Journal Pre-proof

Partial stochastic dominance via optimal transport

Takashi Kamihigashi, John Stachurski

PII: S0167-6377(20)30094-8
DOI: https://doi.org/10.1016/j.orl.2020.07.003
Reference: OPERES 6610

To appear in: *Operations Research Letters*

Received date : 16 June 2020
Revised date : 26 June 2020
Accepted date : 2 July 2020

Please cite this article as: T. Kamihigashi and J. Stachurski, Partial stochastic dominance via optimal transport, *Operations Research Letters* (2020), doi: https://doi.org/10.1016/j.orl.2020.07.003.

Partial Stochastic Dominance via Optimal Transport[1]

Takashi Kamihigashi[a] and John Stachurski[b]

[a]Center for Computational Social Science, Kobe University

[b]Research School of Economics, Australian National University

June 26, 2020

Abstract. In recent years, a range of measures of "partial" stochastic dominance have been introduced. These measures attempt to determine the extent to which one distribution is dominated by another. We assess these measures from intuitive, axiomatic, computational and statistical perspectives. Our investigation leads us to recommend a measure related to optimal transport as a natural default.

Keywords: Stochastic dominance, optimal transport

1. Introduction

(First order) stochastic dominance is one of the most fundamental concepts in social welfare and decision making under uncertainty (see, e.g., [5, 11]). At the same time, stochastic dominance is fragile. For example, two normal distributions can only be ordered by stochastic dominance if their variances are exactly identical, even if one mean is orders of magnitude larger than the other. Such fragility is problematic for quantitative work.

In response, researchers have introduced many different notions of "partial" stochastic dominance. One is "restricted stochastic dominance," which compares the order of cumulative distribution functions (CDFs) only up to some specified point in the domain [1, 2, 3]. Another is "almost stochastic dominance," which was developed in the finance literature [6, 8]. A third kind of measure was proposed in [4], analyzing degree of stochastic dominance in the context of income mobility analysis. Still more measures are considered in [11].

[1]Corresponding author: john.stachurski@anu.edu.au, RSE, College of Business and Economics, Australian National University, ACT, Australia.

Despite the obvious practical relevance of a measure of partial stochastic dominance, none of the above have a clear axiomatic foundation. At the same time, some measures become complex outside of the one-dimensional case. These observations suggest that now is a good time to consider measures of partial stochastic dominance collectively. While doing so, we make the case for what we believe is the most natural measure of partial stochastic dominance. The measure can be understood as the solution to an optimal transport problem.

2. A Measure of Dominance

Let S be Polish with Borel sets $\mathcal{B}$ and closed partial order $\preceq$ (i.e., a partial order on S such that $\{(x,y) \in S \times S : x \preceq y\}$ is closed in the product topology). A function $h \colon S \to \mathbb{R}$ is called *increasing* if $x \preceq y$ implies $h(x) \leq h(y)$, and *decreasing* if $-h$ is increasing. Let $\mathcal{H}$ be the set of increasing Borel measurable h on S with $\operatorname{osc}(h) \leq 1$ (that is, $\sup h - \inf h \leq 1$). Let $\mathcal{M}$ be all finite measures on $(S, \mathcal{B})$ and $\mathcal{P}$ be all $\mu \in \mathcal{M}$ with $\mu(S) = 1$. A measure $\mu \in \mathcal{M}$ is said to be *stochastically dominated* by $\nu \in \mathcal{M}$ and we write $\mu \preceq_{sd} \nu$ if $\mu(S) = \nu(S)$ and $\int h \, d\mu \leq \int h \, d\nu$ for all $h \in \mathcal{H}$. For probability measures $\mu, \nu \in \mathcal{P}$, an alternative characterization is

$$\mu \preceq_{sd} \nu \quad \text{if and only if} \quad \max_{(X,Y) \in \Pi(\mu,\nu)} \mathbb{P}\{X \preceq Y\} = 1, \tag{1}$$

where $\Pi(\mu, \nu)$ is the set of all couplings of μ and ν [12]. (Given $(\mu, \nu) \in \mathcal{P} \times \mathcal{P}$, a pair of S-valued random variables (X, Y) defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *coupling* of (μ, ν) if $\mu(B) = \mathbb{P}\{X \in B\}$ and $\nu(B) = \mathbb{P}\{Y \in B\}$ for all $B \in \mathcal{B}$.) The short summary of our paper is: in light of (1), why not use

$$\tau(\mu, \nu) := \max_{(X,Y) \in \Pi(\mu,\nu)} \mathbb{P}\{X \preceq Y\} \tag{2}$$

as the default measure of partial stochastic dominance? We answer this question in stages. Existence of the maximum for each (μ, ν) pair is verified below.

