ACCSCIENCE PUBLISHING

International Journal of Population Studies

RESEARCH ARTICLE

Used (Lived) versus Offered (Plain) densities of human settlement in space: An instance of the probabilistic consumption model

Fabien Leurent*

International Research Centre on Environment and Development (CIRED), Ecole des Ponts ParisTech, France

Abstract

To people living in areas, the denser is the area, the more numerous are the opportunities of interpersonal and social interaction, of employment and of amenities of all kinds. The spatial density of human settlement is basically studied according to places, that is, area weighted. The notion of population-weighted density, or lived density, puts human density in the perspective of the people that experience it. Considering, respectively, the land units and the people as statistical populations of their own, the article provides a probabilistic model of human density in a geographical space, as a random variable in each statistical population, with specific probability density functions (PDFs) and cumulative distribution functions. The PDF of lived, "Used density" is derived from that of the plain, "Offered density" through a consumption model: Thus, their relationship is a specific instance of a well-established probabilistic model. The average used density is systematically larger than its offered counterpart: The ratio amounts to one plus the squared coefficient of variation of offered density. The relation between the two statistical distributions is illustrated using a Lorenz curve; the associated Gini index constitutes an indicator of population heterogeneity in a geographical space. A case of France's population as of 2019 is studied to demonstrate the methodology.

Keywords: Spatial heterogeneity; Density metrics; Land units; Consumption model; Lorenz curve; Gini index

***Corresponding author:**
Fabien Leurent (fabien.leurent@enpc.fr)

Citation: Leurent, F. (2022). Used (Lived) versus Offered (Plain) Densities of Human Settlement in space: An instance of the probabilistic consumption model. *International Journal of Population Studies,* 8(2):34-50. https://doi.org/10.36922/ijps.v8i2.297

Received: May 23, 2022

Accepted: November 10, 2022

Published Online: November 30, 2022

1. Introduction

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 1000 inhabitants per square km to more than 100,000 in the densest parts of some Asian megacities (e.g., Dhaka). Rural areas, in contrast, exhibit sparse human settlements and low densities: Typical values of rural density range from near zero to some hundred inhabitants per square km (Aliaga *et al.*, 2015; Vorobyev, 2019). 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).

Yet, the basic indicator of human density pertains, essentially to space. As geographical space is likely to exhibit heterogeneity of human occupation, the average plain density (rather than "crude density" as worded by Craig, 1984) does not indicate much about the effective intensity of human occupation that people experience in their everyday life, in the places that they frequent – home, workplace, etc. This is why alternative, population-weighted indicators of density have been developed by pioneering researchers in the mid-1970s. Stairs (1977a; 1977b) introduced the person-average density as "the local density experienced by people, that is, spatial density of human population weighted by the number of people experiencing it." This indicator was called population-weighted (arithmetic mean) density by Craig (1984). The name of "population-weighted density" has popularized among geographers and economists that have used the indicator to understand "density as lived by people" and its effects on urbanization (Eidlin, 2010, Florida, 2012), mobility practices (Barnes, 2001), as well as greenhouse gas (GHG) emissions of travel (Lee & Lee, 2014). As stated by Barnes (2001: p.16), "we are interested in human behavior; what we want to know is what people perceive density to be: This would be more closely captured by giving equal weight to each person, rather than to each square mile of land…. thus (we shall) use a new measure called 'perceived density,' which is defined as a weighted average of traffic zone densities, where each zone is weighted by the number of residents." Ottensmann (2018a) provided a comprehensive literature review of the concept, its applications and the related issues, together with a diachronic application to major US cities.

So far, the concept of "lived density" and "population-weighted density" has been expressed using mathematical formulas of weighted averages involving the numbers of people living in pre-defined zones (Barnes, 2001; Craig, 1984; Ottensmann, 2018a; Stairs, 1977). The relationship between lived and plain density averages has been established by Lewontin & Levins (1989) and Ottensmann (2018a) using lengthy proofs. This article is aimed to state plain density and lived density in the standard probabilistic framework, involving statistical populations, respectively, of land units and people, local human density as a random variable in either statistical population, their respective probability density functions (PDFs), and the general relation between them. This relation consists in a specific instance of a well-known probabilistic model: The "consumption model" that arises in various fields from economics to traffic theory. To emphasize the consumption model that relates plain density to lived density, we consider land units as servers of spatial settlement for people. Then, plain density is also the offered density, whereas lived density is also the "used density," since people are the users of the settlement services. The relation between the PDFs enables one to build average indicators in a straightforward way. It also allows for considering classical indicators of heterogeneity in a simple way: we introduce interquartile ratios and above all the Lorenz curve and the Gini index of human density in a given geographical space.

The rest of the article is organized in four 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 a short conclusion. Four appendices provide more details on (A) the notation table, (B) the consumption model, (C) the lognormal distribution and its use for consumption models, and (D) the linear-log model of used density CDF.

2. Theoretical background

We shall first define land units to deal with a geographical space as a statistical population of such elementary places (§2.1). Then, human density is put as a random variable with its own probabilistic distribution, mean and other statistical moments, and heterogeneity measures (§2.2). The next step is to shift the statistical perspective from the statistical population of land units to that of people: The random variable of used human density inherits its probabilistic features – probability distribution and statistical moments including the average value – from its offered counterpart on the basis of a consumption model (§2.3). We then recall the Lorenz curve and the Gini index as classical tools for inequality measurement and adapt them to the inequality of human population among land units (§2.4). Finally, we recall some previously proposed indicators of heterogeneity for human density and we restate them using our notation (§2.5).

2.1. Geographical space as a statistical distribution of land units

2.1.1. Territory, zoning system, and population

To analyze human density in a geographical space, it is convenient to model that space using a set Z of zones z. Each zone has its own ground area, A_z, and human population, P_z. Its spatial density of human population is simply:

$$x_z = P_z / A_z \tag{1}$$

The territory has total ground area of $A_Z = \sum_{z \in Z} A_z$ and total population of $P_Z = \sum_{z \in Z} P_z$. Its human density averaged over space is therefore

$$\bar{x}_Z = P_Z / A_Z \tag{2}$$

2.1.2. Spatial units as a statistical population

When zones are used to analyze the statistical distribution of some spatial variables in a discrete way, they are often called "spatial units." Here, we shall rather refer to zones as "spatial entities" and define "land units" as elementary places of unit ground area, say a_1. Such land units are more convenientthan 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 land units o to any zone z is an idealization: Thinking of the unit ground area a_1 as 1 square km or 1 hectare, we expect most zones to involve a non-integer number of land units. We nevertheless denote as " $o \in z$" the composition of zone z out of land units o. To every land unit o, with population denoted by p_o, is associated a human density as follows:

$$x_o = \mathrm{p}_o / \mathrm{a}_1 \tag{3}$$

Notionally, the zone area adds up those of the land units in it, and similarly, the zone population adds up those of its land units:

$$\mathrm{A}_z = \sum_{o \in z} \mathrm{a}_1 \tag{4a}$$

$$\mathrm{P}_z = \sum_{o \in z} \mathrm{p}_o \tag{4b}$$

2.2. Human density in the statistical population of land units

Human density x constitutes a random variable in the statistical population of land units, with PDF and CDF denoted by f_{O} and F_{O}, respectively.

2.2.1. Average human density over space

The average human density over space stems from the probability density function f_{O} in the usual way (e.g., Blitzstein & Hwang, 2015):

$$\bar{x}_{\mathrm{O}} \equiv \int x.\mathrm{f}_{\mathrm{O}}(x)dx \tag{5}$$

This general version of average human density over space is equivalent to the discretized one: Denoting as $\bar{\mathrm{O}}$ the total number of land units, it holds that:

$$\bar{x}_{\mathrm{O}} \equiv \frac{\sum_{o \in \mathrm{O}} x_o}{\bar{\mathrm{O}}}$$

As $\bar{\mathrm{O}}.\mathrm{a}_1 = \mathrm{A}_Z$, replacing x_o with p_o/a_1 and $\mathrm{a}_1\bar{\mathrm{O}}$ with A_Z due to (4a) aggregated over zone set Z, it comes out that

$$\bar{x}_{\mathrm{O}} = \frac{\mathrm{P}_\mathrm{Z}}{\mathrm{A}_\mathrm{Z}} \tag{6}$$

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

2.2.2. On the statistical moments of human density

Higher order moments of human density in the statistical population of land units are defined in the usual way (e.g., Blitzstein & Hwang, 2015): At order r,

$$\mathrm{E}_{\mathrm{O}}[x^r] \equiv \int x^r.\mathrm{f}_{\mathrm{O}}(x)dx \tag{7}$$

Under the idealized assumption that human density would be homogenous among the land units composing any zone (i.e., no intra-zonal heterogeneity of density), then $\mathrm{E}_{\mathrm{O}}\left[x^r\right] = \mathrm{E}_{\mathrm{Z}}[x^r]$, wherein the inter-zone average E_{Z} of density moment x^r is defined as

$$\mathrm{E}_{\mathrm{Z}}[x^r] \equiv \sum_{z \in \mathrm{Z}} \frac{\mathrm{A}_z}{\mathrm{A}_\mathrm{Z}} \left(\frac{\mathrm{P}_z}{\mathrm{A}_z}\right)^r \tag{8}$$

The formula enables one to calculate $\mathrm{E}_{\mathrm{O}}\left[x^r\right]$ in an exact way under intra-zonal homogeneity of density, or in an approximate way otherwise.

