Increasing Interdependence of Multivariate Distributions

– Preliminary and Incomplete –

Margaret Meyer Bruno Strulovici

November 9, 2009

Abstract

This paper compares n-dimensional random vectors in terms of their interdependence.

1 Introduction

This paper compares n-dimensional random vectors in terms of their interdependence. We adopt the stochastic dominance approach, relating orderings of interdependence expressed directly in terms of joint probability distributions to orderings expressed indirectly through properties of objective functions whose expectations are used to evaluate distributions. Since the expected values of additively separable objective functions depend only on marginal distributions, attitudes towards interdependence must be represented through non-separability properties. We argue that the property of supermodularity (Topkis, 1978) of an objective function is a natural property with which to capture a preference for greater interdependence. Accordingly, we seek to characterize a partial ordering on joint distributions, with identical marginals, which is equivalent to one distribution's yielding a higher expectation than another for all supermodular objective functions. Following the statistics literature, we refer to this partial ordering as the "supermodular stochastic ordering" (Shaked and Shanthikumar, 1997).

For the special case of two-dimensional random vectors, the economics and statistics literatures have provided a complete characterization of the supermodular ordering. Specifically, Epstein and Tanny (1980) and Tchen (1980), among others, have shown that one bivariate distribution dominates another according to the supermodular ordering if and only if the first distribution dominates the second in the sense of both upper-orthant and lower-orthant dominance. Hu, Xie, and Ruan (2005) have shown that this equivalence continues to hold in three dimensions in the special case of Bernoulli random vectors, but the equivalence breaks down for more than three dimensions (Joe, 1990) and even in three dimensions for larger supports (Muller and Scarsini, 2000). In general, the supermodular ordering is strictly stronger than the combination of upper-orthant and lower-orthant dominance.

Focusing on the case of discrete supports, we are able to make substantial progress in characterizing the supermodular ordering for more than two dimensions.

Comparing two n-dimensional distributions with identical marginals, we first prove that one distribution is preferred to the other by every supermodular objective function if and only if the first distribution can be derived from the other by a sequence of non-negative "elementary transformations". Intuitively, our elementary transformations play a similar role to the Rothschild-Stiglitz (1970) elementary transformations that define mean-preserving spreads for univariate distributions. For multivariate distributions, our elementary transformations provide a local characterization of the notion of "greater interdependence", and they are a natural multivariate generalization of the bivariate "correlation-increasing transformations" defined by Epstein and Tanny (1980). By duality, these transformations can also be interpreted as local test functions for the supermodularity of objective functions. We then develop an algorithm, based on the "double description method" conceptualized by Motzkin et al. (1953) and developed by Avis and Fukuda (1992) to generate, for any n-dimensional discrete support, a set of inequalities that are equivalent to preference by all supermodular functions. These inequalities can be easily checked and can thus be straightforwardly used to determine whether different policies, mechanisms, portfolios, etc., can be ranked according to the supermodular ordering. To generate this inequality-based comparison, we exploit the geometric properties of the set of supermodular functions. Precisely, this set is a cone, which can be described either through its positive dual or through its extreme rays; each extreme ray determines exactly one of the inequalities defining the supermodular ordering. We complement the algorithm with constructive proofs of the results for many of the cases considered.

SAY FINITE SUPPORT ALLOWS to use extreme ray techniques and is reasonable in many practical applications.

2 Applications

Our methods and results are applicable to a wide range of questions in economics and related fields. Consider first some applications in welfare economics. In many group settings where individual outcomes (e.g. rewards) are uncertain, members of the group may be concerned, ex ante, about how unequal their ex post rewards will be (Meyer and Mookherjee, 1987; Ben-Porath et al, 1997; Gajdos and Maurin, 2004; Kroll and Davidovitz, 2003; Adler and Sanchirico, 2006). (This concern is distinct from concerns about the mean level of rewards and about their riskiness.) As argued by Meyer and Mookherjee (1987), an aversion to ex post inequality can be formalized by adopting an ex post welfare function that is supermodular in the realized utilities of the different individuals. We then want to know: Given two mechanisms for allocating rewards (formally, two joint distributions of random utilities), when can we be sure that one mechanism generates higher expected welfare than the other, for all supermodular ex post welfare functions? Our stochastic dominance theorems allow us to answer this question.

Consider a specific illustration. Intuitively, when groups dislike ex post inequality, tournament reward schemes, which distribute a fixed set of rewards among individuals, one to each person, should be particularly unappealing, since they generate a form of negative correlation among rewards: if one person receives a higher reward, this must be accompanied by another person's receiving a lower reward. This intuitive reasoning suggests the conjecture that tournaments should be dominated, in the sense of the supermodular ordering, by reward schemes that provide each individual with the same marginal distribution over rewards but determine rewards independently. Meyer and Mookherjee (1987) proved this conjecture, but only for the special case of a symmetric tournament (one in which each individual has an equal chance of winning each of the rewards), and their method of proof was laborious. Here, we allow tournaments to be arbitrarily asymmetric across individuals, and we apply our three-dimensional characterization result to show that the conjecture is true for the case of three individuals.

A second application in welfare economics concerns comparisons of inequality or poverty when separate data are available on different dimensions of economic status, for example, income, health, and education (Atkinson and Bourguignon, 1982, and Bourguignon and Chakravarty, 2002). Depending on whether the different attributes are regarded as complements or substitutes at the individual level, the function aggregating the attributes into an individual welfare measure will be supermodular or submodular, and our stochastic dominance theorems provide the conditions under which one multidimensional distribution can be ranked above another for all welfare measures in the given class.

Another set of economic applications concerns comparisons of the efficiency of two-sided or many-sided matching mechanisms when the outcomes of the matching process are subject to informational or search frictions. Consider, for example, settings where different categories of workers (e.g. newly-qualified and experienced, or technical and managerial) are matched with firms. Suppose that workers within each category, as well as firms, are heterogeneous and that the production function giving the output of a matched set of workers at a given firm, as a function of the workers' types and the firm's type, is supermodular. In the absence of any frictions, the efficient matching would be perfectly assortative, matching the highest-quality worker in each category with the highest-quality firm, the next-highest-quality workers with the next-highest-quality firm, etc. Such a matching would correspond to a "perfectly correlated" joint distribution of the random variables representing quality in each category (dimension). When, however, matches are formed based only on noisy information, or when search is costly, or when signaling is constrained by market imperfections such as borrowing constraints, perfectly assortative matching will generally not arise. In these settings, our stochastic dominance theorems can be used to assess when one matching mechanism will generate higher expected output than another, for all supermodular production functions. Fernandez and Gali (1999) and Meyer and Rothschild (2003) apply existing two-dimensional results to compare matching institutions, but multi-dimensional applications remain largely unexplored. One exception is Prat (2002), but he compares only a perfectly correlated joint distribution with an independent one, and Lorentz (1953) has shown that the former is preferred to the latter for all supermodular objective functions.

In finance, characterizations of the supermodular ordering can be applied to the comparison of the dependence among assets in a portfolio, and in insurance, to the comparison of the dependence among claim streams (Muller and Stoyan, 2002, and Denuit, Dhaene, Goovaerts, and Kaas, 2005).

3 General Setting

This section introduces the general setting analyzed in the paper.

Distribution Support We consider multivariate distributions with the same number, n, of variables and identical, finite support (these assumptions will discussed later). Formally, let L_i denote the finite, totally ordered set of values taken by the i^{th} random variable, and let L denote the cartesian product of L_i's. For all applications, and in what follows, L_i is a finite subset of $\mathbb{R}$ and L is a finite lattice of $\mathbb{R}^n$ with the following partial order: $x \leq y$ if and only if $x_i \leq y_i$ for all $i \in N = \{1, \ldots, n\}$. If l_i denotes the cardinality of L_i, then L has $d = \prod_{i=1}^{n} l_i$ elements.

As a specific example, let $L_{l_1,\ldots,l_n}$ denote the lattice of $\mathbb{R}^n$ with $L_i = \{0, \ldots, l_i - 1\}$. Thus, for example, $L_{2,2}$ consists of the vertices of the unit square in $\mathbb{R}^2$ based at the origin: $L_{2,2} = \{0,1\}^2$. Similarly, $L_{2,2,2}$ consists of the vertices of the unit cube of $\mathbb{R}^3$ based at the origin: $L_{2,2,2} = \{0,1\}^3$.

For any $x \in L$, let $x + e_i$ denote the element y of L, whenever it exists, such that $y_j = x_j$ for all $j \in N \setminus \{i\}$ and y_i is the smallest element of L_i greater than but not equal to x_i. For example, in $L_{2,2}$, $(0,0) + e_1 = (1,0)$ and $(1,0) + e_2 = (0,0) + e_1 + e_2 = (1,1)$.

