Conjugate partitions in informetrics: Lorenz curves, h-type indices, Ferrers graphs and Durfee squares in a discrete and continuous setting

Peer-reviewed author version

EGGHE, Leo (2010) Conjugate partitions in informetrics: Lorenz curves, h-type indices, Ferrers graphs and Durfee squares in a discrete and continuous setting. In: JOURNAL OF INFORMETRICS, 4(3). p. 320-330.

DOI: 10.1016/j.joi.2010.01.006
Handle: http://hdl.handle.net/1942/10982

Conjugate partitions in informetrics: Lorenz curves, h-type indices, Ferrers graphs and Durfee squares in a discrete and continuous setting

by

L. Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium
leo.egghe@uhasselt.be

ABSTRACT

The well-known discrete theory of conjugate partitions, Ferrers graphs and Durfee squares is interpreted in informetrics. It is shown that partitions and their conjugates have the same h-index, a fact that is not true for the g- and R-index. A modification of Ferrers graph is presented, yielding the g-index.

We then present a formula for the Lorenz curve of the conjugate partition in function of the Lorenz curve of the original partition in the discrete setting.

Ferrers graphs, Durfee squares and conjugate partitions are then defined in the continuous setting where variables range over intervals. Conjugate partitions are nothing else than the

Key words and phrases: conjugate partition, informetrics, Lorenz curve, h-index, g-index, R-index, Ferrers graph, Durfee square.

inverses of rank-frequency functions in informetrics. Also here they have the same h-index and we can again give a formula for the Lorenz curve of the conjugate partition in function of the Lorenz curve of the original partition. Calculatory examples are given where these Lorenz curves are equal and where one Lorenz curve dominates the other one. We also prove that the Lorenz curve of a partition and the one of its conjugate can intersect on the open interval $]0,1[$.

I. Introduction

The attention of this author was drawn by the paper Anderson, Hankin and Killworth (2008) where, for partitions (to be explained below), Ferrers graphs and Durfee squares are used to define a variant of the h-index.

For the sake of completeness we will define these simple concepts here (see also Andrews (1998)). A partition is, simply, a vector of finite dimension T with decreasing coordinates which are elements of $\mathbb{N}$, the positive natural numbers (excluding 0). We will denote such a vector by $C = (c_1, c_2, ..., c_T)$ where $c_1 \geq c_2 \geq ... \geq c_T \geq 1$. The name partition comes from the fact that the natural number $\sum_{i=1}^{T} c_i$ is partitioned by C, whereby each coordinate $c_1, ..., c_T$ indicates the relative size of the coordinate $i = 1, ..., T$.

Informetrically one can interpret C as a description of the number of items c_i in the source i, $i = 1, ..., T =$ the total number of sources. In this interpretation,

$$A = \sum_{i=1}^{T} c_i \qquad (1)$$

is the total number of items in the system (also called an information production process (IPP) in which source i produces (or has) c_i items, $i = 1, ..., T$. In this interpretation, the vector C is nothing else than the rank-frequency function of this IPP. For more on IPPs we refer the reader to Egghe (2005) where many examples are given by interpreting C as an author or a journal and where the T articles of this author or journal are ranked in decreasing order of the number of citations c_i that these articles have received (e.g. until now), $i = 1,...,T$.

In the theory of partitions, there is a handy graphical representation of a partition as a so-called Ferrers graph. Taking $C = (8,5,4,4,3,1)$ as an example, the Ferrers graph of C is depicted as in Fig. 1

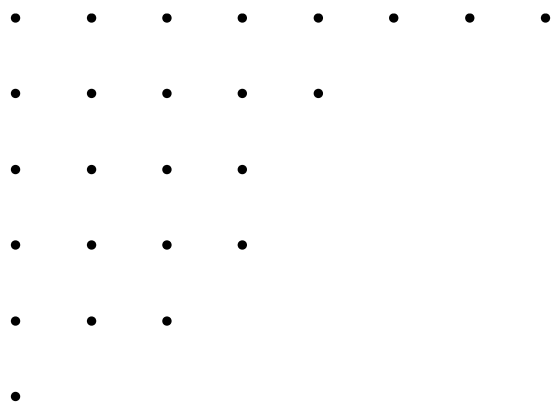

Fig. 1 Ferrers graph of $C = (8,5,4,4,3,1)$

The largest, fully filled, square in this graph (starting from the point on the first row and the first column) is called the Durfee square (after W.P. Durfee). The side of this Durfee square is nothing else than the h-index of this system (here $h = 4$). The h-index was introduced in Hirsch (2005) as the largest rank $r \in \{1,...,T\}$ such that $c_r \geq r$ (hence $c_h \geq h$ and hence $c_i \geq h$ for all $i = 1,...,h$ while $c_{h+1} < h+1$). For more on the use and the (dis)advantages of the h-index we refer to the vast literature and the extensive review Egghe (2008). That the side of the Durfee square has size h was first remarked in Anderson, Hankin and Killworth (2008).