2.2.3. Indicators of offered human density heterogeneity

Local human density is likely to be heterogeneous among land units, even inside every zone z. The intra-zone variance of human density is a metric for that heterogeneity within z. It is defined as $\mathrm{V}_{\mathrm{O}}^{(z)}\left[x\right] \equiv \mathrm{E}_{\mathrm{O}}\left[x^2 | o \in z\right] - (\mathrm{E}_{\mathrm{O}}\left[x | o \in z\right])^2$, and satisfies that

$$\mathrm{V}_{\mathrm{O}}^{(z)}\left[x\right] = \left\{\sum_{o \in z} \frac{\mathrm{a}_1}{\mathrm{A}_z} x_o^2\right\} - x_z^2 \tag{9}$$

Over the territory, the overall variance of human density can be measured using the law of total variance, that is, its decomposition into within-group variance and between-group variance (e.g., Blitzstein & Hwang, 2015):

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

The associated standard deviation (SD) and relative dispersion (coefficient of variation, CV) are therefore

$$\mathrm{SD}_{\mathrm{O}}\left[x\right] \equiv \sqrt{\mathrm{V}_{\mathrm{O}}\left[x\right]} \tag{11a}$$

$$\gamma_{\mathrm{O}}\left[x\right] \equiv \mathrm{SD}_{\mathrm{O}}\left[x\right] / \bar{x}_{\mathrm{O}} \tag{11b}$$

In empirical distributions, the variance and, in turn, the SD and CV are sensitive to outliers, that is, values falling out of the ordinary range of the variable. As the quantile values $\mathrm{F}_{\mathrm{O}}^{(-1)}(\alpha)$ at probability level α neither too small nor too high are less sensitive to extreme values, they are taken as "robust statistics" (Wonnacott & Wonnacott, 1990). Thus, for non-negative real variables, the ratios between corresponding pairs $F_O^{(-1)}(\alpha)$, $F_O^{(-1)}(1-\alpha)$ constitute robust indicators of heterogeneity in the distribution: These include the inter-quartile ratio at level $\alpha = 1/4$ as well as the inter-decile ratio at level $\alpha = 1/10$.

For instance, a log-normal distribution LN (m,s^2) yields the following interquartile ratio at order α (cf. appendix C), with $\Phi^{(-1)}$ the inverse CDF of the reduced Gaussian variable:

$$\frac{F_O^{(-1)}(1-\alpha)}{F_O^{(-1)}(\alpha)} = e^{2s\Phi^{(-1)}(1-\alpha)}$$

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

Any individual u inhabits a zone z_u and inside it a land unit $o(u)$. One defines the "used density" or "lived density" as the human density in the land unit inhabited by the user:

$$x_u \equiv x_{o(u)} \tag{12}$$

Among the population of users, of size $\bar{U} = P_Z$, the user-centric density is a random variable, with specific PDF and CDF denoted, respectively, as f_U and F_U.

2.3.1. Relationship between f_o and f_U

The PDF f_U of used human density is related to that f_O of offered human density by a "consumption model" well-known in probabilistic theory, especially in renewal theory (Kleinrock, 1975):

$$f_U(x) \propto x.f_O(x) \tag{13}$$

The reason is that the land units of which the human density belongs to $[x, x+\delta x]$, in proportion $f_O(x).\delta x$ in their distribution, do contain $x.a_1$ users each: Hence, their total number of users amounts to $a_1.x.f_O(x).\delta x.\bar{O}$. These users are those with user-centric density in $[x, x+\delta x]$ and those users only: Thus, their number is also $f_U(x).\delta x.\bar{U}$. On combining both formulas, as $\bar{O}.a_1 = A_Z$ and $\bar{U} = P_Z$, it comes out that

$$f_U(x) = \frac{A_Z}{P_Z} x.f_O(x)$$

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

$$f_U(x) = \frac{1}{\bar{x}_O} x.f_O(x) \tag{14}$$

2.3.2. Relationship between the statistical moments of the two distributions

From (14) stems a generic relation between the statistical moments of the two distributions (Kleinrock, 1975): At any order r, it holds that

$$E_U[x^r] = \frac{1}{\bar{x}_O} E_O\left[x^{r+1}\right] \tag{15}$$

This relation corresponds to the zone-based formula in Stairs (1977). At order r = 1, the average density as experienced by the users satisfies that

$$\bar{x}_U = \frac{1}{\bar{x}_O} E_O\left[x^2\right] = \bar{x}_O(1+\gamma_O^2) \tag{16a}$$

It is equivalent to a ratio of the average densities (RADs):

$$\mathrm{RAD}_{U|O} \equiv \frac{\bar{x}_U}{\bar{x}_O} = 1+\gamma_O^2 \tag{16b}$$

Both relations correspond to formulas established by Lewontin & Levins (1989) using zones as statistical units and Ottensmann (2018a) using land units. If the human density is homogenous in the territory, then $\gamma_O = 0$ and the average human densities according to either statistical population are equal. However, the larger the heterogeneity (as measured by the relative dispersion), the higher the ratio $\bar{x}_U / \bar{x}_O$ of used to offered average human densities.

Figure 1 illustrates both the used and offered PDFs. Its assumptions are the following: That x_O is distributed log-normal with parameters m_O = 3.24 (mean of $\ln(x_o)$) and s_O = 1.76 (standard deviation of $\ln(x_o)$), stemming from France-like conditions $\bar{x}_O = 120$ p/km² and $\gamma_O = 4.6$. The related x_u is log-normal, too, with parameters m_U = 6.34 and s_U=s_O, hence $\bar{x}_U = 2{,}661$ p/km² and γ_U = 4.6, too. In that particular instance, the ratio of average human densities, $\bar{x}_U / \bar{x}_O$, amounts to 22 – indeed a very large value.

2.3.3. Indicators of used human density heterogeneity

The indicators of heterogeneity recalled in §2.2.3 for the random variable of offered density also apply to that of used density. Its variance stems from moments of the offered density in a specific way: From (15) at order r=2,

$$E_U\left[x^2\right] = \frac{1}{\bar{x}_O} E_O\left[x^3\right]$$

Combining with (16a), it comes out that

$$V_U[x] = \frac{1}{\bar{x}_O} E_O\left[x^3\right] - \left(\frac{E_O\left[x^2\right]}{\bar{x}_O}\right)^2 \tag{17}$$

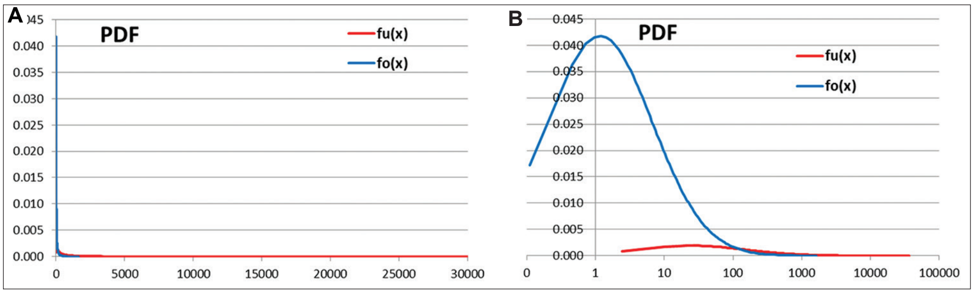

Figure 1. Used and offered PDFs of human density
(A) Standard scales, (B) abscissas in log-scale
Source: author's elaboration

When x_O follows a log-normal distribution, then so does x_U (cf. appendix C), with same variance s^2 of the law of ln x: Then, the interquartile ratios of x_U are identical to their offered counterparts.

2.4. Lorenz curve and Gini index

2.4.1. Recalling the classical theory of inequality measurement

In Gini's classical analysis of income inequalities (Gini, 1912; 1955; Gionanni & Gubbiotti, 2015), the sum of all individual incomes in a group of people is decomposed according to specific subgroups of people. A typical subgroup gathers the people that each earn less than a given level of income. Then, to the proportion $F_P(x)$ of-people whose income is $<x$ is associated the proportion $F_I(x)$ of total income that stems from the aggregation of their individual incomes. As both proportions are increasing functions of x, they depend on each other in a unique way. Their relationship is known as the "Lorenz function" denoted L and defined as follows (Cowell, 2009; Lorenz, 1905):

$$\alpha \mapsto \mathrm{L}(\alpha) \equiv \mathrm{F}_\mathrm{I} \circ \mathrm{F}_\mathrm{P}^{(-1)}(\alpha)$$

The reason is that each value of x gives rise to population proportion $\alpha \equiv \mathrm{F}_\mathrm{I}(x)$, therefore satisfying $x = \mathrm{F}_\mathrm{P}^{(-1)}(\alpha)$, and to income proportion $\mathrm{F}_\mathrm{I}(x)$, which is thus equivalently expressed as $\mathrm{F}_\mathrm{I}(\mathrm{F}_\mathrm{P}^{(-1)}(\alpha))$.

