Using Poset Theory to Compare Fuzzy Multidimensional Material Deprivation Across Regions

Marco Fattore, Rainer Brüggemann, and Jan Owsiński

Abstract In this paper, a new approach to the fuzzy analysis of multidimensional material deprivation data is provided, based on partial order theory. The main feature of the methodology is that the information needed for the deprivation assessment is extracted directly from the relational structure of the dataset, avoiding any kind of scaling and aggregation procedure, so as to respect the ordinal nature of the data. An example based on real data is worked out, pertaining to material deprivation in Italy for the year 2004.

1 Introduction

The aim of this paper is to present new tools for fuzzy analysis of multidimensional material deprivation data and, more generally, for the analysis of multivariate ordinal datasets for evaluation and ranking purposes. The methodology combines fuzzy set theory and partial order theory and can be applied to evaluation problems in several fields, such as assessing quality of life, quality of services or quality of the environment (Brüggemann et al. 2001) to mention a few. The main feature of the approach is that the ordinal nature of the data is fully respected, avoiding any kind of scaling and aggregation procedure. The information needed for the evaluation process is extracted directly from the relational structure of the dataset, without turning ordinal scores into cardinal numbers. This is accomplished by means of partial order theory, a set of algebraic tools that provides the right formal language to tackle ordinal evaluation problems. For the sake of clarity and readability, all the necessary algebraic tools will be presented through a leading example pertaining to material deprivation in Italian macro-regions.

S. Ingrassia et al. (eds.), *New Perspectives in Statistical Modeling and Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, DOI 10.1007/978-3-642-11363-5_6,

2 Material Deprivation Data

Data about household material deprivation in Italian macro-regions are extracted from the EU-SILC survey, for the year 2004. To keep computations simple, only five variables have been considered, namely:

1. HS040 - *Capacity to afford paying for one week annual holiday away from home*;
2. HS050 - *Capacity to afford a meal with meat, chicken, fish (or vegetarian equivalent) every second day*;
3. HS070 - *Owning a phone (or mobile phone)*;
4. HS080 - *Owning a colour TV*;
5. HS100 - *Owning a washing machine*.

All five variables are coded in a binary form: 0 if the household is not deprived on the item and 1 if it is deprived. Each possible sequence of five binary digits defines a different *deprivation state*; for instance, the state $s = 10011$ stands for deprivation on HS040, HS080 and HS100 and non-deprivation on HS050 and HS070. Deprivation states can be ordered in a natural way, according to the following definition:

Definition 2.1 *Let s and t be two material deprivation states. We will write $s \trianglelefteq t$ if and only if $s_i \leq t_i$ $\forall\, i = 1, \ldots, 5$, where s_i and t_i are the $i-th$ digits of the binary representations of s and t respectively. State s is (strictly) less deprived than state t ($s \lhd t$) if and only if $s \trianglelefteq t$ and there exists at least one j such that $s_j < t_j$.*

Clearly, not all the deprivation states can be ordered, based on the previous definition, since there may be incomparabilities among them (e.g. consider states 10000 and 00001). As a result, the set of deprivation states gives rise to a *partially ordered set* (or *poset*, for short). Formally, a poset is a set equipped with a partial order relation, that is a binary relation satisfying the properties of *reflexivity*, *antisymmetry* and *transitivity* (Davey and Priestley 2002). A finite poset P (i.e. a poset defined on a finite set) can be easily depicted by means of a *Hasse diagram*, which is a particular kind of directed graph, drawn according to the following two rules: (1) if $s \lhd t$, then node t is placed above node s; (2) if $s \lhd t$ and there is no other state w such that $s \lhd w \lhd t$ (i.e. if t *covers* s), then an edge is inserted linking node t to node s. By transitivity, $s \lhd t$ in P, if and only if in the Hasse diagram there is a descending path linking the corresponding nodes; otherwise, states s and t are *incomparable* ($s \,||\, t$). Since in the example five binary variables are considered, the poset L of all possible deprivation states is composed of 32 nodes and 27 of them are actually observed in the data pertaining to Italy (referring to a sample of 24,202 households). The Hasse diagram of L is shown in Fig. 1 (observed states are represented as black nodes). The top node ($\top$) and the bottom node ($\perp$) represent the *completely deprived* state (11111) and the *completely non-deprived* state (00000) respectively.

