Working papers series

WP ECON 22.14

An elementary algorithm to make quantitative assessments from multidimensional, unstructured, categorical data

Antonio Villar
Department of Economics, Universidad Pablo de Olavide

Keywords: multidimensional evaluation; categorical data; non-transitive and incomplete preferences; pairwise comparisons.

JEL Classification: C60, D70.

Department of Economics

An elementary algorithm to make quantitative assessments from multidimensional, unstructured, categorical data(*)

Antonio Villar

Universidad Pablo de Olavide & ISEAK

Abstract

This paper proposes and characterizes an elementary algorithm to solve multicriteria evaluation problems when individual judgements are categorical and may fail to satisfy both transitivity and completeness. The evaluation function consists of a weighted sum of the average number of times that each alternative precedes some other, in all pairwise comparisons. It provides, therefore, a quantitative assessment which is well-grounded, immediate to compute, and easy to understand. An application to the evaluation of human development illustrates how this evaluation protocol works.

Keywords: multidimensional evaluation; categorical data; non-transitive and incomplete preferences; pairwise comparisons.

JEL classification numbers: C60, D70

(*) Financial support from the Spanish Ministerio de Ciencia e Innovación (project PID2019-104452RB-I00) and the Junta de Andalucía (projects FEDER UPO-1263503 and P18-RT-2135) is gratefully acknowledged.

1. Introduction

Consider an evaluation problem in which a group of individuals has to provide a comparative assessment of a collection of alternatives with respect to several features or qualities. We focus on those evaluation problems in which the comparison between alternatives is categorical, that is, each individual simply declares if alternative i has more or less of a given quality than alternative j. Moreover, we admit those judgements to lack structure, that is, they need not be complete or acyclic. And we look for an evaluation function that transforms those ordinal judgements into quantitative overall assessments.

Let us consider a familiar evaluation problem of this type, to help the intuition. Think of a panel of experts that has to evaluate a set of research proposals, in order to allocate a given amount of funds, in some scientific field. Here the alternatives are the research proposals, the individuals correspond to the panel of experts, and each alternative must be evaluated regarding several criteria (e.g., originality, relevance of the topic, budget adequacy, and research group performance). Note that, besides finding a global rank of the research projects, it will be most helpful to have some quantitative measure of their relative merit, to provide a guide to allocate the funds.

This type of evaluation problem appears in a multiplicity of situations, such as choosing new headquarters for a firm, rating the quality of diamonds, assessing the impact of medical treatments, valuing social concerns, measuring consumers' satisfaction, or choosing new members of the staff, to name a few. In some of those cases the qualitative nature of the individuals' judgements is transformed into cardinal ratings, by attaching values to the different levels of quality (e.g., the conventional 0-10 pain scale in medical trials). Yet this is problematic not only because the values attached to quality levels are often arbitrary, but also because this process implies a form of interpersonal comparability that may not be justified. That is, there is no guarantee that 4 is twice 2 for everybody (individual rating scales need not be linear), so that number 7 may well have a different meaning for two different evaluators. Yet, cardinal evaluations are relevant in many problems, particularly when they are part of a process involving the allocation of some resources.

In this paper we propose and characterize an elementary evaluation function that provides a cardinal assessment for problems within this multidimensional and unstructured scenario. The approach hinges on two complementary ideas: the recourse to pairwise comparisons, and anonymous head-counting evaluation.

Pairwise comparisons permit valuing the relative merit of a set of alternatives in a very general framework, including the case of individual judgements that fail to satisfy the standard requirements of acyclicity and completeness. Head-counting is a natural way of introducing cardinality in the evaluation, regarding categorical judgements, which incorporates an anonymity principle: all individuals enter the evaluation on an equal foot. Combining these two ideas, which are far from new (some already appear in the XIII Century, as commented below), our evaluation formula relies on counting how many individuals prefer an alternative to another in each pairwise comparison. As for the way of aggregating that information into an overall judgement, we propose and characterize an elementary evaluation function that consists of a weighted sum of the average number of people who considers that an alternative precedes another, within each dimension, when compared pairwise. The weights correspond to the relative importance of those dimensions in the problem. It is, therefore, an algorithm applicable to a wide family of problems, easy to compute and easy to understand.

The properties assumed to pinpoint that function are familiar and transparent: independence (the evaluation of each alternative only depends on the comparison of this alternative with the rest), symmetry (two alternatives with the same support will get the same evaluation), uniformity (the change in the evaluation due to a change in a variable is independent on the level of the variable), neutrality (equal changes in two different pairwise comparisons regarding an alternative, produce the same change in the evaluation of that alternative), and scale (the evaluation of each alternative moves between 0 and 1). Interestingly, this evaluation protocol admits an equivalent formulation in terms of the differences between "wins and losses", in those pairwise confrontations. Moreover, it can be regarded as a multidimensional version of an extended Borda score.

The paper is organized as follows. Section 2 presents and characterizes the evaluation function. Section 3 discusses some relevant features of this protocol. Section 4 provides an application to the evaluation of human development, using the UN data on the components of the human development index. The work concludes with a few final words in Section 5.

2. The evaluation protocol

Let A = {*1, 2, …, m*} be a set of alternatives to be evaluated with respect to a collection $Q = \{1, 2, \dots, q\}$ of qualities or dimensions, by a set $N = \{1, 2, \dots, n\}$ of individuals. Those individuals are endowed with binary relations, defined on A, that express their judgements with respect to each dimension. That is, for each ordered pair (i, j), $i \neq j$, and every $k \in Q$, individual $h \in N$ declares whether an alternative *precedes* the other, or both alternatives *match*, or they are *not comparable*. Here "i precedes j" means that alternative i is regarded by individual h as having more of the quality being evaluated, whereas "i matches j" means that both alternatives are considered as having the same. On those binary relations we only assume that the precedence relation is asymmetric (i precedes j implies that j cannot precede i), whereas the other two are symmetric. We also assume that, for each dimension, there is at least one individual capable of comparing two alternatives. Our target is to obtain a cardinal evaluation of the alternatives in A from the individuals' judgements. We now formalize this idea.

