Information Processing & Management Vol. 29, No. 4, pp. 515-521, 1993
Printed in Great Britain.

0306-4573/93 \$6.00 + .00

A REMARK ON THE GEOMETRY OF EGGHE'S DUAL IPPS

QUENTIN L. BURRELL
Statistical Laboratory, Department of Mathematics, University of Manchester, Manchester M13 9PL, U.K.

(*Received* 7 *April* 1992; *accepted in final form* 12 *June* 1992)

Abstract—Egghe's construction of dual information production processes (IPPs) is considered from a geometric rather than an analytic viewpoint. This reveals more clearly the simple nature of the duality, and certain results then follow immediately. The different notions of self-duality and pure duality are highlighted.

Keywords: Informetrics, Concentration measures, Gini index, Lorenz curve, Duality.

1. INTRODUCTION

In informetric studies, the notion of a collection of "sources" producing distinct "items" is fundamental and has been pointed out by many authors. The single example of a citation index, in which a set of source papers produces a list of cited papers, will suffice for our purposes, but there are many others. In some situations it is of interest to interchange the roles of sources and items; in our example we might wish to focus attention on the cited papers and ask how many original citings each received. In this case the cited papers are viewed as the sources and the citing papers as the items.

To provide a simple mathematical framework, Egghe (1989, 1990a) introduced the idea of an information production process and its dual. These have been further studied by Egghe (1990b,c, 1992) and Rousseau (in press, 1992). We refer the reader to the above, as well as Egghe and Rousseau (1990) for further motivation and merely give the following slightly modified:

Definition I. An information production process (IPP) is a triple (S, I, V) where $S = [0, T]$, $I = [0, A]$, and $V: S \to I$ is strictly increasing and convex with $V(0) = 0$ and $V(T) = A$.

Definition II. The dual IPP of (S, I, V) is the IPP (I, S, U) where, for $y \in I$

$$U(y) = T - V^{-1}(A - y) \tag{1}$$

and V^{-1} denotes the inverse of V.

According to Egghe, S and I represent the original source and item sets, and V can be thought of as giving the cumulation of items as we move over the sources.

Note that in Egghe's work it is further assumed that V is differentiable on S and that the source and item sets are thought of as being continuous. This is not really necessary in most of what follows. Indeed, allowing V to be piece-wise linear covers the important case of discrete sets.

In the following, the aim is not to question the validity or reality of IPPs as informetric models, but merely to point out an alternative way of interpreting them and to seek clarification of the notions of self-duality and pure duality.

2. THE GEOMETRIC APPROACH

The essential part of the 'duality' in the definition is expressed in eqn (1) defining U.

Geometrically, this is illustrated by Fig. 1 for a single point y, from which it follows that the transformation

$$(S, I, V) \to (I, S, U)$$

is as depicted in Fig. 2.

