SENSITIVITY ASPECTS OF INEQUALITY MEASURES

by

L. EGGHE, LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium[*]
and   UIA, Universiteitsplein 1, B-2610 Wilrijk, Belgium

I. INTRODUCTION

It is intuitively easy to understand what is meant by inequality measures. That is because applications occur in very sensitive situations : rich vs. poor persons or countries, high productivity vs. low productivity (in various aspects : economic, scientific, biological, ...). Inequality applies to virtually every aspect of human and non human life. It is, therefore, not surprising that one wishes to measure "degrees" of inequality. Such inequality measures (also called concentration measures) have the property that, the larger the value, the more unequal (uneven, concentrated) the situation is. Usually these measures are "normalized" so that they range between 0 and 1, but this is not necessary.

Sometimes one is not interested in the degree of inequality but in the opposite aspect : how even is a situation, e.g. in biology where one is interested in the diversity of species that are available. Large concentrations of some species imply the rare (or non) occurrence of other spaces and this is considered as an impoverishment of the

situation. However, a special theory on evenness measures or diversity measures is only needed of one wants to study specific problems in specific domains of the sciences (e.g. in biodiversity studies). General aspects of these evenness measures indeed follow immediately from the analogue ones of concentration measures. Indeed, let f be a concentration measure bounded by e.g. 0 and 1. Then it is intuitively clear that

$$g = 1 - f \tag{1}$$

is an evenness measure (e.g. the most concentrated situation gives f = 1 and hence g = 0 and the most even situation gives f = 0 and hence g = 1).

Therefore we will limit ourselves mainly to the study of concentration measures, also because in computer science and technology and in information theory, it is far more important to study inequality than to study equality. The intuitive reason for this is clear : the more unequal a situation is the better we can exploit this in order to optimise certain situations. Let us give an example : the more unequal symbols or words occur in a "text" the better this text can be compressed by giving the shorter codes (e.g. binary codes) to these symbols or words that occur most. An example of application is the Huffman compression technique, which we will discuss briefly in the context of entropy.

Unequal situations are often encountered in information sciences : only a few scientists are prolific (publish a lot) but many publish little; a scientific discipline has only a few "top journals" while many journals occasionally publish in this direction. In linguistics, computer sciences and information sciences one has the law of Zipf stating that the occurrence of words in a text (the tokens as one says in linguistics) is inversely proportional to the rank of this word ("type" in linguistics) (rank according to the occurrence of this word in the text).

Unequal situations can be explained by studying probabilistic properties such as the so-called "Success Breeds Success" principle that states (simplifying) that prolific sources have a higher probability to produce a new item than less prolific ones. In

this article we do not go into these probabilistic aspects : we limit our attention to the problem "how to measure inequality".

For more on these probabilistic aspects of inequality we refer the reader to [1], [2] and references therein.

II. PRINCIPLES OF INEQUALITY

A. Basic Principles

In the sequel we will use the terminology of sources that produce items (e.g. words and their uses in texts, authors producing publications, etc...). We want to measure the degree of inequality i.e. the degree of the differences in production quantities of these sources. In this connection one could mention the so-called 80/20 rule which states : 20 % of the most prolific sources produce 80 % of all items. Of course, in general other numbers might occur. It is clear that a 90/10 result gives a more concentrated situation than a 80/20 result while a 70/25 result is more even. Not directly comparable with 80/20 is e.g. a 90/30 result. We will go into this further on.

These "generalized" 80/20 rules will be encountered in the sequel as an application of our more general framework in which we want to produce a single number that measures concentration (rather than two as above).

Let us fix some notation. Let N be the total number of sources. For source $i \in \{1,...,N\}$, denote by x_i its production $(x_i \in \mathbb{R}^+)$. It is clear that a concentration measure must be a function of the N-tuple $(x_1,...,x_N)$, preferably with positive values.

$$f : (x_1,...,x_N) \rightarrow f(x_1,...,x_N) \in \mathbb{R}^+ .$$

We will assume (although it is not strictly necessary) that the values of f are bounded, i.e. min f and max f exist in $\mathbb{R}$. Most commonly these are 0 and 1 and it will be clear

from the sequel that any good bounded concentration measure can be transformed into one where min f = 0 and max f = 1. These situations must occur in the following cases. Let X = (x,0,...,0), i.e. one source (here the first one) produces everything (x $\in$ $\mathbb{R}^+$), then f(X) = 1. In the other extreme, if Y = (y,y,...,y) (y $\in$ $\mathbb{R}^+$), i.e. all sources produce the same quantity y, then f(Y) = 0. We will call this principle (Z) :

$$f(x,0,...,0) = 1 \qquad (2)$$

(or, more generally, max f(X), where the maximum is taken over all X = $(x_1,...,x_N)$ such that $\sum_{i=1}^{N} x_i = x$)
and

$$f(y,y,...,y) = 0 \qquad (3)$$

(or, more generally, min f(X), where the minimum is taken over all X = $(x_1,...,x_N)$ for which $\sum_{i=1}^{N} x_i = Ny$, hence $\overline{X} = y$, where $\overline{X}$ denotes the average of the values $x_1,...,x_N$).