Even if some possible deprivation states are not realized in the observed population, they have a clear meaning and could indeed be realized in other circumstances (e.g., by different populations or by the same population in different times). So, to get robust results and be able to compare population over time or space, we address

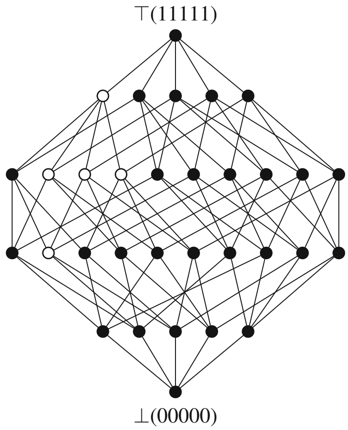

Fig. 1 Hasse diagram for the poset of material deprivation states (black nodes refer to states observed in Italy)

all of the states of L, rather than only the realised ones. Apparently, the partial order induced by Definition 2.1 is poorly informative about the deprivation level of many poset states. For example, states 10000 and 01111 are considered as incomparable, even if the latter seems to be much more deprived than the former. Nevertheless, as will be shown in the following, the different deprivation level of such states is clearly revealed when their different position in the Hasse diagram is considered and when the global relational pattern of the data is analyzed by means of partial order tools.

3 The Material Deprivation Membership Function

The central problem in assessing material deprivation is to assign a degree of deprivation to each state in L, that is to compute the *material deprivation membership function* for each deprivation state. Formally, a deprivation membership function is an *order preserving* map $m(\cdot)$ from L to $[0, 1]$, that is a map

$$\begin{aligned} m : L &\mapsto [0, 1] \\ : s &\rightarrow m(s) \end{aligned} \qquad (1)$$

such that

$$s \trianglelefteq t \Rightarrow m(s) \leq m(t). \qquad (2)$$

Clearly there is no unique criterion to select a deprivation membership function, out of such a class of maps. In this paper, we follow the point of view of 'response structure of the population', as in (Cerioli and Zani 1990) and assume that the degree of deprivation, assigned by $m(\cdot)$ to each state in L, depends upon the combined different assessments given by a population of 'judges'. The first step in view of the definition of $m(\cdot)$ is therefore to make explicit how the set of judges is built.

Linear extensions of a poset. The key idea to identify a suitable set of judges can be explained as follows. Judges produce rankings of deprivation states out of the poset L; when accomplishing this task, they are free to order incomparable pairs as preferred (no ties are allowed), but they cannot violate the constraints given by the original partial order, i.e. if $s \lhd t$ in L, then any judge must rank t above s in his own deprivation ranking. Thus, the set of all possible different judges (i.e. judges not producing the same rankings) coincides with the set of all the *linear extensions* of L. A linear extension of a poset P is a linear ordering of the elements of P which is consistent with the constraints given by the partial order relation. For example, if P is composed of three elements x, y and z, with $y \leq x$, $z \leq x$ and $y \,||\, z$, only two linear extensions are possible, namely $z \leq y \leq x$ and $y \leq z \leq x$, since x is greater than both y and z in P. The set of all the linear extensions of a poset P is denoted by $\Omega(P)$; it comprises all the linear orders compatible with P and identifies uniquely the partial order structure (Neggers and Kim 1998).

Up-sets, down-sets and the deprivation border. In view of a fuzzy assessment of material deprivation, three relevant subsets of L can be identified, namely

- the set D of *certainly deprived* states: $D = \{s \in L : m(s) = 1\}$;
- the set W of *certainly non-deprived* states: $W = \{s \in L : m(s) = 0\}$;
- the set A of *ambiguously deprived* states: $A = \{s \in L : 0 < m(s) < 1\}$.

According to (1) and (2), if $s \in D$ and $s \unlhd t$, then $t \in D$; similarly, if $s \in W$ and $t \unlhd s$, then $t \in W$. In poset theoretical terms, sets like D and W are called *up-sets* and *down-sets*, respectively. When fuzzy poverty is assessed in monetary terms, a threshold τ is usually identified separating certainly poor people from the rest of the population. A similar threshold can also be defined for the poset of material deprivation states. Given the up-set D, there is a unique subset $\underline{d} \subseteq D$ of mutually incomparable elements (a so called *antichain*), such that $s \in D$ if and only $d \unlhd s$ for some $d \in \underline{d}$ (Davey and Priestley 2002). The up-set D is said to be generated by $\underline{d}$ (in formulas, $D = \uparrow\underline{d}$). Excluding the trivial cases of $D = L$ and $D = \top$, any element of the generating antichain is covered only by elements of D and covers only elements of $L \setminus D$, so that it shares the same role that τ has in the monetary case. For this reason, $\underline{d}$ can be called the *material deprivation border*.

