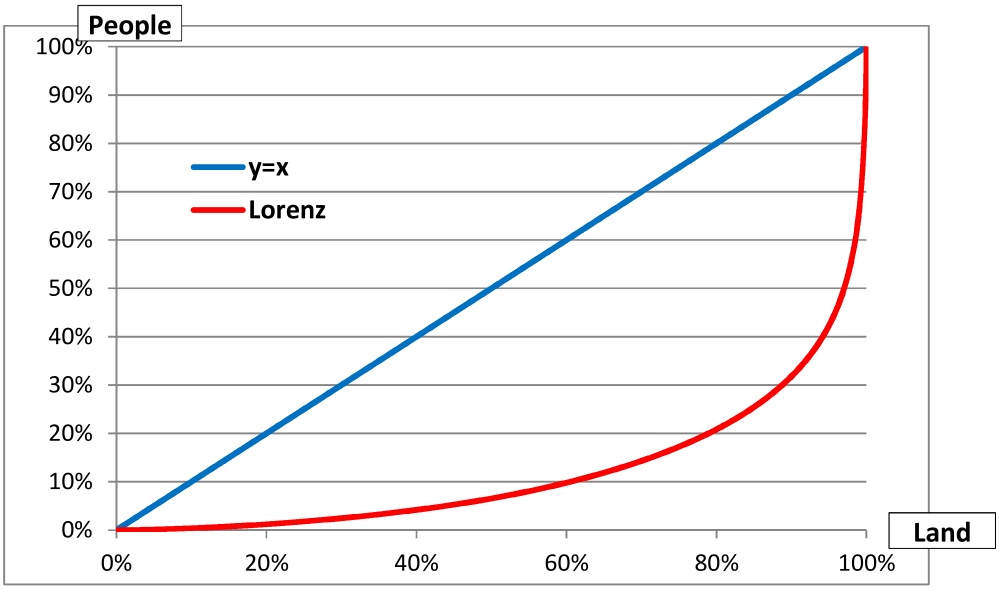

Fig.5. The Lorenz curve of human density in France, 2019.

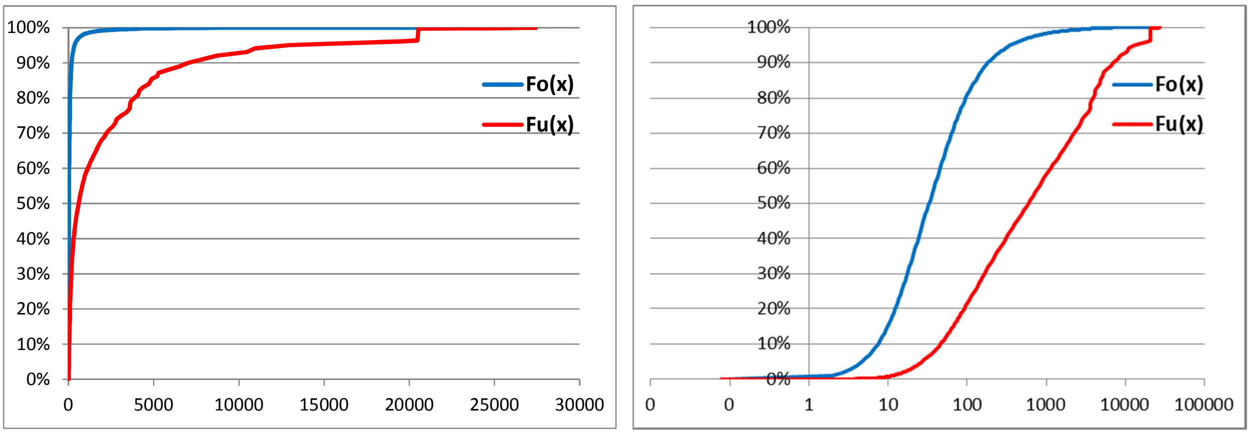

Fig.6. Used & Offered CDFs of Human Density: (a) standard scales, (b) abscissas in log-scale.