The next principle is called the "Permutation Invariance" and is denoted by (P). It states that concentration is not a labeled property, i.e. the concentration value is not changed if e.g. names of persons are interchanged. Formulated exactly we state : for every $(x_1,...,x_N)$ and every permutation π

$$\pi : \{1,2,...,N\} \rightarrow \{1,2,...,N\}$$

of the numbers 1,...,N, we have that

$$f(x_1,...,x_N) = f(x_{\pi(1)},...,x_{\pi(N)}) . \qquad (4)$$

Concentration measures should be "normalized" w.r.t. the total production $\sum_{i=1}^{N} x_i$. Indeed, we need e.g. to compare the concentration of words in texts and this comparison may not be dependent on the size of the text : independent on the size of the database one should indeed be able to determine the optimal compression technique. It is even so that the size of a database is not always known or at least

cannot be predicted e.g. when considering the growth of such databases. An example from econometrics also illustrates this principle very clearly : concentration must be independent from the units, e.g. when measuring wealth, the value of f must be independent of the used currency (U.S. Dollars, yen, Pound, Euro, ...). This principle is called "Scale Invariance", (S) : for every $(x_1,...,x_N)$ and every $c > 0$,

$$f(cx_1,...,cx_N) = f(x_1,...,x_N) .$$ (5)

Last but not least, at the real heart of concentration theory, it should be a fact that "if one takes away from the poor and gives it to the rich, inequality (concentration must increase". For evident reasons this principle is called the "Transfer Principle", (T) : for every $(x_1,...,x_N)$ and if $i,j \in \{1,...,N\}$ are such that $x_i \leq x_j$, and if $0 < h \leq x_i$ then

$$f(x_1,...,x_i-h,...,x_j+h,...,x_N) > f(x_1,...,x_i,...,x_j,...,x_N)$$ (6)

(h is the amount that is taken away from the "poorer" source i and given to the "richer" source j). This is a very natural principle. It also includes several other ones such as

Principle (R) : if the richest gets richer, inequality increases : for every $(x_1,...,x_N)$, if $x_i = \max \{x_1,...,x_N\}$ and if there exists a $k \neq i$ such that $x_k \neq 0$ then, for every $h > 0$

$$f(x_1,...,x_i+h,...,x_N) > f(x_1,...,x_i,...,x_N) .$$ (7)

Principle (N) : the principle of the nominal increase : for every $(x_1,...,x_N)$ where not all x_i are equal, and every $h > 0$:

$$f(x_1+h,...,x_N+h) < f(x_1,...,x_N) .$$ (8)

All these principles are in accordance with our intuitive feeling about inequality. Of course, independent of all these priniciples, one is not yet sure that such functions f exist. That they exist (abundantly) will be clear in the sequel. As said $(T) \Rightarrow (R)$ and $(T) \Rightarrow (N)$. Therefore we only consider the set of principles :

$$(B) = \{(Z),(P),(S),(T)\} \tag{9}$$

and consider (B) as a minimum requirement for a good concentration measure. Further refinements of principle (T) will lead to finer requirements for concentratioin measures and hence to a reduction of their quantity. This will be the main purpose of this survey article : feeling the sensitivity of concentration measures w.r.t. diverse transfer principles such as (T).

B. The Lorenz Curve and Lorenz Order

The Lorenz curve and order are two of the most important findings in contemporary mathematics with basic applications in concentration theory. The concept is simple and beautiful. Let us consider one situation $X = (x_1,...,x_N)$: N sources where source i has a production of x_i items. Consider the "normalized" vector

$$A_X = (a_1,...,a_N)$$

where, for every $i = 1,...,N$
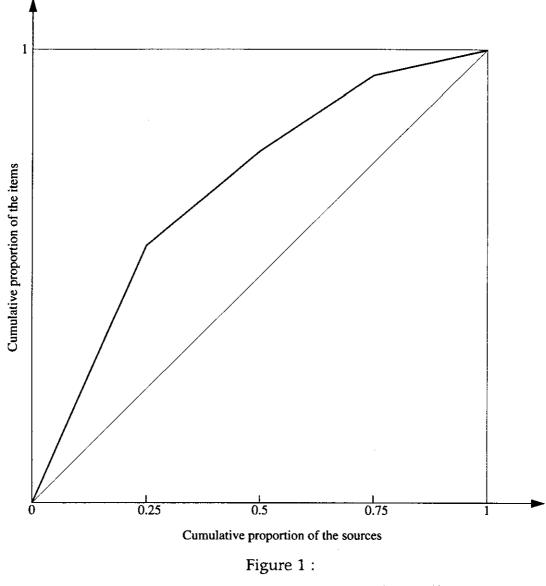
$$a_i = \frac{x_i}{\sum_{k=1}^{N} x_k} . \tag{10}$$

Then consider the set of points

$$\left(\frac{i}{N}, \sum_{j=1}^{i} a_j \right) \tag{11}$$

i.e. the cumulative fraction of the items in the first i sources. Let us suppose that X is decreasing (if it is not a rearrangement of the coordinates in X will yield a decreasing vector, still called X). In this case, by interconnecting the points in (11), and linking the first point with (0,0) yields a concavely increasing polygonal curve. For example, Fig.1 shows the Lorenz curve for the vector X = (1,4,3,10), rearranged as to yield a decreasing vector (10,4,3,1). We could, of course, also consider increasing

vectors. In this case our Lorenz curve is convexely increasing, being the mirror image of the first one over the diagonal (connecting (0,0) and (1,1)) of the unit square.

Figure 1 :

The Lorenz curve of the vector X = (1,4,3,10)

That the Lorenz curve has something to offer to concentration theory is already clear from its very definition. It contains all applicable forms of the so-called 80/20-rule. Indeed, every vertical line indicates an abscis x which corresponds with an ordinate y given by the Lorenz curve. Hence a 100 y/100 x-rule is obtained and this for every value of x ∈ [0,1].

