
Volume 146 Number 6 November 2011 ISSN 0022-0531
ELSEVIER
JOURNAL OF
Economic
Theory
Editors: Christian Hellwig and Karl Shell
Associate Editors
Marco Battaglini
Dirk Bergemann
Alberto Bisin
Yeon-Koo Che
In-Koo Cho
Olivier Compte
Philip H. Dybvig
Larry G. Epstein
Yossi Feinberg
Christian Ghiglino
Jacob Goeree
Piero Gottardi
Jean-Michel Grandmont
Hugo Hopenhayn
Johannes Horner
Vijay Krishna
Felix Kubler
Ricardo Lagos
Massimo Marinacci
Thomas Mariotti
David Martimort
Kiminori Matsuyama
Richard McLean
Giuseppe Moscarini
Efe Ok
Alessandro Pavan
James Peck
Lasse Heje Pedersen
Fabrizio Perri
Benjamin Polak
Stephen A. Ross
Hamid Sabourian
William H. Sandholm
Manuel S. Santos
Chris Shannon
Marciano Siniscalchi
Lones Smith
William L. Thomson
Juuso Välimäki
Randall Wright
Administrative Editors: Edward McDevitt and Susan Witherow Schulze
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal of Economic Theory 146 (2011) 2440–2454

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Income distributions and decomposable divergence measures ☆

Brice Magdalou *, Richard Nock

CEREGMIA, Université des Antilles et de la Guyane, Campus de Schoelcher, B.P. 7209, F-97275 Schoelcher Cedex, France

Received 22 December 2009; final version received 23 April 2011; accepted 16 June 2011

Available online 12 October 2011

Abstract

Inequality indices evaluate the divergence between the income distribution and the hypothetical situation where all individuals receive the mean income, and are unambiguously reduced by a Pigou–Dalton progressive transfer. This paper proposes a new approach to evaluate the divergence between any two income distributions, where the latter can be a reference distribution for the former. In the case where the reference distribution is perfectly egalitarian – and uniquely in this case – we assume that any progressive transfer reduces the divergence, and that the divergence can be additively separated into inequality and efficiency loss. We characterize the unique class of decomposable divergence measures consistent with these views. We derive the associated relative and absolute subclasses, and we illustrate the applicability of our results. This approach extends the generalized entropy studied in inequality measurement.

JEL classification: D31; D63

Keywords: Inequality measures; Mobility; Responsibility; Heterogeneous households; Generalized entropy; Bregman divergences

☆ This paper forms part of the research projects The Multiple Dimensions of Inequality (Contract No. ANR-BLANC-1808-03) and Computational Information Geometry and Applications (Contract No. ANR-07-BLAN-0328-01) of the French National Agency for Research whose financial support is gratefully acknowledged. We are indebted to an associate editor and two anonymous referees whose comments have helped us to substantially improve the paper. We also would like to thank Éric Briys, David Gray, Karen Jackson, Stephen Jenkins and Patrick Moyes for very useful discussions, comments or suggestions.

* Corresponding author. Fax: +33 (0)596 727 403.

E-mail addresses: brice.magdalou@martinique.univ-ag.fr (B. Magdalou), richard.nock@martinique.univ-ag.fr (R. Nock).

0022-0531/$ – see front matter
doi:10.1016/j.jet.2011.06.017

1. Introduction and overview

When individuals are identical in every aspect other than their respective incomes, theories of justice agree that an egalitarian distribution might be the best outcome for society as a whole. In this context, there is a consensus in the literature to use inequality indices, or dominance criteria such as the Lorenz quasi-ordering, for making normative judgments about the fairness of the income distribution. In practice however, individuals differ in many respects and an equally distributed income is no longer a social norm. As an immediate consequence conventional inequality indices become meaningless, unless other evaluation tools are introduced.

The aim of this paper is to provide a unified framework for the economic analysis of income distributions, when objectives other than the strict equality of incomes are considered. We build upon previous works on inequality measurement, by rethinking or extending some usual normative views. We then identify, through an axiomatic characterization, a large class of measures. The conditions we impose are fairly reasonable, and not very demanding. More importantly, we claim that such an approach may shed new light on important issues in inequality measurement.

The cornerstone of the inequality measurement theory is the *Pigou–Dalton principle of transfers*. This principle states that *any* progressive transfer from an individual to one poorer than him – transfer that does not modify the respective positions on the income scale – always reduces the inequality. Even if this principle is well established, it is not immune to criticism, and indeed not universally approved [5]. The principle of transfers actually encapsulates two normative views. Other things being equal, it first defines a *path* which characterizes an unambiguous improvement of the social welfare. Then, it describes a strictly egalitarian distribution as a *social objective*, since the equalizing process is completed when all individuals have the mean income.

These two dimensions have been separately investigated and criticized in the literature. First, whereas the income inequality is unambiguously reduced among the individuals involved in a progressive transfer, it is not so obvious that the inequality is also decreased in society as a whole [15]. Some combinations of progressive transfers can modify the distribution in a questionable direction, resulting for example in an increase in *polarization* [22,38]. Then, a strictly egalitarian distribution does not necessarily appear as a reference point for the social planner. Some income inequality, for example stemming from differences in personal responsibility – such as effort – may be viewed as fair, and might not be compensated [6,10,24,35].

In this paper we assume as a first normative requirement that, for any given income distribution (denoted by $\boldsymbol{x}$), there exists a *representative, reference or objective distribution* (denoted by $\boldsymbol{y}$). This view significantly weakens the second feature of the principle of transfers. Depending on the situation in which the measure is applied, the reference distribution can be, for example, fair according to the ethics of responsibility. We do not characterize this reference. We only assume its existence. Hence, other literature has to be invoked to complement our approach. There are now several possibilities to define what exactly is meant by *improvement*, holding total income constant, to get closer to the reference distribution. The approach we embrace in this paper is not really innovative, even if it slightly weakens the standard view of the principle of transfers – according to which a progressive transfer is *always* a suitable transformation. We assume that a progressive transfer is always an admissible path if, in the reference distribution, every individual has the mean income of the initial distribution. But we also assume that the effect of such a transfer may be ambiguous if the reference distribution is not fully egalitarian. This property is simply called *principle of transfers*, even if our definition is weaker than the standard one.

The second normative requirement we impose, relatively new in the literature, involves a situation where the mean income of the actual distribution $\boldsymbol{x}$ and the reference distribution $\boldsymbol{y}$ may be different. We assume that, if the reference distribution is egalitarian – any income c for all – the measures can be additively decomposed into two components. The first component evaluates the *inequality* within the distribution $\boldsymbol{x}$, or equivalently, the divergence between $\boldsymbol{x}$ and the hypothetical situation where all individuals have the mean. The second component evaluates the *efficiency loss*, due to the divergence between the hypothetical situation where all individuals have the mean and the reference distribution, characterized by the income c for all. This property, called *judgment separability*, is undoubtedly the key element of our approach. Finally, if the actual distribution corresponds exactly to the distribution the social planner wants to achieve, it seems fair to recognize that only the status quo is acceptable. This is a consequence of the two first requirements.

Adding some reasonable conditions to the normative judgments described above, we characterize a large class of decomposable measures which evaluate the divergence between any two income distributions, where the second one can be a reference for the first. As explained later, we focus on *divergence measures*, as opposed to *distance measures*. What follows is a brief overview of the measures we obtain

$$D_\phi^n(\boldsymbol{x} \parallel \boldsymbol{y}) = \sum_{i \in \mathcal{N}} \left[\phi(x_i) - \phi(y_i) - (x_i - y_i)\phi'(y_i)\right],$$

where $\mathcal{N}$ denotes the considered population, consisting of n individuals. ϕ is a twice differentiable and strictly convex function. Traditional inequality indices can be replaced in this context. As already mentioned, an inequality index implicitly evaluates the divergence between the income distribution under consideration and the hypothetical situation where all individuals have the mean income. We show that our divergence measures extend the usual decomposable families of inequality indices: Under some restrictive conditions, or normalizations, they boil down to the relative and absolute versions of the *generalized entropy* initiated by [16,18,36,37].