Lattice vs. Vector Structures. The lattice structure of the support L and its corresponding order is used to compare distributions. In particular, supermodularity of objective functions is defined with respect to that partial order. One may label the d elements (or "nodes") of L and view real functions on L as vectors of $\mathbb{R}^d$, where each coordinate of the vector corresponds to the value of the function at a specific node of L. This representation will prove particularly important for dual characterizations of interdependence relations. A multivariate distribution whose support is L (or a subset of L) can be represented as an element of the unit simplex Δ_d of $\mathbb{R}^d$.

Orderings of Multivariate Distributions. For any function $w : L \to \mathbb{R}$ and distribution $f \in \Delta_d$, the expected value of w given f is the scalar product of w with f, seen as vectors of $\mathbb{R}^d$:

$$E[w|f] = \sum_{x \in L} f(x)w(x) = w \cdot f,$$

where $\cdot$ denotes the scalar product of w and f in $\mathbb{R}^d$. To any class $\mathcal{W}$ of functions on L

corresponds an ordering of multivariate distributions:

$$f \prec_{\mathcal{W}} g \quad \Leftrightarrow \quad \forall w \in \mathcal{W}, \quad E[w|f] \leq E[w|g] \tag{1}$$

The main purpose of this paper is to better understand the orders defined according to such classes of functions, starting with the stochastic supermodular ordering, which is based on supermodular functions.

4 The Stochastic Supermodular Ordering

Supermodular Functions and Elementary Transformations For any $x, y \in L$, denote by $x \wedge y$ the component-wise minimum (or "meet") of x and y, i.e. the element of L such that $(x \wedge y)_i = \min\{x_i, y_i\} \in L_i$ for all $i \in N$. Let $x \vee y$ similarly denote the component-wise maximum (or "join") of x, y. A function w is said to be *supermodular* (on L) if $w(x \wedge y) + w(x \vee y) \geq w(x) + w(y)$ for all $x, y \in L$. Supermodular functions are characterized by the following property (see Topkis, 1968):

$$w \in \mathcal{S} \quad \Leftrightarrow \quad w(x + e_i + e_j) + w(x) \geq w(x + e_i) + w(x + e_j) \tag{2}$$

for all $i \neq j$ and x such that $x + e_i + e_j$ is well-defined (i.e. such that x_i is not the upper bound of L_i and x_j is not the upper bound of L_j). For any $x \in L$ such that $x + e_i + e_j$ is well-defined, let $t^x_{i,j}$ denote the function on L such that

$$t^x_{i,j}(x) = t^x_{i,j}(x + e_i + e_j) = -t^x_{i,j}(x + e_i) = -t^x_{i,j}(x + e_j) = 1 \tag{3}$$

and $t^x_{i,j}(y) = 0$ for all other nodes y of L. We call these functions the *elementary transformations* on L. Let $\mathcal{T}$ denote the class of all elementary transformations.

For example, for $L_{2,2}$, there is a single elementary transformation, which is defined by $t(1,1) = t(0,0) = 1$ and $t(1,0) = t(0,1) = -1$. For $L_{2,2,2}$, there are six elementary transformations, one corresponding to each face of the unit cube. For $L_{3,3}$, there are four elementary transformations, corresponding to the four values of x , namely $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$, such that $x + e_i + e_j$ is well defined. Observe that our definition of elementary transformations confines attention to transformations that i) affect only *two* of the n dimensions (as illustrated by the example of $L_{2,2,2}$) and ii) affect values only at four *adjacent* points in the lattice, x, $x + e_i$, $x + e_j$, and $x + e_i + e_j$ (as illustrated by the example of $L_{3,3}$).

With this notation, (2) can be re-expressed as

$$w \in \mathcal{S} \Leftrightarrow w \cdot t \geq 0 \quad \forall t \in \mathcal{T}. \tag{4}$$

Now that we have a formal characterization of the class of supermodular functions, we can formally define the (stochastic) supermodular ordering:

$$f \prec_{\mathcal{S}} g \Leftrightarrow \forall w \in \mathcal{S}, \quad E[w|f] \leq E[w|g] \tag{5}$$

If $f \prec_{\mathcal{S}} g$, we will say that distribution g is *more interdependent* than distribution f.

Dual Characterization When does a random vector Y, distributed according to g, exhibit more interdependence among its components than another random vector X, distributed according to f? What modifications to the distribution of a random vector increase interdependence among the random variables composing it? The answer is given in the following theorem.

THEOREM 1 (SUPERMODULAR ORDERING) *$f \prec_{\mathcal{S}} g$ if and only if there exist* nonnegative *coefficients $\{\alpha_t\}_{t\in\mathcal{T}}$ such that, with f, g, and t seen as vectors of* $\mathbb{R}^d$,

$$g = f + \sum_{t\in\mathcal{T}} \alpha_t t. \tag{6}$$

Proof. Equation (6) holds if and only if $g - f$ belongs to the convex cone $\mathcal{T}^C$ generated by $\mathcal{T}$, i.e. defined by $\mathcal{T}^C = \{\sum_{t\in\mathcal{T}} \alpha_t t : \alpha_t \geq 0 \quad \forall t \in \mathcal{T}\}$. From (4), $\mathcal{S}$ is the dual cone of $\mathcal{T}^C$. Since $\mathcal{T}^C$ is closed and convex, this implies (see Luenberger, 1969, p. 215) that $\mathcal{T}^C$ is the dual cone of $\mathcal{S}$. That is,

$$\delta \in \mathcal{T}^C \Leftrightarrow w \cdot \delta \geq 0 \quad \forall w \in \mathcal{S}.$$

By definition of the stochastic supermodular ordering (see (5)), the above equation exactly means that $f \prec_{\mathcal{S}} g$ if and only if $g - f \in \mathcal{T}^C$, which shows the result. ■

Coarsening For many applications, the choice of a particular support seems somewhat arbitrary. For example, when comparing several empirical distributions of inequality across various components (such as income, health, and education), the distribution depends on the way data has been aggregated into discrete categories. It is natural, then, to ask whether our notion of greater interdependence is robust with respect to further aggregation. Theorem 1 provides a way to answer this question.

Define a *coarsening* M of some support L by a partitioning of each L_i into M_i, consisting of $m_i \leq l_i$ components of consecutive elements of L_i. For example, if $L = \{0, 1, 2, 3\} \times \{0, 1, 2\}$, one possible coarsening of L is $M = \{\{0, 1\}, \{2, 3\}\} \times \{\{0\}, \{1, 2\}\}$. To any coarsening M of L corresponds a surjective map $\phi : L \to M$ such that $\phi(x) = \phi(x')$ if and only if x_i and x'_i belong to the same element y_i of M_i for all i. Each element of M represents a hyper-rectangle resulting from slicing L along (possibly) each dimension. For any distribution f on L and any coarsening M of L, let f^M denote the "coarsened version" of f, which is defined by

$$f^M(y) = \sum_{x \in L : \phi(x) = y} f(x).$$

To indicate dependence with respect to the chosen support, let $\mathcal{S}(L)$ denote the set of all supermodular functions with domain L.

Theorem 2 (Coarsening Invariance) *If $f \prec_{\mathcal{S}(L)} g$, then for any coarsening M of L, $f^M \prec_{\mathcal{S}(M)} g^M$.*

Proof. Suppose that $f \prec_{\mathcal{S}(L)} g$. By Theorem 1, this implies the existence of nonnegative coefficients α_t such that

$$g = f + \sum_{t \in \mathcal{T}(L)} \alpha_t t, \tag{7}$$

where $\mathcal{T}(L)$ is the set of elementary transformations on L. Let Φ denote the operator which to any function w on L associates the function on M defined by $\Phi(w)(y) = \sum_{x \in L : \phi(x) = y} w(x)$. Φ is a linear operator, and by construction, $f^M = \Phi(f)$. Applying Φ to (7) yields

$$g^M = f^M + \sum_{t \in \mathcal{T}(L)} \alpha_t \Phi(t).$$

Now observe that for $t = t^x_{i,j} \in \mathcal{T}(L)$, $\Phi(t)$ belongs to $\mathcal{T}(M)$ if $\phi(x)$, $\phi(x + e_i)$, $\phi(x + e_j)$, and $\phi(x + e_i + e_j)$ are all distinct, and $\Phi(t)(y) = 0$ for all $y \in M$ otherwise. Therefore,

$$g^M = f^M + \sum_{t \in \mathcal{T}(M)} \alpha_t t,$$

for some nonnegative coefficients α'. Another application of Theorem 1 then implies that $f^M \prec_{\mathcal{S}(M)} g^M$, which concludes the proof. ■

Thus, if distribution g is more interdependent than distribution f on a given support L, then on any coarsening M of L, the coarsened version of g, g^M, is more interdependent than the coarsened version of f, f^M.