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

$$\dot{\mathrm{L}}(\alpha) = \frac{\mathrm{f}_\mathrm{I}\left(x_\mathrm{P}^{(\alpha)}\right)}{\mathrm{f}_\mathrm{P}\left(x_\mathrm{P}^{(\alpha)}\right)} = \frac{1}{\bar{x}_\mathrm{P}} x_\mathrm{P}^{(\alpha)} \tag{18}$$

It is non-negative and increasing with α since $\mathrm{F}_\mathrm{P}^{(-1)}$ is increasing: This makes L an increasing and convex function. Thus, in the diagram of F_I versus F_p in [0,1]×[0.1], the graph of function L, called the "Lorenz curve," lies below the straight line from point (0,0) to point (1,1) (Figure 2). The area between the straight line and the

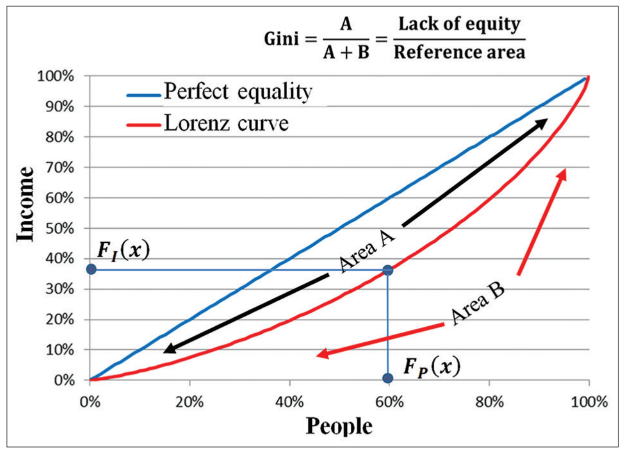

Figure 2. Lorenz function and Gini index
x designates an individual income
Source: Author's adaptation from https://en.wikipedia.org/wiki/Gini_coefficient [Last accessed 8.11.2022]

Lorenz curve, divided by the area below the straight line, that is, 0.5, is known as the Gini index, with mathematical formula as follows (Cowell, 2009):

$$G \equiv 2\int_0^1 (\alpha - \mathrm{L}(\alpha))d\alpha \tag{19}$$

The Gini index takes its value in [0,1]. Between different income distributions, the larger the heterogeneity, the larger the Gini index: It is a metric of inequality (Cowell, 2009). In appendix C, a log-normal instance is addressed to give insight in the consumption model and illustrate the properties of relative dispersions and the Gini index.

2.4.2. Adaptation to human density

Gini's line of reasoning applies to the distributions of human density: To the $\mathrm{F}_\mathrm{O}(x)$ share of space with density $<x$ corresponds the $\mathrm{F}_\mathrm{U}(x)$ share of people each experiencing individual density $<x$. Here, the Lorenz function is $\mathrm{L} \equiv \mathrm{F}_\mathrm{U} \circ \mathrm{F}_\mathrm{O}^{(-1)}$. The resulting Gini index constitutes another metric of density heterogeneity, along with γ_O and γ_U. Figure 3 exhibits a Lorenz curve of population density,

 https://doi.org/10.36922/ijps.v8i2.297

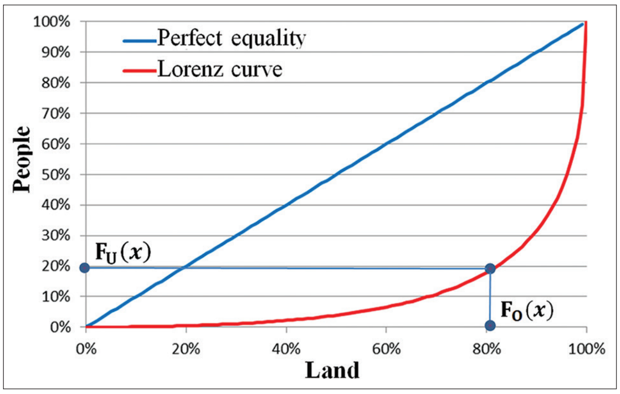

Figure 3. Lorenz function of human density
Source: Author's elaboration

assuming the same used and offered distributions of human density, as shown in Figure 1. This particular instance exhibits a Gini index of 0.79, which is a very high value for that kind of index (Cowell, 2009).

The Lorenz curve depicts the relation between land and people as it relates a proportion of people, on the vertical axis, to the proportion of land that accommodates them, on the horizontal axis. The relation pertains to the spatial density of human settlement: Both the land units (horizontal axis) and the individuals (vertical axis) are ranked in increasing order of density x. Given a specific value x of human density, the proportion $F_o(x)$ of land accommodates the proportion $F_U(x)$ of people, all of them at density lower than x. Conversely, the residual $1-F_O(x)$ share of land accommodates the residual $1-F_U(x)$ share of people, all of them at density greater than x. Thus, the point $(F_O(x), F_U(x))$ splits the diagram in two parts: Lower density space and people on the left side, higher density space and people on the right side. The average densities on the lower and higher sides are proportional to $\frac{F_U(x)}{F_O(x)}$ and $\frac{1-F_U(x)}{1-F_O(x)}$, respectively, with coefficient of $P_Z / A_Z = \bar{x}_O$. The ratio of higher and lower average densities (RADs) is thus equal to

$$\mathrm{RAD}_{H|L} \equiv \frac{1-F_U(x)}{F_U(x)} \frac{F_O(x)}{1-F_O(x)} \tag{20}$$

For instance, in Figure 3, it appears that about 20% of people are accommodated in 80% of space. The used offered ratio of $\frac{20\%}{80\%}$ on the lower side, compared to $\frac{80\%}{20\%}$ on the higher side, imply that the average density in the higher part is about 16 times that in the lower part.

2.5. On alternative average indicators of density

Stairs (1977) introduced generic weighting systems for averaging the population-weighted density and measuring the heterogeneity of human density. Denoting a weighting system as a function $x \mapsto w(x)$ of density level x, the associated average density is stated as follows:

$$x_w \equiv \frac{E_U[w]}{E_O[w]} \tag{21}$$

Stairs (1977) also considered "generalized population density" as the r-th order root of the ratio of moments at orders $k + r$ and k:

$$x_k^r \equiv \left(\frac{E_O\left[x^{k+r}\right]}{E_O\left[x^k\right]} \right)^{\frac{1}{r}} \tag{22}$$

Average lived density $\bar{x}_U$ is a particular instance associated to $(k,r) = (1,1)$.

It is shown in appendix C that, when the offered density follows a log-normal distribution with median M_o and relative dispersion γ_O, then this average indicator satisfies that,

$$x_k^r = M_O(1+\gamma_O^2)^{k+r/2} = M_O(\bar{x}_U / \bar{x}_O)^{k+r/2}$$

Therefore, it is basically a power function of the ratio of used and offered average densities, $\bar{x}_U / \bar{x}_O$.

Complementarily to the population-weighted arithmetic mean density, Craig (1984) also considered the logarithm of density as an indicator of the magnitude of density. He argued that magnitude-based density indicators would be especially relevant to assess the variations of local density over time by laying the emphasis on local significance, since the local meaning of a given change of density depends on the initial, local density level. Using our notation, the average indicator of density magnitude states as $E_U[\ln x]$ and gives rise to the geometric mean of used density,

$$\tilde{x} \equiv \exp E_U[\ln x] \tag{23}$$

Craig (1984) related the geometric mean to "the ideas of information gain and entropy" and insisted on its property of decomposability along spatial sub-divisions. It is also akin to the Theil index of heterogeneity (Cowell, 2009). When the offered density follows a log-normal distribution $LN(m,s^2)$, then the used density is log-normal, too, with identical relative dispersion γ (cf. appendix C) and the average log value $E_U[\ln x]$ is equal to $m+s^2$, that is, to

$E_O\left[\ln x\right]+\frac{1}{2}\ln(1+\gamma^2)$. Therefore, the related geometric mean $\tilde{x}$ satisfies the following equalities which makes it a geometric midpoint between $\bar{x}_O$ and $\bar{x}_U$:

$$\tilde{x}=\exp(m+s^2)=\bar{x}_O\exp\left(\frac{1}{2}s^2\right)=\bar{x}_O\sqrt{1+\gamma^2}=\frac{\bar{x}_U}{\sqrt{1+\gamma^2}} \quad (24)$$

Thus, based on the case of log-normal distributions, it may be conjectured that alternative indicators of average density do not yield much gain beyond considering the $(\bar{x}_O, \bar{x}_U)$ pair of average densities.

3. A case study of France as of 2019

3.1. The territory under study

Metropolitan France comprises about 34,750 municipalities called "communes" (INSEE, 2021a). 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 (Aliaga *et al*., 2015).