Membership function definition. To define the membership function, we need (i) to determine how linear extensions (i.e. the 'judges') $\omega \in \Omega(L)$ assign the degrees of deprivation $Dep_{\omega}(s)$ to a state $s \in L$ and (b) to decide how to combine all such degrees into the 'final' degree $m(s)$. First of all, let us assume that an antichain $\underline{d}^*$ is chosen as the deprivation border (this means that all judges agree to assign degree of deprivation one to all elements in $\uparrow\underline{d}^*$). At a purely illustrative level[1] we can put

[1] The choice of a meaningful deprivation border is subjective and requires experts judgments. In this methodological paper, we do not deal with this fundamental issue and take the border as given.

$$\underline{d}^* = (10100, 11000). \tag{3}$$

(i) Once the border has been identified, the simplest way we can define the degree of deprivation of a state s in a linear extension ω is to put

$$Dep_\omega(s) = \begin{cases} 1 & if\ \exists\, d \in \underline{d}^* \text{such that } d \le s \text{ in } \omega; \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Explicitly, ω assigns degree of deprivation 1 to each element of $\underline{d}^*$ and to each element of L that is ranked by the ω above *at least one* element of $\underline{d}^*$; conversely, ω assigns the degree of deprivation 0 to those states that are ranked, in ω, below *all* the elements of $\underline{d}^*$.

(ii) The membership function for state s (that we write as $m(s|\,\underline{d}^*)$ since it depends upon the choice of the material deprivation border) is then computed as

$$m(s|\,\underline{d}^*) = \frac{|\{\omega : Dep_\omega(s) = 1\}|}{|\Omega(L)|}, \tag{5}$$

i.e. as the fraction of linear extensions ranking state s as deprived, given $\underline{d}^*$.
Before turning to the problem of the computation of $m(s|\,\underline{d}^*)$, some comments to definitions (4) and (5) are in order.

1. Each linear extension classifies the states of L in binary terms; this means that the fuzziness we are dealing with is due to different responses by different judges, while the single judge acts in a crisp way.
2. Taking into account all linear extensions, our measure of fuzziness takes care of both comparable and not comparable, i.e. non commensurable, deprivation states.
3. Given the deprivation border (3), one easily sees that the set D is composed of the states 10100, 10101, 10110, 10111, 11000,11001, 11010, 11011, 11100, 11101, 11110, 11111. In fact, each of these states belongs to the up-set of an element of the deprivation border $\underline{d}^*$ (for example, state 10101 is in the up-set of 10100), so that each linear extension assigns to them the deprivation degree of 1.
4. Formula (5) implies that the set of certainly non-deprived states is composed of all those states that are less deprived than any element of $\underline{d}^*$. In our specific example, it turns out that $W = \{00000, 10000\}$.

Membership function computation. To compute the fraction of linear extensions assigning degree of deprivation 1 to a state s, we will have to list all the elements of $\Omega(L)$ and select those where s is ranked above d for some $d \in \underline{d}^*$. Unfortunately, listing all the linear extensions of a poset is computationally impossible (unless the poset is very small or contains very few incomparabilities), so mutual ranking frequencies must be estimated, based on a sample of linear extensions. The computations presented in this paper are performed by running the Bubley-Dyer

Table 1 Membership function $m(s|10100, 11000)$

State s:	00000	00001	00010	00011	00100	00101	00110	00111
$m(s\|10100, 11000)$	0.00	0.11	0.11	0.65	0.06	0.66	0.66	0.98
State s	01000	01001	01010	01011	01100	01101	01110	01111
$m(s\|10100, 11000)$	0.06	0.66	0.66	0.98	0.67	0.98	0.98	1.00
State s	10000	10001	10010	10011	10100	10101	10110	10111
$m(s\|10100, 11000)$	0.00	0.67	0.67	0.98	1.00	1.00	1.00	1.00
State s	11000	11001	11010	11011	11100	11101	11110	11111
$m(s\|10100, 11000)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

algorithm, which is the most efficient known algorithm for (quasi) uniform sampling of linear extensions (Bubley and Dyer 1999). Particularly, the membership function $m(s\ |10100, 11000)$ has been estimated using a sample of 10^8 linear extensions and is reported in Table 1.

It must be noted that the table actually reports an estimation of the true values of the membership function; this explains why $m(01111\ |10100, 11000) = 1$, even if state 01111 does not belong to D.