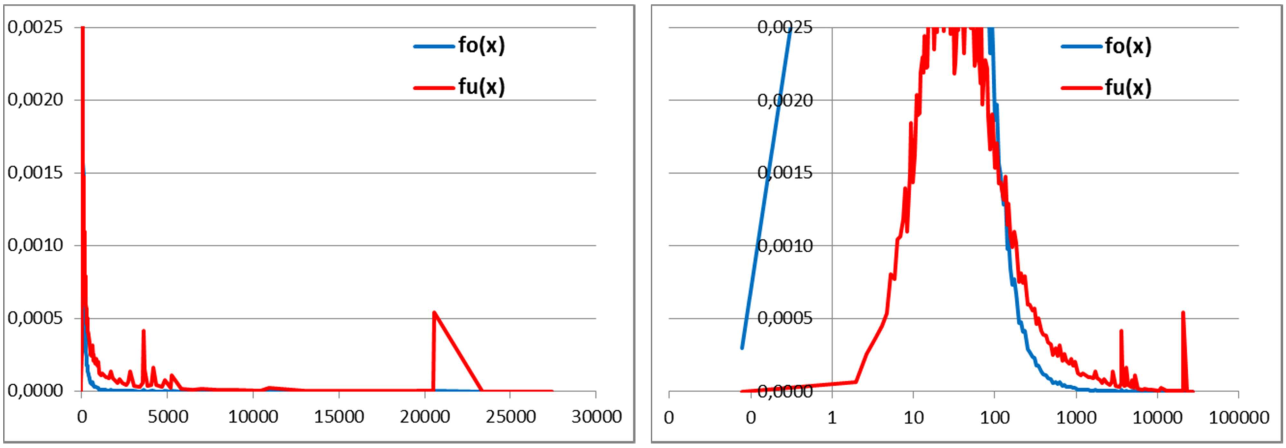

Fig.7. Used & Offered PDFs of Human Density: (a) standard scales, (b) abscissas in log-scale.

3.3/ Comments

From the two CDFs, we recovered the following table of deciles (in inhabitants per km²). The median value of offered space is much lower than the offered mean – its only significant is to state that a major share of France's territory lies under very low density. The 80%-20% shares of low / high density land are associated to 20% - 80% shares of low / high density people. Such contrasts are striking and call for quantitative metrics to complement density maps in territorial analysis.

	10%	20%	30%	40%	50%	60%	70%	80%	90%
xo(a)	7	12	17	24	33	46	64	96	186
xu(a)	46	92	167	306	600	1146	2213	3872	6999

The deciles pave the way to the qualitative assessment of low to high levels of density. With respect to people living in France, the median used density i.e. 600 persons per square kilometer may be taken as "medium level of density", low densities for the bottom 20% i.e. below 92 p/km², high densities for the top 20% i.e. above 4000 p/km². These people-based values are close to the values selected by the Regional and Urban General Directorate of the European Commission (Eurostat, 2019). The land-based deciles have little relevance to depict urban conditions. The average offered density is just a ratio to summarize the intensity of human occupation over a given stretch of land – nothing less, nothing more, especially not about the used density of population.

The average densities are meaningful metrics. The standard deviation of offered density makes little sense to people: and not much more for land, in fact. The Gini index is much more meaningful and so is the relative dispersion.

Concerning people, it is definitely meaningful to them to consider the density of the space in which they live. The notion of x_{u} and its probabilistic features from $\mathrm{f_u}$ to $\bar{x}_{\mathrm{u}}$ and γ_{u} constitute a simple statistical model to analyze human population according to human spatial density. Some hints of that are mentioned in previous studies such as Aliaga et al (2015) for France: these authors mentioned that 90% of French communes contain 35% of French people: up to the difference between spatial entities and land units, the mention is analogous to one point on the Lorenz curve. The full Lorenz curve contains much more information. Beyond the shares of land and people, the magnitude of human density is indicated by the mean value together with the relative dispersion.

4/ Discussion and conclusion

The gist of the article is to put the spatial density of human settlement in the perspective of the people that experience it. The perspective shift from the spatial population of unit places to the statistical population of people, i.e. the human population, constitutes a relativistic theory of density. The probabilistic framework is straightforward since the relationship between people and unit places is typical of a consumption model. Here the original contributions consist in (i) the identification of the relativistic effect, (ii) the transition from zones to unit places so as to constitute an explicit statistical population of places. Indeed, the statistical consideration of territorial spaces is more often implicit than explicit, as zoning systems are especially purported to analyze spatial variables in a discretized way.