As of 2019, the French metropolitan population amounts to about 65 M inhabitants (INSEE, 2021b). Thus, the average commune population is 1800 people only and the overall spatial density of population is about 120 persons per km². Figure 4 exhibits the map of French communes colored according to density level in the way of either the INSEE (Aliaga *et al*., 2015), the French National Institute for Economic Statistics, or Eurostat, the statistical body of the European Union (Eurostat, 2019). It shows that most of the country area has low population density.

3.2. Used density versus 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 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 F_u CDF is obtained at point x_z.

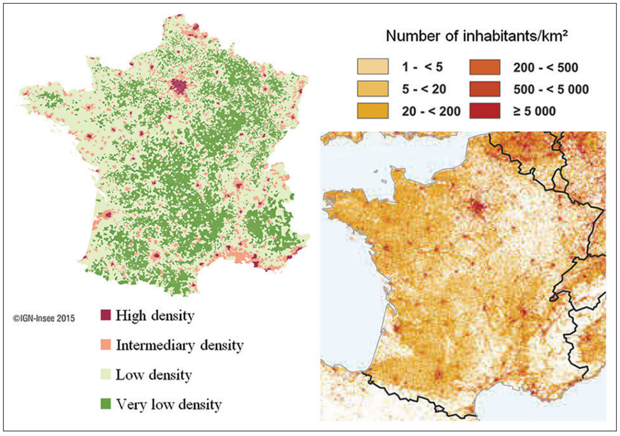

Figure 4. Human density of French communes as of 2015 and 2018
Sources: Aliaga *et al.* (2015), Eurostat (2019)

The next step is to draw the diagram of F_u versus F_o, that is, the Lorenz curve (Figure 5A). The Gini index is easy to calculate, by accumulating $2\left(F_O\left(x_z\right)-F_U\left(x_z\right)\right).(F_O\left(x_z\right)-F_O\left(x_{z-1}\right))$ over communes z. The outcome is 0.76, again a very large value. Furthermore, 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: $E_O\left[x\right]=120$ and $SD_O\left[x\right]=548$, yielding $\gamma_O=4.58$ (dimensionless).
- for x_U: $E_U\left[x\right]=2{,}628$ and $SD_U\left[x\right]=4{,}691$, yielding $\gamma_U=1.79$ (dimensionless).

Figure 5B and C depicts the empirical CDFs F_O and F_U, together with log-normal approximations that mimic the mean and variance of each distribution. To obtain PDFs (Figure 5D and E), we discretized the CDFs and derived the respective PDFs as the average value between two successive points. On trying to model the offered density as a lognormal distribution, a close match was obtained: Yet, a perfect lognormal model would entail identical relative dispersions between the offered and used distributions – a conclusion definitely not supported by the data. Looking for alternative conventional distributions to fit the data, Singh-Maddala CDFs were found appropriate (Figure 6).

More simply, it turned out that the used density CDF, F_U, is about an affine linear function of ln$x$ from its first to 8th deciles. A straightforward consequence is that, on the interval between the two deciles, $\check{f}_U(x)\propto x^{-1}$. In turn, the offered density is about inverse quadratic, $\check{f}_O(x)\propto x^{-2}$. Both approximations are well supported by the data (Figure 7). Some related analytical properties are provided in appendix D.

3.3. Decile values and the heterogeneity of human density

The decile values of the offered and used distributions of human density were derived from their respective

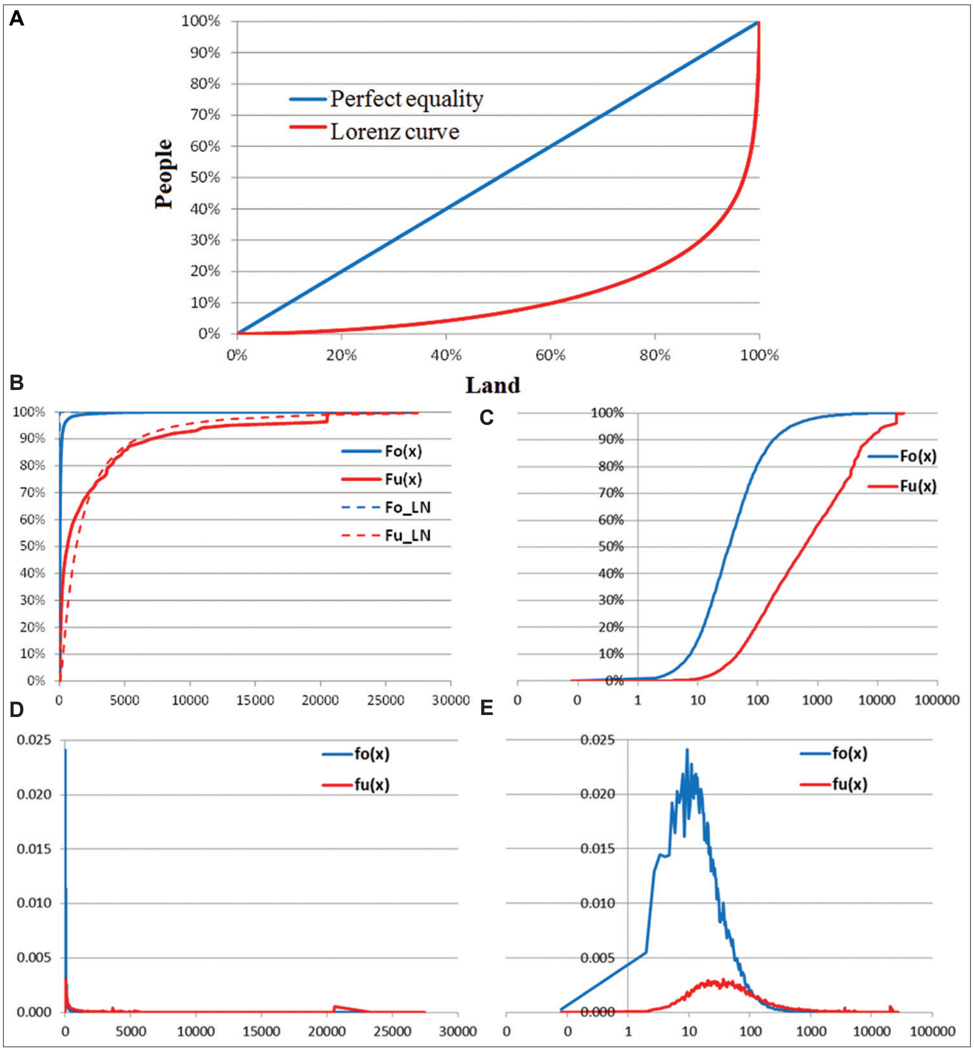

Figure 5. Human density in France, 2019
(A) Lorenz curve, (B) used and offered CDFs of human density, (C) same with abscissas in log-scale, (D) used and offered PDFs of human density, and (E) same with abscissas in log-scale
Source: Author's calculations based on INSEE data (2021a; 2021b).

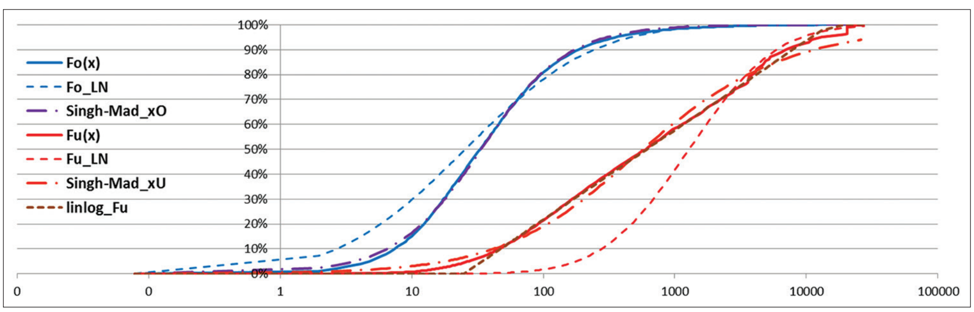

Figure 6. Used and offered CDF, with log-normal, Singh-Maddala, and linear-log approximations
Abscissas in log-scale
Source: Author's estimations based on INSEE data (2021a; 2021b)

CDF (Table 1). The median value of offered density is much lower than the offered mean – its significance is only to state that a major share of France's territory lies under very low density. The 80 – 20% shares of low versus.