In the theory of partitions one defines the notion of "conjugate" of a partition $C = (c_1,...,c_T)$, cf. Andrews (1998), Definition 1.8. We define the conjugate of partition C as $C' = (c_1',...,c_m')$ where for $j = 1,...,m$, c_j' equals the number of coordinates in C that are larger than or equal to j. In the example above: $C = (8,5,4,4,3,1)$ we see that its conjugate is $C' = (6,5,5,4,2,1,1,1)$. It is easily seen (cf. Andrews (1998)) that C' is obtained from C by mirroring the Ferrers graph of

C over the main diagonal (i.e. the line connecting the points for which the column number equals the row number) – see Fig. 2

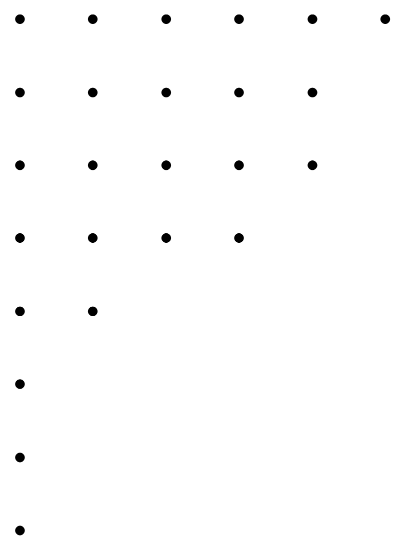

Fig. 2 Ferrers graph of C', the conjugate of C

Note that dim C' (the dimension of C') does not always equal dim C. Only if $c_1 =$ dim C we have that dim $C' = c_1$ and hence dim C = dim C'. Example: $C = (6,5,5,4,2,1)$. Now $C' = (6,5,4,4,3,1)$.

Note also that

$$\sum_{i=1}^{T} c_i = \sum_{j=1}^{m} c_j' \tag{2}$$

In the informetric terminology of sources and items in IPPs we can define the conjugate of the rank-frequency function C as the function yielding for every $j = 1,...,m$ the number of sources with $\geq j$ items.

Ferrers graphs yield a trivial proof of the following important proposition.

<u>Proposition 1</u>: The h-index of a partition equals the h-index of its conjugate.

Proof: In mirroring Ferrers graph of C over the main diagonal, the size of the Durfee square is not changed, hence it has the same side-length (hence h-index) as the one of C. Consequently $h_C = h_{C'}$. □

One can indeed verify in the above example that $h_C = h_{C'} = 4$. Note that Ferrers graphs are handy in finding the conjugate C' of a partition C as well as in the proof of Proposition 1.

The above proposition is false for the g-index and the R-index as we will show below but first a short introduction to the g-index and the R-index. The h-index has the disadvantage of not counting the actual number of citations to papers in the h-core (the h-core is defined as the set of the first h papers; the term is only a convenient definition but does not always constitute a real core set of papers for the author or the journal). Indeed, once a paper is in the h-core, it does not matter how many citations $(\geq h)$ this paper actually receives. This was remarked in Egghe (2006) where e.g. the example was given of equal h-indices for E. Garfield and F. Narin (early 2006), namely $h = 27$ while Garfield had 14 papers with more than 100 received citations and Narin had only 1 paper with more than 100 received citations (namely 112). Furthermore, Garfield had 1 paper with 625 received citations.

In order to improve the h-index for this (almost) insensibility to the actual number of received citations, we defined in Egghe (2006) the g-index as follows: in the same ranking as in $C = (c_1,...,c_T)$, $r = g$ is the highest ranking such that the first g papers together received at least g^2 citations. Equivalently : the g-index is the highest ranking $r = g$ such that the first g papers received, on average, at least g citations. Obviously $g \geq h$ since the first h papers received, together, at least $h.h = h^2$ citations and since $r = g$ is the largest rank with this property. The more discriminatory power of the g-index was illustrated by the fact that, while Garfield and Narin had the same h-index, the g-index of Garfield was 59 while the one of Narin was 40. The greater discriminatory power of the g-index above the h-index was also proved in Schreiber (2008a,b), Tol (2008) and Costas and Bordons (2008).

The R-index, introduced in Jin, Liang, Rousseau and Egghe (2007) serves a similar goal as the g-index and is defined as

$$R = \sqrt{\sum_{i=1}^{h} c_i} \tag{3}$$

, i.e. the square root of the sum of the received citations by the papers in the h-core (hence, the definition uses the h-index itself).

Consider the example $C = (8,6,6,5,1)$. Then $C' = (5,4,4,4,4,3,1,1)$, hence $h_c = h_{c'} = 4$ as predicted in Proposition 1. But $g_C = 5 \neq g_{C'} = 4$ and $R_C = \sqrt{8+6+6+5} = 5 \neq R_{C'} = \sqrt{5+4+4+4} = \sqrt{17}$. Here $g_C > g_{C'}$ and $R_C > R_{C'}$ but opposite inequalities are also possible. This follows from the fact that the conjugate of the conjugate of a partition C equals C: $(C')' = C$ as follows readily from the fact that Ferrers graph of the conjugate of C is the mirror of Ferrers graph of C over the main diagonal (cf. Fig. 2). Note that $R < g$ and $R > g$ are possible: for $(6,4,4,2,1)$ we have $h = 3$, $g = 4$, $R = \sqrt{14} < g$ while for $(13,4,4,2,1)$ we have $h = 3$, $g = 4$, $R = \sqrt{21} > g$.