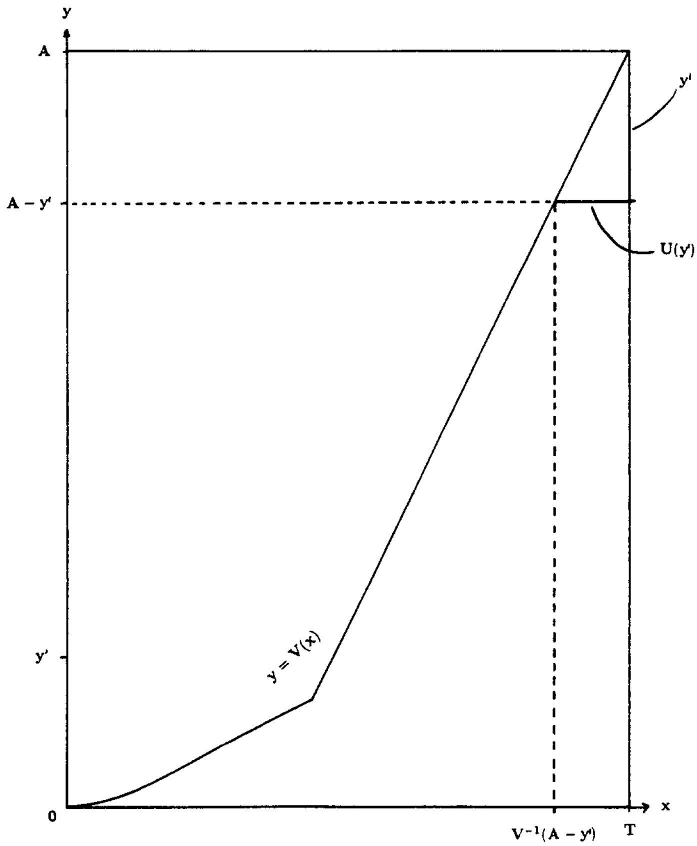

Fig. 1. Graphical form of Egghe's transformation $(S, I, V) \to (I, S, U)$.

Thus the graph of V is first "pushed on its back" (a planar rotation of 90°) and then "flipped over" (a reflection about a vertical axis).

For most purposes there is no real loss of generality in taking $T = A = 1$, since this merely corresponds to changing the scale of measurement, and we shall make this assumption in all that follows. Thus the defining eqn (1) for U becomes:

$$U(y) = 1 - V^{-1}(1 - y), \quad 0 \leq y \leq 1. \tag{2}$$

Note that now the graph of V on the unit square is just the Lorenz curve for the IPP (S, I, V), and similarly for U; see Egghe (1992). (That the Lorenz curve is invariant under scale transformations is well known; see for example Burrell (1990).) Note also that the double transformation to get from the graph of V to the graph of U is achieved simply as

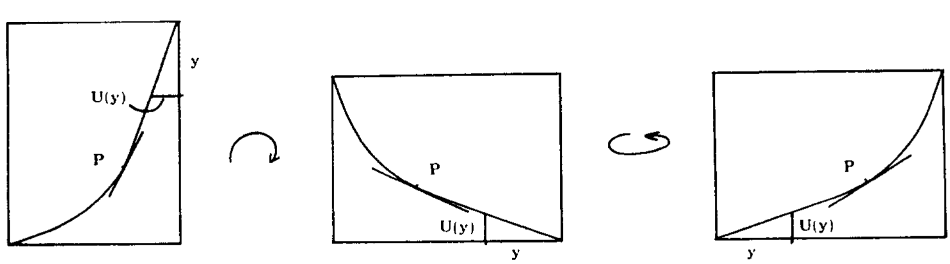

Fig. 2. Geometric visualisation of the construction of the dual IPP.

a single reflection in the diagonal line $x + y = 1$, as in Fig. 3. This feature of the Lorenz curves of dual IPPs was noted, in the discrete case, by Rousseau (1992). It then follows immediately that the LOR measure (i.e., the length of the Lorenz curve) is the same for both, thus generalizing Theorem IV of Rousseau (1992).

Also, since the Gini index is found from the Lorenz curve as Gini index = 1 − 2× area beneath Lorenz curve (see, e.g., Stuart & Ord (1987, p. 60–61)), a simple inspection of Fig. 3 reveals the equality of the Gini index of an IPP and its dual. Thus, Theorem III of Rousseau (1992) for the discrete case and Theorem II.3.3 of Egghe (1992) for the continuous case are immediate.

3. SELF-DUALITY

Following Rousseau (1992), it seems natural to describe an IPP as being self-dual if, in the standardised form we are considering with $S = I = [0,1]$, the dual IPP is identical to the original. Hence we have self-duality if