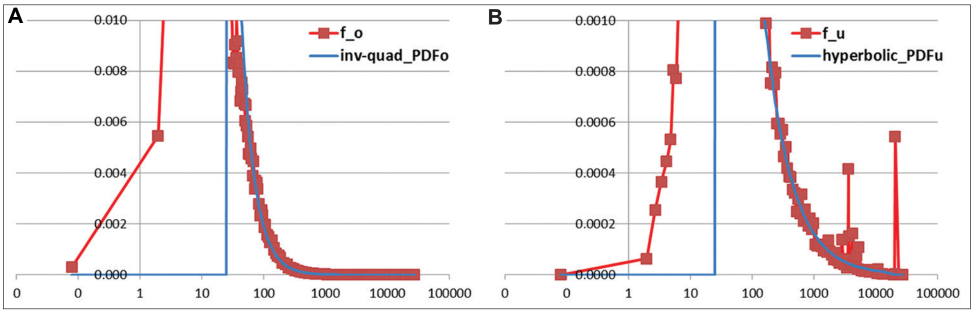

Figure 7. Approximations of (a) offered density PDF and (b) used density PDF
Abscissas in log-scale.
Source: Author's calculations based on INSEE data (2021a; 2021b).

Table 1. Deciles of offered density versus used density (France, 2019), in persons per km²

α	10%	20%	30%	40%	50%	60%	70%	80%	90%
$x_O(\alpha)$	7	12	17	24	33	46	64	96	186
$x_U(\alpha)$	46	92	167	306	600	1,146	2,213	3,872	6,999

Source: Author's calculations based on INSEE data (2021a; 2021b).

high density land are associated to 20 – 80% shares of low versus high density people. Such striking contrast calls for quantitative metrics to complement density maps in spatial analysis.

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, that is, 600 persons per square kilometer may be taken as "medium level of density," low densities for the bottom 20%, that is, below 92 p/km², high densities for the top 20%, that is, 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 are the relative dispersions of used and offered density.

As for interquartile ratios to measure distribution heterogeneity, the outcomes are:

- For offered density, inter-decile ratio (RID) of 186/7 = 27 and inter-quartile ratio (RIQ) of 80/15 = 5.3.
- For used density, inter-decile ratio (RID) of 7000/46= 152 and inter-quartile ratio (RIQ) of 3000/120 = 25.

The RID and RIQ of x_O are quite high for their kinds of indicators. As for x_U, the RID and RIQ values are still much higher: they reveal the very large heterogeneity of used density, that is, of human density as lived by the people.

We utilized the dataset to calculate the alternative indicators of average density recalled in §2.5. Craig's geometric mean of the used density has a value of $\tilde{x} =$ 590 p/km²: Thus, it is close to the geometric midpoint between $\bar{x}_O$ and $\bar{x}_U$, since $\tilde{x} \approx 4.45\bar{x}_O$ and $\bar{x}_U \approx 4.93\tilde{x}$. Both ratios are close to value $\sqrt{1+\gamma_O^2} = 4.69$ that was expected using a lognormal approximation of x_o.

The Stair's generalized indicators x_k^r in (22) were computed for indices k from 0 to 9 and r from 1 to 5. Figure 8 exhibits the reduced indicators $\ln(x_k^r / \mathrm{M_O})/(k+\frac{1}{2}r)$ as functions of k depending on r. The salient values are those associated to pairs $(k,r)=(0,1)$ and $(1,1)$, that is, to $\bar{x}_O$ and $\bar{x}_U$, while the other pairs yield an overall pattern about value 1 that corresponds to lognormal distributions.

4. Discussion

4.1. On the statistical populations and random variables of human density

The history of statistics began with early censuses of the human population in a couple of countries, that is, with human populations as statistical populations – hence, the very name of the latter concept. In such a historical perspective, the used density lends itself to be modeled as a random variable in the human population. However, as geography and cartographic methods have been well developed long before the advent of computers, the plain, offered density has been introduced long before the used, lived density. In the pioneering contributions of Craig (1975; 1979; 1980; 1984) and Stairs (1977a), it has been clearly stated that the average plain density is area weighted, whereas the average lived density is a population-weighted

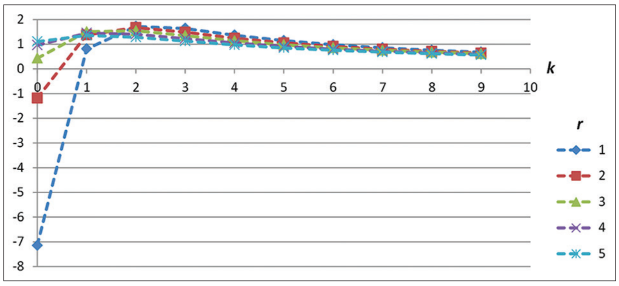

Figure 8. Stair's generalized density indicators under reduced form $\ln(x_k^r / \mathrm{M_O}) / (k + \frac{1}{2} r)$
Source: author's calculations based on INSEE (2021a; 2021b)

one: This points to the respective statistical individuals. Lewontin & Levins (1989) went further in the probabilistic representation of human density by considering spatial cells of unit ground area as the statistical individuals in geographical space as a statistical population: The cell attribute of local human density is inherited by the people inhabiting the cells, therefore constituted the used density as an RV in the human population. Yet, the probabilistic framework of Ottensmann (2018a) is limited to the two statistical populations, the human density RVs on both sides and the relations between the average values.

In the present article, we introduced the respective PDFs of human density in either statistical population and we related them using a consumption model. This affiliation to (simple) probabilistic theory yields several benefits, from the mathematical statement of the PDF relationship which is a fundamental one, to the relation between statistical moments, and up to the consideration of the Lorenz function and the related Gini index. The relation between the respective averages of used versus offered density is a prominent instance of the relation between statistical moments. The reference to theory provides the general relation between moments in a straightforward way, compared to the former derivations of the relation between average densities (Lewontin & Levins, 1989; Ottensmann, 2018a). Stairs (1977a) had stated the relation between the statistical moments in a concise and general way, yet with no explicit consideration of PDFs.

In the application to France as of 2019, the PDFs of offered and used densities were easy to study on a standalone basis. Yet, to visualize both functions jointly, the standard diagram (Figure 5D) depicted mainly the quasi-disjunction of their supporting sets – lower values of x giving most of the probability weight to x_o versus much higher values giving significant probability weight to x_U. The recourse to log-scale on the x-axis (Figure 5E) was instrumental to exhibit the PDF values on the y-axis and enable for some visual comparison. In such a diagram, however, it is less easy to visualize that the area below the PDF curve amounts to 1. The CDF functions are easier to depict jointly than their PDF counterparts. Using conventional abscissas (Figure 5B), function F_U increases in a smooth way paved by the decile values, while function F_O increases in a one-shot way at low values of x, making the decile values hardly legible. The recourse to log-scale on the x-axis (Figure 5C) enables one to visualize the magnitude of the deciles. The graphical practicality of density magnitudes, that is, of the logarithms of the human density, provides another reason to utilize them, beside the point made by Craig (1984) that relative changes in local population densities are more significant than absolute ones to assess the variations over time of human density in a geographical space.

4.2. On land units and the heterogeneity of land-use

Ottensmann (2018a) studied and discussed the definition of spatial cells, in other words, the zoning system to consider for local density and its assignment to people according to their zones of residence. Craig (1984) studied the effects of three zoning systems on the values of the used density in Great Britain as of 1971: He stated that "both the means (of offered and used density) increase as the (spatial) units are progressively subdivided." This is a straightforward consequence of the law of total variance applied progressively to finer and finer subgroups. It emphasizes that the notion of used density strongly depends the underlying system of residence zones. This must be kept in mind in the consideration of any set of numerical values of used density indicators. Such sensitivity also pertains to any indicator of heterogeneity in offered densities, since the zoning sensitivity of the average used density, on the left side of (16a), comes from that of the relative dispersion of the offered density, on the right side of (16a). In fact, the zoning system to estimate both offered density heterogeneity and average used density has to satisfy a twofold condition on zone sizes: A trade-off between, on the one hand, zones small enough to capture the spatial heterogeneity finely and, on the other hand, zones large enough to grasp the "life basins" where people live.

4.3. On indicators of heterogeneity in human density

For positive real variables, the relative dispersion is a heterogeneity indicator derived from the standard deviation divided by the mean. The explicit modeling of the PDFs of human densities, that is, f_o and f_U, induces their respective CDFs f_o and f_U: From these stem, the quantiles of their respective distribution, and in turn robust statistics, including the median as a middle value and also interquartile ratios as measures of heterogeneity, including the inter-quartile ratio and the inter-decile one. Not only are the interquartile ratios simpler than the alternative indicators of average density put forward by Craig (1984) and Stairs (1977a), but also they are more informative in the case of France as of 2019.

The Gini index constitutes a heterogeneity indicator of its own kind. It is a statistical summary of the full Lorenz curve, which is a more comprehensive tool but also more disaggregate to apprehend heterogeneity in both x_O and x_U. Some hints of that appeared in previous studies such as Aliaga *et al.* (2015) who mentioned that 35% of French people are living in even 90% of French communes: 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. In the 2019 France case, it was found that 80% of people live in 20% of space and conversely that the remaining 80% of space accommodate the remaining 20% of people. Thus, it constitutes yet another instance of the Pareto principle that arises in many distributions from incomes in populations to the sizes of firms. It is consistent with the linear-log model of the used density CDF, since this model makes the used density PDF a limit Pareto law and the offered density a truly Pareto law.