We close this introduction by indicating how the g-index can be derived from a modified Ferrers graph. Let $C = (c_1,...,c_T)$ with Durfee square of side = h (cf. above). Enlarge this Durfee square to squares of side $h+1, h+2,...,g$ by using, in the Ferrers graph of C, the elements outside the Durfee square: $c_1 - h$ elements of the first row, $c_2 - h$ elements of the second row,…, $c_h - h$ elements of the h^{th} row, (if $g > h$) c_{h+1} elements of the $(h+1)^{th}$ row,…, until a number of elements $\leq c_g$ of the g^{th} row in the Ferrers graph of C. The newly found Ferrers graph has Durfee square of side size equal to g.

In the next section we will calculate the Lorenz-curve $L_{C'}$ of C' (the conjugate of C) in function of the Lorenz-curve L_C of C and the elements c_i. Examples are given.

In the third section we present a theory of conjugate partitions where r ranges in an interval $[0,T]$ instead of $r \in \{1,...,T\}$ for c_r. It will turn out that C is replaced by the rank-frequency function γ of a continuous IPP and that its conjugate is nothing else than γ^{-1}, the inverse of this function. Ferrers graphs and Durfee squares are discussed in this continuous setting. We also present a formula for the Lorenz-curve $L_{\gamma^{-1}}$ of the conjugate of γ in function of the Lorenz-curve L_γ of γ and of numbers $\gamma(r)$, $r \in [0,T]$. Practical calculations show the value of this formula for the easy calculation of $L_{\gamma^{-1}}$ and we will give examples showing that $L_\gamma = L_{\gamma^{-1}}$, $L_\gamma > L_{\gamma^{-1}}$ or $L_\gamma < L_{\gamma^{-1}}$ are possible (the inequality refers to the existence of a point in the open interval $]0,1[$ on which this inequality is valid). For definitions of Lorenz-curves (in the discrete as well as in the continuous setting) we refer to Egghe (2005a,b) or to the definitions given in the next two sections.

The paper closes with conclusions and some open problems.

II. The Lorenz-curve of the conjugate of a partition in function of the Lorenz-curve of the original partition

Let $C = (c_1,...,c_T)$ be the original partition: a decreasing vector of natural numbers (nonzero). The Lorenz-curve of C, denoted L_C maps the cumulative fraction of sources $\frac{i}{T}$ $(i = 1,...,T)$ onto the corresponding cumulative fraction of items: $\frac{\sum_{j=1}^{i} c_j}{A}$, where A is given by (1). So the Lorenz-curve connects (0,0) to (1,1) via the points

$$\left(\frac{i}{T}, \frac{\sum_{j=1}^{i} c_j}{A}\right).$$

Hence we have the concavely increasing function

$$L_C\left(\frac{i}{T}\right) = \frac{\sum_{j=1}^{i} c_j}{A} \tag{4}$$

(concave because C is decreasing). We refer to Egghe and Rousseau (1990), Egghe (2005a,b) for more (well-known) aspects of discrete Lorenz concentration theory.

To be able to construct the Lorenz-curve of the conjugate partition C' of C we must model the general form of the vector C' itself, in function of the elements in C itself. By the very definition of conjugate partition of C we have

$$C' = \left(\underbrace{T,...,T}_{c_T}, \underbrace{T-1,...,T-1}_{c_{T-1}-c_T}, \underbrace{T-2,...,T-2}_{c_{T-2}-c_{T-1}}, ..., \underbrace{1,...,1}_{c_1-c_2} \right) \qquad (5)$$

Note that this representation is also correct in case some $c_i - c_{i+1}$ are zero.

Denoting $C' = \left(c_1', ..., c_m'\right)$ we should have (cf. Andrews (1998)) that

$$\sum_{j=1}^{m} c_j' = \sum_{i=1}^{T} c_i = A \qquad (6)$$

This will be verified now. By (5) we have

$$\sum_{j=1}^{m} c_j' = Tc_T + (T-1)(c_{T-1} - c_T) + (T-2)(c_{T-2} - c_{T-1}) + (T-3)(c_{T-3} - c_{T-2}) + ... + c_1 - c_2$$

$$= Tc_T + Tc_{T-1} - Tc_T - c_{T-1} + c_T + Tc_{T-2} - Tc_{T-1} - 2c_{T-2} + 2c_{T-1} +$$

$$Tc_{T-3} - Tc_{T-2} - 3c_{T-3} + 3c_{T-2} + ... + c_1 - c_2$$

$$= \sum_{i=1}^{T} c_i = A$$

by (1). □

For the Lorenz-curve $L_{C'}$ of C' it is sufficient that we calculate $L_{C'}$ in the nods $\frac{j}{c_1}$ where

$$j = c_T + c_{T-1} - c_T + \ldots + c_{T-i+1} - c_{T-i+2}$$

$$j = c_{T-i+1}, \tag{7}$$

i.e. $\frac{j}{c_1}$ versus the corresponding cumulative number of coordinates in (5), divided, by (6), by A. Hence

$$L_{C'}\left(\frac{j}{c_1}\right) = L_{C'}\left(\frac{c_{T-i+1}}{c_1}\right)$$

$$= \frac{1}{A}\Big(Tc_T + (T-1)(c_{T-1} - c_T) + (T-2)(c_{T-2} - c_{T-1}) + (T-3)(c_{T-3} - c_{T-2}) + \ldots + (T-i+1)(c_{T-i+1} - c_{T-i+2})\Big)$$