Let n_{ij}^k denote the number of individuals who consider that alternative i precedes alternative j in dimension k, $e_{ij}^k = e_{ji}^k$ the number of individuals who consider that alternative i matches alternative j, in that dimension, and $z_{ij}^k = z_{ji}^k$ the number of those unable or unwilling to compare them (don't know/no answer). By construction, $n = n_{ij}^k + n_{ji}^k + e_{ij}^k + z_{ij}^k$, for all $i, j \in A, i \neq j, \forall\, k \in Q$.

We define the *precedence score* of alternative i with respect to alternative j, for $i \neq j$, relative to dimension k, as follows:

$$p_{ij}(k) = \frac{1}{n}\left(n_{ij}^k + \frac{e_{ij}^k}{2}\right) \qquad [1]$$

This is simply the average number of individuals who consider that i precedes j in dimension k, including one-half of those who consider that the quality of both alternatives match. The ***average precedence score*** of alternative i in dimension k is given by:

$$p_i(k) = \frac{1}{m-1}\sum\nolimits_{j \neq i} p_{ij}(k) \qquad [2]$$

Let $\mathbf{P}^{(k)} = \{p_{ij}(k)\}_{i,j=1}^{m}$ denote the square m-matrix whose (i,j) entry is $p_{ij}(k), \forall i \neq j, k \in Q$, with $p_{ii}(k) = 0, \forall i, \forall k$, by convention. This matrix summarizes the relevant information on the evaluation problem with respect to dimension k and can be regarded as an extension of Young's (1974) vote matrix. Note that we implicitly assume an anonymity principle, as all individual judgements count the same.

Let $\wp^{(k)} = \{\mathbf{P}^{(k)} \in \mathbb{R}_+^{2m}\ with\ p_{ii}(k) = 0, \forall\ i\}$ the set of all non-negative matrices of order $m \times m$, with zeroes in the $p_{ii}(k)$entries, and define:

$$\wp = \prod_{k=1}^{q} \wp^{(k)} \subset \mathbb{R}_+^{2m\times q}\ , \qquad \mathbf{P} = \left(\mathbf{P}^{(1)} \times \mathbf{P}^{(2)} \times \dots \times \mathbf{P}^{(q)}\right)$$

An ***evaluation function*** can now be defined as a mapping $\boldsymbol{v}{:}\,\wp \to \mathbb{R}_+^m$ that associates a number $v_i(\mathbf{P})$ to each alternative $i \in A$, as a function of how many individuals give precedence to this alternative over the others in the different dimensions.

We now introduce five properties that characterize our evaluation function: independence, symmetry, uniformity, neutrality, and scale. Those properties are described informally below. All bear the idea that all judgements have an equal impact on the evaluation and that all alternatives are treated alike, within each dimension. The formal statement of those properties and the proof of the corresponding characterization result can be found in the Appendix.

The first property, *independence*, establishes that the evaluation of alternative i only depends on the comparison of this alternative with the rest, across the q dimensions, that is, $v_i(\mathbf{P}) = f\left(\mathbf{p}_{(i)}\right)$, where $\mathbf{p}_{(i)}$ is the ith row of matrix $\mathbf{P}$.

The second property, *symmetry*, says that if in a given problem two alternatives have the same precedence scores, then they should get the same evaluation. That is, if $\mathbf{P}$ is such that $\mathbf{p}_{(i)} = \mathbf{p}_{(j)}$, then $v_i(\mathbf{P}) = v_j(\mathbf{P})$.

The next two properties, uniformity and neutrality, refer to the sensitivity of the evaluation of an alternative, to changes in the judgment of some other. To motivate the property of *uniformity*, suppose that, after all judgements have been submitted, an individual reconsiders the evaluation of alternative i relative to alternative j in dimension k (e.g., form j precedes i shifts to i precedes j). This would change matrix $\mathbf{P}$ to $\mathbf{P}$' and produce a change in the evaluation of i, from $v_i(\mathbf{P})$ to $v_i(\mathbf{P}')$. Suppose now that, after that change, another individual also reconsiders and makes exactly the same move. We shall have a new matrix $\mathbf{P}$'' and, associated with it, a new evaluation of alternative i, $v_i(\mathbf{P}'')$. What we require is that the impact of both changes be the same, namely: $v_i(\mathbf{P}') - v_i(\mathbf{P}) = v_i(\mathbf{P}'') - v_i(\mathbf{P}')$. In this way, we ensure that the impact of the change of judgements on the evaluation, does not depend on who reconsiders first.

Neutrality introduces the idea that all alternatives count the same in the evaluation, within each dimension. That is, the change in the evaluation of an alternative derived from an increment x in its precedence score, in detriment of another, in dimension k, is the same no matter with respect to which other alternative this change occurs.

Finally, the property of *scale* determines that 1 is the maximum value that an alternative can obtain, and that this happens when all individuals consider that this alternative precedes all others in all dimensions. Complementarily, if all individuals consider that an alternative is preceded by all others in all dimensions, then this alternative is valued zero.

The next result shows that assuming those properties amounts to choose a simple and intuitive evaluation function: Attach to each alternative a convex combination of its average precedence scores. The coefficients of that convex combination correspond to the relative importance of the different dimensions.