In *informetrics* one has the so-called Bradford curves (cf.[1]). They are nothing else than the Lorenz curves but where the fractions are replaced by their counts (i.e. i versus $\sum_{j=1}^{i} x_j$) and where the ordinates are given in a logarithmic scale. The use of these so-called semi-logarithmic scales yielded additional information on the type of bibliography (here sources are journals and items are the articles in these journals) under study. For more on this we refer the reader to [1].

We are now in a position to introduce the Lorenz order between two vectors $X = (x_1,...,x_N)$ and $X' = (x'_1,...,x'_N)$. We suppose that both X and X' are decreasing and that

$$\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i' \ .$$

Then we say that X' dominates X, in notation $X \prec X'$ or $X' \succ X$ if

$$\sum_{i=1}^{j} x_i \leq \sum_{i=1}^{j} x_i' \qquad (12)$$

for every $j = 1,...,N\text{-}1$. Since

$$\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i'$$

(12) is also valid for $j = N$. Because of this equality and by (10) we have now that X $\prec$ X' if and only if the Lorenz curve of X is always below the one of X', i.e. the Lorenz curve of X never intersects the one of X'. The order expressed by $\prec$ (a partial order) is called the Lorenz order.

The Lorenz order is very important in mathematics e.g. in the study of convex function inequalities, geometric inequalities, combinatorial analysis and matrix theory. For more on this see [3]. The importance for the topic discussed in this article will become apparent from the next definition and theorem.

Let f be a function as above. We say that f satisfies principle (L) if it satisfies principles (Z), (P) and (S) and whenever $X = (x_1,...,x_N) \prec X' = (x'_1,...,x'_N)$ and $X \neq X'$ we have

$$f(X) < f(X') . \tag{13}$$

Functions for which $X \prec X'$ implies $f(X) \leq f(X')$ are sometimes called Schur convex functions.

Theorem : (L) is equivalent with (B).

In words, the preservation by f of the Lorenz dominance order is equivalent with the transfer principle. The transfer principle was hinted at by Pigou ([4]) in 1912 but was first exactly described by Dalton in 1920 ([5]). The Lorenz curve was introduced in [6] in 1905. The Lorenz order was introduced (as notation and terminology) by Hardy, Littlewood and Polya in [7]. It is therefore a bit surprising to state that the above theorem was essentillay known (of course in other terminology) by Muirhead in 1903, see [8] and also [9].

The above theorem offers a geometric "visualisation" of what is required of a good concentration measure. It summarizes in one graph and one notation $(\prec)$ of what more or less concentration is all about.

Of course all the definitions and results mentioned so far are empty if we cannot give concrete examples of good inequality measures.

III. CONCENTRATION MEASURES

Let $X = (x_1,...,x_N)$ be an arbitrary vector.

A. The coefficient of variation V

This is the most well-known concentration measure. This measure V is defined as the quotient of the standard deviation σ of X and of the average (mean) μ of X :

$$V = \frac{\sigma}{\mu} \, . \tag{14}$$

Recall that

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

and that

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \mu^2 \, .$$

B. The Yule characteristic K

This measure is derived from V :

$$K = \frac{V^2}{N} \, . \tag{15}$$

The measure is used in linguistics.

C. Pratt's measure and the Gini index

We suppose here that X is decreasing.

Denote

$$q = \sum_{i=1}^{N} i \, a_i$$

with a_i as in (10). Then Preatt's measure C is defined as

$$C = \frac{2\left(\frac{N+1}{2} - q\right)}{N - 1} \qquad (16)$$

and Gini's index by

$$G = \frac{N-1}{N} C \; . \qquad (17)$$

It must, however, be remarked that Gini's index has been introduced in 1909 in econometrics (cf. [10]), long before Pratt's measure was, namely in 1977 (cf. [11]), the latter one in informetrics. Originally, the Gini index G is defined as twice the area between the Lorenz curve and the straight line joining (0,0) and (1,1) (i.e. the Lorenz curve of a uniform distribution). That they are essentially the same (cf. (17) with N usually very high) was only seen in 1979 - see [12]. This is a typical example of "re-inventing the wheel", not so surprising with this highly interdisciplinary subject!

The measures V, C and G are generalized as follows. Define, for $r \geq 1$

$$P(r) = \frac{\left(\frac{1}{2N(N-1)} \sum_{i,j=1}^{N} |x_i - x_j|^r\right)^{\frac{1}{r}}}{\mu} \; , \qquad (18)$$

the so-called generalized Pratt measure, essentially introduced in [13]. It can be shown (cf. [14]) that

$$P(1) = C \qquad \text{and} \qquad P(2) = \sqrt{\frac{N}{N-1}} \, V \; .$$

D. Theil's measures Th and L

These measures were introduced in 1967 in [15] as follows

$$Th = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i}{\mu}\right) \ell n \left(\frac{x_i}{\mu}\right) \tag{19}$$

and

$$L = -\left(\ell n \ N + \frac{1}{N} \sum_{i=1}^{N} \ell n \ (a_i)\right) \tag{19'}$$

again with a_i as in (10).

We close this non-exhaustive list by

E. The Atkinson Indices A(e)

For $0 < e < 1$, define

$$A(e) = 1 - \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i}{\mu}\right)^{1-e}\right]^{\frac{1}{1-e}} \tag{20}$$

and for $e = 1$

$$A(1) = 1 - \frac{\sqrt[N]{x_1 \cdots x_N}}{\mu} \ . \tag{21}$$

cf. [16]. As is clear from the above, most of these measures originate from econometrics or sociometrics.

We refer to [14] for the proofs that all these measures satisfy principle (L), hence are to be considered as good concentration measures.