In the next several sections, we develop a range of methods for determining, given a pair of distributions f and g, whether g is more interdependent than f. These methods apply the characterization result of Theorem 1 and are greatly facilitated by two aspects of our approach. The first is our restriction to a *finite* support L. The second is the manner in which we have defined the elementary transformations on L, requiring that the transformations affect only two of the n dimensions and affect values at only adjacent points in the lattice. These two features of our approach imply that it is very straightforward, either manually or algorithmically, to list the entire set $\mathcal{T}$ of elementary transformations on any given L. Furthermore, given a pair of distributions f, g, when we search for a representation of $g - f$ as a nonnegative weighted sum $\sum_{t \in \mathcal{T}} \alpha_t t$, we can be certain that none of the elementary transformations in $\mathcal{T}$ is redundant, as demonstrated by the following result:

THEOREM 3 *All elements of* $\mathcal{T}$ *are extreme rays of* $\mathcal{T}^C$*, the convex cone generated by* $\mathcal{T}$*.*

Proof. Without loss of generality, we prove the claim for $L = L_{l_1,\ldots,l_n}$ (other cases are treated with an obvious modification of the function w below). Consider a point $x \in L$ and a pair of dimensions i, j such that $x - e_i - e_j$ is well-defined, so that the elementary transformation $t^* \equiv t_{i,j}^{x-e_i-e_j}$ is well-defined. Suppose that, contrary to the claim, there exist nonnegative coefficients α_s such that

$$t^* = \sum_{s \in \mathcal{T} \setminus \{t^*\}} \alpha_s s. \tag{8}$$

Let us define the function w on L by $w(x) = \frac{3}{4} 2^{\sum_k x_k}$ and, for $y \neq x$, $w(y) = 2^{\sum_k y_k}$. It is easy to check that w is supermodular. Moreover, w makes a nonnegative scalar product with all elementary transformations and a positive scalar product with all elementary transformations except for those whose highest corner is x. Since t^* is one of the elementary transformations whose highest corner is x, taking the scalar product of w with both sides of (8) implies that

$$0 = \sum_{s \in \mathcal{T} \setminus \{t\}} \alpha_s (w \cdot s).$$

This equation in turn implies that $\alpha_s = 0$ for all transformations s except possibly those whose highest corner is x. However, t^* cannot be a positive linear combination of only elementary transformations whose highest corner is x. To see this, observe that any elementary transformation s (other than t^*) whose highest coner is x must take value 0 at $x - e_i - e_j$, whereas t^* evaluated at $x - e_i - e_j$ equals 1. ∎

For the special case of two dimensions, a stronger result is easily shown: It is impossible to write any elementary transformation $t \in \mathcal{T}$ as a sum, with weights of *arbitrary* sign, of other elementary transformations in $\mathcal{T}$. However, for three or more dimensions, this stronger condition does not hold, as the following example demonstrates: Let $n = 3$ and $L = \{0,1\}^3$. Then $t_{13}^{(0,0,0)} = t_{13}^{(0,1,0)} - t_{23}^{(1,0,0)} + t_{23}^{(0,0,0)}$.

The constructive methods we develop for determining whether a distribution g is more interdependent than a distribution f also exploit an important implication of the relation $f \prec_{\mathcal{S}} g$, namely that f and g have identical univariate marginal distributions. To see why this holds, note that for any dimension $i \in \{1, \ldots, n\}$ and any $k \in L_i$, the functions $\overline{w}(x) = I_{\{x_i \geq k\}}$ and $\underline{w}(x) = I_{\{x_i < k\}}$ are both supermodular. Therefore $f \prec_{\mathcal{S}} g$ implies that, for all $i \in \{1, \ldots, n\}$ and any $k \in L_i$,

$$\begin{aligned}
0 \leq E[\overline{w}|g] - E[\overline{w}|f] &= \sum_{x:\, x_i \geq k} g(x) - \sum_{x:\, x_i \geq k} f(x) \\
\text{and} \quad 0 \leq E[\underline{w}|g] - E[\underline{w}|f] &= \sum_{x:\, x_i < k} g(x) - \sum_{x:\, x_i < k} f(x),
\end{aligned} \tag{9}$$

and these inequalities together imply that f and g have identical univariate marginal distributions. This conclusion also follows from the characterization of Theorem 1, given that for any elementary transformation $t \in \mathcal{T}$ and for any α, $f + \alpha t$ and f have the same marginal distributions.

5 Two Dimensions

Theorem 1 tells us that, given two distributions f, g, determining whether $f \prec_{\mathcal{S}} g$ is equivalent to determining whether the difference vector $\delta = g - f$ can be decomposed into a nonnegative weighted sum of elementary transformations. For the special case of bivariate distributions ($n = 2$), we now show that, given how we have defined elementary transformations, this determiniation is very simple. Given f, g with identical marginal distributions and defined on $L = L_{l_1,l_2} \equiv \{0, \ldots, l_1 - 1\} \times \{0, \ldots, l_2 - 1\}$, the difference vector δ is fully described by its values at $(l_1-1)\times(l_2-1)$ points (the remaining values being pinned down by the condition of identical marginals), and there are exactly $(l_1-1)\times(l_2-1)$ (linearly independent) elementary transformations defined as in (3). Therefore, there is a *unique* decomposition of δ into a weighted sum of elementary transformations $t \in \mathcal{T}$, where the weights α_t can have arbitrary signs. Since the decomposition is unique, $f \prec_{\mathcal{S}} g$ if and only if every elementary transformation has a nonnegative weight in the decomposition.

It is also straightforward to identify the weight on each elementary transformation in the unique decomposition, as a function of the difference vector δ. To simplify notation, note that with only two dimensions, given an arbitrary $z \in L$, we can write t^z instead of $t^z_{i,j}$ for the elementary transformation defined in (3). Also, let $\alpha(z)$ denote α_{t^z}. The elementary transformation t^z is well-defined for $z \in \{0, \ldots, l_1 - 2\} \times \{0, \ldots, l_2 - 2\} \equiv L_{(l_1-1),(l_2-1)}$. With only two dimensions, for any given $z \in L_{(l_1-1),(l_2-1)}$, there are at most four elementary transformations $t \in \mathcal{T}$ that take on non-zero values at z: t^z, $t^{(z-e_1)}$, $t^{(z-e_2)}$, and $t^{(z-e_1-e_2)}$. If $z = (z_1, 0)$, then $z - e_2$ is not well-defined; it is convenient in this case to say that $t^{(z-e_2)}$ is identically 0. Similarly, if $z = (0, z_2)$, then $z - e_1$ is not well-defined, and in this case we say that $t^{(z-e_1)}$ is identically 0. With these conventions, it follows that for any $z \in L_{(l_1-1),(l_2-1)}$,

$$\begin{aligned}\delta(z) &= \alpha(z)t^z(z) + \alpha(z-e_1)t^{(z-e_1)}(z) + \alpha(z-e_2)t^{(z-e_2)}(z) + \alpha(z-e_1-e_2)t^{(z-e_1-e_2)}(z) \\ &= \alpha(z) - \alpha(z-e_1) - \alpha(z-e_2) + \alpha(z-e_1-e_2), \qquad (10)\end{aligned}$$

where the second line follows from the definition of elementary transformations in (3).