$$U(x) = V(x) \quad \text{for } 0 \leq x \leq 1.$$

In view of eqn (2), this is equivalent to

$$V(x) = 1 - V^{-1}(1 - x)$$

or

$$V^{-1}(1 - x) = 1 - V(x), \tag{3}$$

so that

$$1 - x = V(1 - V(x)) \tag{4}$$

(cf. Rousseau (1992, Definition V)). Note that if we write

$$H(x) = V(1 - x), \tag{5}$$

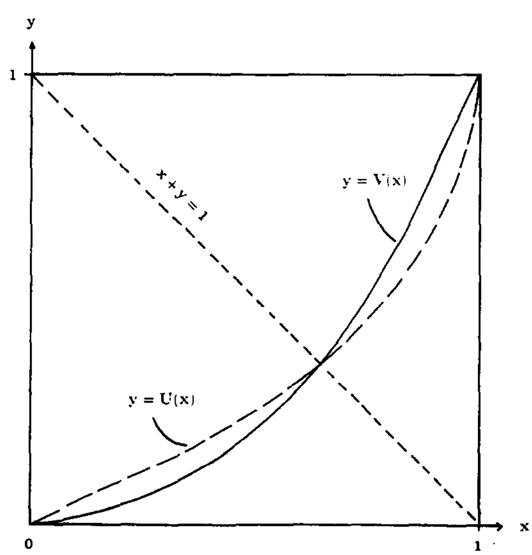

Fig. 3. Geometrical construction of the dual IPP if $A = T = 1$.

then H is a decreasing convex function on $[0,1]$ with $H(0) = 1$, $H(1) = 0$ and, from eqn (4) we have

$$\begin{aligned} x &= V(1 - V(1 - x)) \\ &= V(1 - H(x)) \\ &= H(H(x)), \quad \text{that is,} \end{aligned}$$

$$H(x) = H^{-1}(x),$$

so that H is its own inverse. A simple example of such a "self-inverse" function is given by

$$H(x) = \frac{c(1-x)}{c+x}$$

where c is a positive constant. This corresponds to the family of hyperbolae

$$\alpha xy + (x + y) = 1$$

and is equivalent to a generalization of Ponjaert's example reported by Rousseau (1992, eqn (10)). In fact, the general construction of such self-inverse functions, and hence of self-dual IPPs, is trivial, as can be seen from Fig. 4. Provided $H(0) = 1$, the continuous function H can be defined arbitrarily until it hits the line $x = y$, the only restrictions being that H is decreasing, convex, and with $H'(x) < -1$ on that region, with the rest of the curve being determined by reflection in the line $x = y$. However, as acknowledged by Rousseau, although such constructions may be pleasing from a mathematical point of view, there is no reason to believe that they have any practical relevance in the field of informetrics.

4. PURE DUALITY

In Egghe (1989, 1990a) we find the notion of pure duality. For this it is assumed that V, and hence U, are differentiable on S, I respectively, and we write ρ, σ respectively for their derivatives. Convexity demands that these be increasing. The analytic relationship between ρ and σ is given by differentiating eqn (2):

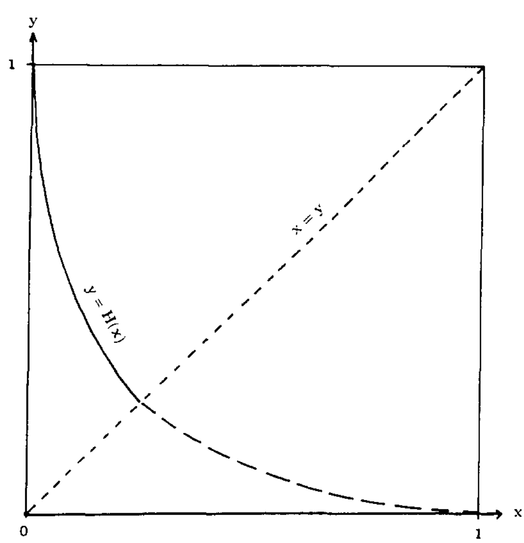

Fig. 4. General construction of 'self-inverse' functions.

$$\begin{aligned}\sigma(y) &= U'(y)\\ &= \frac{d}{dy}(1 - V^{-1}(1-y))\\ &= \frac{1}{V'(V^{-1}(1-y))}\\ &= \frac{1}{\rho(V^{-1}(1-y))}\end{aligned} \tag{6}$$

or equivalently

$$\rho(x) = \frac{1}{\sigma(U^{-1}(1-x))}.$$

These are easily seen from Fig. 2 where, if P is any point on the V—curve at which V is differentiable, if the slope is θ then the slope at the corresponding point after the first rotation is $-1/\theta$ and finally, after the reflection, is $1/\theta$.