Note 1 :

The Theil measures are also called Theil's entropy measures. Indeed they resemble the well-known average entropy formula - known in information theory -

$$\overline{H} = - \sum_{i=1}^{N} a_i \log_2 a_i \ .$$

(22)

Stated in this form, $\overline{H}$ is a valuable measure but is not a concentration measure but a measure of evenness. Of course, upon normalization and the application of the transformation (1) we obtain a concentration measure. Don't be mistaken : the measures Th and L $\underline{are}$ concentration measures.

Note 2 :

So far we have always supposed that N is fixed. This, of course, excludes dynamic studies of concentration (or evenness or diversity) since, if time is variable, N is variable too in most cases. Especially in biology the variability of N has an implication on possibilities of studying diversity of species. We assume that the N-variability is less important in computer science and technology as well as in informetrics : there one can still compare concentration values of e.g. databases of different sizes and draw professional conclusions based on these comparisons. An example of this is given in the next section.

N-dependent theories can be developed as follows. It is clear that it is possible to draw a Lorenz curve for any vector $X = (x_1,...,x_N)$ and any N. Now, we generalize the Lorenz order, introduced earlier, cf. (12) by stating that the N-vector X dominates the N'-vector X' if the Lorenz curve of X' lies never above the Lorenz curve of X. This will be denoted as $X' \prec\!\!\prec X$. This requirement generalizes the partial order introduced for fixed N.

A good concentration measure should respect this generalized Lorenz order. This means that such a concentration measure f should satisfy the following relation :

$$X' \neq X \quad \text{and} \quad X' \prec\!\!\prec X \rightarrow f(X') < f(X)$$

(23)

Related to this, we introduce the cell-replication axiom (introduced by Dalton [5]). First, we define the mathematical operator REPEAT_c with c a non-zero natural number, as

$$\text{REPEAT}_c(x_1,...,x_N) = (x_1,...,x_1,x_2,...,x_2,...,x_N,...,x_N) \tag{24}$$

where every x_i is repeated c times. Then a function f satisfies the cell-replication axiom if

$$f(\text{REPEAT}_c(X)) = f(X) . \tag{25}$$

Note that the Lorenz curve of X and of $\text{REPEAT}_c(x)$ coincide, so that this is a very natural requirement.

It can be shown [17] that if a function is a good concentration measure for fixed N, and, moreover, satisfies the cell-replication axiom, then it satisfies relation (23).

Now, the coefficient of variation, V, the Gini index, G, Theil's measures Th and L, and all Atkinson indices A(e) satisfy this additional requirement. Hence, they can be used to compare situations where the number of sources is different. The Yule characteristic, K and Pratt's measure, C, on the other hand, do not satisfy this additional axiom.

Finally, we note that it is possible to introduce other requirements to compare situations with a different number of sources. These requirements take the number of sources explicitly into account [17].

IV. Some Applications in Computer Science, in Information Science and in Linguistics

A. Huffman compression of codes

It is an obvious fact that the occurrence of words in texts is very dependent on the words : some words occur very often (e.g. the, a, of, for, ...) for grammatical reasons and some only occur a few times. It is then logical (and it is so in most cases) that the words that occur more often should be the shortest ones. This is often the case in natural langauge, although not in an optimal way. It is for this reason that one often uses abbreviations, acronyms etc. We do not go further into this.

In the same way, codes for symbols should be shorter if the symbols occur more often than others. Let us give an example with an alphabet of only eight symbols A, B, C, D, E, F, G, H. As is always the case in practise, we can also here suppose that these symbols do not occur the same number of times. Suppose we have already ordered the symbols in decreasing order of occurrence. Suppose their fraction of occurrence is as follows : A : 0.338, B : 0.203, C : 0.135, D : 0.088, E : 0.081, F : 0.061, G : 0.054, H : 0.040.

For the sake of simplicity we only look at binary coding of these eight symbols but any n-ary coding can be studied in the same way. It is clear that binary codes of length 3 suffice, since $2^3 = 8$. We could then code these eight symbols as follows : A = $\phi\phi\phi$, B = $\phi\phi1$, C = $\phi1\phi$, D = $\phi11$, E = $1\phi\phi$, F = $1\phi1$, G = 11ϕ, H = 111. Hence the average length per symbol is 3. There is however the following theorem.

Theorem (Huffman [18]) :
There exists a non-fixed length decodable coding which is optimal in the sense that no shorter average length per symbol is possible. The average length per symbol of this coding is $\overline{H}$ as given by formula (22).

This coding is called the Huffman coding. In the above example one can calculate (using $a_1 = 0.338$, $a_2 = 0.203$, and so on, and $N = 8$ of course) that $\overline{H} = 2.65 < 3$. One cannot give the "shortest" binary coding $A = \phi$, $B = 1$, $C = \phi\phi$, $D = \phi 1$, $E = 1\phi$, $F = 11$, $G = 1\phi\phi$, $H = 1\phi 1$ since this coding is not decodable. The (up to a permutation of ϕ and 1) unique solution is $A = \phi\phi$, $B = 1\phi$, $C = \phi 11$, $D = 111$, $E = \phi 1\phi\phi$, $F = \phi 1\phi 1$, $G = 11\phi\phi$, $H = 11\phi 1$, with average length per symbol equal to 2.7, very close to 2.65 and a serious improvement (i.e. compression) w.r.t. the fixed length coding with length 3. There is a simple algorithm to find this solution. The reader is referred to [19] for this as well as for much more on compression of texts. Note that the Huffman coding is decodable since no code is the beginning of a larger code. So is it clear that the text $\phi 111111111\phi 11\phi\phi\phi\phi 11\phi 111\phi\phi$ is uniquely decodable as CDDCAAHG.