Now consider $z = (0, 0)$. Since the only elementary transformation that takes on a non-zero value on $(0, 0)$ is $t^{(0,0)}$ (i.e. none of $(z - e_1)$, $(z - e_2)$, or $(z - e_1 - e_2)$ is well-defined), (10) reduces to $\delta(0, 0) = \alpha(0, 0)$. Thus the weight $\alpha(0, 0)$ on $t^{(0,0)}$ in the unique decomposition of δ is $\delta(0, 0)$. Proceed now to $z = (1, 0)$. Since the only two elementary transformations that take on non-zero values on $(1, 0)$ are $t^{(1,0)}$ and $t^{(0,0)}$, (10) reduces to $\delta(1, 0) = \alpha(1, 0) - \alpha(0, 0)$, and hence $\alpha(1, 0) = \delta(0, 0) + \delta(1, 0)$. Straightforward induction arguments then show that for $z = (z_1, 0)$, $\alpha(z_1, 0) = \sum_{i=0}^{z_1} \delta(i, 0)$; for $z = (0, z_2)$, $\alpha(0, z_2) = \sum_{j=0}^{z_2} \delta(0, j)$; and finally for $z = (z_1, z_2)$, $\alpha(z_1, z_2) = \sum_{i=0}^{z_1} \sum_{j=0}^{z_2} \delta(i, j)$. If we define G and F as the cumulative distribution functions corresponding to g and f, respectively, then we have $G(z_1, z_2) - F(z_1, z_2) = \sum_{i=0}^{z_1} \sum_{j=0}^{z_2} \delta(i, j)$. Thus, in the unique decomposition of $\delta = g - f$ into a weighted sequence of elementary transformations, the weight $\alpha(z)$ on the transformation t^z is the difference $G(z) - F(z)$. Since $f \prec_{\mathcal{S}} g$ if and only if every elementary transformation has a nonnegative weight in the decomposition, it follows that for two dimensions,

$$f \prec_{\mathcal{S}} g \quad \Leftrightarrow \quad G(z) - F(z) \geq 0 \quad \forall z \in L. \qquad (11)$$

Note that (11) is written for all $z \in L$ and not just for all $z \in L_{(l_1-1),(l_2-1)}$, because identical marginals is a necessary condition for $f \prec_{\mathcal{S}} g$ and ensures that for $z = (l_1 - 1, 0)$ or $z = (0, l_2 - 1)$, $G(z) - F(z) = 0$.

For random variables $(Y_1, \ldots, Y_n)$ and $(X_1, \ldots, X_n)$ with distribution g and f, respectively, define the survival functions $\overline{G}$ and $\overline{F}$ by $\overline{G}(z) = P(Y \geq z)$ and $\overline{F}(z) = P(X \geq z)$. In the special case of two dimensions, if g and f have identical marginal distributions, then $\overline{G}(z) - \overline{F}(z) = G(z - e_1 - e_2) - F(z - e_1 - e_2)$, so

$$G(z) - F(z) \geq 0 \quad \forall z \in L \quad \Leftrightarrow \quad \overline{G}(z) - \overline{F}(z) \geq 0 \quad \forall z \in L. \tag{12}$$

Joe (1990) has defined a notion of greater interdependence for multivariate distributions which he terms the "concordance order": g dominates f according to the concordance order, written $f \prec_{\mathcal{C}} g$, if for all $z \in L$, both $G(z) - F(z) \geq 0$ and $\overline{G}(z) - \overline{F}(z) \geq 0$ hold. For bivariate distributions, by combining (11) and (12) we can conclude that

$$f \prec_{\mathcal{S}} g \quad \Leftrightarrow \quad f \prec_{\mathcal{C}} g. \tag{13}$$

The equivalence between the supermodular order and the concordance order for bivariate distributions is well known and has been proved by Levy and Parousch (1974), Epstein and Tanny (1980), and Tchen (1980).

Not yet finished

Footnote to the effect that there is an analogous simple proof of Rothschild and Stiglitz's results, when we define a special type of MPS.

This also means that that δ must make a positive scalar product with the supermodular function that equals 1 at (l_1, l_2) and $(l_1, l_2 - 1)$ and zero everywhere else, or equivalently that the expected value of this function is higher under g than under f. This function is also an extreme ray of the cone of supermodular functions, as is easily checked. In fact proceeding as above, one can read off all weights of elementary transformations, and find that nonnegativity of these weights is equivalent to the requirement that all extreme supermodular functions have a higher expectation under g than under f. That is, each elementary transformation corresponds to an extreme ray of the cone of supermodular functions. That property, however, does not hold for three or more dimensions.

6 Constructive Methods for Comparing Distribution Interdependence

Given two distributions f and g, how can one determine whether g is more interdependent than f? We provide several answers to this question based on the characterization of Theorem 1, which says that $f \prec_{\mathcal{S}} g$ if and only if the vector $g - f$ is decomposable into a non-negative weighted sum of elementary transformations.

The simplest answer consists of building explicit weights and deriving explicit conditions on $g-f$, as we did for the two dimensional case. As previously mentioned, however, there is not a one-to-one mapping between elementary transformations and extreme rays of the cone of supermodular functions, and so it is impossible to apply the simple algorithm described in the two-dimensional case. Nevertheless, direct constructive methods can be used for other special cases which provide a number of insights concerning the structure of the supermodular ordering in higher dimensions. We have characterized the supermodular ordering in several such cases, and present three of them. The first, simplest example is the cube, that is, the case where $L = \{0,1\}^3$. The second example is the case where $L = \{0,1\}^4$ and where we confine attention to distributions satisfying a symmetry property that we term "top-to-bottom symmetry" (defined precisely below). The third example is the case where $L = \{0,1,2\}^3$ and where we impose a different form of symmetry, symmetry across dimensions. We defer discussion of this example until Section XXX, where we analyze the symmetric supermodular ordering in detail.

A second approach to determining whether g is more interdependent than f is to formulate a linear program, based on the set of elementary transformations on L, such that the optimum value of the program is zero if and only if there exist non-negative coefficients $\{\alpha_t\}_{t\in\mathcal{T}}$ such that $g - f = \sum_{t\in\mathcal{T}} \alpha_t t$. As we will see, this method has the advantage of constructing an explicit sequence of elementary transformations that, added to f, result in g. However, it also has the drawback that one has to solve a different linear program for each pair of distributions to be compared.

A third method, based on Minkowski's and Weyl's representation theorems for polyhedral cones, allows one to compute once and for all, for any given support L, a minimal set of inequalities that characterize the stochastic supermodular ordering, such that $f \prec_{\mathcal{S}}$ g if and only if the vector $g - f$ satisfies these inequalities. This method can be used for optimization problems such as mechanism design or analysis of optimal policy, where each mechanism or policy generates a multivariate distribution and the set of mechanisms or policies is large. In such settings, one must compare many distributions, and this so-called "double description method" may significantly reduce computations.

6.1 Supermodular Ordering on the Three-Dimensional and Four-Dimensional Cubes

Consider the case of three dimensions, where each dimensions has two points in its support, i.e. $L = \{0,1\}^3$. The difference vector $\delta = g - f$ is represented on Figure XXX. There are six elementary transformations: three upper ones and three lower ones. Let X_{ij} (resp. Y_{ij}) denote the upper transformation involving b_i and b_j (resp. c_i and c_j). In contrast to the two-dimensional case, the difficulty is that one cannot simply read off the weights of elementary transformations. Indeed, consider first the top corner, i.e. $\delta(1,1,1) = a$. That corner is sensitive to the *three* elementary transformations X_{ij}. Indeed, if δ can be decomposed at all on these six elementary transformations, then necessarily $a = \sum x_{i,j}$ where $x_{i,j}$ is the weight of ET X_{ij}. Nonnegative of these weights then implies that a must be nonnegative or, put differently, that the expectation of the supermodular function that equals 1 at $(1,1,1)$ and zero everywhere else, is higher under g than under f. Now consider the the node $(0,1,1)$. By decomposition, the value b_3 must equal $-x_{13} - x_{12} + y_{23}$. Proceeding as above, Theorem 1 says that $f \prec_{\mathcal{S}} g \iff \exists\{x_{ij} \geq 0, y_{ij} \geq 0\}$ s.t.

$$x_{12} + x_{13} + x_{23} = a,\ b_1 + x_{12} + x_{13} = y_{23},\ b_2 + x_{12} + x_{23} = y_{13}, \text{ and } b_3 + x_{13} + x_{23} = y_{12}. \quad (*)$$

Set $x_{ij} = a(\frac{a+b_k}{3a+\sum_{i=1}^3 b_i})$ and $y_{ij} = (2a + \sum_{i=1}^3 b_i)(\frac{a+b_k}{3a+\sum_{i=1}^3 b_i})$. Then, equations (*) are satisfied, and the inequalities $a \geq 0$, $2a + \sum_{i=1}^3 b_i \geq 0$, and $a + b_i \geq 0 \quad \forall i = 1,2,3$, which are equivalent to $f \prec_{\mathcal{C}} g$ in this context, ensure that $\{x_{ij} \geq 0, y_{ij} \geq 0\}$, so $f \prec_{\mathcal{S}} g$. Necessity of these inequalities can be easily shown by using a basis of supermodular functions. For example, the function that puts a 2 at the top node, a 1 at its three adjacent nodes, and zero everywhere else, is supermodular. The requirement that this function should make a positive scalar product with δ immediately implies that $2a + \sum_i b_i \geq 0$. This supermodular function is "equivalent" to the supermodular function that puts a one at the bottom corner of the cube and zero everywhere else, in the sense that their difference is a modular function or, put differently, they make the same scalar product with each ET. In fact one may check that in the case of the cube, extreme rays of the cone of supermodular functions are the set of indicator functions of upper and lower orthants, which implies that the supermodular order and the concordance order are equivalent for the cube. For three or more dimensions, this property does not hold if there is at least one dimension i for which L_i has cardinality greater than 2. Nevertheless, the insight behind the constructive characterization for $L = \{0,1\}^3$ can be extended to larger supports, as we now illustrate for the case where $L = \{0,1\}^4$.