The evaluation of the divergence between any two income distributions is not really new in the literature. Cowell [17] characterizes a large class of divergence measures, called *measures of distributional change*. Nevertheless our approach is conceptually different, on two main features, namely, (i) the properties required for the measures, and (ii) the measures obtained. (i) Since divergence measures generalize inequality indices, Cowell [17] proposes to generalize the principle of transfers to a property called *monotonicity in distance*. Whereas it represents an appropriate extension of the principle of transfers in a more general framework, this property is quite demanding. The property we impose, called simply principle of transfers, is weaker. Moreover, the measures identified by [17] are not consistent with the other main property we assume, called judgment separability. (ii) The divergence measures obtained by [17] and the divergence measures characterized in the present paper are different. Surprisingly enough, both classes of measures are widely used in *information theory* and *information geometry*. The relative measures identified by [17] are *Csiszár f-divergences*, independently introduced by [19] and [2]. The measures we obtain are *Bregman divergences* [12]. A well-known result in information theory is that the classes are distinct, but coincide in one specific case: This divergence – which is a single measure – is called *Kullback–Leibler divergence* [29]. Also, Cowell [17] characterized a class of absolute measures. There is only one divergence at the intersection of this class and our class, namely the *squared Euclidean distance*, a generalization of the variance.

This paper is only a first step towards the new direction we propose. Its distinctive feature lies in the fact that we use one distribution as a reference. Although the characteristics of this distribution have to be specified, this is not really a weakness. Indeed, it leaves the application of the measures open to a wide number of fields where income distributions are involved. The real question is to preserve the standard principle of transfers, as accepted in this paper, when the income is equally distributed in the reference distribution. Accordingly, some modifications of the distributions, such as polarization, cannot be tracked. That seems however to be a necessary prelude. Based on the results derived here, many extensions can be undertaken, by redefining the path towards the reference distribution.

The following section sets out the notation and illustrates our framework in the context of inequality measurement. We introduce in Section 3 the main conditions, namely smoothness, judgment separability, anonymity and the principle of transfers, which will be imposed on all divergence measures. We isolate a class of measures compatible with them. The class is characterized by a single evaluation function. The consistency with the principle of transfers, combined with the anonymity requirement, is captured by the *strict Schur-convexity* of this function. We then focus on decomposable measures, consistent with a non-negativity requirement. The implication is an additive structure. The entire class, compatible with all the properties, is finally identified. Section 4 hints at some directions to refine the general class by way of normalizations. We also discuss the relationship with the related literature. Section 5 concludes the paper and illustrates how the measures can be applied. We focus on the analysis of households with differing needs, the ethics of responsibility, and mobility measurement.

2. The framework

We consider a population $\mathcal{N} := \{1, 2, \dots, n\}$ consisting of $n \geqslant 1$ individuals. An income distribution for population $\mathcal{N}$ is a list $\boldsymbol{x} := (x_1, x_2, \dots, x_n)$, where $x_i \in \mathscr{D}$ is the income of individual $i \in \mathcal{N}$, and $\mathscr{D}$ is a closed and bounded interval of the real line $\mathbb{R}$. Given a distribution $\boldsymbol{x} \in \mathscr{D}^n$, the mean income is indicated by $\mu(\boldsymbol{x}) := \sum_{i=1}^{n} x_i/n$. We will write $\mu(\boldsymbol{x})$ or sometimes μ for the mean of $\boldsymbol{x}$ if no ambiguity arises. We denote the normalized distribution of $\boldsymbol{x} \in \mathscr{D}^n$ as $\hat{\boldsymbol{x}} := (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$ and the centered distribution as $\tilde{\boldsymbol{x}} := (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$, where respectively $\hat{x}_i := x_i/\mu$ and $\tilde{x}_i := x_i - \mu$ for all $i \in \mathcal{N}$. We let $\mathbf{1}_n := (1, \dots, 1)$ represent a list where 1 is repeated n times. Then, we define $D^n : \mathscr{D}^n \times \mathscr{D}^n \longrightarrow \mathbb{R}$ as a function which evaluates the divergence between two income distributions in $\mathscr{D}^n$ of the same size.[1] A *divergence measure* consists of a countable sequence $\{D^n\colon\ n = 1, 2, \dots\}$, which contains exactly one function D^n for every population size $n \geqslant 1$. In order to simplify notation we let, in the following, D^n indicate a measure where n is treated as a parameter. Mathematically speaking, a divergence measure is different from a *distance measure*, as it needs to neither be symmetric nor satisfy the triangle inequality. To emphasize this point, we write $D^n(\boldsymbol{x} \parallel \boldsymbol{y})$, instead of $D^n(\boldsymbol{x}, \boldsymbol{y})$, in what follows for any divergence D^n. At this stage, no assumptions are imposed on D^n.

In this paper, $\boldsymbol{y}$ is considered as a *representative, reference or objective distribution* of $\boldsymbol{x}$, from which the latter distribution is compared. Traditional inequality indices can be viewed in this framework, letting the reference distribution explicit. For example, Shorrocks [36,37] characterizes the following additive divergence:

[1] The requirement of a common size for both distributions can be perceived as a severe restriction to make a divergence measure applicable in practice, since it is unusual to observe two empirical distributions with such a common characteristic – unless one distribution is explicitly constructed from the other. Nevertheless empirical distributions are traditionally divided in *quantiles*, and each quantile is represented by the mean income within the quantile. Then the empirical distributions are replaced by the distributions of *quantile means*, and the common size of the distributions corresponds to the fixed number of such points.

$$D^n(\boldsymbol{x} \parallel \mu \mathbf{1}_n) = \sum_{i \in \mathcal{N}} \bigl[\phi(x_i) - \phi(\mu)\bigr], \tag{1}$$

where the reference distribution is the mean income of $\boldsymbol{x}$ for all the individuals, and $\phi : \mathscr{D} \longrightarrow \mathbb{R}$ is a strictly convex function. The author then deduces, by an appropriate normalization, a large family of inequality indices called in this paper *generalized relative entropies*. If the population can be partitioned into disjoint subgroups, the reference income of a subgroup becomes the subgroup mean. The mean income for all as reference distribution is not the only possibility. Other reference distributions have been studied in the literature, with the common characteristic to define a unique reference income for all the individuals within the population (or the subgroup). For example, Blackorby et al. [8] argue that the subgroup mean has to be replaced by the equally-distributed-equivalent income of the subgroup, and Foster and Shneyerov [25,26] justify the use of the q-order mean. We propose in this paper to weaken this view. We let flexible the choice of the reference distribution, so that each individual can have his own reference income.

3. A class of decomposable divergence measures

We characterize a large class of divergence measures, by extending the usual framework of the inequality measurement theory. The main feature of our approach is to let $\boldsymbol{y}$ be the reference distribution of $\boldsymbol{x}$ in the majority of the properties required for the measures.

Property 1 *(Smoothness).* For all $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{D}^n$, $D^n(\boldsymbol{x} \parallel \boldsymbol{y})$ has continuous first-order partial derivatives.

It corresponds to a regularity condition which implies that small changes in the distributions $\boldsymbol{x}$ or $\boldsymbol{y}$ result in small changes in the divergence. We assume the differentiability of D^n, essentially for mathematical convenience. This condition could be weakened, but at the cost of openness in the characterization. Note that the differentiability is not useful until Theorem 1.