4.4. On the applications of used density

The local density of a place is a specific attribute of individual people living there. The notion of x_U and its probabilistic features from f_U to $\bar{x}_U$ and γ_U constitute a simple statistical model to analyze human population according to human spatial density.

Ottensmann (2018a) pointed out to the existing areas of application of the used density: (i) Primarily as "a descriptive measure of distribution, often in comparison with conventional density," (ii) urbanization patterns and their dynamic variations, including urban sprawl, (iii) mobility analysis relating residential and job densities to the modal share of transit modes of transportation (Barnes, 2001), (iv) agglomeration economies, relating used density to total urban factor productivity (Rappaport, 2008a) and to urban consumption amenities (Rappaport, 2008b), and (v) energy use and GHG emissions from household residential and travel patterns (Lee & Lee, 2014).

In recent years, the concept of population-weighted density has disseminated, notably through web online contributions showing its descriptive power (Bradford, 2008a-c; Florida, 2012; NENAD, 2021; Ottensmann, 2016) and above all in the academic literature of its various fields of applications: (i) Geographical analysis (Hanberry, 2022) including studies on the COVID epidemiology (Pascoal & Rocha, 2022), (ii) urbanization patterns (Townsend & Ellis-Young, 2018) and urban sprawl (Ottensmann, 2018b), (iii) mobility analysis (An *et al.*, 2022), (iv) economics (Albouy & Stewart, 2012; Faberman & Freedman, 2016; Krugman, 2013), and (v) environmental impacts (Huang & Brown, 2021) and potentials (Lu *et al.*, 2022). However, as of end 2022, population-weighted density still had not an entry of its own in the English-speaking Wikipedia, where "living density" is just mentioned as an alternative measure of human density (Wikipedia, 2022).

4.5. Further developments

The notion of population-weighted density was introduced in the mid-1970s by a demographer (Craig) and a chemistry scientist (Stairs), perhaps because quantitative socioeconomic analysis is a science of composition as is chemistry. Since then it has been adopted by geographers, economists interested in regional science and urban economics, and transport scientists.

Lived density may well be viewed as a simple form of spatial accessibility, as theorized by Hansen (1959), Poulit (1974) and Koenig (1974; 1980): The Hansen accessibility index, taken at a given zone as origin of trip-making, aggregates opportunities of a given kind over a larger territory, with numbers weighted by a declining function of origin-destination distance – or travel time or cost in the Poulit-Koenig formulation. This indicator has taken a central position in agglomeration economics (Fujita & Thisse, 2002) and geography economics (Krugman, 1997; Fujita *et al.*, 1999); it is also called the effective density in economic geography (Graham & Gibbons, 2019). It still remains to be considered as a property of the people residing in the origin zone, for all origin zones and all people, and to be analyzed as a random variable among the statistical population of people.

5. Concluding Remarks

The gist of the article is to model human density in geographical space using basic probability theory: (i) Statistical populations of land units and of people, respectively, (ii) putting human density as a random variable in both statistical populations, with specific PDFs and CDFs, (iii) relating the used density PDF to the offered density PDF through the probabilistic consumption model, (iv) deriving the statistical moments of the used density from those of the offered density, and (v) considering the Lorenz curve and the Gini index. The original contribution is 3-fold: The formal statement as random variables, the identification of the consumption model, and the consideration of heterogeneity indicators for human density in space (interquartile ratios and Gini index).

All of the concepts are well established in their own field, geography, or probabilistic modeling: The article provides a fresh perspective to relate the two fields – casting one more bridge between them. 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 (i) Gini's analysis of income inequality, but also (ii) the queuing theory of waiting times (e.g., Kleinrock, 1975), (iii) Wardrop's model of temporal vs. spatial distributions of vehicular speeds on roadways (Wardrop, 1952), and (iv) a model of transit vehicle loads and transit users' exposure to crowding conditions (Leurent *et al*., 2012; 2017).

Acknowledgments

The article has benefited greatly from the wise advice and recommendations of two anonymous referees, leading to substantial improvement on the initial version.

Funding

None.

Conflict of interest

The author has no conflicts of interest to declare.

Author contributions

This is a single-authored paper.

Ethics approval and consent to participate

All human data from publicly available dataset of selected census outcomes.

Consent for publication

Not applicable.

Availability of data

In sources INSEE (2021a) and INSEE (2021b).

References

Albouy, D., & Stuart, B. (2012). Urban quantities and amenities: The neoclassical model of location. *International Economic Review*, 61(1): 127-158.

https://doi.org/10.1111/iere.12419

Aliaga, C., Eusebio, P., & Levy, D. (2015). A novel approach to spaces with low or high density. In: *France and its territories*, 2015 edition, collection *INSEE References*. Available from https://www.insee.fr/fr/statistiques/fichier/1373022/FST15.pdf [Last accessed on 2022 Nov 09] [Book in French].

An, R., Wu, Z., Tong, Z., Qin, S., Zhu, Y., & Liu, Y. (2022). How the built environment promotes public transportation in Wuhan: A multiscale geographically weighted regression analysis. *Travel Behaviour and Society*, 29: 186-199.

https://doi.org/10.1016/j.tbs.2022.06.011

Barnes, G. (2001). Population and Employment Density and Travel Behavior in Large U.S. Cities. Final Report 2001-24. St. Paul: Minnesota Department of Transportation. Available from: https://www.lrrb.org/pdf/200124.pdf [Last accessed on 2022 Oct 14].

Blitzstein, J.K., & Hwang, J. (2015). Introduction to Probability. 2nd ed. United States: CRC Press.

Bradford, C. (2008a). On Weighted Density. Available from: https://www.austinzoning.typepad.com/austincontrarian/2008/03/weighted-densit.html [Last accessed on 2022 Oct 14].

Bradford, C. (2008b). Density Calculations for U.S. Urbanized Areas, Weighted by Census Tract. Available from: https://www.austinzoning.typepad.com/austincontrarian/2008/03/weighted-densit.html [Last accessed on 2022 Oct 14].

Bradford, C. (2008c). The Association between Density and Mode of Commute. Available from: https://www.austinzoning.typepad.com/austincontrarian/2008/09/the-association-between-density-and-mode [Last accessed on 2022 Oct 14].

Cowell, F.A. (2009). Measuring Inequality. LSE Perspectives in Economic Analysis. United Kingdom: Oxford University Press. p.241. Available from: https://darp.lse.ac.uk/papersdb/Cowell_measuringinequality3.pdf [Last accessed on 2022 Nov 9].

Craig, J. (1975). Population Density and Concentration in Great Britain 1931, 1951 and 1961. Studies on Medical and Population Subjects No 35: Office of Population Censuses and Surveys. London: HMSO.

Craig, J. (1979). Population density: Changes and patterns. *Population Trends,* 17: 12-16.

Craig, J. (1980). Population Density and Concentration in Great Britain 1951, 1961 and 1971. Studies on Medical and Population Subjects No 42: Office of Population Censuses and Surveys. London: HMSO.

Craig, J. (1984). Averaging Population Density. *Demography,* 21(3): 405-412.

https://doi.org/10.2307/2061168

Cramer, J.S. (1962). The Ownership of Major Consumer Durables. Cambridge, United Kingdom: Cambridge University Press.

Dijkstra, L., & Poelman, H. (2014). A Harmonised Definition of Cities and Rural Areas: The New Degree of Urbanisation. Working Papers in 01/2014, European Commission General Directorate for Regional and Urban Policy. Available from: https://ec.europa.eu/regional_policy/sources/docgener/work/2014_01_new_urban.pdf [Last accessed on 2022 Oct 14].

Eidlin, E. (2010). What density doesn't tell us about sprawl. *Access Magazine*, 37 (Fall): 2-9.

Eurostat. (2019). Methodological Manual on Territorial Typologies, 2018 edition. Available from: https://ec.europa.eu/eurostat/fr/web/products-manuals-and-guidelines/-/KS-GQ-18-008 [Last accessed on 2022 Oct 14].

Faberman, R.J., & Freedman, M. (2016). The urban density premium across establishments. *Journal of Urban Economics*, 93:71-84.

https://doi.org/10.1016/j.jue.2016.03.006

Florida, R. (2012). America's Truly Densest Metros. Available from: https://www.bloomberg.com/news/articles/2012-10-15/america-s-truly-densest-metro-areas [Last accessed on 2022 Oct 20].

Fujita, M., Krugman, P.R., & Venables, A.P. (1999). The Spatial Economy: Cities, Regions and International Trade. United States: MIT Press.

Fujita, M., & Thisse, J.F. (2002). The Economics of Agglomeration: Cities, Industrial Location and Regional Growth. Cambridge, United Kingdom: Cambridge University Press.

Gini, C. (1912). On variability and mutability. Reprinted in E. Pizetti and T. Salvemini (eds.) (1955) *Italy's Annals of Statistical Methodology*. Roma: Libreria Eredi Virgilio Veschi, [In Italian].