Proposition: *An evaluation function* $\boldsymbol{v}: \wp \to \mathbb{R}_+^m$ *satisfies the properties of independence, symmetry, uniformity, neutrality, and scale, if and only if it is given by:*

$$v_i(\mathbf{P}) = \sum_{k=1}^{q} \alpha_k p_i(k)\,, \qquad \forall\, i \in M \qquad [3]$$

Where $\alpha_k \geq 0, \forall k, \sum_{k=1}^{q} \alpha_k = 1$*. Moreover, these properties are independent.*
(The proof is given in the Appendix)

We shall refer to vector function $\boldsymbol{v}: \wp \to \mathbb{R}_+^m$ as the ***precedence function***. Observe that this function in monotone in $p_{ij}(k)$.

To give a closed form to the precedence function [3], we need to determine the weights α_k. There may be cases in which those weights are given from outside or agreed upon before the partial evaluations take place. In other cases, though, the evaluators may also be asked to give weights to the dimensions. A particular way of approaching that problem is to choose the weights *consistently* with the evaluation process. By that we mean that each dimension is compared pairwise with each other, and the number of individuals who give precedence to one another is computed. Let $\mathbf{P}^{(0)}$ denote the square semipositive q-matrix that summarizes those evaluations when the set of qualities plays the role of the set of alternatives, in a single-dimensional evaluation problem. That is, $v_k\left(\mathbf{P}^{(0)}\right) = \frac{1}{q-1}\sum_{j \neq k} p_{kj}(0)$. Then we can define:

$$\alpha_k\left(\mathbf{P}^{(0)}\right) = \frac{v_k\left(\mathbf{P}^{(0)}\right)}{\sum_{r=1}^{q} v_r\left(\mathbf{P}^{(0)}\right)}$$

We can define the ***closed precedence function*** as that given by:

$$v_i(\mathbf{P}) = \sum_{k=1}^{q} \frac{\sum_{j\neq k} p_{kj}(0)}{\sum_{r=1}^{q} \sum_{j\neq r} p_{rj}(0)} p_i(k) \qquad [4]$$

3. Discussion

3.1 Welcome back to the XVIII Century

The ideas presented here can be traced back to the works of Borda (1784) and Condorcet (1785), regarding voting procedures. The name of Condorcet is associated to the evaluation of alternatives in terms of pairwise comparisons, or tournaments (even though this method had already been proposed by Ramón Llull at the end of the XIII Century). The key idea is that the best candidate is the one that beats all others in those pairwise confrontations. In terms of our formulation, a Condorcet winner will be that alternative that satisfies $p_i(k) > p_j(k), \forall\, j, \forall\, k$. The precedence score of an alternative relative to another, $p_{ij}(k)$, is sometimes called the *Condorcet number* (Moulin, 1988). It is well known that a Condorcet winner may not exist, and that the Condorcet approach permits having weak orderings (indifferences) and admit non-transitive and non-complete preferences.

Borda proposed a different venue, which can be summarized as follows. Each candidate is given a score that reflects *how much* support it accrues, rather than *how many* supporters back this candidate, as in Condorcet's approach. The "amount of support" of candidate i is given by the total number of candidates that are below i in the rankings of all voters. Besides producing a complete ranking of candidates, the Borda approach also provides a cardinal way of rating them, adding up the number of defeated candidates in each voter's ranking.

In spite of the different principles behind Borda and Condorcet approaches, which yield different outcomes, both methods are closely related. Indeed, the average precedence score, $p_i(k)$, is precisely the Borda score relative to dimension k (Moulin, 1988).[1] So, the average precedence score of an alternative may be regarded as the normalized Borda score in a more general scenario, and the Proposition can be interpreted as a straightforward characterization of the multidimensional Borda score, in a general context. Implicit in our approach is the idea of "one person, one vote", which is a specific form of anonymity that conveys cardinality to the comparisons and focus the evaluation on the Condorcet numbers.

Needless to say, the Borda rule has been characterized by different authors, in the more standard setting of social choice (e.g. Young, 1974, Hansson & Sahlquist, 1976, Nitzan & Rubinstein, 1981, and Mihara, 2017). In a recent paper, Barberà & Bossert

[1] Also note that approaching the Borda score from Condorcet's pairwise comparison is what permits applying the Borda score in a much more general scenario (Young, 1974).

(2022) provide a characterization of the Borda rule in a very general setting, similar to the one presented here. They focus, though, on the Borda ranking rather than on the Borda scores.[2] Besides the general framework, the key element in their work consists of interpreting the Borda rule in terms of the differences between favourable and unfavourable opinions, as an implication of the cancellation axiom, which we discuss next.

3.2 Wins and losses and the cancellation principle

Barberà & Bossert (2022) argue forcefully that the right approach to understand Borda and Condorcet choice procedures is in terms of differential outcomes, the differences between "wins and losses", so to speak. In our framework that would correspond to taking $d_{ij}(k) = p_{ij}(k) - p_{ji}(k)$ as the reference variable. The use of differential outcomes as the reference variables derives from the *cancellation axiom*, a property that establishes that two opposite judgements cancel each other, so that they would not affect the evaluation (e.g., Young, 1974).