Property 2 *(Judgment separability).* For all $\boldsymbol{x} \in \mathscr{D}^n$ and all $c \in \mathscr{D}$, we have

$$D^n(\boldsymbol{x} \parallel c\mathbf{1}_n) = D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) + D^n(\mu\mathbf{1}_n \parallel c\mathbf{1}_n). \tag{2}$$

Judgment separability is the key property for our analysis. Suppose that the reference distribution $\boldsymbol{y} \in \mathscr{D}^n$ is equally distributed among the individuals, such that $y_i = c$ for all $i \in \mathcal{N}$. This property assumes that the divergence between the actual distribution $\boldsymbol{x}$ and the reference distribution $c\mathbf{1}_n$ is additively separable into two components. The first component evaluates the *inequality within the distribution* $\boldsymbol{x}$, which corresponds to the divergence between $\boldsymbol{x}$ and the hypothetical situation where all individuals have the mean income μ. The second component evaluates the *efficiency loss* which results from the divergence between the mean income μ and the social objective c. Put precisely, the second component measures the divergence between the hypothetical situation where all individuals have μ and the hypothetical situation where all individuals have c.

We complete the framework by incorporating two standard assumptions of normative economics, in order to define the *preference for equality*.

Property 3 *(Anonymity).* For all $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{D}^n$ and all $n \times n$ permutation matrix Π, we have $D^n(\boldsymbol{x}\Pi \parallel \boldsymbol{y}\Pi) = D^n(\boldsymbol{x} \parallel \boldsymbol{y})$.

Anonymity means that the evaluation of the divergence between the actual distribution $\boldsymbol{x}$ and the reference distribution $\boldsymbol{y}$ is not affected by a permutation of the identity of the individuals. Note that a measure D^n, consistent with Property 3, is invariant to a simultaneous and identical permutation for both distributions $\boldsymbol{x}$ and $\boldsymbol{y}$. A stronger version of this property – which does not make sense in our framework – should be to require that $D^n(\boldsymbol{x}\Pi \parallel \boldsymbol{y}\Pi') = D^n(\boldsymbol{x} \parallel \boldsymbol{y})$ for two, possibly different, $n \times n$ permutation matrices Π and Π'. Then, it is typically assumed that inequality is reduced by a transfer of income from a richer to a poorer individual. Precisely, given two distributions $\boldsymbol{x}, \boldsymbol{x}' \in \mathscr{D}^n$, we will say that $\boldsymbol{x}'$ is obtained from $\boldsymbol{x}$ by means of a *progressive transfer* if there exist an income amount $\Delta > 0$ and two individuals $h, k \in \mathcal{N}$ such that:

$$x_i' = x_i, \quad \forall i \neq h, k, \quad \text{and} \tag{3a}$$

$$x_h' = x_h + \Delta \leqslant x_k - \Delta = x_k'. \tag{3b}$$

By analogy with the inequality literature, we assume that a progressive transfer unambiguously reduces the divergence between $\boldsymbol{x}$ and the reference distribution, *whenever* this last distribution is the mean income of $\boldsymbol{x}$ for all the individuals, *but* only in this case.

Property 4 *(Principle of transfers).* For all $\boldsymbol{x}, \boldsymbol{x}' \in \mathscr{D}^n$, if $\boldsymbol{x}'$ is obtained from $\boldsymbol{x}$ by means of a *progressive transfer*, then we have $D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) > D^n(\boldsymbol{x}' \parallel \mu\mathbf{1}_n)$.

As noted earlier, the notion of progressive transfer does not make sense in the general situation where the reference distribution is not egalitarian. In accordance with this, our definition of the principle of transfers does not require that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{D}^n$, if $\boldsymbol{x}'$ is obtained from $\boldsymbol{x}$ by means of a progressive transfer, then $D^n(\boldsymbol{x} \parallel \boldsymbol{y}) > D^n(\boldsymbol{x}' \parallel \boldsymbol{y})$. We say that a function $\phi^n : \mathscr{D}^n \longrightarrow \mathbb{R}$ is *strictly Schur-convex* if for all distributions $\boldsymbol{x} \in \mathscr{D}^n$ and an $n \times n$ *bistochastic matrix* B, we have $\phi^n(\boldsymbol{x}) > \phi^n(\boldsymbol{x}B)$ whenever $\boldsymbol{x}B$ is not a permutation of $\boldsymbol{x}$, and $\phi^n(\boldsymbol{x}) = \phi^n(\boldsymbol{x}B)$ otherwise (see [32]). If we add anonymity and the principle of transfers to Properties 1 and 2 we obtain the following proposition. All the proofs are in Appendix A.

Proposition 3.1. *If a measure D^n satisfies Properties 1 to 4, then there exists a continuous and strictly Schur-convex function $\phi^n : \mathscr{D}^n \longrightarrow \mathbb{R}$, such that for all $\boldsymbol{x} \in \mathscr{D}^n$, we have*

$$D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) = \phi^n(\boldsymbol{x}) - \phi^n(\mu\mathbf{1}_n). \tag{4}$$

By the Schur-convexity of ϕ^n, we have $D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) \geqslant 0$ for all $\boldsymbol{x} \in \mathscr{D}^n$. Hence, Proposition 3.1 isolates an *extended class of entropy measures*, possibly non-additive, which admits as a special case the measure presented in (1) and proposed by [36,37]. This extension is familiar from the measurement of social welfare. In particular, the well-known utilitarian model, with a strictly concave utility function, is an additive subclass of the general family of the strictly Schur-concave social welfare functions.

Property 5 *(Non-negativity).* For all $c, c' \in \mathscr{D}$, we have $D^1(c \parallel c') \geqslant 0$.

Non-negativity might be imposed for all the values $n \geqslant 1$ of the measure D^n. Nevertheless, non-negativity for $n = 1$ is sufficient to obtain our main characterization result.

In this paper, we only focus on *decomposable measures*. To clarify this, assume that the whole population $\mathcal{N}$ can be decomposed into G subgroups $\mathcal{N}_g$ consisting of n_g individuals each. Precisely, we let $\{\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_{\mathrm{G}}\}$ be a partition of $\mathcal{N}$. The income distribution of subgroup g is denoted by $\boldsymbol{x}_g := (x_{g,1}, x_{g,2}, \ldots, x_{g,n_g}) \in \mathscr{D}^{n_g}$, with mean income $\mu_g \equiv \mu(\boldsymbol{x}_g)$. The following property was introduced by [23]:

Property 6 *(Decomposability).* For all $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{D}^n$ with $n \geqslant 2$, and a partition $\{\mathcal{N}_1, \mathcal{N}_2\}$ of $\mathcal{N}$, we have

$$D^n(\boldsymbol{x} \parallel \boldsymbol{y}) = F^n\big(D^{n_1}(\boldsymbol{x}_1 \parallel \boldsymbol{y}_1), D^{n_2}(\boldsymbol{x}_2 \parallel \boldsymbol{y}_2)\big), \tag{5}$$

where $F^n : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ is a continuous function.

It is not required here that the function F^n is strictly increasing in its arguments. Yet, this is a consequence of the consistency with the principle of transfers and the non-negativity assumption, as observed in the following result.

Proposition 3.2. *If a measure D^n satisfies Properties 1 to 6, then the function F^n in (5) is symmetric and it is strictly increasing on $\mathbb{R}_+ \times \mathbb{R}_+$. Furthermore, $F^n(0,0) = 0$, so $F^n(\mathbb{R}_+ \times \mathbb{R}_+) \subseteq \mathbb{R}_+$.*

Therefore, for all distributions $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{D}^n$, under Properties 1 to 6, $D^n(\boldsymbol{x} \parallel \boldsymbol{y}) \geqslant 0$ and the equality holds if and only if $\boldsymbol{x} = \boldsymbol{y}$. Thus D^n attains its minimum value if the distribution $\boldsymbol{x}$ under consideration is exactly equal to the reference distribution $\boldsymbol{y}$. The following result demonstrates that under Property 6, the measure D^n has an additive structure.