which yields, after some calculation

$$L_{C'}\left(\frac{c_{T-i+1}}{c_1}\right) = \frac{1}{A}\left(c_T + c_{T-1} + \ldots + c_{T-i+2} + (T-i+1)c_{T-i+1}\right) \tag{8}$$

$$= \frac{1}{A}\left(\sum_{i=1}^{T} c_i - (c_1 + c_2 + \ldots + c_{T-i+1}) + (T-i+1)c_{T-i+1}\right) \tag{9}$$

$$L_{C'}\left(\frac{c_{T-i+1}}{c_1}\right) = 1 - L_C\left(\frac{T-i+1}{T}\right) + \frac{T-i+1}{A}c_{T-i+1} \tag{10}$$

for $i = 1,\ldots,T$. In decreasing order of the abscissa of $L_{C'}$, but in increasing order of the abscissa of L_C, we get

$$L_{C'}\left(\frac{c_i}{c_1}\right) = 1 - L_C\left(\frac{i}{T}\right) + \frac{i}{A}c_i \tag{11}$$

which looks a bit simpler. We have proved the following proposition:

Proposition 2:

Let L_C denote the Lorenz-curve of partition $C = (c_1,...,c_T)$ and $L_{C'}$ denote the Lorenz-curve of the conjugate C' of C, then, for all $i = 1,...,T$ we have

$$L_{C'}\left(\frac{c_i}{c_1}\right) = 1 - L_C\left(\frac{i}{T}\right) + \frac{i}{A}c_i$$

Note that, when i varies, the abscissa of L_C and $L_{C'}$ go in opposite direction (since C decreases). We can verify that, if $i = 1$, we have

$$L_{C'}(1) = 1 - L_C\left(\frac{1}{T}\right) + \frac{c_1}{A} = 1$$

by (4) and that, if $i = T$, we have

$$L_{C'}\left(\frac{c_T}{c_1}\right) = 1 - L_C(1) + \frac{Tc_T}{A}$$

$$= \frac{Tc_T}{A} \qquad (12)$$

which is ok according to (5). Note also that the number of nods of L_C is equal to the number of nods of $L_{C'}$ and is equal to the number of different values c_i in C, although dim C can be different from dim C'.

The valid remark (see above) that $(C')' = C$ yields in (11) that

$$L_C\left(\frac{i}{T}\right) = 1 - L_{C'}\left(\frac{c_i}{c_1}\right) + \frac{c_i}{A}i$$

which is the same equation as (11) (as it should).

We also note that, in the first nod of L_C we have that the slope of the line connecting this point $\left(\frac{1}{T}, L_C\left(\frac{1}{T}\right)\right)$ with $(0,0)$ equals (by (4))

$$\frac{L_C\left(\frac{1}{T}\right)}{\frac{1}{T}} = \frac{\frac{c_1}{A}}{T} = \frac{c_1 T}{A} \tag{13}$$

For the first nod of $L_{C'}$ we have that the slope of the line connecting this point $\left(\frac{c_T}{c_1}, L_{C'}\left(\frac{c_T}{c_1}\right)\right)$ with $(0,0)$ equals (by (12))

$$\frac{L_{C'}\left(\frac{c_T}{c_1}\right)}{\frac{c_T}{c_1}} = \frac{\frac{Tc_T}{A}}{\frac{c_T}{c_1}} = \frac{c_1 T}{A} \tag{14}$$

, hence equal to (13). This implies that, if $\frac{1}{T} < \frac{c_T}{c_1}$, for certain values $> \frac{1}{T}$, $L_{C'}$ is above L_C (since L_C has a nod in $\frac{1}{T}$) and, analogously, if $\frac{1}{T} > \frac{c_T}{c_1}$ that, for certain values $> \frac{c_T}{c_1}$, L_C is above $L_{C'}$ (since $L_{C'}$ has a nod in $\frac{c_T}{c_1}$, if $c_T \neq c_{T-1}$). Further arguments on $L_C \geq L_{C'}$ or $L_{C'} \geq L_C$ on $]0,1[$ are left as an open problem.

<u>Examples</u>:

1. $C = (3,2,1)$, $T = 3$, $A = 6$. Then $C' = (3,2,1) = C$. We can verify that $L_C = L_{C'}$:

   $L_C\left(\frac{1}{3}\right) = \frac{1}{2}$ and $L_C\left(\frac{2}{3}\right) = \frac{5}{6}$ and the same for C'. Using (11) we find

$$L_{C'}\left(\frac{c_2}{c_1}\right) = L_{C'}\left(\frac{2}{3}\right) = 1 - L_C\left(\frac{2}{3}\right) + \frac{2}{6}c_2$$

$$= \frac{5}{6}$$

and

$$L_{C'}\left(\frac{c_3}{c_1}\right) = L_{C'}\left(\frac{1}{3}\right) = 1 - L_C(1) + \frac{3}{6}c_3$$

$$= \frac{3}{6} = \frac{1}{2}$$

2. $C = (3,3,1)$, $T = 3$, $A = 7$, hence $C' = (3,2,2)$. We have $L_C\left(\frac{1}{3}\right) = \frac{3}{7}$, $L_C\left(\frac{2}{3}\right) = \frac{6}{7}$. Using $C' = (3,2,2)$ or (11) we find (there is only one nod for $L_{C'}$), e.g. via (11)

$$L_{C'}\left(\frac{c_3}{c_1}\right) = L_{C'}\left(\frac{1}{3}\right)$$

$$= 1 - L_C(1) + \frac{3}{7} = \frac{3}{7}$$

In $\frac{2}{3}$ as abscissa (which is not a nod for $L_{C'}$) we have $L_{C'}\left(\frac{2}{3}\right) = \frac{5}{7}$ (half way between $\frac{3}{7}$ and 1), hence here $L_C > L_{C'}$ since $L_C\left(\frac{2}{3}\right) = \frac{6}{7}$.