Let us now consider our evaluation function from the viewpoint of computing wins and losses. The *average differential score* of alternative i is the difference between the average number of times that this alternative precedes some other, and the number of times that some other alternative precedes i. That is:

$$d_i(k) = \frac{1}{m-1}\left(\sum\nolimits_{j\neq i} p_{ij}(k) - \sum\nolimits_{j\neq i} p_{ji}(k)\right)$$

Note that,

$$\frac{1}{m-1}\left(\sum\nolimits_{j\neq i} p_{ij}(k) + \sum\nolimits_{j\neq i} p_{ji}(k) + \frac{1}{n}\sum\nolimits_{j\neq i} z_{ij}^{k}\right) = 1$$

so that we can write $d_i(k) = 2p_i(k) + z_i(k) - 1$, where $z_i(k)$ is the average number of times that alternative i is not comparable, that is, $z_i(k) = \frac{1}{m-1}\sum_{j\neq i}\frac{z_{ji}^{k}}{n}$. Therefore,

$$p_i(k) = \frac{1}{2}\left(d_i(k) - z_i(k) + 1\right)$$

From this relationship we can provide the following alternative version of the Proposition:

[2] Other relevant differences between their framework and ours are: (1) They do not assume that judgements can be identified with individuals, as we do, which implies that the precedence relation need not be asymmetric. (2) They move in a single-dimensional context, whereas here we consider judgements over different qualities simultaneously.

Proposition (alternative version): *An evaluation function* $\boldsymbol{v}: \wp \to \mathbb{R}_+^m$ *satisfies the properties of independence, symmetry, uniformity, neutrality, and scale, if and only if it is given by:*

$$v_i(\mathbf{P}) = \frac{1}{2}\left(1 + \sum\nolimits_{k=1}^{q} \alpha_k\big(d_i(k) - z_i(k)\big)\right) \quad [5]$$

Where $\alpha_k \geq 0, \forall k, \sum_{k=1}^{q} \alpha_k = 1$. *Moreover, the three properties are independent.*

There are three implications of this way of reframing the Proposition worth considering. (i) When judgements are complete (i.e., $z_i(k) = 0, \forall k$), the precedence function corresponds to a linear transformation of the differences between "wins and losses", so that the rankings associated with the $p_i(k)$ and the $d_i(k)$ variables, coincide. (ii) This expression illustrates the different role played by indifferences and non-comparabilities in the evaluation. Indifferences cancel out in $d_i(k)$, whereas non-comparabilities penalize the evaluation. (iii) Equation [5] also shows that the cancellation principle does not hold. This is so because we may have $d_i(k) = d_j(k), \forall k$, and yet, $v_i(\mathbf{P}) \neq v_j(\mathbf{P})$, due to the role of non-comparabilities. So, in a general framework, the ranking associated with de precedence function and that associated with Barberà & Bossert (2022) version of Borda's ranking, may differ.

3.3 A probabilistic interpretation

The precedence score, $p_{ij}(k) = \frac{1}{n}\left(n_{ij}^k + \frac{e_{ij}^k}{2}\right)$, corresponds to the probability that the alternative i precedes j in dimension k, when choosing randomly an individual $h \in N$. Similarly, the average precedence score, $p_i(k) = \frac{1}{m-1}\sum_{j \neq i} p_{ij}(k)$, is the probability that alternative i precedes some other alternative, when facing a pairwise comparison in which both the individual and the other alternative are chosen randomly. And, by the same token, $d_i(k) = \frac{1}{m-1}\left(\sum_{j\neq i} p_{ij}(k) - \sum_{j \neq i} p_{ji}(k)\right)$ is the difference between the probability that i beats some other alternative in a pairwise confrontation, and the probability that i be beaten by some other alternative.

Consequently, the precedence function tells us the probability that each alternative has of preceding some other, when the individual, the other alternative, and the corresponding dimension are randomly chosen. Note that regarding individuals and alternatives the associated probabilities exhibit a uniform distribution, whereas the probability of each dimension is given by the corresponding coefficient α_k.

We can thus rephrase the Proposition as follows:

Proposition (second alternative version): *An evaluation function* $\boldsymbol{v}: \wp \to \mathbb{R}_+^m$ *satisfies the properties of independence, symmetry, uniformity, neutrality, and scale, if and only if, with each alternative,* $i \in A$*, it associates the probability of beating some other alternative, in some dimension.*

This interpretation permits comparing the precedence function with the Borda-Condorcet rule, in Herrero & Villar (2021). In a similar framework, the Borda-Condorcet rule is derived from the probabilities of each alternative to beat another in an infinite sequence of random matchings that defines a Markov chain. The associated evaluation formula, restricted to the single-dimensional case to facilitate the discussion, is given by:

$$f_i(\mathbf{P}) = \frac{\sum_{j \neq i} p_{ij} f_j(\mathbf{P})}{\sum_{j \neq i} p_{ij}}$$

In this simplified context the precedence function would be $v_i(\mathbf{P}) = \frac{1}{m-1} \sum_{j \neq i} p_{ij}$. There are, therefore, two clear differences between both evaluation protocols. The first difference refers to the use of the "losses" to evaluate the alternatives. The precedence rule uses implicitly the *difference* between wins and losses, as discussed above, whereas the Borda-Condorcet computes explicitly the *ratio* $r_{ij}(\mathbf{P}) = \frac{p_{ij}}{\sum_{j \neq i} p_{ij}}$. That is, the probability that i beats j, divided by the probability that i be beaten by some other alternative. The second difference is that the precedence rule directly aggregates the precedence scores, whereas the Borda-Condorcet rule aggregates those relative precedence scores, $r_{ij}(\mathbf{P})$, weighted by the value of the corresponding alternatives.[3]

So, the key difference between both rules can be summarized as follows. The precedence function associates with each alternative, the one-shot probability of beating some other alternative, in some dimension. The Borda-Condorcet rule applies the same idea in terms of the probability that derives from an indefinite sequence of pairwise encounters, which can be interpreted in terms of the fraction of time that each alternative keeps the floor in this process. Which implies considering the strength of the different

[3] This implies that the evaluation of each alternative depends on the evaluation of all others and thus all have to be calculated simultaneously. This can be obtained as the dominant eigenvector of the square non-negative m-matrix $\mathbf{R}$ whose elements are those $r_{ij}(\mathbf{P})$, with $r_{ii}(\mathbf{P}) = 0$. That is, $\mathbf{R}f(\mathbf{P}) = \lambda f(\mathbf{P})$.

alternatives and so pondering the (relative) precedence scores by the worth of those competing alternatives.