Gionanni, G., & Gubbiotti, S. (2015). On Corrado Gini's 1932 paper "About the concentration curve": A selection of translated excerpts. Metron 73(1):1-24. Available from https://www.researchgate.net/publication/273515333 [Last accessed on 2022 Oct 14].

https://doi.org/10.1007/s40300-015-0062-7

Graham, D.J., & Gibbons, S. (2019). Quantifying wider economic impacts of agglomeration for transport appraisal: Existing evidence and future directions. *Economics of Transportation* 19: 100121

https://doi.org/10.1016/j.ecotra.2019.100121

Hanberry, B.B. (2022). Imposing consistent global definitions of urban populations with gridded population density models: Irreconcilable differences at the National Scale. *Landscape and Urban Planning* 226, 104493.

https://doi.org/10.1016/j.landurbplan.2022.104493

Hansen, W.G. (1959). How accessibility shapes land use. *Journal of the American Institute of planners*, 25(2): 73-76.

https://doi.org/10.1080/01944365908978307

Huang, G., & Brown, P.E. (2021). Population-weighted exposure to air pollution and COVID-19 incidence in Germany. *Spatial Statistics* 41:100480

https://doi.org/10.1016/j.spasta.2020.100480

INSEE. (2021a). "*The Observatory of territories*" 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 [Last accessed on 2022 Jan 19].

INSEE. (2021b). Dataset of Communal Populations Up To 2019. Available from: https://www.insee.fr/fr/statistiques/fichier/3698339/base-pop-historiques-1876-2019.xlsx [Last accessed on 2022 Jan 19].

Kleinrock, L. (1975). Queuing Systems. vol. 1. New Jersey: Wiley Interscience. p. 169-173.

Koenig, J.G. (1974). An economic theory of urban accessibility. France's Economic review 25(2):275-297. Available from; http://www.persee.fr/web/revues/home/prescript/article/reco_0035-2764_1974_num_25_2_408144 [Last accessed on 2022 Oct 14] [Article in French].

Koenig, J.G. (1980). Indicators of urban accessibility: Theory and application. *Transportation,* 9(2): 145-172.

https://doi.org/10.1007/BF00167128

Krugman, P.R. (1997). Development, Geography, and Economic Theory. United States: MIT Press.

Krugman, P. (2013). Density. The Conscience of a Liberal. New York: New York Times (website). Available from: https://krugman.blogs.nytimes.com/2013/04/16/density [Last accessed on 2022 Oct 20].

Lee, S., & Lee, B. (2014). The influence of urban form on GHG emissions in the U.S. household sector. *Energy Policy*, 68: 534-549.

https://doi.org/10.1016/j.enpol.2014.01.024

Leurent, F., Benezech, V., & Combes, F. (2012). A stochastic model of passenger generalized time along a transit line. *Procedia Social and Behavioral Sciences*, 54: 785-797.

https://doi.org/10.1016/j.sbspro.2012.09.795

Leurent, F., Pivano, C., & Poulhes, 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.

https://doi.org/10.1016/j.trpro.2017.12.076

Lewontin, R.C., & Levins, R. (1989). On the characterization of density and resource availability. *The American Naturalist* 134(4): 513-524.

Lorenz, M.O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 9(70): 209-219.

Lu, Y., Zhang, Y.Y., & Ma, K. (2022). The effect of population density on the suitability of biomass energy development. *Sustainable Cities and Society* 87: 104240.

https://doi.org/10.1016/j.scs.2022.104240

NENAD (2021). Available from: https://gmnenad.com/2021/05/analysis-of-real-population-density-per-countries [Last accessed on 2022 Oct 14].

Ottensmann, J.R. (2016). Density of Large Urban Areas in the U.S., 1950-2010. Working Paper. p. 25. Available from: https://urbanpatternsblog.files.wordpress.com/2016/12/density-of-large-urban-areas.pdf [Last accessed on 2022 Nov 09].

Ottensmann, J.R. (2018a). On Population-Weighted Density. Working Paper. p.33. Available from: https://urbanpatternsblog.files.wordpress.com/2018/02/on-population-weighted-density.pdf [Last accessed on 2022 Nov 09].

Ottensmann, J.R. (2018b). An Alternative Approach to the Measurement of Urban Sprawl. Working Paper. p. 39.

Pascoal, R., & Rocha, H. (2022). Population density impact on COVID-19 mortality rate: A multifractal analysis using French data. *Physica A*, 593: 126979.

https://doi.org/10.1016/j.physa.2022.126979

Poulit, J. (1974) *On land-use and transport: Criteria for accessibility and urban development*. SETRA – Urban Division, French Ministry of Infrastructures [Book in French].

Rappaport, J. (2008a). A productivity model of city crowdedness. Journal of Urban Economics, 63(2):715-722.

https://doi.org/10.1016/j.jue.2007.04.008

Rappaport, J. (2008b). Consumption amenities and city population density. *Regional Science and Urban Economics*, 38/6: 533-552.

https://doi.org/10.1016/j.regsciurbeco.2008.02.001

Stairs, R.A. (1977a). The concept of population density: A suggestion. *Demography,* 14: 243-244.

https://doi.org/10.2307/2060580

Stairs, R.A. (1977b). Erratum to: The concept of population density: A suggestion. *Demography*, 14(3): 378.

https://doi.org/10.2307/2060795

Townsend, C., & Ellis-Young, M. (2018). Urban population density and freeways in North America: A Re-assessment. *Journal of Transport Geography*, 73: 75-83.

https://doi.org/10.1016/j.jtrangeo.2018.10.008

Vorobyev, A.N. (2019). The mapping of population density in a sparsely populated region: A case study of the Irkutsk Region. *IOP Conference Series Earth Environment Science*, 381: 012095.

https://doi:10.1088/1755-1315/381/1/012095

Wardrop, J.G. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3): 325-362.

https://doi.org/10.1680/ipeds.1954.11628

Wikipedia. (2022). Available from: https://en.wikipedia.org/wiki/Population_density [Last accessed on 2022 Oct 20].

Wonnacott, T.H., & Wonnacott, R.J. (1990). *Introductory Statistics*. 5th ed. New Jersey: Wiley.

Appendix

Appendix A: Nomenclature

z zone, a spatial entity in set Z covering the territory under study

A_z, ground area of zone z

P_z, population in zone z

A_z, overall ground area of territory

P_z, overall population in territory

a_1, unit ground area

o· a land unit

P_o, population in o

$\bar{O}$, total number of land units (equal to A_Z / a_1)

$\bar{U}$, total number of people in territory (equal to P_Z,)

x, density level

f_O & F_O, PDF & CDF of x regarding land units, with mean $\bar{x}_O$ and relative dispersion γ_O

f_U & F_U, PDF & CDF of x regarding people, with mean $\bar{x}_U$ and relative dispersion γ_U

L, Lorenz function

G, Gini index

Appendix B: Consumption model

A consumption model can be stated in a generic fashion as follows. It relies on a consumption function say $c : x \mapsto c(x)$, which takes nonnegative real values and measures the amount of consumption made by an individual with attribute x.

Let then 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 f_U that satisfies the following relation:

$$f_U(x) \propto c(x).f_O(x) \quad \text{(B-1)}$$

Postulating that the consumption function is monotonous, then eqn. (B-1) can be demonstrated using the same proof as for Equation (13). The proportionality coefficient is the reciprocal of $\bar{c}_O \equiv \int c(x) f_O(x) dx$. Thus

$$f_U(x) = \frac{1}{\bar{c}_O} c(x).f_O(x) \quad \text{(B-2)}$$

Appendix C: Log-normal distributions and their basic properties

The log-normal distribution

The log-normal distribution is especially well-suited to consumption models of two kinds: Power laws, on the first hand (e.g., Cowell, 2009), 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:

$$c(x) = c_1.x^r \quad \text{(C-1)}$$

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

Basic properties of log-normal distributions

Let us recall the definition of a unidimensional 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, that is, $\ln(X) \sim \mathrm{N}(m,s^2)$. Denoting as Φ the CDF of a reduced Gaussian variable and $\varphi(t) \equiv \exp(-t^2/2)/\sqrt{2\pi}$ the associated PDF, and letting $t_x \equiv (\ln(x) - m)/s$, the following outcomes are derived successively in a straightforward way (e.g., Cowell, 2009):

$$F_O(x) = \Phi(t_x)$$

$$F_O^{(-1)}(\alpha) = \exp(m + s.\Phi^{(-1)}(\alpha))$$

$$f_O(x) = \frac{1}{s.x}\varphi(t_x)$$

$$E_O[x] = \exp(m + \frac{1}{2}s^2)$$

$$V_O[x] = (E_O[x])^2 (\exp(s^2) - 1)$$

$$\gamma_O = \sqrt{\exp(s^2) - 1}$$

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

Furthermore, any derived random variable $Y \equiv c_1.X^r$ with $c_1 > 0$ is a log-normal variable, too. This is because $Y \geq 0$ and $\ln(Y) = \ln(c_1) + r.\ln(X)$, implying that $\ln(Y) \sim \mathrm{N}(\ln(c_1) + r.m, (rs)^2)$, making Y an LN variable with parameters $\ln(c_1) + r.m$ and $(rs)^2$.