Note that we can rewrite τ as $1 - \min_{(X,Y) \in \Pi(\mu,\nu)} \mathbb{E}\, c(X, Y)$, where $c(x, y) = \mathbb{1}\{x \npreceq y\}$. The infimum is a standard optimal transport problem. In particular, c is nonnegative, bounded and lower-semicontinuous (since $\preceq$ is a closed partial order), so, by Theorem 4.1 of [13], a solution exists. In particular, the "max" in (2) is justified. From now on, we call a pair (X, Y) that attains the maximum in (2) an *optimal coupling*.

It now follows from Kantorovich duality that $\min_{(X,Y) \in \Pi(\mu,\nu)} \mathbb{E}\, c(X, Y)$ can be replaced by a supremum over supporting "prices." In particular,

$$\tau(\mu, \nu) = 1 - \max_{g \in \mathcal{D}} \left\{ \int g d\nu - \int g d\mu \right\}. \tag{3}$$

where $\mathcal{D}$ is all $g\colon S \to \mathbb{R}$ with $g(y) - g(x) \leq \mathbb{1}\{x \npreceq y\}$. This is (iii) of Theorem 5.10 of [13]. (See "Particular case 5.16" on p. 60.) The existence of a maximizer is guaranteed by the same theorem. Clearly $\mathcal{D}$ is just the set of decreasing functions on S with $\operatorname{osc}(h) \leq 1$. Hence we can rewrite (3) as

$$\tau(\mu, \nu) = 1 - \max_{h \in \mathcal{H}} \left\{ \int h d\mu - \int h d\nu \right\}. \tag{4}$$

These expressions have additional representations that can be useful in some settings. For example, if $\mu = F$ and $\nu = G$ are CDFs, then one can show τ has the simple representation

$$\tau(F, G) = 1 - \sup_x \{G(x) - F(x)\}. \tag{5}$$

The representation in (4) is very similar to measures of partial stochastic dominance studied in [11]. They use essentially the same measure but replace $\mathcal{H}$ with the set of all increasing functions satisfying a Lipschitz bound with respect to the underlying metric. Such a measure is more attuned to topology but lacks the some of the advantages of (4) described below.

3. Alternative Measures

Here and below, a *measure of degree of stochastic dominance* is any function

$$\delta\colon \mathcal{P} \times \mathcal{P} \to [0, 1] \quad \text{such that } \mu \preceq_{sd} \nu \implies \delta(\mu, \nu) = 1. \tag{6}$$

Clearly τ is a measure of degree of stochastic dominance in the sense of (6). Another is the quantile ratio measure proposed in [4]. Let $S = [a, b]$ and let $\preceq$ be the usual order $\leq$ on $\mathbb{R}$. Letting F and G be CDFs on $[a, b]$, define

$$q(F, G) := \frac{m}{n} \quad \text{where} \quad m := \sum_{i=1}^{n} \mathbb{1}\{G(x_i) \leq F(x_i)\}. \tag{7}$$

Here $\{x_i\}$ is a grid of n specified values in $[a, b]$, typically corresponding to some quantile points. Thus $q(F, G)$ measures the fraction of times that $G(x_i) \leq F(x_i)$ is observed over specified test points. Clearly q satisfies $q(F, G) = 1$ when $G \leq F$ pointwise (i.e., $F \preceq_{sd} G$), which corresponds to (6).

Another version of partial stochastic dominance is *restricted* stochastic dominance (see, e.g. [1, 3]). Let F and G be one dimensional CDFs on an interval $[a, b]$. Given $c \in [a, b]$, distribution F is said to be dominated by G in the restricted sense if $G(x) \leq F(x)$ for all $x \leq c$. We can turn this into a measure by considering the largest such c, defined by

$$c^*(F, G) := \sup\{c \in [a, b] \,:\, G(x) \leq F(x), \ \forall\, x \leq c\}. \tag{8}$$

After normalizing we get

$$r(F,G) := \frac{c^*(F,G) - a}{b - a}. \tag{9}$$

Evidently r is a measure of degree of stochastic dominance in the sense of (6).