4 Fuzzy Material Deprivation in Italian Macro-Regions

The membership function computed in the previous paragraph has been used to compute the amount of fuzzy poverty in Italy and in five Italian macro-regions, namely North–West, North–East, Centre, South and Islands, based on the data described at the beginning of the paper. Two normalized indicators have been considered. The first indicator FH is the fuzzy extension of the classical Head Count Ratio H; it is defined as

$$FH = \frac{\sum_{s \in L} m(s \mid \underline{d}^*) \cdot |s|}{\sum_{s \in L} |s|} \tag{6}$$

where $|s|$ is the number of people occupying state s. FH is simply the ratio between the fuzzy cardinality of the set of people having a non-null degree of deprivation and the cardinality of the entire population. The second indicator C is defined as

$$C = \frac{\sum_{s \in L} m(s|\underline{d}^*) \cdot |s|}{\sum_{s \in L} \delta(s) \cdot |s|}, \quad \delta(s) = \begin{cases} 1 & \text{if } m(s|\underline{d}^*) > 0 \\ 0 & \text{if } m(s|\underline{d}^*) = 0. \end{cases} \tag{7}$$

If $\sum_{s \in L} \delta(s) \cdot |s| = 0$, C is set to 0. C measures the mean level of ‘deprivation certainty’ characterizing the subpopulation of people occupying states with a non-null degree of deprivation. The results are reported in Table 2.

Although the present application has an illustrative goal, the results obtained are consistent with the territorial differences in Italy. Southern regions show a much

Table 2 Values of FH and C for Italian macro-regions and the whole country, year 2004

	FH	*C*
North-West	0.04	0.76
North-East	0.05	0.76
Centre	0.06	0.88
South	0.15	0.93
Islands	0.15	0.94
Italy	0.07	0.86

greater incidence of material deprivation than regions in the North and in the Centre. Also index C is markedly higher in the Southern regions, revealing a possible social polarization: not only is material deprivation greater in the Southern part of Italy, but in those regions deprived people are also ‘almost’ certainly deprived, i.e. in a fuzzy sense they belong definitely to the set of deprived people.

5 Conclusion

In this paper, we have shown how poset theory provides an effective setting for fuzzy modeling of multidimensional material deprivation data and, more generally, of multidimensional ordinal datasets. The main advantage of the methodology is that, differently from other approaches (Cerioli and Zani 1990, Lemmi and Betti 2006), it relies only on the ordinal nature of the data, without supposing any quantitative model behind them. This way, a sound numerical evaluation of the deprivation degrees is computed out of qualitative ordinal information, preserving those ambiguities that are co-essential to the concepts that are dealt with. This is consistent with Sen’s point of view, that if there is some ambiguity in a concept, ‘a precise representation of that ambiguous concept must preserve that ambiguity’ (Sen 1992). Here, partial order theory plays a role similar to that of linear algebra in quantitative multivariate data analysis: it makes it possible to represent and exploit the structure of the data and to extract information directly out of it. There are indeed some open issues. The combinatoric approach used in the paper is in fact particularly suitable when the number of states is not too large. To extend the proposed methodology to situations where many variables or states are considered, different approaches are being developed. They combine (1) techniques for the clustering of states, based on the algebraic tools of congruences (Cheung and Vogel 2005); (2) the development of better performing software procedures to shorten the computation time in the estimation of mutual ranking frequencies; (3) the identification of analytical formulas yielding approximated values of such frequencies, directly out of the poset topology. In conclusion, partial order theory paves the way to a new approach for studying multidimensional systems of ordinal data; at the same time, it calls for further methodological research, so as to extend and tune partial order concepts and techniques towards the needs of applied statistical analysis and modeling.

References

Brüggemann, R., Halfon, E., Welzl, G., Voigt, K., & Steinberg C. (2001). Applying the concept of partially ordered sets on the ranking of near-shore sediments by a battery of tests. *Journal of Chemical Information and Computer Science*, *41*, 918–925.

Bubley, R., & Dyer, M. (1999). Faster random generation of linear extensions. *Discrete mathematics*, *201*, 81–88.

Cerioli, A., & Zani, S. (1990). A fuzzy approach to the measurement of poverty. In C. Dagum, & M. Zenga (Eds.), *Income and wealth distribution, inequality and poverty* (pp. 272–284). Berlino Heidelberg: Springer-Verlag.

Cheung, K. S. K., & Vogel, D. (2005). Complexity reduction in lattice-based information retrieval. *Information Retrieval, 8*, 285–299.

Davey, B. A., & Priestley B. H. (2002). *Introduction to lattices and order*. Cambridge: Cambridge University Press.

Lemmi, A., & Betti, G. (2006). *Fuzzy set approach to multidimensional poverty measurement*. New York: Springer.

Neggers, J., & Kim, H. S. (1988). *Basic posets*. Singapore: World Scientific.

Sen, A. K. (1992). *Inequality reexamined*. Oxford: Clarendon Press.