6.2 The Linear Programming Approach: Comparing Two Specific Distributions

From Theorem 1, $f \prec_{\mathcal{S}} g$ if and only there exist nonnegative coefficients $\{\alpha_t\}_{t\in\mathcal{T}}$ such that $g - f = \sum_{t\in\mathcal{T}} \alpha_t t$. Given a specific pair of distributions f and g, we can formulate the problem of determining whether such a set of coefficients exists as a linear programming problem. Let $\tau = |\mathcal{T}|$ denote the number of elementary transformations on L, and let E denote the $d \times \tau$-matrix whose columns are the d-dimensional vectors consisting of all elementary transformations of L. Theorem 1 can be re-expressed as $f \prec_{\mathcal{S}} g$ if and only if there exists $\alpha \in \mathbb{R}^\tau$ such that i) $\alpha \geq 0$ and ii) $E\alpha = g - f$. Now define the d-dimensional vector δ^+ such that $\delta_i^+ = |(g-f)_i|$, and let E^+ denote the matrix whose i^{th} row, denoted E_i^+, satisfies $E_i^+ = (-1)^{\varepsilon_i} E_i$, where $\varepsilon_i = 1$ if $(g-f)_i < 0$ and 0 otherwise. The condition $E\alpha = g - f$ can be re-expressed as $E^+\alpha = \delta^+$. Now consider the following[1] linear program (A):

$$\min_{(\alpha,\beta)\in\mathbb{R}^\tau\times\mathbb{R}^d} \sum_{i=1}^{d} \beta_i$$

subject to

$$E^+\alpha + \beta = \delta^+, \quad \alpha \geq 0, \quad \beta \geq 0.$$

Theorem 4 (Pairwise Comparison) *The linear program (A) always has an optimal solution. $f \prec_{\mathcal{S}} g$ if and only if the optimum value is zero, and in that case $g = f + \sum_{t\in\mathcal{T}} \alpha_t^* t$, where (α^*, β^*) is any minimizer of (A) and $\beta^* = 0$.*

Proof. There always exists a feasible vector (α, β), namely $(\alpha, \beta) = (0, \delta^+)$. Moreover, the value function is nonnegative since the feasibility constraints require that β have nonnegative components, and therefore the optimum is nonnegative. If $f \prec_{\mathcal{S}} g$, there exists $\alpha^* \geq 0$ such that $E^+\alpha^* = \delta^+$, so the optimum value of program (A) must indeed be

[1] This corresponds to the auxiliary program for the determination of a basic feasible solution described in Bertsimas and Tsitsiklis (1997, Section 3).

zero, since that value is achieved by $(\alpha, \beta) = (\alpha^*, 0)$. Reciprocally, if there exists (α^*, β^*) such that the value of the program is zero, then necessarily $\beta^* = 0$ and $E^+\alpha^* = \delta^+$. ■

COMMENT ON THE MEANING OF ALPHA - ADD OBJECTIVE FUNCTION - MEANING OF BASIC FEASIBLE SOLUTION - SHADOW VALUES - Interpretation of alpha to locate the causes of interdependence differential.

6.3 The Double Description Method

The linear programming approach just described has the drawback of requiring a new program to be solved each time a new pair of distributions is to be compared. When many distributions are to be compared, for example as part of a larger optimization problem, it is more convenient to have an explicit representation of the stochastic supermodular ordering for the common support of these distributions. We now provide such a representation in the form of a list of inequalities that are satisfied by the vector $g - f$ if and only if $f \prec_{\mathcal{S}} g$. For any given finite support L, these inequalities are computed once and for all, a computation which is made possible by the support's finiteness.

Recall that $f \prec_{\mathcal{S}} g$ if $g - f$ makes a nonnegative scalar product with all supermodular functions on L, seen as vectors of $\mathbb{R}^d$. This condition can be reduced to a finite set of inequalities by exploiting the geometric properties of $\mathcal{S}$. $\mathcal{S}$ is a convex cone characterized by the fact that w is supermodular (i.e. belongs to $\mathcal{S}$) if and only if it makes a nonnegative scalar product with all elementary transformations on L. In matrix form, $\mathcal{S} = \{w \in \mathbb{R}^d : Aw \geq 0\}$, where $A = E'$ is the matrix whose rows consist of all elementary transformations (i.e. the transpose of the matrix E introduced earlier). A is called the *representation matrix* of the polyhedral cone $\mathcal{S}$. Minkowski's theorem states that to any representation matrix corresponds a *generating matrix* R such that

$$Ax \geq 0 \quad \Leftrightarrow \quad x = R\lambda \quad \text{for some } \lambda \geq 0.$$

The columns of the matrix R are the extreme rays of the cone $\mathcal{S}$. There exists a finite number of such extreme rays. The stochastic supermodular ordering is entirely determined by the extreme rays:

$$E[w|f] \leq E[w|g] \quad \forall w \in \mathcal{S} \quad \Leftrightarrow \quad R'(g - f) \geq 0.$$

Minkowski's theorem thus proves the existence of a finite list of inequalities that entirely characterize the stochastic supermodular ordering. How to determine the extreme rays of the cone of supermodular functions? The *double description method*, conceived by Motzkin et al. (1953) and implemented by Fukuda et al. and Fukuda (2004) builds on Minkowski's and Weyl's representation theorems for polyhedral cones (REFERENCE NEEDED). A polyhedral cone can be either represented by a set of inequalities (i.e. by the intersection of a number half-spaces) or by extreme rays. The double description method provides an algorithm to determine one description from the other. Luckily, the set of elementary transformations is trivially computable, and can be automatically generated for any given support L. From this input, the double description method can compute the set of extreme supermodular functions. Using Fukuda's algorithm for the double description method, we have computed the stochastic supermodular order for a range of problems that are intractable by hand. In the Appendix, we illustrate the method for the case where $L = \{0,1\}^4$ and no symmetry assumptions of any sort are imposed.

Complexity of the Double Description Method Although the double description method is very useful in theory, its computational complexity is unsurprisingly exponential in the size of L. Keeping in mind the potential applications of the stochastic supermodular ordering (WHAT DOES THIS MEAN EXACTLY?), we now provide an exact computation of the algorithm's complexity.

Avis and Bremner (1995) (CHECK PRECISE STATEMENT) show that the double description algorithm described by Motzkin et al. (1953) has complexity $O(p^{\lfloor d/2 \rfloor})$ where d is the dimension of the space and p is the number of inequalities defined by the representation matrix. Given a finite lattice $L = \times_{i=1}^n L_i$ of $\mathbb{R}^n$ with $|L_i| = l_i$, the dimension of the vector space generated by associating a dimension to each node of L is $d = \prod_{i=1}^n l_i$. To compute the number p of inequalities, first recall Theorem 3, which states that all of the elementary transformations $t \in \mathcal{T}$ are extreme, so it is impossible to reduce the number of inequalities required to check supermodularity by removing redundant elementary transformations. Therefore, p equals the number of elementary transformations on L, which it is straightforward to calculate:

$$p = \sum_{1 \le i < j \le n} (l_i - 1)(l_j - 1)\Pi_{k \notin \{i,j\}} l_k.$$

Suppose, for example, that l_i is exactly l for each of the n dimensions. Then $p = \frac{n(n-1)}{2}(l-1)^2 l^{n-2} \sim \frac{n(n-1)}{2} l^n$ and $d = l^n$. Therefore, the complexity of the double description method is $O(\exp(l^n(n \log l + 2 \log n)))$. In practice, therefore, the stochastic supermodular ordering can only be computed via this method for "small-size" problems. However, the "size" of a problem can be reduced by aggregating data into coarser categories. As Theorem 2 showed, aggregation of data preserves the supermodular ordering. Therefore, despite its potential complexity, the double description method can in practice easily be used in conjunction with data coarsening to achieve a tractable comparison of distributions.

7 Symmetric Supermodular Ordering

In some applications, it is interesting to focus on objective functions that are symmetric. The resulting order is called the symmetric supermodular ordering, denoted $SSPM$.

When the lattice is symmetric, i,e. each variable has the same number of points in its support, the symmetric supermodular order is equivalent to imposing the supermodular order on symmetrized distributions. The symmetrized distribution of a distribution f is the one that obtains if, with equal probability, $1/n!$,for each permutation of the variables, one applies f to the permuted vector of variables.