III. Theory of continuous partitions, their conjugates and Lorenz-curves

Since discrete partitions $C = (c_1,...,c_T)$ can be interepreted, informetrically, as $c_r =$ the number of items in the source on rank r $(r = 1,...,T)$ (where the c_r are ranked in decreasing order), we have an interpretation of C as a rank-order function. Rank-order functions of continuous variables are well-known in informetrics : it are strictly decreasing functions $\gamma : r \in [0,T] \to \gamma(r)$ where $\gamma(r)$ is the density of the items in the rank-density r. As in (1) we denote by A the total number of items, hence

$$\int_0^T \gamma(r)dr = A \qquad (15)$$

We suppose γ to be differentiable (hence also continuous) so that $\gamma(r)$ also ranges in an interval that we suppose (for reasons which will be given later) to start in 0 : let us denote this interval by $[0,m]$. This is almost the same as the rank-frequency functions described in Egghe (2005a) except that these functions range in an interval starting in 1 (this was needed for developing a Lotkaian informetrics theory); here we need the interval $[0,m]$ to develop the theory of the conjugate of γ (see further). Note that in Egghe (2005a) this function is denoted by g but, to avoid confusion with the in this paper used g-index, we use the notation γ .

Fig. 3 gives an example of such a rank-frequency function. In this graph we clearly see the continuous version of the Durfee square. If we turn this graph clockwise over 90° we obtain the continuous version of Ferrers graph (and again the Durfee square). As in the discrete case the side of this Durfee square is the h-index of γ

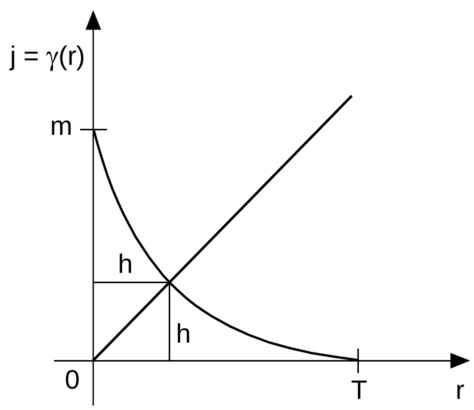

Fig. 3. Example of the function γ , Durfee square and the h-index.

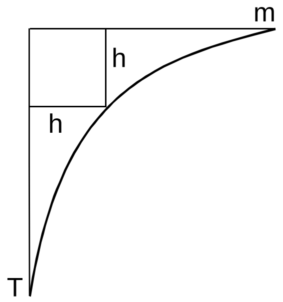

Fig. 4. Ferrers graph, Durfee square
and h-index of γ .

Now we will define the size-frequency function φ (denoted in Egghe (2005a) by f) as follows (see also Egghe and Rousseau (2003)) : for $j \in [0, m]$,

$$\varphi(j) = -\frac{1}{\gamma'(\gamma^{-1}(j))} \tag{16}$$

Note that γ^{-1}, the inverse function of γ, exists since γ is strictly decreasing.

Equivalently, we have, if $j = \gamma(r)$, $r \in [0, T]$, $j \in [0, m]$,

$$r = r(j) = \gamma^{-1}(j) = \int_j^m \varphi(j')dj' \tag{17}$$

From this it follows that, for $j = \gamma(r)$,

$$\int_0^r \gamma(r')dr' = \int_j^m j'\varphi(j')dj' \tag{18}$$

, being the total number of items in the sources on rank densities $r' \in [0, r]$. Formula (18) is an exercise in Egghe (2005a) (Exercise II.1.2.3, p. 110) for which the proof was not given. Because of this and for the sake of completeness, we present here the short proof.

Proof of formula (18) :

From formula (16) we have, for all $j \in [0, m]$:

$$-\frac{j}{\gamma'\left(\gamma^{-1}(j)\right)} = jf(j)$$

, hence

$$\int_j^m \frac{-j'dj'}{\gamma'\left(\gamma^{-1}(j')\right)} = \int_j^m j'f(j')dj' \tag{19}$$

For $j = \gamma(r)$ we have $r = \gamma^{-1}(j)$, $dj = \gamma'(r)dr$ and, since $\gamma(0) = m$ we have by (19) :

$$\int_j^m j'f(j')dj' = \int_{r'=r}^{r'=0} \frac{-\gamma(r')d\gamma(r)}{\gamma'\left(\gamma^{-1}(\gamma(r))\right)}$$

$$= \int_0^r \gamma(r')dr'. \qquad \square$$

So, this result is not different from the one in Egghe (2005a) where $j \geq 1$ while here $j \geq 0$.

In the discrete case, the conjugate of the partition $C = (c_1, ..., c_T)$ is, by definition, $C' = (c_1', ..., c_m')$, where for each $j = 1, ..., m$, c_j' equals the number of coordinates in C that are larger than or equal to j. Hence, as mentioned above, in the informetric terminology, the conjugate of the rank-frequency function C is "the function yielding for each $j = 1, ..., m$ the number of sources with $\geq j$ items.