Another measurement for partial stochastic dominance is *almost* stochastic dominance [7, 6]. Once again the context is $S = [a,b]$ with the usual order $\leq$. For CDFs F and G on $[a,b]$, the measure can be expressed as

$$\alpha(F,G) := \frac{\int (F(x) - G(x))_+ \, dx}{\int |F(x) - G(x)| \, dx} \tag{10}$$

where $v_+ := \max\{v, 0\}$ for any $v \in \mathbb{R}$. Intuitively, if F is almost dominated by G, then $G \leq F$ on most of its domain, and $(F(x) - G(x))_+ = |F(x) - G(x)|$ for most x. Hence $\alpha(F,G)$ is close to 1. In order to ensure that the measure is defined for all pairs F, G, we adopt the convention that $\alpha(F,G) = 1$ when $F = G$.

4. Axioms

Next we propose two axioms for degree of stochastic dominance $\delta(\mu,\nu)$. The first says that $\delta(\mu,\nu)$ should not be large unless the distributions are nearly ordered. To state it, we write $\mu \leq \nu$ to indicate pointwise ordering on $\mathcal{B}$ and define, for each $\mu, \nu \in \mathcal{P}$,

$$\Phi(\mu,\nu) := \{(\mu',\nu') \in \mathcal{M} \times \mathcal{M} : \mu' \leq \mu, \; \nu' \leq \nu, \; \mu' \preceq_{sd} \nu'\}.$$

Think of $\Phi(\mu,\nu)$ as the set of "ordered component pairs" corresponding to (μ,ν). If μ is "almost" dominated by ν, then we can choose relatively large components. (Recall that, to admit the ordering $\mu' \preceq_{sd} \nu'$, we insist that total mass is equal, so $\mu'(S) = \nu'(S)$ must hold.)

Axiom 4.1. For each $(\mu,\nu) \in \mathcal{P} \times \mathcal{P}$ and $\epsilon > 0$, there exists a (μ',ν') in $\Phi(\mu,\nu)$ such that $\delta(\mu,\nu) \leq \mu'(S) + \epsilon$.

Figure 1 helps to illustrate the axiom, with μ and ν represented by densities. In the top subfigure, μ is in no sense dominated by ν, so we wish to enforce $\delta(\mu,\nu) = 0$. Axiom 4.1 does enforce this, since, for this pair (μ,ν), we cannot extract an ordered component pair (μ',ν') with positive mass. Hence $\mu'(S) = \nu'(S)$ is always zero.

In the lower subfigure, there is some overlap in probability mass, so we should permit $\delta(\mu,\nu) > 0$. Inspection shows this to be true. For example, we could take both μ' and ν' to be the function enclosing the shaded region (the pointwise infimum $\mu \wedge \nu$ of the density representations of μ and ν), so that, as measures again, $\mu'(S) = \nu'(S) =$ the area of the shaded region. Since this is positive, $\delta(\mu,\nu)$ can be positive.

Axiom 4.1 adds the converse implication to (6), so δ identifies ordered pairs:

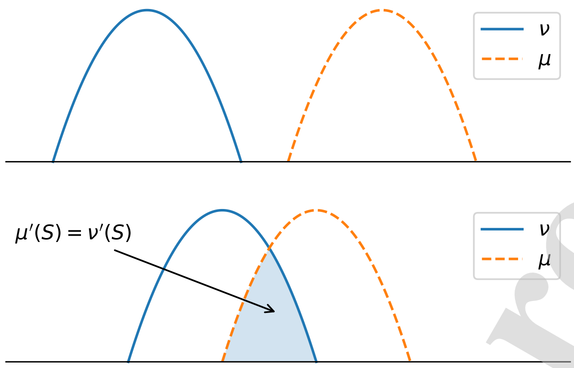

FIGURE 1. Densities and ordered component pairs

Proposition 4.1. *If δ is a measure of degree of stochastic dominance that satisfies Axiom 4.1, then $\delta(\mu, \nu) = 1$ if and only if $\mu \preceq_{sd} \nu$.*

Proof. Suppose that δ satisfies Axiom 4.1 and that $\delta(\mu, \nu) = 1$. Then $\mu'(S) = \nu'(S) = 1$ for some (μ', ν') in $\Phi(\mu, \nu)$. From this we immediately have $\mu' = \mu$ and $\nu' = \nu$, and hence $\mu \preceq_{sd} \nu$, which is all we need to show. □

The second axiom plays the opposite role. It implies that $\delta(\mu, \nu)$ is close to 1 when μ is "nearly" dominated by ν.

Axiom 4.2. Given $\mu_a, \mu_b, \nu_a, \nu_b \in \mathcal{P}$ and $\lambda \in [0, 1]$, the ordering $\mu_a \preceq_{sd} \nu_a$ implies $\delta(\mu, \nu) \geq \lambda$ for $\mu := \lambda\mu_a + (1 - \lambda)\mu_b$ and $\nu := \lambda\nu_a + (1 - \lambda)\nu_b$.