Proposition 3.3. *If a measure D^n satisfies Properties 1 to 6, then $D^n(\boldsymbol{x} \parallel \boldsymbol{y}) = \sum_{i\in\mathcal{N}} D^1(x_i \parallel y_i)$, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{D}^n$.*

Decomposability has a rather strong implication: It automatically excludes non-separable measures that belong to the class of rank-dependent inequality indices, such as the *Gini index*.[2] Nevertheless, the class of decomposable divergence measures is sufficiently large to encompass many inequality indices, such as the generalized relative and absolute entropies, or the – ordinally equivalent – *Atkinson–Kolm–Sen* and *Kolm–Pollak* families of normative indices.

We now present the main result of this paper:

Theorem 1. *A measure D^n satisfies Properties 1 to 6 if and only if $D^n \equiv D^n_\phi$ where, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{D}^n$:*

$$D^n_\phi(\boldsymbol{x} \parallel \boldsymbol{y}) := \sum_{i\in\mathcal{N}} \big[\phi(x_i) - \phi(y_i) - (x_i - y_i)\phi'(y_i)\big], \tag{6}$$

and $\phi : \mathscr{D} \longrightarrow \mathbb{R}$ is a twice differentiable and strictly convex function.

The measure D^n_ϕ is not really new: It is well known and widely used in information theory; It is called *Bregman divergence* [12,4]. An alternative characterization is proposed by [7] in information theory.

We have proposed in this section an intuitive and relatively undemanding extension of standard inequality measures, which can be written in the form (1). The only new requirement is

[2] Property 6 is, actually, the strong version of decomposability in [23].

Property 2, namely judgment separability. We will discuss this property in Section 4.2. Moreover, Property 2 implies an asymmetric structure for D^n_ϕ, which is discussed in Section 5 (application to mobility measurement).

4. Subclasses and relationship with the literature

4.1. Normalized divergence measures

This section aims at refining the measures D^n_ϕ characterized in Theorem 1. It clarifies the relationship with earlier work on information theory and distributional changes. To apply the measures to real data, we need to introduce some parametric functions ϕ. Consider for example the strictly convex function, associated to the generalized relative entropies [36]. Letting $\mathscr{D} \subseteq \mathbb{R}_{++}$, we have

$$\phi_r(c) := \begin{cases} \frac{1}{r(r-1)} c^r, & \text{if } r \neq 0, 1, \\ c \ln c, & \text{if } r = 1, \\ -\ln c, & \text{if } r = 0. \end{cases} \tag{7}$$

Hence, it follows that:

$$D^n_{\phi_r}(\boldsymbol{x} \parallel \boldsymbol{y}) = \begin{cases} \frac{1}{r(r-1)} \sum_{i \in \mathcal{N}} [x_i^r + (r-1) y_i^r - r x_i y_i^{r-1}], & \text{if } r \neq 0, 1, \\ \sum_{i \in \mathcal{N}} [x_i \ln(x_i/y_i) + y_i - x_i], & \text{if } r = 1, \\ \sum_{i \in \mathcal{N}} [x_i/y_i - \ln(x_i/y_i) - 1], & \text{if } r = 0. \end{cases} \tag{8}$$

This subclass is known in information theory and statistics as *Bregman–Csiszár divergences* [20]. It generalizes the *Kullback–Leibler divergence* (if $r = 1$) and *Itakura–Saito divergence* (if $r = 0$). We can also consider the following function, that has been used implicitly by [9] to characterize the generalized absolute entropies ($\mathscr{D} \subseteq \mathbb{R}$):

$$\phi_a(c) := \begin{cases} e^{ac}, & \text{if } a \neq 0, \\ c^2, & \text{if } a = 0, \end{cases} \tag{9}$$

so that:

$$D^n_{\phi_a}(\boldsymbol{x} \parallel \boldsymbol{y}) = \begin{cases} \sum_{i \in \mathcal{N}} [e^{a x_i} - e^{a y_i} - a(x_i - y_i) e^{a y_i}], & \text{if } a \neq 0, \\ \sum_{i \in \mathcal{N}} (x_i - y_i)^2, & \text{if } a = 0. \end{cases} \tag{10}$$

When $a = 0$, $D^n_{\phi_a}$ corresponds to the *squared Euclidean distance*. Without constraint on a, to the best of our knowledge, no similar family exists in information theory.

The measures presented in (8) and (10) are of particular interest. It is common practice, in the literature, to supplement the axiomatic framework by normalization requirements. It is usually assumed that an identical replication of the population, or unilateral scale transformations of the distributions $\boldsymbol{x}$ and $\boldsymbol{y}$, should not impact the measure. These properties are respectively called *principle of populations* and *scale invariance*. Magdalou and Nock [31] have established that a measure is consistent with Properties 1 to 6, and these new properties if and only if it can be written as $I_r(\boldsymbol{x} \parallel \boldsymbol{y}) = D^n_{\phi_r}(\hat{\boldsymbol{x}} \parallel \hat{\boldsymbol{y}})/n$, up to an increasing transformation. If translation invariance is substituted for scale invariance, then $I_a(\boldsymbol{x} \parallel \boldsymbol{y}) = D^n_{\phi_a}(\tilde{\boldsymbol{x}} \parallel \tilde{\boldsymbol{y}})/n$, up to an increasing transformation.

Our approach is closely related to the literature on distributional changes, initiated by [17]. In order to simplify the comparison, let $\mu(\boldsymbol{x}) = \mu(\boldsymbol{y})$. The author characterizes two classes of additively decomposable measures. Depending on the invariance axiom which has to be imposed, the unnormalized Cowell's measures can be written as, respectively ($\mathscr{D} \subseteq \mathbb{R}_{++}$):

$$K_f^n(\boldsymbol{x} \parallel \boldsymbol{y}) := \sum_{i \in \mathcal{N}} y_i f\left(\frac{x_i}{y_i}\right), \quad \text{or} \quad K_g^n(\boldsymbol{x} \parallel \boldsymbol{y}) := \sum_{i \in \mathcal{N}} g(x_i - y_i), \tag{11}$$

where $f = \phi_r$ and $g = \phi_a$ as defined, respectively (and up to a linear transformation), in (7) and (9). Measures K_f^n are known in information theory as *Csiszár f-divergences* [19,2] and coincide with $D_{\phi_r}^n$ if and only if $f(c) = \phi_r(c) = c\ln(c)$. In that case one obtains the Kullback–Leibler divergence. Equivalently, the measures K_g^n coincide with $D_{\phi_a}^n$ if and only if $g(c) = \phi_a(c) = c^2$. One obtains the squared Euclidean distance. We stress the consistency of our class to the Cowell's condition of *monotonicity in distance*.

4.2. Inequality measurement

If, for any given distribution, the reference distribution is that obtained by allocating the mean income to each individual, then divergence reduces to inequality. Noting that $\sum_{i\in\mathcal{N}}(x_i - \mu)\phi'(\mu) = 0$, it follows from Theorem 1 that:

$$D_\phi^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) = \sum_{i \in \mathcal{N}} \left[\phi(x_i) - \phi(\mu)\right], \tag{12}$$

where ϕ is a strictly convex function. This class was independently introduced in information theory by [13] – now known as *Burbea–Rao divergences* – and in inequality measurement by [36,37]. Due to the anteriority of Shorrocks' paper, these may be called *Shorrocks–Burbea–Rao divergences*. If we consider the above-mentioned normalized measures $I_r(\boldsymbol{x} \parallel \mu\mathbf{1}_n)$, we obtain the generalized relative entropies [16,18,36,37]. If I_a is substituted for I_r, we deduce the generalized absolute entropies [14,9].