Hence, in our continuous case, the conjugate of the rank-frequency function γ can only be the inverse function

$$\gamma^{-1} : j \in [0, m] \rightarrow \gamma^{-1}(j) = r \in [0, T] \tag{20}$$

For a conjugate partition, in the discrete case, we required that the sum of the coordinates should be the same as the sum of the coordinates of the original partition, denoted A (see formulae (1) and (2)). This should also be the case here (e.g. also for simplicity in the construction of the Lorenz-curves, to be introduced further on). It will now become clear why we assumed $j \in [0, m]$. We have the following proposition.

Proposition 3 :

$$\int_0^m \gamma^{-1}(j)dj = \int_0^T \gamma(r)dr = A \tag{21}$$

Proof : By formula (17) :

$$\int_0^m \gamma^{-1}(j)dj = \int_{j=0}^{j=m} \int_{j'=j}^{j'=m} \varphi(j')dj'dj$$

which is equal, by Fubini's theorem to

$$= \int_{j'=0}^{j'=m} \int_{j=0}^{j=j'} \varphi(j')djdj'$$

$$= \int_{j'=0}^{j'=m} j'\varphi(j')dj' = A$$

by (15) and (18) and the fact that $\gamma(T) = 0$. □

For γ^{-1} we can make similar graphs as for γ in Figs. 3 and 4. As in Figs. 1 and 2, Ferrers graph of γ^{-1} is a mirroring of the one of γ (Fig. 4) over the diagonal (as is the graph of γ^{-1} itself with respect to the one of γ (Fig. 3)).

Proposition 1 for discrete partitions is also valid in the continuous case.

Proposition 4 : The h-index of a partition equals the h-index of its conjugate.

Proof :
By the definition, the h-index is the unique rank $r = h$ for which $\gamma(h) = h$ (see Egghe and Rousseau (2006) for a proof of the existence of h). Obviously $h = \gamma^{-1}(h)$ and γ^{-1} is the conjugate partition of γ. □

Continuous Lorenz theory is described in Egghe (2005a,b). We will denote the Lorenz-curve of γ by L_γ and the one of its conjugate by $L_{\gamma^{-1}}$. By definition, we have for all $r \in [0,T]$ and all $j \in [0,m]$:

$$L_\gamma\left(\frac{r}{T}\right) = \frac{\int_0^r \gamma(r')dr'}{A} \tag{22}$$

$$L_{\gamma^{-1}}\left(\frac{j}{m}\right) = \frac{\int_0^j \gamma^{-1}(j')dj'}{A} \tag{23}$$

Note that the denominator in both cases is A because of Proposition 3. The relation between L_γ and $L_{\gamma^{-1}}$ is as given in Proposition 5.

Proposition 5 : For all $j = \gamma(r)$:

$$L_{\gamma^{-1}}\left(\frac{\gamma(r)}{m}\right) = L_{\gamma^{-1}}\left(\frac{j}{m}\right) = 1 - L_\gamma\left(\frac{r}{T}\right) + \frac{r\gamma(r)}{A} \tag{24}$$

Proof : Using (17) and (23) we have

$$L_{\gamma^{-1}}\left(\frac{j}{m}\right) = \frac{1}{A}\int_{j'=0}^{j'=j}\left(\int_{j''=j}^{j''=m}\varphi(j'')dj''\right)dj' \tag{25}$$

But

$$\int_{j'=0}^{j'=j}\int_{j''=j'}^{j''=m}\varphi(j'')dj''dj' = \int_{j''=0}^{j''=j}\int_{j'=0}^{j'=j''}\varphi(j'')dj'dj'' + \int_{j''=j}^{j''=m}\int_{j'=0}^{j'=j}\varphi(j'')dj'dj''$$

by Fubini's theorem. This equals

$$\int_{j''=0}^{j''=j} j''\varphi(j'')dj'' + \int_{j''=j}^{j''=m} j\varphi(j'')dj''$$

$$= \int_{j''=0}^{j''=m} j''\varphi(j'')dj'' - \int_{j''=j}^{j''=m} j''\varphi(j'')dj'' + j\int_{j''=j}^{j''=m}\varphi(j'')dj''$$

$$= A - AL_{\gamma}\left(\frac{r}{T}\right) + r\gamma(r),$$

by (18) (and the fact that $\gamma(T) = 0$), (17) (and the fact that $j = \gamma(r)$) and by (22). This proves (24), using (25). □

So we found, essentially, the same formula (11) as in the discrete case. Note that for the h-index h (of γ or of γ^{-1} – see Proposition 4) we have, from (24),

$$L_{\gamma^{-1}}\left(\frac{h}{m}\right) = 1 - L_{\gamma}\left(\frac{h}{T}\right) + \frac{h^2}{A} \tag{26}$$

and for the g-index g of γ we have ($j = \gamma(g)$ here)

$$L_{\gamma^{-1}}\left(\frac{\gamma(g)}{m}\right) = 1 - \frac{g^2}{A} + \frac{g\gamma(g)}{A} \tag{27}$$

$$L_{\gamma^{-1}}\left(\frac{\gamma(g)}{m}\right) = 1 - \frac{g}{A}\left(g - \gamma(g)\right) \tag{28}$$

(note that $g > \gamma(g)$ since $g > h = \gamma(h)$ and since γ is decreasing).