Axiom 4.2 uses the natural convexity of $\mathcal{P}$ to implement "continuity near 1" while avoiding being tied to a particular topology.

The axioms we have listed are strong enough to give uniqueness:

Theorem 4.1. *The only measure of degree of stochastic dominance satisfying Axioms 4.1–4.2 is τ.*

Proof. First we claim that τ satisfies Axiom 4.1. To see this, let (X, Y) be an optimal coupling for (μ, ν), attaining the maximum in (2). Set $\mu'(B) = \mathbb{P}\{X \in B,\ X \preceq Y\}$ and $\nu'(B) = \mathbb{P}\{Y \in B,\ X \preceq Y\}$. This pair satisfies $(\mu', \nu') \in \Phi(\mu, \nu)$ and $\mu'(S) = \tau(\mu, \nu)$. To see that τ satisfies Axiom 4.2, fix $\mu, \nu \in \mathcal{P}$ and let the decompositions in Axiom 4.2 be given, with $\lambda \in [0, 1]$ and $\mu_a \preceq_{sd} \nu_a$. Let (X_i, Y_i) be an optimal coupling of (μ_i, ν_i) for $i = a, b$. Let $X = \ell X_a + (1 - \ell)X_b$ where ℓ is independent and binary with $\mathbb{P}\{\ell = 1\} = \lambda$. Let $Y = \ell Y_a + (1-\ell) Y_b$. Then $(X, Y) \in \Pi(\mu, \nu)$. Hence $\tau(\mu, \nu) \geq \mathbb{P}\{X \preceq Y\} \geq \lambda \mathbb{P}\{X_a \preceq Y_a\} = \lambda$, and Axiom 4.2 holds.

Now let δ be an arbitrary measure of degree of stochastic dominance satisfying Axioms 4.1–4.2. Fix (μ, ν) in $\mathcal{P} \times \mathcal{P}$ and let (X, Y) be an optimal coupling of (μ, ν). Define $\lambda := \mathbb{P}\{X \preceq Y\}$ and the probabilities

$$\mu_a(B) = \frac{\mathbb{P}\{X \in B,\ X \preceq Y\}}{\lambda}, \quad \nu_a(B) = \frac{\mathbb{P}\{Y \in B,\ X \preceq Y\}}{\lambda},$$
$$\mu_b(B) = \frac{\mathbb{P}\{X \in B,\ X \npreceq Y\}}{1-\lambda}, \quad \nu_b(B) = \frac{\mathbb{P}\{Y \in B,\ X \npreceq Y\}}{1-\lambda}.$$

We have $\mu_a \preceq_{sd} \nu_a$, since, when $I \in \mathcal{B}$ is increasing,

$$\mu_a(I) = \frac{\mathbb{P}\{X \in I,\ X \preceq Y\}}{\lambda} \leq \frac{\mathbb{P}\{Y \in I,\ X \preceq Y\}}{\lambda} = \nu_a(I).$$

For μ we have $\mu(B) = \mathbb{P}\{X \in B\}$, which can be decomposed as $\mathbb{P}\{X \in B,\ X \preceq Y\} + \mathbb{P}\{X \in B,\ X \npreceq Y\} = \lambda \mu_a(B) + (1-\lambda)\mu_b(B)$. Since δ satisfies Axiom 4.2, we have $\delta(\mu, \nu) \geq \lambda = \mathbb{P}\{X \preceq Y\} = \tau(\mu, \nu)$.

For the reverse inequality, recall from the arguments above that we can obtain an ordered component pair (μ', ν') satisfying $\nu'(S) = \mu'(S) = \tau(\mu, \nu)$. Since δ satisfies Axiom 4.1, we have $\delta(\mu, \nu) \leq \tau(\mu, \nu) + \epsilon$ for all $\epsilon > 0$. Hence $\delta(\mu, \nu) \leq \tau(\mu, \nu)$. □

5. Measures vs Axioms

Let us reconsider measures of stochastic dominance other than τ. By Theorem 4.1, they fail at least one of the axioms. For example, regarding Axiom 4.2, note that q fails whenever the grid $\{x_i\}$ has at least three points. To see this let F', F'' and G'' be any CDFs on $[a, b]$ such that $F'' < G''$ on (a, b). Let $\lambda \in (0, 1)$ and let

$$F = \lambda F' + (1-\lambda) F'', \quad G = \lambda F' + (1-\lambda) G''.$$

Since $F' \preceq_{sd} F'$, Axiom 4.2 implies that $q(F, G) \geq \lambda$. On the other hand, $G(x_i) > F(x_i)$ on any interior point x_i. Hence m in (7) is at most 2, and $q(F, G) \leq 2/n$. Since λ can be arbitrarily close to 1 this contradicts Axiom 4.2.