3.4 Some practical features

We now discuss some practical features that this evaluation protocol exhibits. To start with, there are two minor variants in the way of applying this formula, which might be useful in some problems.

The first variant consists of setting a lower bound, greater than zero, to the evaluations. This amounts to defining $\tilde{v}_i(\mathbf{P}) = max\{v_i(\mathbf{P}), c\}$, for some scalar $c > 0$, suitably chosen. Think of the case in which the alternatives to be evaluated refer to the importance of the different dimensions, discussed above (see equation [4]). Even if all individuals agree that one of the qualities is the less important, we might be willing to give it a minimum role in the overall evaluation, rather than ignore this quality. This variant would solve this problem. As for the value of the parameter c, we can choose a number greater than zero and smaller than the minimum positive evaluation, e.g., $c = \frac{1}{2} min_i\{v_i(\mathbf{P})\ with\ v_i(\mathbf{P}) > 0\}$. Yet we can also opt for a larger value, so that all alternatives below that threshold are treated alike (this might be reasonable when using the evaluation to share a given amount of some asset and want to ensure that each alternative receives a minimum).

The second variant that may be of interest and also uses a lower bound, even though with an opposite purpose, is the following. Let $L > 0$ and set all evaluations below that threshold equal to zero. That is, define:

$$\hat{v}_i(\mathbf{P}) = \begin{cases} v_i(\mathbf{P})\ , if \geq L \\ 0, otherwise \end{cases}$$

This variant may be of interest when the evaluation is the basis for the distribution of an asset, and we want to exclude from the allocation those alternatives with low profiles. Think of the example of evaluating research projects, referred to in the Introduction. The lower bound L represents an excellence threshold that excludes those proposals that fail to achieve a minimum quality level. We can think of $L = \gamma Med\{v_i(\mathbf{P})\}_{i \in A}$, where Med stands for median, and $\gamma > 0$ is a parameter of preference for quality.

Another obvious practical feature of this evaluation formula is that it is decomposable by population subgroups. This might be important when there is a large number of evaluators with different characteristics. Think for instance of the evaluation of the efficacy of different painkillers in a medical trial. Besides rating globally those painkillers, it might be relevant to know their effects on different groups of patients (e.g., classified by age, gender, clinical record, and occupation).

To see this, let $N = \bigcup_{g=1}^{G} N^g$, where N^g is population subgroup g with cardinal n^g, so that $\sum_{g=1}^{G} n^g = n$. In this scenario, we define

$$p_i^g(k) = \frac{1}{m-1}\sum_{j\neq i}\frac{1}{n^g}\left(n_{ij}^k(g) + \frac{n_{ij}^k(g)}{2}\right)$$

Then, evaluation of this alternative in terms of the judgements of population subgroup g is given by: $v_i^g(\mathbf{P}) = \sum_{k=1}^{q} \alpha_k\, p_i^g(k)$.

Now observe that:

$$\begin{aligned} v_i(\mathbf{P}) &= \frac{1}{n}\sum_{g=1}^{G}\sum_{k=1}^{q}\alpha_k\frac{1}{(m-1)}\sum_{j\neq i}\left(n_{ij}^k(g) + \frac{n_{ij}^k(g)}{2}\right) \\ &= \sum_{g=1}^{G}\frac{n^g}{n}\sum_{k=1}^{q}\alpha_k\sum_{j\neq i}p_{ij}^g(k) \end{aligned}$$

Therefore,

$$v_i(\mathbf{P}) = \sum_{g=1}^{G}\frac{n^g}{n}v_i^g(\mathbf{P}) \qquad [6]$$

An interesting context in which this decomposability property applies refers to those problems in which different subsets of alternatives in A are evaluated by separate subsets of individuals in N. The evaluation of research projects, mentioned in the Introduction may serve to illustrate this case. Typically, each evaluator receives a subset of the proposals and then a decision is made from those partial evaluations. This can be regarded as a case of different population subgroups, each of which exhibits incomplete evaluations.

4. An application: assessing human development

Let us illustrate the working of the precedence function in an empirical application: the study of human development, according to the United Nations Development Programme's approach. We compare the evaluation provided by this rule and the UN Human Development Index (HDI), using the same data, regarding those countries with "very high human development" (see United Nations Development Programme, 2022).

The Human Development Index is a multidimensional indicator that approaches human development in terms of three key dimensions: health (H), education (E), and material wellbeing (MW). Achievements in health are measured by the variable life expectancy at birth (the number of years that a new-born is expected to live). Educational achievements are computed by a composite indicator that consists of the (geometric) mean of the mean years of schooling and the expected years of schooling. Finally, material wellbeing is associated with the logarithm of the per capita gross national income. Variables are normalized with respect to specific goalposts (a maximum and a minimum), to obtain a common range in the interval [0, 1]. The HDI consists of the geometric mean of those normalized variables, under the assumption that all dimensions are equally important. The HDI is probably the most successful and popular alternative to the GDP, with a large impact in the mass media because of its intuitive character and the large number of participating countries.