If we now define ρ^* on $I = [0,1]$ by

$$\rho^*(y) = \rho(V^{-1}(y)), \tag{7}$$

then we have pure duality in Egghe's sense if

$$\sigma(y) = c\rho^*(y) \tag{8}$$

where $c > 0$ is a constant. In view of eqn (6) we have equivalently

$$\sigma(y)\sigma(1-y) = c, \quad 0 \le y \le 1. \tag{9}$$

A simple example is provided by

$$\sigma(y) = \alpha e^{\lambda y} \tag{10}$$

for suitable choices of α, λ, but note that the general construction of σ satisfying eqn (9) is trivial. Indeed, σ can be defined arbitrarily, but still positive and increasing, on $[0,\frac{1}{2}]$ and then extend to $[0,1]$ by setting

$$\sigma(y) = \frac{[\sigma(\frac{1}{2})]^2}{\sigma(1-y)}, \quad \frac{1}{2} \le y \le 1.$$

This then satisfies eqn (9) with $c = \sigma(\frac{1}{2})^2$. A simple example is provided by:

$$\sigma(y) = \begin{cases} y & \text{if } 0 \le y \le \frac{1}{2} \\ [4(1-y)]^{-1} & \text{if } \frac{1}{2} \le y \le 1. \end{cases}$$

It is of some interest to consider the example given by eqn (10) in a little more detail. Note first that it is just, in a slightly different notation, what Egghe terms the *group free Bradford function*. Note that

$$\begin{aligned}U(y) &= \int_0^y \sigma(u)\,du\\ &= \frac{\alpha}{\lambda}(e^{\lambda y} - 1), \quad 0 \le y \le 1\end{aligned}$$

so that $U(1) = 1$ implies $\alpha = \lambda/(e^{\lambda} - 1)$, and hence

$$\sigma(y) = \frac{\lambda e^{\lambda y}}{(e^{\lambda} - 1)}, \quad 0 \le y \le 1.$$

Upon solving $x = U(y)$ in terms of y we find

$$\begin{aligned} y &= U^{-1}(x) \\ &= \frac{1}{\lambda} \log\left(1 + \frac{\lambda x}{\alpha}\right) \end{aligned} \tag{11}$$

and then

$$\begin{aligned} V(x) &= 1 - U^{-1}(1 - x) \\ &= 1 - \frac{1}{\lambda} \log\left(1 + \frac{\lambda}{\alpha}(1 - x)\right) \end{aligned}$$

so that

$$\begin{aligned} \rho(x) &= V'(x) \\ &= \frac{1}{\alpha + \lambda(1 - x)} \end{aligned} \tag{12}$$

where $\alpha = \lambda/(e^{\lambda} - 1)$. Again adopting Egghe's terminology, eqn (11) determines *Leimkuhler's function*, and eqn (12) gives *Mandelbrot's function*.

Note that in Egghe's work the function ρ^* defined on I by eqn (7) is somewhat confusingly denoted by ρ. However, it is important to distinguish between ρ^* on I and ρ on S, even in the current situation in which $I = S = [0,1]$. (For instance, in the preceding example, ρ is given by eqn (12) while ρ^* is proportional to σ, given by eqn (10).) Because of this confusion it is easy to misunderstand the definition of pure duality as given by Egghe (1990a, Definition II.3.1) or Egghe and Rousseau (1990, Definition IV.3.3.1), where it is given as

$$\sigma(i) = c\rho(i) \quad \text{for } i \in I,$$

rather than as at eqn (8). Indeed, this might seem to imply that we are back in the situation of self-duality, an impression that is compounded when we read:

> ". . . suppose we have an IPP (S, I, V) [having the pure duality property] then the informetric 'calculus' ρ in (S, I, V) is the same as the informetric 'calculus' σ in the dual IPP (I, S, U). This means, for instance, that if (S, I, V) is a [pure dual IPP] of citation data . . . then the 'cited' triple (I, S, U) satisfies the same informetric laws with the same proportional parameters" (Egghe, 1990a; Egghe & Rousseau, 1990, p. 317).

In view of our earlier remarks, it is hard to justify use of the word '*same*' in its three occurrences in the above quotation. Surely '*is equivalent to*' would be a much better usage, where the equivalence is provided by the original duality construction, there being no purity involved. It is interesting to note that the original author only uses pure duality to justify consideration of his group free Bradford function (i.e., with σ as at eqn (10)), which is then held to demonstrate "*the unique place of Bradford's law in informetrics*" (Egghe, 1990)!