Precisely, we have the following result. Suppose that L is a symmetric lattice, and let f and g denote two distributions. Say that $f \prec_{SSPM} g$ if and only if $E[w|f] \leq E[w|g]$ for all w that are symmetric and supermodular, where symmetric means that $w(x) = w(\sigma(x))$ for all vector x of the lattice and permutation $\sigma(x)$ thereof. The symmetrized distribution of f, f^s, is defined as follows: for any x,

$$f^s(x) = \frac{1}{n!} \sum_{\sigma \in \Sigma(n)} f(\sigma(x)),$$

where $\Sigma(n)$ is the set of all permutations of $\{1, \ldots, n\}$.

THEOREM 5 *The three following statements are equivalent:*

- $f \prec_{SSPM} g$
- $f^s \prec_{S} g^s$
- $f^s \prec_{SSPM} g^s$

Proof. To Do.

In words, Theorem 5 states that one can characterize the symmetric supermodular order in terms of the supermodular order applied to symmetric distributions. Furthermore, when considering symmetric distributions, the supermodular order is equivalent to the symmetric supermodular one Theorem 5 can be used to simplify the analysis of the symmetric supermodular ordering by focusing on symmetric distributions.

This theorem is also important with respect to some economic applications of the theory, particularly with respect to welfare analysis. Indeed, focusing on symmetric objective functions amounts to assuming a form of ex post anonymity across individuals: one does not care whether 1 got the high prize and 2 the low prize, or vice versa. However, it does not impose anything on the ex ante fairness of mechanisms. For example: a mechanism that randomizes with equal probability between giving the high prize to 1 and low prize to 2 or vice versa yields the same expectation as a mechanism that always gives 1 a high prize and 2 a low one. A justification for the focus on symmetric distributions is precisely that we can make any mechanism fair ex ante by randomizing equally across all possible player permutations before applying the initial mechanism. In that sense, symmetry provides both an ex post anonymous and ex ante anonymous mechanism.

In the analysis to follow, we will mostly focus on the symmetric supermodular ordering, keeping in mind the interpretation in terms of symmetrized distributions provided by Theorem 5.

We start with two simple cases, which illustrate how symmetry combined with supermodularity relates to convexity.

7.1 Binary variables, n dimensions

Consider the hypercube $L = \{0,1\}^n$. With symmetric objective functions, only the number of 1's, $c(x) = \sum_{i=1}^n I_{\{x_i=1\}}$, contained in any x matter for the objectives. Thus, an equivalent representation of $L$is $\tilde{L} = \{0,1,\ldots,n\}$. To any distribution f on L we can associate a distribution $\tilde{f}$ on $\tilde{L}$ defined by $\tilde{f}(k) = \sum_{x:c(x)=k} f(x)$ for each $k \in \tilde{L}$.

Similarly, to any symmetric function $w : L \to \mathbb{R}$, corresponds another function $\tilde{w} : \tilde{L} \to \mathbb{R}$ such that $w(x) = \tilde{w}(c(x))$.

Moreover, w is supermodular over L if and only if $\tilde{w}$ is convex on $\tilde{L}$. Indeed, supermodularity of w is equivalent to $w(x) + w(x + e_i + e_j) \geq w(x + e_i) + w(x + e_j)$, for all nodes x with zero i^{th} and j^{th} components. Symmetry of w then allows us to write the inequality as $\tilde{w}(k) + \tilde{w}(k+2) \geq 2\tilde{w}(k+1)$, where $k = c(x)$. Since this is true everywhere, this shows convexity of $\tilde{w}$. Proceeding in the reverse direction shows equivalence of the results. Another simple way to see the result, which will be applied in the section to follow, is to use duality: recall from Theorem 1 supermodular functions are characterized by the dual cone of elementary transformations, $t^x_{i,j}$. Each transformation forms an elementary square in L with ones at the bottom left and top right, and minus ones along the off diagonal. When such transformation is projected on $\tilde{L}$, the result is an elementary transformation of the form $(0, \ldots, 0, 1, -2, 1, 0, \ldots, 0)$, where the first nonzero entry corresponds to the number of 1's in x. Such function is an elementary transformation characterizing convexity. Since their dual cones are equivalent, symmetric supermodular functions on L are equivalent to convex functions on $\tilde{L}$.

This shows a key relation between supermodularity and convexity:

PROPOSITION 1 *On $L = \{0,1\}^n$, $f \prec_{SSPM} g$ if and only if $\tilde{g}$ dominates $\tilde{f}$ according to the convex ordering on $\tilde{L}$.*

[DO WE NEED TO ASSUME IDENTICAL MARGINALS, as in the SLIDES? I Can't remember the argument for why we do...There seem to be no reason that f and g have identical marginals if they are ranked according to the symmetric supermodular ordering. Only their symmetrized versions do. But does that prevent us from comparing them and/or apply the above argument?]

7.2 Three-point support

Now consider the case $L = \{0,1,2\}^n$. In that case, symmetries implies that only the numbers of zeros, ones, and twos matter for the objective function. Accordingly, and similarly to the case of a two-point support, we will represent outcomes on a simpler domain. Let $\tilde{L}$ denote the two- dimensional integer simplex that counts the number of twos along the horizontal axis and zeros along the vertical axis. Since L has n dimensions, $\tilde{L}$ is limited by the equation $h+b \leq n$, where $h = H(x)$ is the number of twos and $b = B(x)$ is the number of zeros.

Given $w : L \to \mathbb{R}$ symmetric, define $\tilde{w} : \tilde{L} \to \mathbb{R}$ by $w(x) = \tilde{w}(H(x), B(x))$. The distribution $\tilde{f}$ on $\tilde{L}$ formed from some distribution f on L is defined as in the case of two-point support. In contrast to the case of two-point supports, convexity of $\tilde{w}$ does not characterize supermodularity of w. Instead, w is supermodular if and only if $\tilde{w}$ is supermodular and componentwise-convex on $\tilde{L}$. Since $\tilde{L}$ is not a lattice, we should clarify what is meant by supermodularity of $\tilde{w}$: whenever k and k' belong to $\tilde{L}$ and are such that $k \wedge k'$ and $k \vee k'$ also belong to $\tilde{L}$, where the meet and join operate on $\mathbb{R}^2$, the supermodularity relation must hold, i.e. $\tilde{w}(k \wedge k') + \tilde{w}(k \vee k') \geq \tilde{w}(k) + \tilde{w}(k')$. To show that symmetry and supermodularity on L is equivalent to supermodularity and componentwise convexity on $\tilde{L}$, we use the dual approach, by showing that any "supermodular" elementary transformation $t^x_{i,j}$ of L maps into either a supermodular elementary transformation of $\tilde{L}$, or into an elementary transformation characterizing convexity. Thus consider any $t^x_{i,j}$. By construction, $x_i \in \{0,1\}$ and $x_j \in \{0,1\}$. First suppose that $x_i = x_j$. Then the points $x + e_i$ and $x + e_j$ of L map into the same point in $\tilde{L}$. Therefore, $t^x_{i,j}$ maps to a function containing the sequence $(1, -2, 1)$ along either the horizontal or the vertical axis in $\tilde{L}$, and zeros everywhere else. This type of elementary transformation characterizes componentwise convexity, and it is easy of all them are achieve by the projection. Now suppose that $x_i \neq x_j$, and without loss of generality, that $x_i = 0$ and $x_j = 1$. In that case the nonzero entries of $t^x_{i,j}$ maps into four distinct points of $\tilde{L}$. Indeed, $x + e_i$ and $x + e_j$ do not have the same numbers of zeros, since $B(x + e_j) = B(x)$ but $B(x + e_i) = B(x) + 1$. Moreover, these four distinct points of $\tilde{L}$ for an elementary square of the plane with ones along the main diagonal and zeros along the off diagonal. This shows the following result.

Proposition 2 *On $L = \{0, 1, 2, \}^n$, $f \prec_{SSPM} g$ if and only if $\tilde{g}$ dominates $\tilde{f}$ according to the supermodular and componentwise-convex ordering on $\tilde{L}$.*

[DO WE NEED TO ASSUME IDENTICAL MARGINALS??]

For $L = \{0, 1, \ldots, l-1\}^n$ and for f and g with identical marginals, the symmetric supermodular ordering is equivalent to the supermodular and componentwise-convex ordering of appropriately derived distributions f^{l-1} and g^{l-1} on an $(l-1)$-dimensional support.