B. Zipf's law and Mandelbrot's law

One can ask the question : how do we obtain the probabilities (fractions) for the occurrence of the symbols as in the above example. More generally, how can we know the fraction of occurrence of the different words or letters in a text? The general answer is that this can be done statistically by sampling and then calculate confidence intervals for these sampled fractions. We must warn the reader that calculating confidence intervals for a large number of fractions is requiring very high sample sizes (cf. [20]). If this cannot be executed we can put all our hopes to known distributions of these multinomial fractions. They will give a good approximation of the concrete situation of the text and hence (also since the Huffman technique is very stable w.r.t. the initial input parameters, i.e. the fractions of occurrence) will yield an optimal or at least a quasi-optimal compression of the text. We will discuss two "classical" distributions.

1. The law of Zipf

The law of Zipf ([21], [22]) is a distribution that is encountered in linguistics. It states that if one orders the words in a text in decreasing order of their occurrence

in this text, then the number of occurrence is inversely proportional to its rank. In mathematical notation, if r denotes the rank of the word and y(r) the fraction of the number of times this word occurs in the text, then

$$y(r) = \frac{C}{r} \tag{26}$$

where C is a constant for the text. It is clear that formula (26) expresses the unequal occurrence of words in the text. In [23] I calculated the measure of Pratt, expressing the degree of concentration and found, e.g. for a text of 10,000 words a value of C = 0.80.

2. The law of Mandelbrot

Even in linguistics it is clear that the law of Zipf is not always valid. Often one finds better approximations (of the data derived from the text) by using an extra parameter, e.g. r^a instead of r in formula (26). Of a different nature is the law of Mandelbrot ([24], [25]). Stated more generally in terms of sources and items it can be formulated as : if we rank the sources in decreasing order of the number of items they produce and if r denotes the rank of the source and y(r) its fraction of the total production, then

$$y(r) = \frac{A}{1 + Br} \tag{27}$$

where A and B are constants. This law is encountered very often in informetrics, e.g. when describing inequality in publication patterns by authors or journals in a certain field. For this law one finds the value C = 0.61 for Pratt's measure of inequality (again for 10,000 sources).

We see that this value is much lower than the one found for Zipf's law. Hence Zipf's law expresses a more concentrated situation than Mandelbrot's. This result is somewhat surprising but can be explained - philosophically - as follows. In all situations where Zipf's or Mandelbrot's law applies we clearly deal with an unequal situation : few sources produce a lot of items and many sources produce not many

items. But in cases where Mandelbrot's law applies, the most prolific sources refrain a bit their high production : this is so in the case of author-publication patterns - here the most prolific authors will only publish their best work; this is also the case in the journal-article pattern - the more important a journal becomes in a certain discipline the more strict (by applying heavy refereeing procedures) it will be in accepting papers for publication. This is not so in e.g. texts (where Zipf's law usually applies) : the most heavily used words are continued to be used in that pace, no matter how long the text is, due to grammatical reasons. This creates the more concentrated situation, which is not a bad result from the point of view of compression of the text : use of many small words and possibly of abbreviations (e.g. prof., Dr., ...) of heavily used words that are somewhat longer.

In the same way one also finds the following result on the generalized 80/20 rule : the higher the average number of items per source, the smaller fraction of the most prolific sources one needs in order to have a fixed cumulative fraction of the items. This has e.g. applications in libraries : the higher the average number of circulations per book (say per year) in a library, the smaller the fraction of the most heavily used books is needed to cover say 95 % of all circulations. In other words (as is the case in public libraries as compared to scientific libraries) the weeding problem is easier in such libraries since there a larger fraction (now of the least used books) can be taken that accounts for a very low fixed percentage of the circulations. In general one could also say that the higher the average number of items per source, the higher the Lorenz curve is situated in the unit square.

In the next section we will study some refinements of the transfer principle (T) as it is the basis in concentration theory. Most of these refinements go in the direction of the study of sensitivity properties of concentration measures w.r.t. transfers.

V. OTHER TRANSFER PRINCIPLES AND SENSITIVITY WITH RESPECT TO TRANSFERS

The classical transfer principle (T) deals with "taking away from one poor and giving it to one rich". Of course, as is also reflected in the theorem in section II.B, consecutive compositions of bilateral transfers are possible, yielding the same inequality for our concentration measure f. But real life is more complicated! It can be that - to stay within the appealing econometric terminology - so many transfers happen, giving an overall bonus to the richer sources but in which there are also bonuses for some poor sources. The next transfer principle, stronger than (T) goes in this direction.

A. The Transfer principles E(p)

It is immediately verified that the change of $X = (x_1,...,x_i,...,x_j,...,x_N)$ into $X' = (x_1,...,x_i-h,...,x_j+h,...,x_N)$, where $x_i \le x_j$ and where $0 < h \le x_i$ (as in the definition of the transfer principle (T)), implies that the variance of the vector X' is strictly larger than the variance of the vector X. Hence a good generalization of principle (T), hereby possibly affecting all sources, is as follows (cf. [9]).

Let f be as always. We say that it satisfies the principle E(p) $(p \ge 1)$ if for every $X = (x_1,...,x_N)$ and $X' = (x'_1,...,x'_N)$ such that

$$\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i'$$

and such that

$$\sum_{i,j=1}^{N} |x_i - x_j|^p < \sum_{i,j=1}^{N} |x_i' - x_j'|^p \tag{28}$$

we have that

$$f(x_1,...,x_N) < f(x'_1,...,x'_N) \quad . \tag{29}$$

Note that

$$\frac{1}{2\,N^2} \sum_{i,j=1}^{N} |x_i - x_j|^p \tag{30}$$

is the p-variance of the vector X and also that it is equal to σ^2 if $p = 2$. The above definition can even be extended by replacing the function $x \to |x|^p$ by any convex function. In [9] it is shown that $E(p) \Rightarrow (T)$ for every $p \geq 1$, hence we have at our disposition a stronger transfer principle. Hence the concentration measures encountered so far, which also satisfy a principle $E(p)$ ($p \geq 1$) are stronger (better) measures than those which satisfy (T) but no $E(p)$.