The explicit consideration of used density enables for better understanding the spatial occupation. The progress to harvest is the same one as in other instances of consumption models: not only Gini's analysis of income inequality, but also (i) the queuing theory of waiting times (e.g. Kleinrock, 1975), (ii) Wardrop's model of temporal vs. spatial distributions of vehicular speeds on roadways (Wardrop & Charlesworth, 1954), (iii) our own model of transit vehicle loads and transit users' exposure to crowding conditions (Leurent et al. 2012, 2017).

In the probabilistic framework, the ratio of used and offered average density is equal to one plus the squared relative dispersion of the offered density. The Gini index well known to economists is an appropriate metric to assess the inequality of human occupation over space. When the density variable follows a log-normal distribution, then the Gini index and the ratio of averages are equivalent.

The application to France's population and territorial space as of 2020 reveals that the vast majority of places have low human occupation. Conversely, 80% of people live in 20% of space.

As the concept is significant and the application procedure is quite easy, we may expect the statistical analysis of used density to spread out in territorial studies. Its indicators will be useful to summarize the dynamics of urbanization in any territory. It may be adapted to sub-populations, to jobs or categories of jobs, etc.

As for theoretical development, a direction of research is to carry the user-centric perspective in the analysis of (i) accessibility to amenities owing to available transport modes and considering their quality of service, (ii) "effective density" of amenities that stems from accessibility conditions. Both topics lie at the interface of geography and economics.

5/ References

Aliaga C., Eusebio P. and Levy D. (2015) « Une nouvelle approche des espaces à faible et forte densité », in La France et ses territoires, coll. « Insee Références » édition 2015. https://www.insee.fr/fr/statistiques/1372992?sommaire=1373022

Cramer, J.S. (1962) The ownership of major consumer durables. Cambridge University Press.

Dijkstra, Lewis & Poelman, Hugo (2014) "A harmonised definition of cities and rural areas: the new degree of urbanisation". Working papers n° 01/2014, European Commission General Directorate for Regional and Urban Policy, https://ec.europa.eu/regional_policy/sources/docgener/work/2014_01_new_urban.pdf

Eurostat (2019) Methodological Manual on Territorial Typologies, 2018 edition. DOI: 10.2785/930137 https://ec.europa.eu/eurostat/fr/web/products-manuals-and-guidelines/-/KS-GQ-18-008

Gini, C. "Variabilitá e mutabilita." 1912. Reprinted in Memorie di metodologia statistica (Ed. E. Pizetti and T. Salvemini.) Rome: Libreria Eredi Virgilio Veschi, 1955.

Insee (2021a) « Observatoire des territoires" website : file insee_rp_hist_1968 obtained from request https://www.observatoire-des-territoires.gouv.fr/outils/cartographie-interactive/api/v1/functions/GC_API_download.php?type=stat&nivgeo=com2021&dataset=insee_rp_hist_1968&indic=superf_choro. Accessed on Jan. 19, 2022.

Insee (2021b) Dataset of communal populations up to 2019, available at https://www.insee.fr/fr/statistiques/fichier/3698339/base-pop-historiques-1876-2019.xlsx . Accessed on Jan. 19, 2022.

Kleinrock, Leonard (1975) Queuing systems, Vol. I. Wiley Interscience.

Leurent, Fabien, Benezech, V. & Combes, F. (2012) "A Stochastic Model of Passenger Generalized Time Along a Transit Line", Proceedings of EWGT2012. *Procedia - Social and Behavioral Sciences*, 54: 785-797, DOI: 10.1016/j.sbspro.2012.09.795

Leurent, Fabien, Pivano, C. & Poulhès, A. (2017) On passenger traffic along a transit line: a stochastic model of station waiting and in-vehicle crowding under distributed headways. *Transportation Research Procedia*. 27: 1219-1226. http://authors.elsevier.com/sd/article/S2352146517309730

Wardrop, John-Glen & Charlesworth, G. (1954) A method of estimating speed and flow of traffic from a moving vehicle. Proc. Institution of Civil Engineers, 312/3: 158-169. http://dx.doi.org/10.1680/ipeds.1954.11628

6/ Appendix: consumption model with a log-normal distribution

6.1/ Consumption model

Let us first define a consumption model in a generic fashion. It relies upon a consumption function say $c: x \mapsto \mathrm{c}(x)$ that is real valued and monotonous. It measures the amount of consumption made by an individual with attribute x.