Now, we discuss our axiomatic framework, in comparison with the special case (12) characterized by [37]. All our axioms are almost equivalent, setting aside our Properties 2 and 5 (judgment separability and non-negativity).[3] These properties are replaced in Shorrocks by the normalization condition $D^n(\mu\mathbf{1}_n \parallel \mu\mathbf{1}_n) = 0$. In the particular case of inequality measurement, the non-negativity is implied by the principle of transfers. That has to be imposed in the general framework (Property 5). The only significant difference is Property 2. That might be considered as a rather strong requirement, since it imposes an additive and unweighted structure on the trade-off between inequality and efficiency loss.

We first observe that Property 2 is sufficient to have the Shorrocks' normalization, if we are only concerned by inequality. Indeed, it implies that $D^n(c\mathbf{1}_n \parallel c\mathbf{1}_n) = 2D^n(c\mathbf{1}_n \parallel c\mathbf{1}_n) = 0$ for all $c \in \mathscr{D}$, and consequently $D^n(\mu\mathbf{1}_n \parallel \mu\mathbf{1}_n) = 0$. Then, the same trade-off is implicitly assumed in inequality measurement. To bring out this point, consider that the whole population $\mathcal{N}$ can be partitioned into G disjoint subgroups $\mathcal{N}_g$, with mean μ_g. From Property 2, we have $D_\phi^{n_g}(\boldsymbol{x}_g \parallel c\mathbf{1}_{n_g}) = D_\phi^{n_g}(\boldsymbol{x}_g \parallel \mu_g\mathbf{1}_{n_g}) + D_\phi^{n_g}(\mu_g\mathbf{1}_{n_g} \parallel c\mathbf{1}_{n_g})$ for each subgroup $g = 1, 2, \ldots, \mathrm{G}$. Note that c can be equal to the global mean μ. Now, adding decomposability (Property 6), we have

[3] As only difference, we extend the notion of anonymity to a more general framework and we assume, by mathematical convenience, differentiability of the measure.

$$D^n_\phi(\boldsymbol{x} \parallel c\mathbf{1}_n) = \sum_{g=1}^{\mathrm{G}} D^{n_g}_\phi(\boldsymbol{x}_g \parallel \mu_g \mathbf{1}_{n_g}) + \sum_{g=1}^{\mathrm{G}} n_g D^1_\phi(\mu_g \parallel c), \tag{13}$$

or equivalently:

$$\begin{aligned} D^n_\phi(\boldsymbol{x} \parallel c\mathbf{1}_n) &= D^n_\phi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\mathrm{G}} \parallel \mu_1 \mathbf{1}_{n_1}, \ldots, \mu_{\mathrm{G}} \mathbf{1}_{n_{\mathrm{G}}}) \\ &\quad + D^n_\phi(\mu_1 \mathbf{1}_{n_1}, \ldots, \mu_{\mathrm{G}} \mathbf{1}_{n_{\mathrm{G}}} \parallel c\mathbf{1}_n). \end{aligned} \tag{14}$$

Eqs. (13) and (14) state that the divergence between $\boldsymbol{x}$ and $c\mathbf{1}_n$ can be expressed as the sum of a *within-subgroup component* which measures the divergence, within each subgroup, between the incomes and the subgroup mean, and a *between-subgroup component*, which computes the divergence between the subgroup means and c, taking into account the size of the subgroups. Thus, the usual *additive decomposability* condition [11,36] for inequality measures, in its unweighted form, follows from judgment separability and decomposability in our framework. The weights then result from the relative or absolute normalization procedure.

Shorrocks [37] shows that a rather weak aggregation rule (similar to our decomposability condition) implies an additively decomposable structure on the inequality indices. Our results suggest that the judgment separability assumption, implicit in inequality measurement, is also involved. Thus, a weakening of this property – by way of a more general trade-off between inequality and efficiency loss – supplemented by a weaker decomposability assumption, might be a solution in order to not exclude non-separable measures. This question falls beyond the scope of this paper.

5. Applications and discussion

We now briefly illustrate some possible applications of our results. Although this paper is devoted to the analysis of income distributions, the measures we have introduced are clearly applicable to many types of distributions and economic issues. By way of a conclusion, we also mention some other directions.

Households with differing needs. In practice, income distributions are collected from households that differ in many respects. Given the single adult as a reference household type, the usual procedure consists in first deflating the household's income by a scale factor that reflects its needs, and then weighting the resulting income by the number of persons in the household. Once this is done, conventional inequality measures are applied to the so-called household equivalent incomes. Drawing on the works of [27] and [33], Jenkins and O'Higgins [28] proposed an alternative to the above-described two steps procedure, namely a *norm income approach*. First, the social planner determines the income each household should have from an equity point of view, taking into account the differences in equity-relevant characteristics – especially size and needs – between households. The divergence between the actual distribution and the reference distribution can then be evaluated by quasi-orderings (comparable to the Lorenz criterion) or synthetic indices. Our measures appear to be sensible candidates.

Ethics of responsibility. The issue of individual responsibility and the distinction between fair and unfair inequalities have been the subject of much attention in the contemporary theories of justice. These considerations cannot be taken into account by the conventional inequality measures. Indeed, they evaluate a strict egalitarian income distribution as the best outcome. The degree of unfairness can be quantified by the divergence between the actual distribution of individual incomes and the fair distribution, which is obtained applying responsibility-sensitive fairness principles. Devooght [21] and Almas et al. [3] are recent attempts to introduce, and test on real data, measures of unfairness. It might be interesting to analyze the relationship with our measures.

Mobility measurement. All the axioms we impose are consistent with the measurement of income mobility.[4] According to [23], we shall regard income mobility as arising from two sources: transfer of money among individuals with total income held constant, and changes in the total income available. The larger the changes in the individual incomes, the more mobility there is. To give an illustration in our framework, consider a cohort of individuals alive for at least two periods. We define the reference distribution $\boldsymbol{y}$ as the situation in the first period, and $\boldsymbol{y}'$ and $\boldsymbol{y}''$ as two possible distributions for the second period. Using D^n_ϕ, we will say that there is more income mobility from $\boldsymbol{y}$ to $\boldsymbol{y}'$, than from $\boldsymbol{y}$ to $\boldsymbol{y}''$, if and only if $D^n_\phi(\boldsymbol{y}' \parallel \boldsymbol{y}) > D^n_\phi(\boldsymbol{y}'' \parallel \boldsymbol{y})$. We now compare D^n_ϕ to the symmetric measures of [23]. Both classes are equivalent with respect to the first source of mobility, namely *transfer mobility*. This dimension is encapsulated by the consistency with the Cowell's condition of *monotonicity in distance*. In our case, it is captured by the convexity of ϕ. But the classes significantly differ with respect to the second source of mobility, namely *growth mobility*. Consider that $\boldsymbol{y}'$ and $\boldsymbol{y}''$ are obtained from $\boldsymbol{y}$ by a similar increment of income, allocated to only one individual, and assume that the recipient is initially poorer in $\boldsymbol{y}'$ than in $\boldsymbol{y}''$. In the sense of Fields and Ok's measures, the mobility is equivalent from $\boldsymbol{y}$ to $\boldsymbol{y}'$ or from $\boldsymbol{y}$ to $\boldsymbol{y}''$. In our framework, if ϕ is decreasing [resp. increasing], then there is more [resp. less] mobility from $\boldsymbol{y}$ to $\boldsymbol{y}'$ than from $\boldsymbol{y}$ to $\boldsymbol{y}''$. This is a consequence of the *asymmetric structure* of D^n_ϕ.[5] This feature seems sensible in the case where the social planner wants to give *priority to the poor* in the notion of mobility (letting ϕ decreasing).