7.3 Characterization of the Symmetric Supermodular Ordering: Three Dimensions, Three-Point Supports

We now provide an explicit characterization of the symmetric supermodular ordering for the case of three dimensions with three points each in their support. From Theorem 5, we may focus on symmetric distribution, since otherwise the comparison can be made on symmetrized distributions. Thus, consider two symmetric distributions, f and g, and let

$\delta = g - f$. The values taken by δ are represented on Figure XXXXX. Given the symmetry, f and g have identical marginals (see Section 9). This implies the following equalities:

i) $a + 2b + c + 2d + 2e + f = 0$

ii) $b + 2c + g + 2e + 2h + i = 0$

iii) $d + 2e + h + 2f + 2i + j = 0$

Our goal is to determine whether δ may be represented as a positive sum of transformations of the form described in 7.2, i.e. transformations which are symmetric across dimensions. We also construct this sequence simultaneously from the top $(2, 2, 2)$ and bottom $(0, 0, 0)$, since the situation (though not the actual numbers) is similar viewed from the top and from the bottom. We proceed in several steps, representing progression from the top and bottom of L towards its center.

First step Assign to the 3 ET's of the type (a, b, b, c) a value $a/3$ each. Similarly, assign to the 3 ET's of type (j, i, i, h) a weight $j/3$ each. This guarantees that the sum of these weighted ET's will match δ at the top and bottom of L.

For some other nodes, there are several ways, even after imposing symmetry on elementary transformations, of matching δ. Therefore, the introduce a couple of unknowns to match some of these nodes, which will be determined at ulterior steps so as to match the entire δ.

Second step

- 2 ET's on (b, c, d, e): weight λ_b each. [WHAT IS THE NUMBER OF ET's IN FRONT? DOESN'T SEEM TO CORRESPOND TO ACTUAL NUMBER OF ETs!!!]
- 1 ET on (b, c, c, g): weight $b + 2a/3 - 2\lambda_b$ each. [SAME]
- 2 ET's on (i, h, f, e): weight λ_i each. [SAME]
- 1 ET on (i, h, h, g): weight $i + 2j/3 - 2\lambda_i$ each. [SAME]

Third Step

- 3 ET's on (d, e, e, h): weight $d + 2\lambda_b$ each. [WHAT IS THE NUMBER OF ET's IN FRONT?]

- 3 ET's on (f, e, e, c): weight $f + 2\lambda_i$ each.

Fourth step 6 ET's on (c, e, g, h): weight of size $a + 2b + c + d + e - (\lambda_b + \lambda_i)$ each.

We seek necessary and sufficient conditions for the existence of (λ_b, λ_i) such that every ET in sequence above has nonnegative weight. This is achieved by setting

- $\lambda_b = \max\{0, -d/2\} \geq 0$
- $\lambda_i = \max\{0, -f/2\} \geq 0$

Then the following 10 conditions are sufficient for every ET to be nonnegative. It is also easy to check that they are necessary: indeed, each condition corresponds to a supermodular function.

1. $a \geq 0$
2. $j \geq 0$
3. $2a + 3b \geq 0$
4. $2a + 3b + 3d \geq 0$
5. $2j + 3i \geq 0$
6. $2j + 3i + 3f \geq 0$
7. $3a + 6b + 3c \geq 0$
8. $3a + 6b + 3c + 3d \geq 0$
9. $3j + 6i + 3h \geq 0$
10. $3j + 6i + 3h + 3e \geq 0$

The last four conditions come from the more general condition

$$2a + 4b + 2c + 2d + 2e + \min\{0, d\} + \min\{0, f\} \geq 0,$$

along with the fact that distributions have identical marginal distributions.

The ten inequalities correspond to the 10 basis functions of the set of symmetric supermodular functions on L, described in Figures XXXXXX.

7.4 Sufficient Conditions

WHAT DOES IT MEAN THAT VECTORS ARE INDEPENDENT? WHAT IS THE RELATION TO INTERDEPENDENCE???

Let A and B denote two $n \times m$ row-stochastic matrices i.e. such that rows have nonnegative components and sum up to 1. Also suppose that for each $j \leq m$, the j^{th} column of A and B have equal sum.

One should think of each row of A as describing the lottery among m prizes to some individual i, for $i \leq n$. The first column corresponds to the highest prize, the second column to the second highest, etc. Moreover, we are interested in the case where these lotteries are independently distributed across individuals. Thus $a_{i,j}$ is the probability that i receives prize j independently of what others receive.

Indeed we will call X and Y the random vector of prizes that individuals receive under distributions defined by A and B respectively.

QUESTION: WHY DO WE NEED TO SYMMETRIZE THE DISTRIBUTION, SINCE THE OBJECTIVE IS ALREADY SYMMETRIC?

Let $\bar{A}$ denote the *cumulative sum matrix* of A, defined by $\bar{a}_{i,j} = \sum_{k=j}^{m} a_{i,k}$, with a similar definition for $\bar{B}$. There is a one to one mapping between row-stochastic matrices and their cumulative-sum equivalent, so slightly abusing notation we will use $\bar{A} \prec_{SSPM} \bar{B}$, $A \prec_{SSPM} B$, and $X \prec_{SSPM} Y$ equivalently.

Say that A is *stochastically ordered* if for each k, $\bar{A}_{i,k}$ is weakly increasing in i. Intuitively, this means that under A, high-index individuals are more likely to receive low prizes.

Finally, say that A dominates B according to the *cumulative column criterion*, denoted CCM, if for all k, the k^{th} column vector of $\bar{A}$, seen from the bottom up, majorizes the k^{th} column vector of $\bar{B}$, that is, for each k, and i

$$\sum_{l=1}^{i} \bar{A}_{l,k} \geq \sum_{l=1}^{i} \bar{B}_{l,k}.$$

THEOREM 6 *Suppose that A is stochastically ordered and that $A \succ_{CCM} B$. Then $Y \succ_{SSPM} X$.*

In words, A implies a more unequal prize distribution across individuals than B. Notice that only A is required to be stochastically ordered. This result will prove important for the corollary to follow.

The proof of Theorem 6 is based on the two following lemmas.

LEMMA 1 *Suppose that $X \prec_{SSPM} Y$ are two dimensional and that Z is a p-dimensional (p arbitrary) random vector independent of X and Y. Then $(X, Z) \prec_{SSPM} (X, Y)$.*

Proof. We need to check that $Ew(X, Z) \leq Ew(Y, Z)$ for all w symmetric and supermodular. For each z in $\mathbb{R}^p$, let $f(z) = Ew(X, z)$ and $g(z) = Ew(Y, z)$. For each z, the function $w(\cdot, Z)$ is symmetric and supermodular in its two arguments, and so $X \prec_{SSPM} Y$ implies that $f(z) \leq g(z)$ for all z. Taking expectations with respect to Z (and using independence of Z) then shows the result.

Let X and Y be two-dimensional random vectors generated by $2 \times m$-matrices A and B, respectively. Suppose that

$$B = A + \sum_{k=2}^{m} \varepsilon_k E_k,$$

where $\varepsilon_k \geq 0$ and E_k is the matrix with zeros everywhere except for columns $k-1$ and k, where it is defined by

$$(E_k)_{1,k-1} = (E_k)_{2k} = -1$$

and

$$(E_k)_{1,k} = (E_k)_{2,k-1} = 1$$

Intuitively, B is putting, for each pair of consecutive prizes, less probability on the second individual (row) getting the lower of the two prizes and more weight on him getting the better one. Given this, one would expect that B is more equal than A if A was treating individual one (first row) better than the second one. This intuition is captured by the following lemma.

LEMMA 2 *Suppose that for each $k \in \{2, \dots m\}$,*

$$\sum_{k=j}^{m} \alpha_{2j} \geq \sum_{k=j}^{m} \alpha_{1j} + \varepsilon_k.$$

Then,

$$X \prec_{SSPM} Y.$$

The condition in Lemma 2 implies that the one-dimensional distribution generated by the second row of A assigns lower prizes (in the fist-order stochastic dominance sense) than the one generated by the first row of A, and is strictly stronger than that, since the FOSD inequalities must hold by more than ε_k for each k.