Obviously V and K satisfy E(2) and in [14] it was shown that C and G satisfy E(1) (in essence because $P(1) = C$, with $P(1)$ as in (18) with $r = 1$). Obviously $P(r)$ satisfies $E(r)$ for evey $r \geq 1$. That any $E(p)$ is strictly stronger than (T) follows from the fact that Th nor A(e) ($0 < e < 1$) satisfy any $E(p)$, while they all satisfy (T). It is shown in [26] that the validity of one $E(p)$ ($p \geq 1$) excludes the validity of any $E(q)$ ($q \geq 1$, q $\neq$ p) at least if $N \geq 3$, which is always the case in practise. It is also so for $N = 3$ except for $p = 2$, $q = 4$ since here $E(2) = E(4)$. Since E(2) relates to the classical variance we could consider it to be the most important $E(p) =$ principle and hence, we can consider V, K or other from V derived measures to be the most important concentration measures. Also its formulation $V = \sigma/\mu$ is very simple. The principle E(2) is equivalent with the following desirable transfer principle. Let $X = (x_1,...,x_N)$ and $X' = (x'_1,...,x'_N)$ be such that $x'_i = x_i + h_i$ ($i = 1,...,N$) with

$$\sum_{i=1}^{N} h_i = 0$$

and such that

$$\sum_{i=1}^{N} h_i(x_i + x'_i) \geq 0 \quad . \tag{31}$$

If $X \neq X'$, then

$$f(x_1,...,x_N) < f(x_1',...,x_N') \quad .$$

The proof can be found in [9].

B. Sensitivity Aspects of Inequality Measures with respect to Transfers

It is very difficult to express what is meant by "sensitivity" of inequality measures. It is a kind of "second order" aspect, describing how dependent a measure is w.r.t. a transfer (apart from the fact that we have a strict inequality)[*]. Let us go back to the original definition of the transfer principle (T). It states that, when $x_i \leq x_j$ and when $0 < h \leq x_i$,

$$f(x_1,...,x_i,...,x_j,...,x_N) < f(x_1,...,x_i-h,...,x_j+h,...,x_N) \quad . \tag{32}$$

In this formulation it is in no way stated how the difference

$$\Delta f = f(x_1,...,x_i-h,...,x_j+h,...,x_N) - f(x_1,...,x_N)$$

is dependent on x_i or i or x_j or j or even $x_j - x_i$ or $j - i$, etc. (dependence on the indices presupposes, of course, a certain ordering of the x_i's). Discussions around this theme in economemtrics can be found in [16] and for sociometrics in [13]. In [27], Allison pleads for measures that are sensitive for differences (such as $x_j - x_i$). This plea is given in the context of publications and citations and indeed it is our feeling that sensitivity must be studied within the context of the application.

In [14] one has found that ΔV^2 is linearly dependent on $x_j - x_i$ but independent of i or j. Dependence of the latter is found for ΔG and ΔC, while, not surprisingly, ΔTh is dependent on $\ell n(x_j) - \ell n(x_i)$.

---

[*] We do not deal with statistical sensitivity (e.g. the dependence of the sensitivity of statistical methods w.r.t. the sample size).

We will now look at (T) in another way. It is an easy consequence of (T) that (by applying (32) N-1 times) (supposing $x_1 = \max_{i=1,\dots n} x_i$ - using property (P) this can always be realized)

$$f(x_1,\dots,x_N) < f\left(x_1+h,\ x_2 - \frac{h}{N-1},\dots,\ x_N - \frac{h}{N-1}\right) \tag{33}$$

if $0 < h < \min_{i=1,\dots,N} (N-1)\, x_i$.

This principle is related to moonlighting : one source (here the first one) gets richer, e.g. by not paying taxes, which is detremental to all other citizens. In scientometrics we have an example whereby the most important research group in a certain domain is recognized by a subsidizing authority (such as a government) thereby donating an important amount of money to this group, which is then not available any more for all the other groups.

Now it is clear that inequality (33) is rather artificial : what happens if the second largest source grows at the cost of all the others, and so on? It is an intuitive feeling that if x_1 is "above average production", then (33) should be valid. What average can we use? This will become clear in the sequel. Let us fix some notation. Let $X = (x_1,x_2,\dots;x_N)$ rearranged in such a way that the transfer h occurs at the first source (not necessarily anymore the source with maximum production). Denote

$$Y = \left(1,\ -\frac{1}{N-1},\ -\frac{1}{N-1},\dots,\ -\frac{1}{N-1}\right) \tag{34}$$

where there are N-1 coordinates with value - 1/(N-1). Then (33) is equivalent with

$$f(X+hY) - f(X) > 0\ . \tag{35}$$

This only compares two situations : X (before the transfer) and X' = X+hY (after the transfer). In these cases a total transfer of h > 0 is involved and inequalities should change if h < 0 and furthermore, any small |h| can be used. This boils down to calculate (for every "small" h)

$$\frac{f(X+hY) - f(X)}{h} \tag{36}$$

and examining the sign of this number, now irrespective of the sign of h. This leads to the use of the directional derivative

$$f'(X;Y) = \lim_{h \to 0} \frac{f(X+hY) - f(X)}{h}$$

$$= \sum_{i=1}^{N} \frac{\partial f}{\partial x_i} (X) \, y_i \, ,$$

where $Y = (y_1,...,y_N)$ as in (34). Hence

$$f'(X;Y) = \frac{\partial f}{\partial x_1} (X) - \frac{1}{N-1} \sum_{i=2}^{N} \frac{\partial f}{\partial x_i} (X) . \tag{37}$$