Formula (24) is easy to apply as the next examples show.

Examples : For $\alpha > 0$ (a parameter) define

$$\gamma(r) = m - r^{\alpha}\frac{m}{T^{\alpha}} \tag{29}$$

γ is strictly decreasing, $\gamma(0) = m$, $\gamma(T) = 0$. Varying α will lead to an infinite number of examples. We have

$$\int_0^r \gamma(r')dr' = mr - \frac{r^{\alpha+1}}{\alpha+1}.\frac{m}{T^{\alpha}} \tag{30}$$

so that

$$L_\gamma\left(\frac{r}{T}\right) = \frac{mr - \frac{r^{\alpha+1}}{\alpha+1}.\frac{m}{T^{\alpha}}}{A} \tag{31}$$

The requirement $L_\gamma(1) = 1$ leads to the relation (this also follows from (30), requiring $\int_0^T \gamma(r')dr' = A$)

$$\frac{\alpha}{\alpha+1}.\frac{mT}{A} = 1 \tag{32}$$

Now (24) yields, by (29)

$$L_{\gamma^{-1}}\left(\frac{\gamma(r)}{m}\right) = L_{\gamma^{-1}}\left(1 - \frac{r^{\alpha}}{T^{\alpha}}\right)$$

$$= 1 - L_\gamma\left(\frac{r}{T}\right) + \frac{r\gamma(r)}{A}$$

$$L_{\gamma^{-1}}\left(\frac{\gamma(r)}{m}\right) = 1 - \frac{\alpha}{\alpha+1}\frac{r^{\alpha+1}m}{T^\alpha A} \tag{33}$$

Requirement (32) makes sure that $L_{\gamma^{-1}}(0) = 0$, $L_{\gamma^{-1}}(1) = 1$ as it should.

Of course, we want to compare L_γ and $L_{\gamma^{-1}}$ in the same abscissa $\frac{r}{T}$, $r \in [0,T]$. For L_γ we have already (31). For $L_{\gamma^{-1}}$ we put

$$\frac{r}{T} = \frac{\gamma(x)}{m} = 1 - \frac{x^\alpha}{T^\alpha}$$

hence

$$x = \left(T^{\alpha-1}(T - r)\right)^{\frac{1}{\alpha}} \tag{34}$$

So

$$L_{\gamma^{-1}}\left(\frac{r}{T}\right) = L_{\gamma^{-1}}\left(\frac{\gamma\left(\left(T^{\alpha-1}(T-r)\right)^{\frac{1}{\alpha}}\right)}{m}\right)$$

$$= 1 - \frac{\alpha}{\alpha+1}\frac{m}{AT^\alpha}\left(T^{\alpha-1}(T-r)\right)^{\frac{\alpha+1}{\alpha}} \tag{35}$$

Since both Lorenz-curves are defined on the interval $[0,1]$, it is handy to express L_γ and $L_{\gamma^{-1}}$ as function of $\theta = \frac{r}{T}$. We have then, using (31), (32) and (35):

$$L_\gamma(\theta) = \frac{\alpha+1}{\alpha}\theta\left(1 - \frac{\theta^\alpha}{\alpha+1}\right) \tag{36}$$

$$L_{\gamma^{-1}}(\theta) = 1 - (1-\theta)^{\frac{\alpha+1}{\alpha}} \tag{37}$$

Note that for $\alpha = 1$, $L_\gamma(\theta) = L_{\gamma^{-1}}(\theta) = 2\theta - \theta^2$. For $\alpha \neq 1$, we are not able to determine the relations between $L_\gamma(\theta)$ and $L_{\gamma^{-1}}(\theta)$, although some numerical calculations suggest that $L_\gamma > L_{\gamma^{-1}}$ in case $\alpha = 2$ (perhaps this is so for all $\alpha > 1$) and $L_\gamma < L_{\gamma^{-1}}$ in case $\alpha = 0.5$ (perhaps this is so for all $0 < \alpha < 1$).

In general, however, we can show that rank-frequency functions γ exist such that L_γ and $L_{\gamma^{-1}}$ intersect in a point in the open interval $]0,1[$. Indeed, for $\theta \in [0,1]$ we have, by (22) and (23) that

$$L'_\gamma(\theta) = \frac{1}{A}\gamma(\theta T) \tag{38}$$

$$L'_{\gamma^{-1}}(\theta) = \frac{1}{A}\gamma^{-1}(\theta m) \tag{39}$$

A necessary condition for $L_{\gamma^{-1}} < L_\gamma$ is that $L'_{\gamma^{-1}}(0) < L'_\gamma(0)$ and $L'_{\gamma^{-1}}(1) > L'_\gamma(1)$. A necessary condition for $L_{\gamma^{-1}} > L_\gamma$ is that $L'_{\gamma^{-1}}(0) > L'_\gamma(0)$ and $L'_{\gamma^{-1}}(1) < L'_\gamma(1)$. So, if for $\theta \approx 0$, $L'_{\gamma^{-1}}(\theta) < L'_\gamma(\theta)$ and, if for $\theta \approx 1$, $L'_{\gamma^{-1}}(\theta) < L'_\gamma(\theta)$ then L_γ and $L_{\gamma^{-1}}$ must intersect in a point in $]0,1[$. By (38) and (39) we hence require (in order that L_γ and $L_{\gamma^{-1}}$ intersect), that for $\theta \approx 0$, $\gamma^{-1}(\theta m) < \gamma(\theta T)$ and that for $\theta \approx 1$, $\gamma^{-1}(\theta m) < \gamma(\theta T)$. This is the case for γ as in Fig. 5 if $\mathrm{tg}\alpha = \gamma'(0) < \mathrm{tg}\beta = \gamma'(T)$ and $T < m$.

