Computational Statistics and Data Analysis 71 (2014) 868–886

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Stochastic dominance with imprecise information

Ignacio Montes *, Enrique Miranda, Susana Montes

University of Oviedo, Dep. of Statistics and Operations Research, Spain

ARTICLE INFO

Article history:
Received 11 November 2011
Received in revised form 3 July 2012
Accepted 9 July 2012
Available online 31 August 2012

Keywords:
Distribution function
Stochastic dominance
Probability boxes
Possibility measures
Lorenz curves

ABSTRACT

Stochastic dominance, which is based on the comparison of distribution functions, is one of the most popular preference measures. However, its use is limited to the case where the goal is to compare pairs of distribution functions, whereas in many cases it is interesting to compare sets of distribution functions: this may be the case for instance when the available information does not allow to fully elicitate the probability distributions of the random variables. To deal with these situations, a number of generalisations of the notion of stochastic dominance are proposed; their connection with an equivalent p-box representation of the sets of distribution functions is studied; a number of particular cases, such as sets of distributions associated to possibility measures, are investigated; and an application to the comparison of the Lorenz curves of countries within the same region is presented.

1. Introduction

The comparison of random variables is a natural problem that arises in many fields, and for this reason there are many different proposals in stochastic ordering. One of the most popular is stochastic dominance. First degree stochastic dominance considers one random variable greater than another one when the first one is more likely to take greater values. Although this notion has been employed since the 1930s, it has been in the past decades when it has witnessed increasing popularity; this is testified by the applications of stochastic dominance in many different areas, such as economics (Bawa, 1975), social welfare (Atkinson, 1987), agriculture (Islam and Braden, 2006), operational research (Noyan et al., 2006), etc.

It is not uncommon, however, to encounter situations where there is uncertainty about the probability distributions underlying the random variables of interest; we may for instance have vague or conflicting information, or errors in the transmission of the available data. This results in the impossibility of eliciting the probability distribution with certain guarantees. Instead, we may consider more realistic to work with sets of probability distributions which are sure to include the 'true' one. In other situations, we may be able to work with precise probability measures, but we may be interested in comparing the sets of probability distributions induced by random variables that share some common features.

Our goal in this paper is to extend the notion of stochastic dominance to the comparison of sets of distribution functions, following the steps made by Denoeux for the particular case of belief and plausibility measures (Denoeux, 2009). After giving some preliminary concepts in Section 2, we overview in Section 3 the work by Denoeux and generalise his work to arbitrary sets of probability measures. We investigate the relationships between different definitions and study their main properties, detailing their connection with the notions proposed in Denoeux (2009).

In Section 4, we investigate a number of particular cases which are of interest in relation to other works, such as the comparison of possibility measures by means of stochastic dominance or the comparison of two sets of distribution functions with the same lower and upper bounds. Our work is illustrated in Section 5 with an application to the comparison of sets

* Corresponding author. Tel.: +34 985186466; fax: +34 985103354.
E-mail addresses: imontes@uniovi.es (I. Montes), mirandaenrique@uniovi.es (E. Miranda), montes@uniovi.es (S. Montes).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.07.030

of Lorenz curves. Finally, in Section 6 we discuss the main implications of our results and future lines of research. We have gathered a number of examples in the Appendix.

We shall restrict our work to random variables taking values on the unit interval; since this is homeomorphic to any closed interval on the real line, the results extend immediately to distribution functions taking values on any interval $[a, b]$, where $a < b \in \mathbb{R}$. In fact, our work can be easily generalised to any totally ordered space, simply by adding a smallest and a greatest value. On the other hand, we shall work with sets of σ-additive probability measures, to be closer to the usual works on stochastic dominance. This gives rise, however, to a number of additional complications with respect to the usual works with non-additive measures, which are based on sets of finitely additive probability measures. To what extent this assumption makes a difference in the corresponding results shall be discussed in Section 3.

2. Preliminary concepts

2.1. Stochastic dominance

The notion of stochastic dominance between random variables is based on the comparison of their corresponding distribution functions. In this paper, we are going to work with random variables taking values on [0, 1]. The distribution function is thus defined in the following way.

Definition 1. A *cumulative distribution function* (cdf) is a function $F : [0, 1] \to [0, 1]$ satisfying the following properties.

- $x \le y \Rightarrow F(x) \le F(y)\ \forall x, y$ [Monotonicity].
- $F(1) = 1$ [Normalisation].
- $F(x) = \lim_{\epsilon \downarrow 0} F(x + \epsilon)\ \forall x < 1$ [Right-continuity].

When F satisfies the properties of monotonicity and normalisation, it is associated to a finite additive probability distribution, and we shall call it a *finitely additive distribution function*.

One of the most popular methods for the comparison of cdf is stochastic dominance (Levy, 1998; Müller and Stoyan, 2002; Shaked and Shanthikumar, 2006).

Definition 2. Given two cumulative distribution functions F and G, we say that F *stochastically dominates* G, and denote it by $F \succeq_{\text{FSD}} G$, if

$$F(t) \le G(t) \quad \text{for every } t \in [0, 1]. \tag{1}$$

This definition produces a partial order in the space of cumulative distribution functions, from which we can derive the notions of strict stochastic dominance, indifference and incomparability.

- We say that F *strictly* stochastically dominates G, and denote it by $F \succ_{\text{FSD}} G$, if $F \succeq_{\text{FSD}} G$ but $G \succeq_{\text{FSD}} F$. This holds if and only if $F \le G$ and there is some $t \in [0, 1]$ such that $F(t) < G(t)$.
- F and G are *stochastically indifferent* if $F \succeq_{\text{FSD}} G$ and $G \succeq_{\text{FSD}} F$, or equivalently, if $F = G$.
- F and G are *stochastically incomparable* if $F \succeq_{\text{FSD}} G$ and $G \succeq_{\text{FSD}} F$.

Stochastic dominance is commonly used in economics and finance (Denuit et al., 2005; Goovaerts et al., 1990) and can be given the following interpretation: $F \succeq_{\text{FSD}} G$ means that the choice of F over G is rational, in the sense that we prefer the alternative that provides greater probability of having a greater profit. The notion has also been used in other frameworks such as reliability theory, statistical physics, epidemiology, etc. (see Levy, 1998; Müller and Stoyan, 2002; Shaked and Shanthikumar, 2006 for more information).

2.2. P-boxes

Our goal in this paper is to extend the notion of stochastic dominance to the case where we consider sets of probability measures instead of a single one. As a consequence, we shall work within the theory of *imprecise probabilities*. The term imprecise probability (Walley, 1991) refers to uncertainty models applicable in situations where the available information does not allow us to single out a unique probability measure for all random variables involved. Examples of such models include 2- and n-monotone capacities (Choquet, 1953–1954), lower and upper previsions (Walley, 1991), belief functions (Antoine et al., 2012; Shafer, 1976), credal sets (Levi, 1980), possibility and necessity measures (Dubois and Prade, 1988), interval probabilities (Weichselberger, 2001), and coherent risk measures (Artzner et al., 1999).

One such model is considered in this paper: pairs of lower and upper distribution functions, also called *probability boxes*, or briefly, p-boxes (Ferson et al., 2003; Ferson and Tucker, 2006). P-boxes are often used in risk or safety studies, in which cumulative distributions play an essential role. Many theoretical properties and practical aspects of p-boxes have already been studied in the literature. They have been connected to other uncertainty models, such as random sets (Kriegler, 2005) and possibility measures (Troffaes et al., 2011), and have been applied in different contexts (Fuchs and Neumaier, 2008; Oberguggenberger and Fellin, 2008).