This leads us to several new transfer principles (which will lead us to the notion of sensitivity) : Let M(X) denote any average (e.g. arithmetic, geometric, harmonic, median, ...). We say that f satisfies transfer principle (T_M) if $x_1 > M(X)$ implies

$$f'(X;Y) > 0 \tag{38}$$

This transfer principle has been introduced in [28]. Also in [28] it is noted that the value of f'(X;Y) is a measure for the sensitivity of f w.r.t. transfers. Note that the sign of f'(X;Y) changes for evenness measures and hence the value of f'(X;Y) (or rather |f'(X;Y)|) can also be used for evenness measures. Here, however, we restrict ourselves to concentration measures.

Let us fix some notations for the different averages of a vector $X = (x_1,...,x_N)$.

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{39}$$

the arithmetic mean,

$$g = \sqrt[N]{x_1 \ldots x_N} \tag{40}$$

the geometric mean,

$$h = \frac{N}{\sum\limits_{i=1}^{N} \dfrac{1}{x_i}} \tag{41}$$

the harmonic mean, and by m the median, i.e. $x_{\tilde{m}}$ *for which*

$$\oint_{\tilde{m}}^{m} = \frac{N+1}{2} \ . \tag{42}$$

(and in case N is even, take the average of the two x_i *- values)*

In these cases we can talk of the transfer principles (T_μ), (T_g), (T_h) and (T_m).

In [28] the following theorem has been proved, based on (34) and (35).

Theorem :

$$(1) \qquad V'(X;Y) = \frac{\dfrac{1}{\mu \sqrt{N}}}{\sqrt{\sum\limits_{i=1}^{N} (x_i - \mu)^2}} \left[(x_1 - \mu) - \sum\limits_{i=2}^{N} \frac{(x_i - \mu)}{N-1} \right] \ . \tag{43}$$

Hence V satisfies (T_μ). Hence also K satisfies (T_μ).

$$(2) \qquad Th'(X;Y) = \frac{1}{N\mu} \left[\left(\ell n \left(\frac{x_1}{\mu} \right) + \mu \right) - \frac{1}{N-1} \sum\limits_{i=2}^{N} \left(\ell n \left(\frac{x_i}{\mu} \right) + \mu \right) \right] \ . \tag{44}$$

Hence Th satisfies (T_g).

$$(3) \qquad L'(X;Y) = -\frac{1}{N}\left(\frac{1}{x_1} - \frac{1}{N-1}\sum_{i=2}^{N}\frac{1}{x_i}\right) . \qquad\qquad (45)$$

Hence L satisfies (T_h). Also A(1) satisfies (T_h).

$$(4) \qquad G'(X;Y) = -\frac{2(N+1-i)}{\mu N^2} + \frac{2}{\mu N^2(N-1)}\sum_{j\ne i}(N+1-j) . \qquad (46)$$

Hence G satisfies (T_m). Hence also C satisfies (T_m).

The respective values of $f'(X;Y)$ for $f = V$, Th, L or G give the sensitivity of these measures w.r.t. transfers. To the best of our knowledge, in [28] it is the first time that there has been found a connection between these classical measures of concentration and these different averages of a vector $X = (x_1,...,x_N)$. In [28] some other concentration measures have been linked with some other (T_M) transfer properties (for generalized averages M).

This theory of sensitivity w.r.t. transfers is exact in the sense that it uses a clear definition $(f'(X;Y) > 0)$ involving a derivative, which is expected for the study of sensitivity. Yet it allows for differences in nature of the diverse concentration measures (w.r.t. the used average). The appearance of the four basic averages (arithmetic, geometric, harmonic and the median) is remarkable. Overall, this sensitivity theory can be considered as a mathematical solution to some problems raised (with partial solutions) in [13], [16] and [27].

VI. SUMMARY AND CONCLUSION

This review paper was dealing with various ways of describing inequality (or concentration) as is occurring in virtually every aspect of life. The emphasis here is on computer science, linguistics and information science. We underline the importance of unequal situations (as opposed to evenness studies as e.g. in biology) in these domains, e.g. in applications such as compression of databases and texts.

The Lorenz curve is introduced that gives a visual method of studying concentration. The relation with the so-called 80/20-rule is given. We also give basic principles ("axioms") that good concentration measures should satisfy, amongst which the appealing transfer principle. The paper then continues by giving examples of good concentration measures such as V (the coefficient of variation), Pratt's measure C and the Gini index G and their generalizations, Theil's measures Th and L and, finally, the Atkinson indices A(e) (0 < e < 1). In this connection, the entropy formula from information theory is described and applied to aspects of compression (cf. the Huffman compression of codes). Also the concentration aspects of the laws of Zipf and Mandelbrot are discussed. Some remarks are made about the N-dependence of these theories (i.e. the dependence on the number of sources).