Let then $\mathrm{f_o}$ denote the PDF of attribute x in the statistical population of such individuals. The consumed units of all the individuals make up a statistical population of their own, with PDF function $\mathrm{f_u}$ that satisfies the following relation:

$$\mathrm{f_u}(x) \propto c(x).\mathrm{f_o}(x). \tag{A1}$$

Postulating that the consumption function is monotonous, then eqn. (A1) can be demonstrated using the same proof as for (13). The proportionality coefficient is the reciprocal of $\bar{\mathrm{c}}_{\mathrm{o}} := \int c(x)\mathrm{f}_{\mathrm{o}}(x)dx$. Thus

$$\mathrm{f}_{\mathrm{u}}(x) = \frac{1}{\bar{\mathrm{c}}_{\mathrm{o}}} c(x).\mathrm{f}_{\mathrm{o}}(x). \tag{A2}$$

6.2/ The log-normal distribution

The log-normal distribution is especially well-suited to consumption models of two kinds: power laws, on the first hand, and log-normal CDFs, on the other hand. The latter kind has been used by Cramer (1962) to study the diffusion of car motorization among a population of households. Here we shall focus on the former kind, with some power r that needs not be an integer:

$$c(x) = \mathrm{c}_1.x^r. \tag{A3}$$

Of course factor c_1 needs be nonnegative to make some sense.

6.3/ Basic properties of log-normal distributions

Let us recall the definition of a uni-dimensional log-normal distribution: a real random variable X is said to be distributed $\mathrm{LN}(m, s^2)$ if it is positive and its natural logarithm is Gaussian, i.e. $\ln(X) \sim \mathrm{N}(m, s^2)$. Denoting as Φ the CDF of a reduced Gaussian variable and $\varphi(t) = \exp(-t^2/2)/\sqrt{2\pi}$ the associated PDF, and letting $t_x := (\ln(x) - m)/s$, it holds that

$$\mathrm{F}_{\mathrm{o}}(x) = \Phi(t_x)$$

$$\mathrm{F}_{\mathrm{o}}^{(-1)}(\alpha) = \exp(m + s.\Phi^{(-1)}(\alpha))$$

$$\mathrm{f}_{\mathrm{o}}(x) = \frac{1}{s.x}\varphi(t_x)$$

$$\mathrm{E}_{\mathrm{o}}[x] = \exp(m + \frac{s^2}{2})$$

$$\mathrm{V}_{\mathrm{o}}[x] = (\mathrm{E}_{\mathrm{o}}[x])^2(\mathrm{e}^{s^2} - 1)$$

$$\gamma_{\mathrm{o}} = \sqrt{\mathrm{e}^{s^2} - 1}$$

Hence $s = \sqrt{\ln(1 + \gamma_{\mathrm{o}}^2)}$.

Furthermore, any derived random variable $Y := \mathrm{c}_1.X^r$ with $\mathrm{c}_1 > 0$ is a log-normal variable, too. This is because $Y \geq 0$ and $\ln(Y) = \ln(\mathrm{c}_1) + r.\ln(X)$, implying that $\ln(Y) \sim \mathrm{N}(\ln(\mathrm{c}_1) + r.m, (rs)^2)$, making Y an LN variable with parameters $\ln(\mathrm{c}_1) + r.m$ and $(rs)^2$.

6.4/ The "Truncated Moments" formula

Coming to the population of consumed units in a consumption model with power function, we can avail ourselves of the "truncated moment" formula, namely:

$$\int_a^b x^r d\mathrm{F}_\mathrm{o}(x) = e^{r(m+rs^2/2)}\left\{\Phi\left(\frac{\ln(b)-m}{s} - rs\right) - \Phi\left(\frac{\ln(a)-m}{s} - rs\right)\right\}. \tag{A4}$$

The reason is that $\int_a^b x^r d\mathrm{F}_\mathrm{o}(x) = \int_{t_a}^{t_b} e^{r(m+st)}\varphi(t)dt = e^{r(m+rs^2/2)}\int_{t_a}^{t_b}\varphi(t-rs)dt$, in which we replace $\int_{t_a}^{t_b}\varphi(t-rs)dt$ with $\Phi(t_b - rs) - \Phi(t_a - rs)$.