To illustrate geometrically the difference between self-duality and pure duality, note for the latter we have

$$\sigma(y)\sigma(1 - y) = c, \quad 0 \le y \le 1 \text{ from eqn (9)}$$

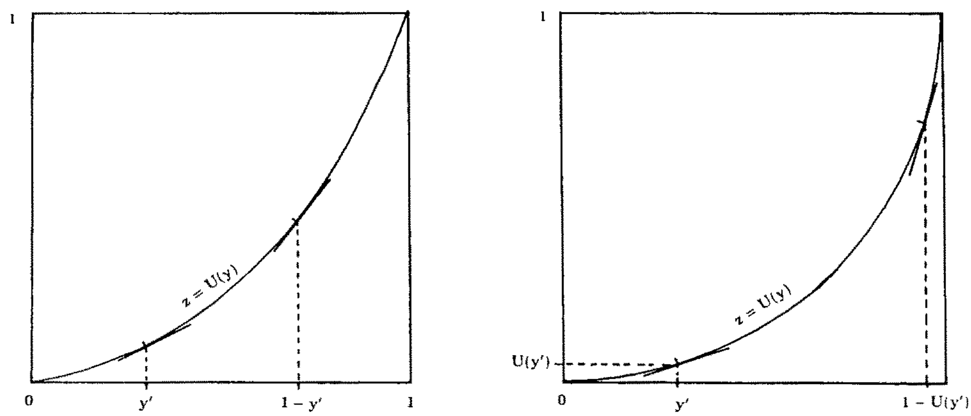

Fig. 5. Geometric comparison of pure and self-duality.

whereas for self-duality, differentiating both sides of eqn (4) yields

$$\rho(x)\rho(1 - V(x)) = 1$$

or, since $\rho = \sigma$ and $V = U$ in this case,

$$\sigma(y)\sigma(1 - U(y)) = 1, \quad 0 \leq y \leq 1. \tag{13}$$

The geometric interpretation of eqns (9) and (13) is illustrated in Fig. 5. In both cases the product of the gradients of the U curve at pairs of points as marked is constant.

5. CONCLUSION

It has been shown that the construction of dual IPPs can be given a simple graphical formulation. Not only does this give an alternative presentation, which may be more appealing to the less mathematically inclined, but certain results become transparent. It has also been shown that both the notions of self-duality and pure duality are purely mathematical constructs with, in the author's opinion, little practical relevance to the field of informetrics.

REFERENCES

Burrell, Q.L. (1990). The non-equivalence of the Bradford, Leimkuhler and Lotka 'laws'. Research Report 134/QLB, Statistical Laboratory, Department of Mathematics, University of Manchester, U.K.

Egghe, L. (1989). The duality of informetric systems with applications to the empirical laws. Doctoral dissertation, City University, London.

Egghe, L. (1990a). The duality of informetric systems with applications to the empirical laws. *Journal of Information Science*, *16*, 17–27.

Egghe, L. (1990b). A note on different Bradford multipliers. *Journal of the American Society for Information Science*, *41*, 204–209.

Egghe, L. (1990c). New Bradfordian laws, equivalent with old Lotka laws, evolving from a source-item duality argument. In L. Egghe & R. Rousseau (Eds.), *Informetrics 89/90: Proceedings of the Second International Conference on Bibliometrics, Scientometrics and Informetrics*, London, Ontario, 1989 (pp. 97–137). Amsterdam: Elsevier.

Egghe, L. (1992). Duality aspects of the Gini index for general information production processes. *Information Processing & Management*, *28*, 35–44.

Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science.* Amsterdam: Elsevier.

Rousseau, R. (in press). Category theory and information production processes. *Scientometrics.*

Rousseau, R. (1992). Two remarks on the preceding paper by L. Egghe. *Information Processing & Management*, *28*, 45–51.

Stuart, A. & Ord, J.K. (1987). *Kendall's advanced theory of statistics, Volume 1: Distribution theory* (5th edition). London: Griffin.