As a final note it is worth emphasizing that the measures we have introduced in this paper do have rich applications outside the field of income inequalities. First, D^n_ϕ can be used as a *goodness-of-fit* measure of a statistical model. For example, Pardo [34] proposes minimum Burbea–Rao divergences (a subclass of our measures) as goodness-of-fit tests and analyzes the relationship with maximum likelihood and khi-squared methods. Another subfield of economics which can be investigated is *decision-making under risk and uncertainty*. Recently, Maccheroni et al. [30] have characterized a class of *ambiguity indices*, which can be written as divergences. In their paper, the authors consider Csiszár f-divergences (see Section 4.1). We claim that the measures D^n_ϕ can also be applied. These questions might be the object of future investigations.

Appendix A. Proofs of the results

Proof of Proposition 3.1. Consider a distribution $\boldsymbol{x} \in \mathscr{D}^n$ and a constant $c \in \mathscr{D}$, and assume that Properties 1 to 4 are satisfied. From Property 2, we deduce that:

$$D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) = D^n(\boldsymbol{x} \parallel c\mathbf{1}_n) - D^n(\mu\mathbf{1}_n \parallel c\mathbf{1}_n). \tag{A.1}$$

Since this holds for all $\boldsymbol{x} \in \mathscr{D}^n$ and all $c \in \mathscr{D}$, and the left-hand side of (A.1) does not depend on c, the right-hand side cannot depend on c. Thus there exists a function $\phi^n : \mathscr{D}^n \longrightarrow \mathbb{R}$ such that $D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) = \phi^n(\boldsymbol{x}) - \phi^n(\mu\mathbf{1}_n)$. From Property 1, ϕ^n is continuous. Now consider two distributions $\boldsymbol{x}, \boldsymbol{x}' \in \mathscr{D}^n$ and suppose that $\boldsymbol{x}'$ is obtained from $\boldsymbol{x}$ by means of a permutation or

[4] In that sense, we emphasize that our anonymity requirement is weaker than the standard one in the theories of inequality, and consistent with mobility measurement.

[5] Note that the symmetry in [23] is not explicitly imposed, but a consequence of other axioms.

a progressive transfer from individual k to h, as described in (3). After simplification, it follows

$$D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) - D^n(\boldsymbol{x}' \parallel \mu\mathbf{1}_n) = \phi^n(\boldsymbol{x}) - \phi^n(\boldsymbol{x}'). \tag{A.2}$$

Applying Property 3, the left-hand side of (A.2) is equal to 0 if $\boldsymbol{x}'$ is a permutation of $\boldsymbol{x}$. From Property 4, the left-hand side of (A.2) is strictly positive if $\boldsymbol{x}'$ is obtained from $\boldsymbol{x}$ by means of a progressive transfer. Hence ϕ^n is strictly Schur-convex. □

Proof of Proposition 3.2. Consider two distributions $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{D}^n$, and assume that Properties 1 to 6 are satisfied. The symmetry of the function F^n in (5) comes from Property 3. If $x_i = y_i = c$ for all $i \in \mathcal{N}$, we know from Property 2 that $D^n(c\mathbf{1}_n \parallel c\mathbf{1}_n) = 2D^n(c\mathbf{1}_n \parallel c\mathbf{1}_n) = 0$ whenever $n \geqslant 1$. Also, applying Property 6, $D^n(c\mathbf{1}_n \parallel c\mathbf{1}_n) = F^n(D^{n_1}(c\mathbf{1}_{n_1} \parallel c\mathbf{1}_{n_1}), D^{n_2}(c\mathbf{1}_{n_2} \parallel c\mathbf{1}_{n_2}))$. Thus $F^n(0,0) = 0$. The rest of the proof consists in two steps. In the first step we show that $F^n(\mathbb{R}_+ \times \mathbb{R}_+) \subseteq \mathbb{R}_+$, under the assumption that $F^n(u,v)$ is strictly increasing in u and v. This assumption is proved in the second step.

Step 1. In order to simplify the notation, let $u_i := D^1(x_i \parallel y_i)$ for all $i \in \mathcal{N}$ such that $D^2(x_1, x_2 \parallel y_1, y_2) = F^2(u_1, u_2)$. We know from Property 5 that $u_1, u_2 \in \mathbb{R}_+$. Thus the domain of F^2 is $\mathbb{R}_+^2$. Now suppose that $F^2(u_1, u_2)$ is strictly increasing in u_1 and u_2. In that case, with $F^2(0,0) = 0$, the range of F^2 is $\mathbb{R}_+$. Moreover we have $D^3(x_1, x_2, x_3 \parallel y_1, y_2, y_3) = F^3(F^2(u_1, u_2), u_3)$. If F^3 is strictly increasing in its arguments then, with $F^3(0,0) = 0$, we have $F^3(\mathbb{R}_+ \times \mathbb{R}_+) \subseteq \mathbb{R}_+$. By successive decompositions and using the same reasoning, we deduce that $F^n(\mathbb{R}_+ \times \mathbb{R}_+) \subseteq \mathbb{R}_+$.

Step 2. It remains to prove that $F^n(u,v)$ is strictly increasing in u and v, for all $u, v \in \mathbb{R}_+$. Let $\mu \equiv \mu(\boldsymbol{x})$. From Proposition 3.1, we have $D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) = \phi^n(\boldsymbol{x}) - \phi^n(\mu\mathbf{1}_n)$ with ϕ^n a continuous and strictly Schur-convex function, such that $D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) \geqslant 0$. Thus we can pick any arbitrary $u, v \in \mathbb{R}_+$ and choose $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{D}^n$ such that $\boldsymbol{x} := (\boldsymbol{x}_1, \boldsymbol{x}_2)$, $\boldsymbol{y} := \mu\mathbf{1}_n$, $D^{n_1}(\boldsymbol{x}_1 \parallel \mu\mathbf{1}_{n_1}) = u$ and $D^{n_2}(\boldsymbol{x}_2 \parallel \mu\mathbf{1}_{n_2}) = v$. It follows from Property 6 that $D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) = F^n(u,v)$. Now consider the distribution $\boldsymbol{x}'_1$ obtained from $\boldsymbol{x}_1$ by means of a progressive transfer. If we define $u' := D^{n_1}(\boldsymbol{x}'_1 \parallel \mu\mathbf{1}_{n_1})$, we know that $u' \in \mathbb{R}_+$ and it follows from Properties 2 and 4 that $u > u'$. Letting $\boldsymbol{x}' := (\boldsymbol{x}'_1, \boldsymbol{x}_2)$, we have $D^n(\boldsymbol{x}' \parallel \mu\mathbf{1}_n) = F^n(u', v)$. If $\boldsymbol{x}'_1$ is obtained from $\boldsymbol{x}_1$ by means of a progressive transfer, then $\boldsymbol{x}'$ is also obtained from $\boldsymbol{x}$ by means of a progressive transfer. Thus, from Property 4, we have $D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n) > D^n(\boldsymbol{x}' \parallel \mu\mathbf{1}_n)$. We conclude that $u > u'$ implies $F^n(u,v) > F^n(u',v)$, which signifies that F^n is strictly increasing in its first argument. The same reasoning can be applied for the second argument of F^n. □

Proof of Proposition 3.3. The proof consists in three steps. Step 1 and the end of step 3 are, respectively, based on Theorems 1 and 3 in [23].