It follows that $\mathrm{F}_\mathrm{u}(x) = \Phi(\frac{\ln(x)-m}{s} - rs)$ i.e. that in the population of consumed units, level x is distributed $\mathrm{LN}(m + rs^2, s^2)$.

Thus $\mathrm{E}_\mathrm{u}[x] = \exp(m + rs^2 + \frac{1}{2}s^2)$ and $\gamma_\mathrm{u} = \sqrt{\mathrm{e}^{s^2} - 1} = \gamma_\mathrm{o}$.

When $r = 1$, $\mathrm{F}_\mathrm{u}(x) = \Phi(\frac{\ln(x)-m}{s} - s)$ and $\mathrm{E}_\mathrm{u}[x] = \exp(m + \frac{3}{2}s^2)$.

It is then easy to obtain

$$\frac{\bar{x}_\mathrm{u}}{\bar{x}_\mathrm{o}} = 1 + \gamma_\mathrm{o}^2 = \mathrm{e}^{s^2} \tag{A5}$$

6.5/ Lorenz curve and Gini index

The Lorenz function $\mathrm{L} := \mathrm{F}_\mathrm{u} \circ \mathrm{F}_\mathrm{o}^{(-1)}$ here involves $\mathrm{F}_\mathrm{o}^{(-1)}(\alpha) = \exp(m + s.\Phi^{(-1)}(\alpha))$. Thus

$$\mathrm{L}(\alpha) = \Phi(\Phi^{(-1)}(\alpha) - s) \tag{A6}$$

The Gini index, $G := 2\int_0^1(\alpha - \mathrm{L}(\alpha))d\alpha$, can be considered as a function of s.

At point $s = 0$, as $\Phi\left(\Phi^{(-1)}(\alpha)\right) = \alpha$, then $\mathrm{G}(0) = 2\int_0^1(\alpha - \alpha)\, d\alpha = 0$.

Differentiating G with respect to s, we get that: $\dot{\mathrm{G}}(s) := \frac{d\mathrm{G}}{ds} = 2\int_0^1 \varphi(\Phi^{(-1)}(\alpha) - s)d\alpha$

Changing variables according to $t := \Phi^{(-1)}(\alpha)$ hence $d\alpha = \varphi(t)dt$, we get that

$\dot{\mathrm{G}}(s) = 2\int_{-\infty}^{+\infty}\varphi(t-s).\varphi(t)dt$

Rearranging $(t-s)^2 + t^2 = 2t^2 - 2ts + s^2 = 2(t - \frac{1}{2}s)^2 + \frac{1}{2}s^2 = (\sqrt{2}t - \frac{s}{\sqrt{2}})^2 + (\frac{s}{\sqrt{2}})^2$, consequently $\varphi(t-s).\varphi(t) = \varphi\left(\sqrt{2}t - \frac{s}{\sqrt{2}}\right).\varphi(\frac{s}{\sqrt{2}})$ hence

$\dot{\mathrm{G}}(s) = 2\varphi\left(\frac{s}{\sqrt{2}}\right)\int_{-\infty}^{+\infty}\varphi\left(\sqrt{2}t - \frac{s}{\sqrt{2}}\right)dt = \sqrt{2}\varphi(\frac{s}{\sqrt{2}})\int_{-\infty}^{+\infty}\varphi(u)du = \sqrt{2}\varphi(\frac{s}{\sqrt{2}})$.

By integration, $\int_0^s \dot{\mathrm{G}}(v)dv = \sqrt{2}\int_0^s \varphi\left(\frac{v}{\sqrt{2}}\right)dv = 2\int_0^{\frac{s}{\sqrt{2}}}\varphi(w)dw = 2\left[\Phi\left(\frac{s}{\sqrt{2}}\right) - \Phi(0)\right]$.

To sum up, $\mathrm{G}(s) = \mathrm{G}(0) + \int_0^s \dot{\mathrm{G}}(v)dv$ so that

$$\mathrm{G}(s) = 2\Phi\left(\frac{s}{\sqrt{2}}\right) - 1$$