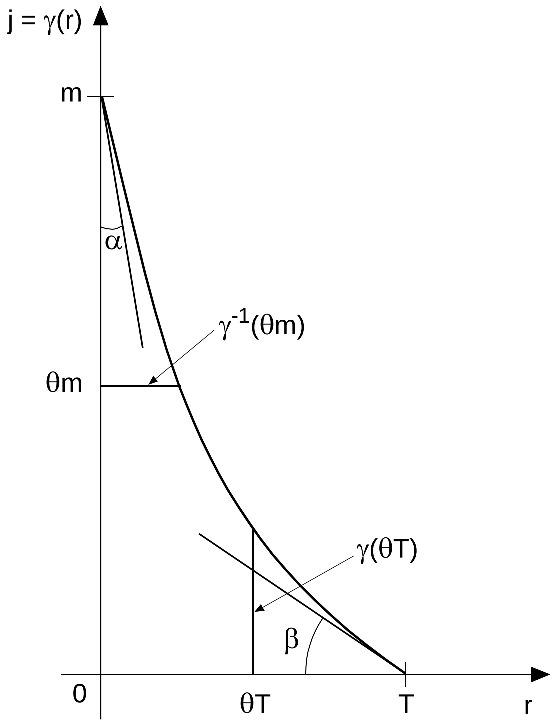

Fig. 5. An example of a function γ such that L_γ and $L_{\gamma^{-1}}$ intersect.

Also an example where $T > m$ and $tg\alpha > tg\beta$ yield intersecting Lorenz-curves L_γ and $L_{\gamma^{-1}}$ (here the roles of γ and γ^{-1} are interchanged).

IV. Conclusions and open problems

We have interpreted the discrete theory of partitions, conjugate partitions, Ferrers graph and Durfee square in informetrics. We showed that the h-index of a partition equals the one of the conjugate partition and we noted that this is false for the g-index and the R-index. A modification of Ferrers graph yielding the g-index has been presented. A formula for the Lorenz-curve of the conjugate partition in function of the one of the original partition is given.

A continuous theory of partitions, conjugate partitions, Ferrers graph and Durfee square has been given. Also here a partition and its conjugate have the same h-index. Also here we proved a formula for the Lorenz-curve of a conjugate partition in function of the one of the original one. Examples are given.

Although we have proved that these Lorenz-curves can intersect in the open interval $]0,1[$, we had to leave open to problem of the characterization of partitions (discrete or continuous) for which certain relations exist between their Lorenz-curves and the ones of the conjugate partitions. In particular we conjecture (in the discrete case) that, if the partition and its conjugate have the same dimension (i.e. where $C = (c_1,...,c_T)$ where $c_1 = T =$ dim C, then dim C = dim C′ where C′ is the conjugate partition of C) that their Lorenz-curves do not intersect (except, of course, in $(0,0)$ and $(1,1)$).

References

T.R. Anderson, R.K. Hankin and P.D. Killworth (2008). Beyond the Durfee square: Enhancing the h-index to score total publication output. Scientometrics 76(3), 577-588.

G.E. Andrews (1998). The Theory of Partitions. Cambridge University Press, Cambridge, UK.

R. Costas and M. Bordons (2008). Is g-index better than h-index ? An exploratory study at the individual level. Scientometrics 77(2), 267-288.

L. Egghe (2005a). Power Laws in the Information Production Process: Lotkaian Informetrics. Elsevier, Oxford, UK.

L. Egghe (2005b). Zipfian and Lotkaian continuous concentration theory. Journal of the American Society for Information Science and Technology 56(9), 935-945.

L. Egghe (2006). Theory and practise of the g-index. Scientometrics 69(1), 131-152.

L. Egghe (2008). The Hirsch-index and related impact measures. Annual Review of Information Science and Technology, to appear.

L. Egghe and R. Rousseau (1990). Introduction to Informetrics. Quantitative Methods in Library Documentation and Information Science. Elsevier, Amsterdam, the Netherlands.

L. Egghe and R. Rousseau (2003). Size-frequency and rank-frequency relations, power laws and exponentials : a unified approach. Progress in Natural Science 13(6), 478-480.

L. Egghe and R. Rousseau (2006). An informetric model for the Hirsch-index. Scientometrics 69(1), 121-129.

B. Jin, L. Liang, R. Rousseau and L. Egghe (2007). The R- and AR-indices : Complementing the h-index. Chinese Science Bulletin 52(6), 855-863.

M. Schreiber (2008a). The influence of self-citations on Egghe's g-index. Scientometrics 76(1),

M. Schreiber (2008b). An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the A-index and the R-index. Journal of the American Society for Information Science and Technology 59(9), 1513-1522.

R.S.J. Tol (2008). A rational successive g-index applied to economics departments in Ireland. Journal of Informetrics 2(2), 149-155.