Step 1. Consider two distributions $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{D}^{n+1}$ and $c \in \mathscr{D}$ such that $\boldsymbol{x} := (x_1, c\mathbf{1}_n)$, $\boldsymbol{y} := (y_1, c\mathbf{1}_n)$. Pick an arbitrary $u \in \mathbb{R}_+$ and choose x_1 and y_1 such that $D^1(x_1, y_1) = u$. From Property 6 one observes that:

$$D^{n+1}(\boldsymbol{x} \parallel \boldsymbol{y}) = F^{n+1}(u, 0) = F^{n+1}(u, F^n(0,0)) = F^{n+1}(F^n(u,0), 0). \tag{A.3}$$

Thus $F^{n+1}(u,0) = F^{n+1}(F^n(u,0),0)$. Since F^{n+1} is strictly increasing in its arguments, we have $F^n(u,0) = u$. Hence it follows

$$F^n(u,v) = F^{n+1}(F^n(u,v), 0) = F^{n+1}(u, F^n(v,0)) = F^{n+1}(u,v), \tag{A.4}$$

or equivalently $F^n = F^{n+1}$ for all $n \geqslant 2$. We conclude that $F(F(u,v),w) = F(u, F(v,w))$ for all $u, v, w \in \mathbb{R}_+$. So F is an *associative function*. The prerequisites of Aczel's theorem on associative functions [1, Theorem, p. 254] are satisfied, hence there exists an invertible function f such that:

$$F(u,v)=f\bigl(f^{-1}(u)+f^{-1}(v)\bigr),\quad \forall u,v\in\mathbb{R}_+. \tag{A.5}$$

The following steps are devoted to prove that f is linear.

Step 2. Consider two incomes $c,c'\in\mathscr{D}$ and let $D^1(c\parallel c')=u\in\mathbb{R}_+$. From Property 6 and Eq. (A.5), we deduce that $D^2(c,c\parallel c',c')=F(u,u)=f(2f^{-1}(u))$. Then $D^3(c,c,c\parallel c',c',c')=F(F(u,u),u)=F(f(2f^{-1}(u)),u)$, or equivalently:

$$D^3\bigl(c,c,c\parallel c',c',c'\bigr)=f\bigl(f^{-1}\bigl[f\bigl(2f^{-1}(u)\bigr)\bigr]+f^{-1}(u)\bigr)=f\bigl(3f^{-1}(u)\bigr). \tag{A.6}$$

By successive applications of this reasoning, one obtains

$$D^n\bigl(c\mathbf{1}_n\parallel c'\mathbf{1}_n\bigr)=f\bigl[nf^{-1}\bigl(D^1\bigl(c\parallel c'\bigr)\bigr)\bigr],\quad \forall n\geqslant 1,\ \forall c,c'\in\mathscr{D}. \tag{A.7}$$

Step 3. Pick any arbitrary $u,v\in\mathbb{R}_+$ and choose $\boldsymbol{x}\in\mathscr{D}^2$ with mean $\mu\equiv\mu(\boldsymbol{x})$ and $c\in\mathscr{D}$, such that $D^2(x_1,x_2\parallel c,c)=u$ and $D^1(\mu\parallel c)=v$. From Property 6 one observes that:

$$D^3(x_1,x_2,\mu\parallel c,c,c)=F(u,v). \tag{A.8}$$

Moreover, from Property 2, it follows that:

$$D^3(x_1,x_2,\mu\parallel c,c,c)=D^3(x_1,x_2,\mu\parallel \mu,\mu,\mu)+D^3(\mu,\mu,\mu\parallel c,c,c). \tag{A.9}$$

The first term of the right-hand side of (A.9) is such that:

$$\begin{aligned}D^3(x_1,x_2,\mu\parallel \mu,\mu,\mu)&=F\bigl(D^2(x_1,x_2\parallel\mu,\mu),D^1(\mu\parallel\mu)\bigr)\\&=F\bigl(D^2(x_1,x_2\parallel\mu,\mu),0\bigr).\end{aligned} \tag{A.10}$$

From step 1, we know that $F(w,0)=w$, thus $D^3(x_1,x_2,\mu\parallel\mu,\mu,\mu)=D^2(x_1,x_2\parallel\mu,\mu)$. From Property 2, we can write $D^2(x_1,x_2\parallel\mu,\mu)=D^2(x_1,x_2\parallel c,c)-D^2(\mu,\mu\parallel c,c)$, from which:

$$D^3(x_1,x_2,\mu\parallel\mu,\mu,\mu)=D^2(x_1,x_2\parallel c,c)-D^2(\mu,\mu\parallel c,c). \tag{A.11}$$

Reintroducing (A.11) into (A.9), it follows

$$\begin{aligned}D^3(x_1,x_2,\mu\parallel c,c,c)={}&D^2(x_1,x_2\parallel c,c)+D^3(\mu,\mu,\mu\parallel c,c,c)\\&-D^2(\mu,\mu\parallel c,c).\end{aligned} \tag{A.12}$$

By using step 2 and (A.7), there exists a function H such that $D^3(\mu,\mu,\mu\parallel c,c,c)-D^2(\mu,\mu\parallel c,c)=H(D^1(\mu\parallel c))$. Hence one concludes that $D^3(x_1,x_2,\mu\parallel c,c,c)=D^2(x_1,x_2\parallel c,c)+H(D^1(\mu\parallel c))$, or equivalently $F(u,v)=u+H(v)$. From Property 3 the function F is symmetric, and also $F(u,v)=H(u)+v$. Using the same arguments as developed in [23, Theorem 3], we deduce that $f^{-1}(u)=Au$ where $A>0$ in (A.5). By successive decompositions, one finally obtains $D^n(\boldsymbol{x}\parallel\boldsymbol{y})=\sum_{i\in\mathcal{N}}D^1(x_i\parallel y_i)$ from (A.5). □

Proof of Theorem 1. Consider two distributions $\boldsymbol{x},\boldsymbol{y}\in\mathscr{D}^n$ and assume that Properties 1 and 6 are satisfied. From Proposition 3.1, we have $D^n(\boldsymbol{x}\parallel\mu\mathbf{1}_n)=\phi^n(\boldsymbol{x})-\phi^n(\mu\mathbf{1}_n)$ where ϕ^n is strictly Schur-convex. Consequently ϕ^n is symmetric and for all $n\geqslant 1$, it follows that $\phi^1(x)=\phi^n(x,0,\dots,0)=\phi^n(0,x,0,\dots,0)=\cdots=\phi^n(0,0,\dots,0,x)$. From Proposition 3.3, we have $D^n(\boldsymbol{x}\parallel\mu\mathbf{1}_n)=\sum_{i\in\mathcal{N}}D^1(x_i\parallel\mu)$. Thus $D^n(\boldsymbol{x}\parallel\mu\mathbf{1}_n)=\sum_{i\in\mathcal{N}}[\phi(x_i)-\phi(\mu)]$, where we let $\phi^1\equiv\phi$. Thus $\phi^n(\boldsymbol{x})=\sum_{i\in\mathcal{N}}\phi(x_i)$ for all $\boldsymbol{x}\in\mathscr{D}^n$. Since ϕ^n is strictly Schur-convex, we know from [32, Proposition C.1.a, p. 64], that ϕ is strictly convex. Reintroducing $D^n(\boldsymbol{x} \parallel \mu\mathbf{1}_n)$ into (2) and using the additive structure of D^n, one obtains

$$\sum_{i\in\mathcal{N}}\left[D^1(x_i \parallel c) - \phi(x_i)\right] = \sum_{i\in\mathcal{N}}\left[D^1(\mu \parallel c) - \phi(\mu)\right]. \tag{A.13}$$

From Property 1, D^1 is differentiable. We let $D^1_{x_k}(x_k \parallel c) := \partial D^1(x_k \parallel c)/\partial x_k$ and $\phi'(x_k) := \partial\phi(x_k)/\partial x_k$ for all $k \in \mathcal{N}$. Differentiating (A.13) with respect to x_i, x_j and subtracting, one deduces that:

$$D^1_{x_i}(x_i \parallel c) - \phi'(x_i) = D^1_{x_j}(x_j \parallel c) - \phi'(x_j), \quad \forall i, j \in \mathcal{N}. \tag{A.14}$$