Moment formulas for log-normal distributions

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:

$$\int_a^b x^r d\mathrm{F}_{\mathrm{O}}(x) = e^{r(m+rs^2/2)} \left\{ \begin{matrix} \Phi\left(\dfrac{\ln(b)-m}{s} - rs \right) \\ -\Phi(\dfrac{\ln(a)-m}{s} - rs) \end{matrix} \right\} \quad \text{(C-2)}$$

An immediate consequence is that

$$\mathrm{E}_{\mathrm{O}}\left[x^r\right] = \exp(rm + \frac{1}{2}r^2 s^2)$$

Proof of (C-2). 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 $\mathrm{F}_{\mathrm{U}}(x) = \Phi(\frac{1}{s}(\ln(x)-m)-rs)$, i.e., that, in the population of consumed units, level x is distributed $\mathrm{LN}(m+rs^2, s^2)$.

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

The case of r = 1

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

It is then easy to obtain

$$\frac{\bar{x}_{\mathrm{U}}}{\bar{x}_{\mathrm{O}}} = 1 + \gamma_{\mathrm{O}}^2 = \exp(s^2) \quad \text{(C-3)}$$

On Stairs' generalized population density

If the offered density is distributed $x_{\mathrm{O}} \sim \mathrm{LN}(m, s^2)$ then we have that

$\mathrm{E}_{\mathrm{O}}\left[x^k\right] = \exp(km + \frac{1}{2}k^2 s^2)$ and similarly

$$\mathrm{E}_{\mathrm{O}}\left[x^{k+r}\right] = \exp((k+r)m + \frac{1}{2}(k+r)^2 s^2)$$

Yielding that

$$\frac{\mathrm{E}_{\mathrm{O}}\left[x^{k+r}\right]}{\mathrm{E}_{\mathrm{O}}\left[x^k\right]} = \exp(rm + \frac{1}{2}(r^2 + 2rk)s^2)$$

And in turn, a Stairs' generalized density indicator of

$$x_k^r \equiv (\frac{\mathrm{E}_{\mathrm{O}}\left[x^{k+r}\right]}{\mathrm{E}_{\mathrm{O}}\left[x^k\right]})^{1/r} = \exp(m + \frac{1}{2}s^2(r+2k))$$

$$= \mathrm{M}_{\mathrm{O}}(\bar{x}_{\mathrm{U}} / \bar{x}_{\mathrm{O}})^{k+r/2}$$

since $\exp(m)$ is the median M_{O} of x_{O} and $s^2 = 1 + \gamma_{\mathrm{O}}^2$.

Lorenz curve and Gini index

Here, the Lorenz function, $\mathrm{L} \equiv \mathrm{F}_{\mathrm{U}} \circ \mathrm{F}_{\mathrm{O}}^{(-1)}$, involves $\mathrm{F}_{\mathrm{O}}^{(-1)}(\alpha) = \exp(m + s.\Phi^{(-1)}(\alpha))$ together with $\mathrm{F}_{\mathrm{U}}(x) = \Phi(\frac{1}{s}(\ln(x)-m)-s)$. It thus is a function of α parameterized by *s*:

$$\mathrm{L}_s(\alpha) = \Phi(\Phi^{(-1)}(\alpha) - s) \quad \text{(C-4)}$$

The Gini index, $G_s \equiv 2\int_0^1(\alpha - \mathrm{L}_s(\alpha))d\alpha$, can be considered as a function of *s*. It holds that

$$\mathrm{G}(s) = 2\Phi\left(\frac{s}{\sqrt{2}}\right) - 1 \quad \text{(C-5)}$$

Proof of (C-5). At point *s* = 0, as $\Phi \circ \Phi^{(-1)}(\alpha) = \alpha$, then $\mathrm{G}_0 = 2\int_0^1(\alpha - \alpha)d\alpha = 0 \cdot$

Differentiating G_s with respect to *s*, we get that:

$$\dot{\mathrm{G}}(s) \equiv \frac{d\mathrm{G}_s}{ds} = 2\int_0^1 \varphi(\Phi^{(-1)}(\alpha) - s)d\alpha$$

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

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

Rearranging

$$(t-s)^2 + t^2 = 2t^2 - 2ts + s^2 = 2\left(t - \frac{1}{2}s\right)^2 + \frac{1}{2}s^2$$

$$= (\sqrt{2}t - \frac{s}{\sqrt{2}})^2 + (\frac{s}{\sqrt{2}})^2$$

It follows that $\varphi(t-s).\varphi(t)=\varphi\left(\sqrt{2}t-\frac{s}{\sqrt{2}}\right).\varphi(\frac{s}{\sqrt{2}})$ and in turn that

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

$$=\sqrt{2}\varphi\left(\frac{s}{\sqrt{2}}\right)\int_{-\infty}^{+\infty}\varphi(u)\,du$$

$$=\sqrt{2}\varphi(\frac{s}{\sqrt{2}})$$

By integration,

$$\int_0^s\dot{G}(v)\,dv=\sqrt{2}\int_0^s\varphi\left(\frac{v}{\sqrt{2}}\right)dv=2\int_0^{\frac{s}{\sqrt{2}}}\varphi(w)\,dw$$

$$=2\left[\Phi\left(\frac{s}{\sqrt{2}}\right)-\Phi(0)\right]$$

Lastly, formula (C-5) is obtained on making use of: $G(s)=G_0+\int_0^s\dot{G}(v)\,dv\cdot$

Appendix D: Linear-log model of used density CDF

Here, it is assumed that, on a range $[A,B]$ of x, the used density CDF, F_u, is an affine linear function of ln x: thus, for $x\in[A,B]$,

$$F_U(x)=F_U(A)+\frac{F_U(B)-F_U(A)}{\ln B-\ln A}(\ln x-\ln A). \quad \text{(D-1)}$$

Differentiating with respect to x, the used density PDF is obtained as a hyperbolic function:

$$f_U(x)=\frac{1}{x}\frac{F_U(B)-F_U(A)}{\ln B-\ln A} \text{ for } x\in[A,B].$$

From the general relationship (14), the offered density is determined as

$$f_O(x)=\frac{\bar{x}_O}{x^2}\frac{F_U(B)-F_U(A)}{\ln B-\ln A} \text{ for } x\in[A,B].$$

From this stem the offered density CDF on the range:

$$F_O(x)=F_O(A)+(\frac{1}{A}-\frac{1}{x})\bar{x}_O\frac{F_U(B)-F_U(A)}{\ln B-\ln A} \quad \text{(D-2)}$$

On the [A,B] range, formulas (D-1) and (D-2) enable one to recover the quantiles of x_U and x_O.

For $\alpha\in[F_U(A),F_U(B)]$, the quantile $x_U^{[\alpha]}$ of x_U at order α satisfies that

$$\frac{\alpha-F_U(A)}{F_U(B)-F_U(A)}=\frac{\ln x-\ln A}{\ln B-\ln A}, \text{ therefore}$$

$$x_U^{[\alpha]}=A(\frac{B}{A})^{\frac{\alpha-F_U(A)}{F_U(B)-F_U(A)}}$$

The quantile $x_O^{[\alpha]}$ of x_O at order $\alpha\in[F_O(A),F_O(B)]$ satisfies that

$$\frac{\alpha-F_O(A)}{F_U(B)-F_U(A)}=\left(\frac{1}{A}-\frac{1}{x}\right)\frac{\bar{x}_O}{\ln B-\ln A}, \text{ therefore}$$

$$x_O^{[\alpha]}=1/\left(\frac{1}{A}-\frac{\ln B-\ln A}{\bar{x}_O}\frac{\alpha-F_O(A)}{F_U(B)-F_U(A)}\right)$$

If the linear-log assumption is valid on the full range of x_U, then $F_U(A)=0$ and $F_U(B)=1$, yielding that

$$x_U^{[\alpha]}=A(\frac{B}{A})^{\alpha}=(A)^{1-\alpha}(B)^{\alpha} \quad \text{(D-3)}$$

$$\bar{x}_U=\frac{B-A}{\ln B-\ln A}$$

From $F_O(A)=0$ and $F_O(B)=1$, $\bar{x}_O=\dfrac{\ln B-\ln A}{\frac{1}{A}-\frac{1}{B}}$

$$\frac{1}{x_O^{[\alpha]}}=\frac{1-\alpha}{A}+\frac{\alpha}{B}.$$

Applying (D-3) to order $\alpha=\frac{1}{2}$, the median M_U of x_U satisfies $M_U=\sqrt{AB}$. Applying it again to orders ¾ and ¼, the interquartile ratio RIQ_U satisfies $RIQ_U=\frac{x_U^{[0.75]}}{x_U^{[0.25]}}=\sqrt{B/A}$.

From the empirical M_U and RIQ_U, the A and B parameters for France as of 2019 are recovered as

$$A=\frac{M_U}{RIQ_U}=24 \text{ p/km}^2,$$

$$B=M_U RIQ_U=14{,}895 \text{ p/km}^2.$$