Formulated in our terms, the human development evaluation problem involves a set of 66 alternatives (countries), a single evaluator (the UNDP), and three dimensions, which are equally weighted (that is, $q = 3$, with $\alpha_H = \alpha_E = \alpha_{MW} = \frac{1}{3}$). Interpreting the values of each normalized variable as categorical, rather than quantitative, the relevant information about the evaluation problem will be summarized in three 66×66 matrices, regarding the three different dimensions, $\mathbf{P}^H, \mathbf{P}^E, \mathbf{P}^{MW}$. Entry p_{ij}^k tells how country i fares with respect to country j in dimension k (with values of either 0, 0,5 or 1). By letting $\mathbf{1}$ be the column vector all whose 66 components are equal to 1, we would have:

$$\boldsymbol{v}(\mathbf{P}) = \frac{1}{65} \times \frac{1}{3} \times (\mathbf{P}^H\mathbf{1} + \mathbf{P}^E\mathbf{1} + \mathbf{P}^{MW}\mathbf{1})$$

Table 1 provides the values and the rankings derived from the HDI and the precedence function, for the 66 "very high human development" countries. Values have been normalized so that the highest country is set to 100 for both evaluations. There are two main features that emerge from this comparison. On the one hand, that there are many changes in the ranking of the countries, with several countries shifting by more than ten positions (15 in the case of the USA) in different directions. On the other hand, we observe a much larger variability in the evaluations with the precedence function than with the HDI, with coefficients of variation of 0.462 and 0.283, respectively, and ranges of variation of 85.6 points versus 16.8 points.

Table 1: Countries with Very High Human Development Index (2021)

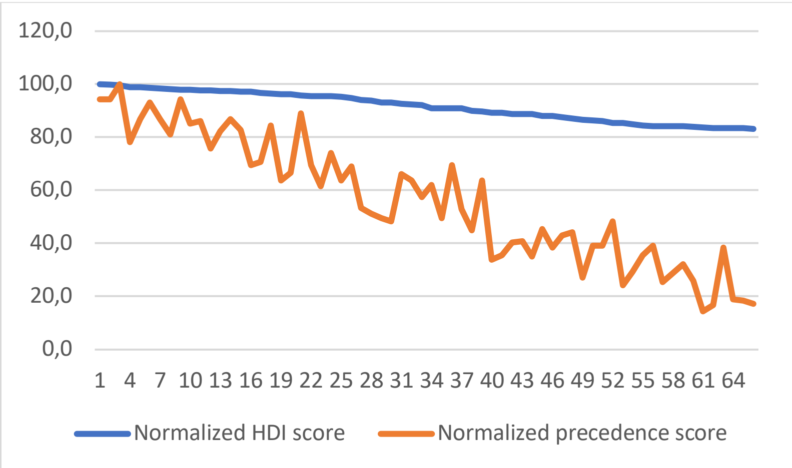
Country	HDI ranking	Normalized HDI score	Precedence Ranking	Normalized precedence score
Andorra	40	89,2	54	33,9
Argentina	47	87,5	43	43,1
Australia	5	98,9	7	86,8
Austria	25	95,2	26	63,8
Bahamas	55	84,4	51	35,6
Bahrain	35	91,0	36	49,4
Belarus	60	84,0	58	25,9
Belgium	13	97,4	14	82,2
Brunei Darussalam	51	86,2	46	39,1
Canada	15	97,3	13	82,8
Chile	42	88,9	45	40,2

Costa Rica	58	84,1	57	28,7
Croatia	40	89,2	52	35,6
Cyprus	29	93,1	37	49,4
Czechia	32	92,4	27	63,8
Denmark	6	98,5	5	93,1
Estonia	31	92,5	25	66,1
Finland	11	97,7	10	86,2
France	28	93,9	35	51,1
Georgia	63	83,4	49	38,5
Germany	9	97,9	2	94,3
Greece	33	92,2	32	57,5
Hong Kong, China (SAR)	4	99,0	16	78,2
Hungary	46	87,9	50	38,5
Iceland	3	99,7	1	100,0
Ireland	8	98,2	15	81,0
Israel	22	95,5	20	69,5
Italy	30	93,0	38	48,3
Japan	19	96,2	28	63,8
Kazakhstan	56	84,3	47	39,1
Korea (Republic of)	19	96,2	24	66,7
Kuwait	50	86,4	48	39,1
Latvia	39	89,7	29	63,8
Liechtenstein	16	97,2	21	69,5
Lithuania	35	91,0	22	69,5
Luxembourg	17	96,7	19	70,7
Malaysia	62	83,5	65	16,7
Malta	23	95,4	31	61,5
Mauritius	63	83,4	62	19,0
Montenegro	49	86,5	58	27,0
Netherlands	10	97,8	11	85,1
New Zealand	13	97,4	8	86,8
Norway	2	99,9	3	94,3
Oman	54	84,8	56	29,3
Panama	61	83,7	66	14,4
Poland	34	91,1	30	62,1
Portugal	38	90,0	41	44,8
Qatar	42	88,9	44	40,8
Romania	53	85,3	61	24,1
Russian Federation	52	85,4	39	48,3
San Marino	44	88,7	53	35,1
Saudi Arabia	35	91,0	34	52,9
Serbia	63	83,4	63	18,4
Singapore	12	97,6	17	75,9
Slovakia	45	88,1	40	45,4
Slovenia	23	95,4	18	74,1
Spain	27	94,1	33	53,4

Sweden	7	98,4	9	86,8
Switzerland	1	100,0	4	94,3
Thailand	66	83,2	64	17,2
Trinidad and Tobago	57	84,2	60	25,3
Türkiye	48	87,1	42	44,3
United Arab Emirates	26	94,7	23	69,0
United Kingdom	18	96,6	12	84,5
United States	21	95,7	6	89,1
Uruguay	58	84,1	55	32,2

Figure 1 gives a visual appraisal of the differences in the evaluations provided by those two protocols. The countries are arranged on the horizontal axis in decreasing order according to the HDI, so that the HDI line is a monotonously decreasing function. The line that describes the evaluation made with the precedence function is always below that of the HDI (except for the top value, by construction) and exhibits jumps that mark the changes in the ranking. In this way both the increase in the spread and the differences in the rankings are easily spotted.