So, there exists a function $\psi : \mathscr{D} \longrightarrow \mathbb{R}$, such that:

$$D^1_{x_k}(x_k \parallel c) - \phi'(x_k) = \psi(c). \tag{A.15}$$

From Property 1, ψ is continuous. From Proposition 3.2, we know that $D^1(x_k \parallel c) \geqslant D^1(c \parallel c) = 0$. Thus $D^1_{x_k}(c \parallel c) = 0$. Letting $x_k = c$ in (A.15), one deduces that $\psi(c) = -\phi'(c)$. Because (A.15) holds for all $c \in \mathscr{D}$, we can let $c = y_k$. One obtains

$$D^1_{x_k}(x_k \parallel y_k) - \phi'(x_k) + \phi'(y_k) = 0. \tag{A.16}$$

By integrating it follows that:

$$D^1(x_k \parallel y_k) - \phi(x_k) + x_k\phi'(y_k) = \varphi(y_k). \tag{A.17}$$

Also we know that $D^1(y_k \parallel y_k) = 0$. Thus, letting $x_k = y_k$ in (A.17), we have $\varphi(y_k) = -\phi(y_k) + y_k\phi'(y_k)$. One concludes

$$D^1(x_k \parallel y_k) = \phi(x_k) - \phi(y_k) - (x_k - y_k)\phi'(y_k). \tag{A.18}$$

From Property 1, D^1 is differentiable in its arguments. Hence ϕ is twice differentiable. Recalling that $D^n(\boldsymbol{x} \parallel \boldsymbol{y}) = \sum_{k\in\mathcal{N}} D^1(x_k \parallel y_k)$, one obtains the desired result. The sufficiency part of the proof does not present difficulties and is left to the reader. □

References

[1] J. Aczél, Lectures on Functional Equations and Their Applications, Academic Press, New York, London, 1966.

[2] S. Ali, S. Silvey, A general class of coefficients of divergence of one distribution from another, J. Roy. Statistical Society B 28 (1966) 131–142.

[3] I. Almas, A. Cappelen, J. Lind, E. Sørensen, B. Tungodden, Measuring unfair inequality, J. Public Econ. 95 (2011) 488–499.

[4] S. Amari, H. Nagaoka, Methods of Information Geometry, Oxford University Press, Oxford, 2000.

[5] Y. Amiel, F. Cowell, Measurement of income inequality: Experimental test by questionnaire, J. Public Econ. 47 (1992) 3–26.

[6] R. Arneson, Equality and equal opportunity for welfare, Philos. Stud. 56 (1989) 77–93.

[7] A. Banerjee, X. Guo, H. Wang, On the optimality of conditional expectation as a Bregman predictor, IEEE Trans. Inform. Theory 51 (2005) 2664–2669.

[8] C. Blackorby, D. Donaldson, M. Auersperg, A new procedure for the measurement of inequality within and among population subgroups, Can. J. Econ. 14 (1981) 665–685.

[9] K. Bosmans, F. Cowell, The class of absolute decomposable inequality measures, Econ. Letters 109 (2010) 154–156.

[10] W. Bossert, Redistribution mechanisms based on individual characteristics, Math. Soc. Sci. 29 (1995) 1–17.

[11] F. Bourguignon, Decomposable income inequality measures, Econometrica 47 (1979) 901–920.

[12] L. Bregman, The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming, USSR Comput. Math. Math. Phys. 7 (1967) 200–217.

[13] J. Burbea, C. Rao, On the convexity of some divergence measures based on entropy functions, IEEE Trans. Inform. Theory 28 (1982) 489–495.
[14] S. Chakravarty, S. Tyagarupananda, The subgroup decomposable absolute indices of inequality, in: S. Chakravarty, D. Coondo, R. Mukherjee (Eds.), Quantitative Economics: Theory and Practice, Essays in Honor of Professor N. Bhattacharya, Allied Publishers Limited, New Delhi, 1998, pp. 247–257.
[15] A. Chateauneuf, P. Moyes, Measuring inequality without the Pigou–Dalton condition, in: M. McGillivray (Ed.), Inequality, Poverty and Well-Being, Palgrave Macmillan, Basingtoke, Hampshire, New York, 2006, pp. 22–65.
[16] F. Cowell, On the structure of additive inequality measures, Rev. Econ. Stud. 47 (1980) 521–531.
[17] F. Cowell, Measures of distributional change: An axiomatic approach, Rev. Econ. Stud. 52 (1985) 135–151.
[18] F. Cowell, K. Kuga, Additivity and the entropy concept: An axiomatic approach to inequality measurement, J. Econ. Theory 25 (1981) 131–143.
[19] I. Csiszár, Eine informationstheorische Ungleichung und Ihre Anwendung auf den Beweis der ergodizität von markoffschen Ketten, Publ. Math. Inst. Hungar. Acad. Sci. 8 (1963) 85–107.
[20] I. Csiszár, Why least square and maximum entropy? An axiomatic approach to inference for linear inverse problems, Ann. Statist. 19 (1991) 2032–2066.
[21] K. Devooght, To each the same and to each his own: A proposal to measure responsibility-sensitive income inequality, Economica 75 (2008) 280–295.
[22] J. Esteban, D. Ray, On the measurement of polarization, Econometrica 62 (1994) 819–851.
[23] G. Fields, E. Ok, The meaning and measurement of income mobility, J. Econ. Theory 71 (1996) 349–377.
[24] M. Fleurbaey, Equality and responsibility, Europ. Econ. Rev. 39 (1995) 683–689.
[25] J. Foster, A. Shneyerov, A general class of additively decomposable inequality measures, Econ. Theory 14 (1999) 89–111.
[26] J. Foster, A. Shneyerov, Path independent inequality measures, J. Econ. Theory 91 (2000) 199–222.
[27] G. Garvy, Inequality of income: Causes and measurement, in: Eight Papers on Size Distribution of Income, in: Studies in Income Wealth, vol. 15, National Bureau of Economic Research, New York, 1952, pp. 25–48.
[28] S. Jenkins, M. O'Higgins, Inequality measurement using 'norm incomes': Were Garvy and Paglin onto something after all?, Rev. Income Wealth 35 (1989) 265–282.
[29] S. Kullback, R. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1951) 79–86.
[30] F. Maccheroni, M. Marinacci, A. Rustichini, Amibiguity aversion, robustness and the variational representation of preferences, Econometrica 74 (2006) 1447–1498.
[31] B. Magdalou, R. Nock, Income distributions and decomposable divergence measures, CEREGMIA Working Paper 2010-01, Université des Antilles et de la Guyane, 2010.
[32] A. Marshall, I. Olkin, Inequalities: Theory of Majorization and Its Applications, Academic Press, New York, 1979.
[33] M. Paglin, The measurement and trend of inequality: A basic revision, Amer. Econ. Rev. 65 (1975) 598–609.
[34] M. Pardo, On Burbea–Rao divergence based goodness-of-fit tests for multinomial models, J. Multivariate Anal. 69 (1999) 65–87.
[35] J. Roemer, Theories of Distributive Justice, Harvard University Press, Cambridge, 1996.
[36] A. Shorrocks, The class of additively decomposable inequality measures, Econometrica 48 (1980) 613–625.
[37] A. Shorrocks, Inequality decomposition by population subgroups, Econometrica 52 (1984) 1369–1385.
[38] M. Wolfson, When inequalities diverge, Amer. Econ. Rev.: Papers and Proceedings 84 (1994) 353–358.