The rest of the paper deals with sensitivity of concentration measures, w.r.t. transfers. New transfer principles, stronger than the classical one are introduced and we indicate which concentration measures satisfy which generalized transfer principles. A nice class of transfer principles can be defined as follows. Let f be a concentration measure and denote by $f'(X;Y)$ the directional derivative in X (in the direction Y). Let

$$Y = \left(1, -\frac{1}{N-1}, ..., -\frac{1}{N-1}\right) \in \mathbb{R}^N \quad .$$

Then $x_1 >$ a certain average of X should lead to $f'(X;Y) > 0$. This is the case for V with the arithmetic average, for Th with the geometric average, for L with the harmonic

average and for G and C for the median of X. The values of f'(X;Y) can be used as a measure of sensitivity of the measure f w.r.t. transfers.

The merit of this sensitivity theory for concentration measures is that it presents a clear and exact definition of what sensitivity means, yet it allows for differences in nature of the diverse concentration measures (as revealed by the use of four different averages of X).

REFERENCES

[1] EGGHE, L. and ROUSSEAU, R.

Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science, Elsevier, Amsterdam, 1990.

[2] EGGHE, L. and ROUSSEAU, R.

Mathl. Comput. Modelling, 23 : 93-104 (1996).

[3] MARSHALL, A.W. and OLKIN, I.

Inequalities : Theory of Majorization and its Applications, Academic Press, New York, 1979.

[4] PIGOU, L.

Wealth and Welfare, Macmillan, New York, 1912.

[5] DALTON, H.

Econom. J., 30 : 348-361 (1920).

[6] LORENZ, M.O.

J. Amer. Stat. Assoc., 9 : 209-219 (1905).

[7] HARDY, G.H.,LITTLEWOOD, J.E. and POLYA, G.

Inequalities, 1st ed., 2nd. ed., Cambridge University Press, Cambridge, 1934, 1952.

[8] MUIRHEAD, R.F.

Proc. Edinburgh Math. Soc., 21 : 144-157 (1903).

[9] EGGHE, L. and ROUSSEAU, R.

J. Amer. Soc. Inf. Sci., 42 : 479-489 (1991).

[10] GINI, C.

Giornale degli Economisti, 11 : 37 (1909).

[11] PRATT, A.D.

J. Amer. Soc. Inf. Sci., 28 : 285-292 (1977).

[12] CARPENTER, M.P.

J. Amer. Soc. Inf. Sci., 30 : 108-110 (1979).

[13] ALLISON, P.D.

Amer. Sociol. Rev., 43 : 865-880 (1978).

[14] EGGHE, L. and ROUSSEAU, R.

Elements of Concentration Theory. In : Informetrics 89/90 (L. Egghe and R. Rousseau, eds.), Elsevier, 1990, pp.97-137.

[15] THEIL, H.

Economics and Information Theory, North-Holland, Amsterdam, 1967.

[16] ATKINSON, A.B.

J. Econ. Theory, 30 : 244-263 (1970).

[17] ROUSSEAU, R.

Belgian J. of Operations Res., Stat. and Comp. Sci., 32 : 99-126 (1992).

[18] HUFFMAN, D.A.

Proc. of the IRE, 40 : 1098-1101 (1952).

[19] STORER, J.A.

Data Compression. Methods and Theory, Computer Science Press, Rockville, 1988.

[20] EGGHE, L. and VERAVERBEKE, N.

Int. J. of Scientometrics and Informetrics, 1 : 183-193 (1995).

[21] ZIPF, G.K.

Selected Studies of the Principle of relative Frequency in Language, Harvard University Press, Cambridge, 1932.

[22] ZIPF, G.K.

The Psycho-Biology of Language : an Introduction to dynamic Philology, Houghton Mifflin, 1935. Reprinted by MIT-Press, Cambridge, 1965.

[23] EGGHE, L.

J. Amer. Soc. Inf. Sci., 38 : 288-297 (1987).

[24] MANDELBROT, B.

Word, 10 : 1-27 (1954).

[25] MANDELBROT, B.

The fractal Geometry of Nature, Freeman, New York, 1977.

[26] EGGHE, L.

Scientometrics, 25 : 167-191 (1992).

[27] ALLISON, P.D.

Soc. Stud. of Sci., 10 : 163-179 (1980).

[28] EGGHE, L. and ROUSSEAU, R.

Inf. Proc. and Manag., 31 : 511-523, 1995.

SUGGESTIONS FOR BASIC READINGS

ALLISON, P.D.

Measures of Inequality. American Sociological Review, 43 : 865-880 (1978).

ARNOLD, B.C.

Majorization and the Lorenz Order : a Brief Introduction, Springer-Verlag, Berlin, 1987.

EGGHE, L. and ROUSSEAU, R.

Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science, Elsevier, Amsterdam, 1990.

GINI, C.

Memorie di Metodologia Statistica, Vol.1 : Variabilità e Concentrazione, Libreria Eredi Virgilio Veschi, Rome, 1955.

HARDY, G.H., LITTLEWOOD, J.E. an,d POLYA, G.

Inequalities, 1st ed., 2nd ed., Cambridge University Press, Cambridge, 1934, 1952.

LAMBERT, P.J.

The Distribution and Redistribution of Income : a Mathematical Analysis; sec.ed., Manchester University Press, Manchester, 1993.

MAGURRAN, A.

Ecological Diversity and its Measurement, Chapman and Hall, London, 1991.

MARSHALL, A.W. and OLKIN, I.

Inequalities : Theory of Majorization and its Applications, Academic Press, New York, 1979.

SEN, A.K.

On Economic Inequality, Clarendon Press, Oxford, 1973.

STORER, J.A.

Data Compression. Methods and Theory, Computer Science Press, Rockville, 1988.

THEIL, H.

Economics and Information Theory, North-Holland, Amsterdam, 1967.

ZIPF, G.K.

Human behavior and the principle of least effort, Addison-Wesley, Cambridge, 1949. Reprinted by Hafner, New York, 1965.