Figure 1: Normalized scores for the HDI and the precedence rule for the 66 "very high HDI" countries



One may wonder how a criterion that disregards the quantitative differences in the variables involved can produce an evaluation that discriminates more among the countries than the HDI, which is a geometric mean and computes all those differences in magnitude. One of the reasons is the normalization process adopted by the UNDP when building the HDI. Firstly, note that the per capita GNI, which is by far the variable with greatest variability, is much flattened by taking logs. Secondly, observe that the health and material wellbeing variables are also ironed out by deducting from each value the minimum value and dividing by the range. This makes all variables move in the [0, 1] interval, but alters both the original rankings and their relative values, compressing their differences (see Herrero, Martínez & Villar, 2012, for a detailed discussion).

From this it follows that it would have been more sensible to use the values of the variables before normalization, rather than after as we did (taking logs does not affect the rankings, so that this operation would not alter the outcome). Yet we have preferred using exactly the values utilized in the calculation of the HDI, to facilitate the comparison between both evaluation protocols.

5 Wrapping up

We have presented an elementary way of summarizing quantitatively the qualitative assessments of a set of individuals, regarding multidimensional evaluation problems, without requiring those individual judgements to be complete or transitive.

Non-transitive or incomplete judgments appear in many evaluation problems. They may derive from several factors, including the presence of many alternatives, which makes coherence difficult; the insufficient knowledge on the nature or significance of some of those alternatives; the existence of costs when making comparisons; the allocation of subsets of alternatives to subsets of individuals; the use of several principles by the individuals when rating each quality; tiredness or insufficient attention; or the presence of focal alternatives, which concentrates the attention on some of them and makes fuzzy the rest. None of those factors imply "irrational" behaviour on the individuals who judge the alternatives, and yet they often appear in real life decision problems. It is interesting, therefore, having evaluation protocols that permit handling those situations.

Pairwise comparisons enable the evaluation of alternatives in this scenario in terms of the precedence scores, that is, the number of times that each alternative beats the others in those pairwise confrontations. The precedence function may thus be regarded as an extension of the Borda score, which can be easily characterized in terms of simple properties, it is straightforward to compute, and transparent. Moreover, this elementary algorithm helps dealing with many different families of evaluation problems, well beyond the realm of voting or social choice.

REFERENCES

Barberà, S. & Bossert, W. (2022), Opinion aggregation: Borda and Condorcet revisited, **BSE working paper** 1262 (revised).

Borda, J.C. de (1784). **Mémoire sur les Elections au Scrutin**. Paris, Historie de lÁcademie Royale des Sciences.

Condorcet, M.J.A.N. C, Marquis de (1785). **Essai sur l'Application de l'Analyse a la Probabilité des Décisions Redues à la Pluralité del Voix.** Paris: De l'Imprimerie Royale.

Hansson B, Sahlquist H (1976) A proof technique for social choice with variable electorate, **Journal of Economic Theory**, 13 : 193–200.

Herrero, C., Martínez, R. & Villar, A. (2012), A Newer Human Development Index, **Journal of Human Development and Capabilities**, 13 : 247-268.

Herrero, C. & Villar, A. (2021), Group decisions from individual rankings: The Borda-Condorcet rule, **European Journal of Operational Research**, 291: 757-765.

Mihara, H. R. (2017). Characterizing the Borda ranking rule for a fixed population, MPRA paper, nº 78093.

Moulin, H. (1988). **Axioms of Cooperative Decision Making**. Cambridge, Cambridge University Press.

Nitzan, S., & Rubinstein, A. (1981), A further characterization of Borda ranking method, **Public choice**, 153-158.

United Nations Development Programme (2022), **Human Development Report (2021/22): Uncertain times, unsettled lives: Shaping our future in a transforming world**, UN, New York.

Young, H.P. (1974), An axiomatization of Borda's rule, **Journal of Economic Theory**, 9 : 43–52.

APPENDIX: Chracterization of the Precedence function

To formalize the property of independence, let $\mathbf{p}_{(i)}$ denote the ith row of matrix $\mathbf{P}$, and $\mathbf{P}_{(-i)}$ the remaining $(m-1)$ rows. Then:

Independence: Let $\mathbf{P} = \left(\mathbf{p}_{(i)}, \mathbf{P}_{(-i)}\right), \mathbf{P}' = \left(\mathbf{p}_{(i)}, \mathbf{P}'_{(-i)}\right) \in \wp$. Then, $v_i(\mathbf{P}) = v_i(\mathbf{P}')$.

Similarly, to define the property of symmetry, let us describe matrix $\mathbf{P}$ as $\left(\mathbf{p}_{(i)}, \mathbf{p}_{(j)}, \mathbf{P}_{(-i,j)}\right)$, where $\mathbf{p}_{(i)}, \mathbf{p}_{(j)}$ correspond to the ith and jth rows of matrix $\mathbf{P}$, and $\mathbf{P}_{(-i,j)}$ to the remaining $(m-2)$ rows.

Symmetry: Let $\mathbf{P} = \left(\mathbf{p}_{(i)}, \mathbf{p}_{(j)}, \mathbf{P}_{(-i,j)}\right) \in \wp$, with $\mathbf{p}_{(i)} = \mathbf{p}_{(j)}$. Then, $v_i(\mathbf{P}) = v_j(\mathbf{P})$.