For this same pair F, G, the fact that $G > F$ on (a, b) implies that $c^*(F, G) = 0$ for c^* defined in (8), and hence $r(F, G) = 0$ for the restricted stochastic dominance measure defined in (9). Likewise, for the same pair, $\alpha(F, G) = 0$, where α is the almost stochastic dominance measure. Hence r and α also fail to satisfy Axiom 4.2.

Regarding Axiom 4.1, the measures q, r and α all fail. To see this, let $S = [0, 1]$ and, given some positive number ϵ, let μ put mass ϵ on 0 and $1 - \epsilon$ on 1, and let ν put all mass on $1 - \epsilon$. Let F be the CDF of μ and let X be a draw from μ. Let G and

Y be the CDF of and a draw from ν respectively. Since Y is certainly $1-\epsilon$ we have $X \leq Y$ if and only if $X = 0$, and hence $\mathbb{P}\{X \leq Y\} = \epsilon$ for all couplings. On the other hand, $F(x) > G(x)$ iff $0 \leq x < 1-\epsilon$, and hence $r(F, G) = 1-\epsilon$. If $\epsilon < 1/2$ then $1-\epsilon > \epsilon$, and hence r fails Axiom 4.1. The measures q and α also give values larger than ϵ when ϵ is small, although the details are omitted.

6. Final Comments

We end with some avenues for future research. One is that an estimation theory for τ should be straightforward to construct. For example, if we replace the CDFs F and G in (5) with empirical counterparts F_n and G_n, then $\tau(F_n, G_n)$ is a simple transform of the statistic used in the one-sided two-sample Kolmogorov-Smirnov test. This allows for construction of confidence intervals and hypotheses tests related to the value of τ.

Another topic of interest is computation. In higher dimensional settings, computation is difficult for all measures of partial stochastic dominance, but τ at least has an interpretation as the solution to an optimal transport problem. Computing solutions of optimal transport problems is an active research area [10].

Third, there is some connection between the measure τ, which takes values in the interval $[0, 1]$ and represents the continuum between no dominance and complete first order dominance, and the measure proposed in [9], which represents the continuum between first order and second order stochastic dominance. Clarifying this connection and investigating the preceding two topics are left for future work.

7. Acknowledgements

We gratefully acknowledge JSPS KAKENHI Grant 15H05729 and Australian Research Council Grant DP120100321.

References

[1] Anthony B Atkinson. On the measurement of poverty. *Econometrica*, 55(4):749–764, 1987.

[2] Russell Davidson and Jean-Yves Duclos. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, 68(6):1435–1464, 2000.

[3] Russell Davidson and Jean-Yves Duclos. Testing for restricted stochastic dominance. *Econometric Reviews*, 32(1):84–125, 2013.

[4] Gary S Fields, Jesse B Leary, and Efe A Ok. Stochastic dominance in mobility analysis. *Economics Letters*, 75(3):333–339, 2002.

[5] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time.* De Gruyter Textbook Series. De Gruyter, 2011.

[6] Moshe Leshno and Haim Levy. Preferred by "all" and preferred by "most" decision makers: Almost stochastic dominance. *Management Science*, 48(8):1074–1085, 2002.

[7] Haim Levy. Stochastic dominance and expected utility: survey and analysis. *Management Science*, 38(4):555–593, 1992.

[8] Moshe Levy. Almost stochastic dominance and stocks for the long run. *European Journal of Operational Research*, 194(1):250–257, 2009.

[9] Alfred Müller, Marco Scarsini, Ilia Tsetlin, and Robert L Winkler. Between first- and second-order stochastic dominance. *Management Science*, 63(9):2933–2947, 2017.

[10] Gabriel Peyré and Marco Cuturi. *Computational optimal transport.* Now Publishers, 2019.

[11] Stoyan V Stoyanov, Svetlozar T Rachev, and Frank J Fabozzi. Metrization of stochastic dominance rules. *International Journal of Theoretical and Applied Finance*, 15(02), 2012.

[12] Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, pages 423–439, 1965.

[13] Cédric Villani. *Optimal transport: old and new.* Springer Science, 2008.