The next two properties, uniformity and neutrality, refer to the sensitivity of the evaluation of an alternative, to changes of size x in the precedence scores. [4]

Uniformity: Let $\mathbf{P}, \mathbf{P}', \mathbf{P}'' \in \wp$ be such that: (a) $p'_{ij}(k) = p_{ij}(k) + x, p'_{ji}(k) = p_{ji}(k) - x$, $p''_{ij}(k) = p'_{ij}(k) + x, p''_{ji}(k) = p'_{ji}(k) - x$; (b) $p''_{st}(k) = p'_{st}(k) = p_{st}(k), \forall (s,t) \neq (i,j)$, $(s,t) \neq (j,i)$ all k. Then, $v_i(\mathbf{P}') - v_i(\mathbf{P}) = v_i(\mathbf{P}'') - v_i(\mathbf{P}')$.

Neutrality: Let $\mathbf{P}, \mathbf{P}', \mathbf{P}'' \in \wp$ be such that: (a) $p'_{ij}(k) = p_{ij}(k) + x, p'_{ji}(k) = p_{ji}(k) - x$, $p''_{ir}(k) = p_{ir}(k) + x, p''_{ri}(k) = p_{ri}(k) - x$; (b) $p''_{st}(k) = p'_{st}(k) = p_{st}(k), \forall (s,t) \neq (i,j), (j,i), (i,r), (r,i)$, all k. Then, $v_i(\mathbf{P}') - v_i(\mathbf{P}) = v_i(\mathbf{P}'') - v_i(\mathbf{P})$.

Finally,

Scale: $p_{ij}(k) = 1, \forall j \neq i, \forall k \implies v_i(\mathbf{P}) = 1$, and $p_{ij}(k) = 0, \forall j \neq i, \forall k \implies v_i(\mathbf{P}) = 0$.

Proposition: *An evaluation function* $\boldsymbol{v}: \wp \to \mathbb{R}^m_+$ *satisfies the properties of independence, symmetry, uniformity, neutrality, and scale, if and only if it is given by:*

$$v_i(\mathbf{P}) = \sum_{k=1}^{q} \alpha_k p_i(k), \qquad \forall\, i \in M \qquad [3]$$

Where $\alpha_k \geq 0, \forall k, \sum_{k=1}^{q} \alpha_k = 1$. *Moreover, these properties are independent.*

[4] Note that, as $\mathbf{P}, \mathbf{P}', \mathbf{P}'' \in \wp$, the change x is bound to be small enough to avoid yielding negative precedence scores.

Proof.-

Clearly, this function satisfies those properties. Let us consider the converse implication.

First observe that independence establishes that $v_i(\mathbf{P}) = f\left(\mathbf{p}_{(i)}\right)$. Uniformity and independence together imply that the impact of a given change in the jth variable is independent of the level of the variable, that is, $\frac{\Delta v_i}{p_{ij}(k)+x} = a_{ij}(k)$, for some constant $a_{ij}(k)$, all x. We can, therefore, write:

$$v_i(\mathbf{P}) = \sum_{k=1}^{q} \sum_{j\neq i} a_{ij}(k)\, p_{ij}(k) + b_{ij}(k)$$

It now follows from independence and symmetry that $\mathbf{p}_{(i)} = \mathbf{p}_{(j)}$ implies $v_i(\mathbf{P}) = v_j(\mathbf{P})$, so that $a_{ij}(k) = a_{rj}(k)$. Moreover, neutrality and independence imply that $\frac{\Delta v_i}{p_{ij}(k)+x} = \frac{\Delta v_i}{p_{ir}(k)+x}$. By combining those four properties, we get: $a_{ij}(k) = a(k), \forall i,j \in A, \forall k$. Therefore,

$$v_i(\mathbf{P}) = \sum_{k=1}^{q} \sum_{j\neq i} a(k)\, p_{ij}(k) + b(k)$$

As $p_{ij}(k) = 0, \forall j \neq i, \forall k \implies v_i(\mathbf{P}) = 0$ (scale), it follows that $b(k) = 0$, $\forall k$. Moreover, $p_{ij}(k) = 1, \forall j \neq i, \forall k \implies v_i(\mathbf{P}) = 1$ so that $(m-1)\sum_{k=1}^{q} a(k) + 0 = 1, \forall i,j$. Therefore, by letting $\alpha_k = (m-1)a(k), \forall k$, it follows that:

$$v_i(\mathbf{P}) = \sum_{k=1}^{q} \alpha_k \frac{1}{(m-1)} \sum_{j\neq i} p_{ij}(k) = \sum_{k=1}^{q} \alpha_k p_i(k)$$

with $\sum_{k=1}^{q} \alpha_k = 1$.

To check that all those properties are independent, consider the evaluation functions, whose ith component is given by:

(i) $v_i(\mathbf{P}) = \sum_{k=1}^{q} \alpha_k \frac{1}{2} \sum_{j\neq i}\left(p_{ij}(k) + p_{mj}(k)\right)$

(ii) $v_i(\mathbf{P}) = \sum_{k=1}^{q} \sum_{j\neq i} a_i(k)\, p_{ij}(k)$

(iii) $v_i(\mathbf{P}) = \sum_{k=1}^{q} \alpha_k [p_i(k)]^{1/2}$

(iv) $v_i(\mathbf{P}) = \sum_{k=1}^{q} \sum_{j\neq i} a_j(k)\, p_{ij}(k)$

(v) $v_i(\mathbf{P}) = 1 + \sum_{k=1}^{q} \alpha_k\, p_i(k)$

Function (i) satisfies all properties but independence. Function (ii) satisfies all the properties except symmetry. Function (iii) satisfies all properties but uniformity. Function (iv) satisfies all properties but neutrality. Finally, function (v) satisfies all properties except scale.

Q.e.d.