Proof. With two dimensions the symmetric supermodular ordering is characterized by the following symmetric supermodular functions:

$$w^k(X) = 1_{X_1 \geq c_k, X_2 \geq c_k}$$

for each $k \geq 2$ and, for $k \neq l$ greater than 2,

$$w^{kl}(X) = 1_{X_1 \geq c_k, X_2 \geq c_l} + 1_{X_1 \geq c_l, X_2 \geq c_k},$$

where $c_1 < c_2 < \cdots < c_m$ is an arbitrary vector of indices decreasing with prize values (so that the first prize has the lowest index, etc.). The reason why indices are greater than 2 is that for $k = 1$ the indicator-based conditions above are always satisfied, since all prizes have indices above c_1. For each k, $Ew^k(X) \leq Ew^k(Y)$ is equivalent to

$$0 \leq \left(\sum_{j=k}^{m} \beta_{1j}\right)\left(\sum_{j=k}^{m} \beta_{2j}\right) - \left(\sum_{j=k}^{m} \alpha_{1j}\right)\left(\sum_{j=k}^{m} \alpha_{2j}\right),$$

where α's and β's are the entries of matrices A and B, respectively. Since all ε_j's simplify in the above β sums except for ε_k, this condition becomes, after simplification,

$$\varepsilon_k \left[\sum_{j=k}^{m} \alpha_{2j} - \left(\sum_{j=k}^{m} \alpha_{2j} + \varepsilon_k\right)\right],$$

which is nonnegative by assumption. For each $k \neq l$ greater than 2, $E^{kl}(X) \leq Ew^{kl}(Y)$ is equivalent to, using the more compact notation of cumulative matrices $\bar{A}$ and $\bar{B}$ with entries $\bar{\alpha}$ and $\bar{\beta}$,

$$0 \leq \bar{\beta}_{1k}\bar{\beta}_{2l} - \bar{\alpha}_{1k}\bar{\alpha}_{2l} + \bar{\beta}_{1l}\bar{\beta}_{2k} - \bar{\alpha}_{1l}\bar{\alpha}_{2k}.$$

Since by construction $\bar{\beta}_{1k} = \bar{\alpha}_{1k} + \varepsilon_k$ and $\bar{\beta}_{2k} = \bar{\alpha}_{2k} - \varepsilon_k$ for all $k \geq 2$, the condition simplifies to

$$0 \leq \varepsilon_k \left[\bar{\alpha}_{2l} - (\bar{\alpha}_{1l} + \varepsilon_l)\right] + \varepsilon_l \left[\bar{\alpha}_{2k} - (\bar{\alpha}_{1k} + \varepsilon_k)\right],$$

both terms of which are nonnegative by assumption.

We can now conclude the proof of Theorem 6. We first show that $\bar{A} \prec_{SSPM} \bar{B}^{so}$ and then than $\bar{B}^{so} \prec_{SSPM} \bar{B}$, where $\bar{B}^{so}$ is the matrix obtained from $\bar{B}$ by reordering each of its column from the smallest to the greatest element. This will then prove the result, by transitivity. Notice that $\bar{B}^{so}$ is essentially a stochastic reordering of the matrix B so as to systematically put more probability of lower prizes to high index individuals. With this interpretation, it is not surprising that $\bar{B}^{so} \prec_{SSPM} \bar{B}$. Since A is already assumed to be stochastically ordered the comparison assumed on A and B carries over to a comparison between A and B^{so}, an so it is not surprising either that $A \prec_{SSPM} B^{so}$.

Proof that $\bar{A} \prec_{SSPM} \bar{B}^{so}$. We use the following algorithm: We start by transforming the last column of $\bar{A}$ into the last column of $\bar{B}$ by applying to $\bar{A}$ a sequence of elementary transformations $\varepsilon_m E_m$ of the type described in Lemma 2, only involving the last column of $\bar{A}$ and only one pair of rows at each time, and such that, after each step, the resulting matrix is still stochastically ordered.[2] Such construction is given by Hardy et al. (XXXX). At each step, the last column of the resulting matrix is stochastically ordered, and remaining columns are untouched, so Lemma 2 can be applied. Lemma 2 combined with Lemma 1 ensures that at each step the new matrix SSPM dominates the previous and, by transitivity, $\bar{A}$. Once the last column of $\bar{A}$ has been transformed into that $\bar{B}^{so}$, one proceed to do the same for the second to last column of $\bar{A}$, etc. Once the second column has been transformed, the resulting matrix is $\bar{B}^{so}$ itself, which shows by transitivity, that $\bar{A} \prec_{SSPM} \bar{B}^{so}$.

Proof that $\bar{B}^{so} \prec_{SSPM} \bar{B}$. Columns of $\bar{B}^{so}$ and $\bar{B}$ have the same entries, only in a different order, since $\bar{B}^{so}$'s entries are increasing with the row index, for fixed columns. Without loss of generality, reset the entries in each column of $\bar{B}^{so}$ as $1, 2, \ldots, n$, with the same correspondence for $\bar{B}$. The goal is to find an algorithm that rearranges these entries to match $\bar{B}$'s. Resetting entries is for convenience only in order to emphasize the workings of the algorithm. In practice, the elementary transformations used will match actual entries of $\bar{B}^{so}$. Starting from the last row, n, of $\bar{B}^{so}$, whose entries are equal to n after relabeling, we will move these 'n'-labeled entries upwards, gradually, so as to position them as in $\bar{B}$. We will do this by a sequence of entry permutations between rows n and i for i starting from $n-1$ until i reaches 1. We will do this so that, at each step i, the rows above n remain stochastically ordered, and the n^{th} row remains stochastically higher than rows above i. This guarantees that applications of Lemma 2, at each step, is valid and so that the transformed matrix always SSPM dominates the previous one and, by transitivity, $\bar{B}^{so}$. Thus, starting with rows n and $n-1$, flip entries of $\bar{B}^{so}$ for each column

[2] In terms of A, these transformations involve only the last two columns of A. Note that E_m's have impact on cumulative sums for $k < m$ so they only affect $\bar{A}$ through its last column. For convenience, we state the result in terms of the cumulative matrix $\bar{A}$.

j in which $\bar{B}_{nj} \neq n$. The result is that some entries of in the last row of $\bar{B}^{so}$ are now equal to $n-1$, while entries in its $(n-1)$ row are equal to n, for exactly those columns where $\bar{B}_{nj} \neq n$. The result is that now the n and $n-1$ rows of $\bar{B}^{so}$ are no more stochastically ordered, but both rows still dominate all rows with indices less than $n-2$. The next step is to flip entries between the n and $n-2$ rows of the resulting matrix, for columns where its n^{th}-row entry does not match that of $\bar{B}$. As a result, the n^{th} row now contains (possibly) entries labeled '$n-2$' while the $n-2$ two contains $n-1$ entries. Notice that, i) the n, $n-1$, and $n-2$ rows still dominate all rows with indices less than $n-3$, and ii) the $n-1$ row dominates the $n-2$ row. The reason for the last point is that the $n-2$ row inherited an $n-1$ only if the $n-1$ row inherited an n entry. Proceeding systematically by decreasing the row index each time, the result is that the n^{th} row now has the same entries as $\bar{B}$'s, and that the first $n-1$ rows of the resulting matrix are still stochastically ordered. Applying next to the $n-1$ row what was done to the n row, we can transform it into the $n-1$ row of $\bar{B}$ while preserving at each step the stochastic ordering of the first $n-2$ rows and guaranteeing that the $n-1$ row dominates rows with which it has not yet been flipped. Applying this larger algorithmic loop to each row, in decreasing index order, one eventually transform $\bar{B}^{so}$ into $\bar{B}$ through a sequence of steps that increase in the SSPM sense, which proves the result.

We have examples showing that Theorem 6 holds no longer if one relaxes either assumption that A is stochastically ordered or that $A \succ_{CCM} B$. Theorem 6 has the following corollary, which generalizes the notion of tournament and shows that this generalized tournament does worse than any other scheme in terms of the symmetric supermodular order. Equivalently, this sort of tournament is the optimal mechanism for all objective functions that are symmetric and submodular.

Corollary 1 *For any n and m-dimensional probability vector p, there exists a unique $n \times m$ row-stochastic matrix A whose j^{th} column, for each j, sums up to np_j and such that, $A \prec_{SSPM} B$ for all $n \times m$ row-stochastic matrix B with the same column sums as A.*

The matrix A generating (among all row-stochastic matrices with matching column sums) the worst distribution with respect to SSPM dominance is constructed as follows. For any real number x, let $\lfloor x \rfloor$ denote the largest integer below x. Set $a_{i,1} = 1$ for all $i \leq i_1 = \lfloor np_1 \rfloor$, $a_{i_1+1,1} = np_1 - i_1$, and $a_{i,1} = 0$ for all $i > \lfloor i_1 + 1$. In effect, we have maximized the entries of the low-index rows of the first column, subject to A's row-stochasticity constraint and to the sum of entries in the first column being equal to np_1. Put differently, the first column of A, seen from top down, majorizes all vectors with entries less than one and summing to np_1. Remaining vectors are defined similarly: the second column vector is the vector that majorizes all vectors that respect A's row-stochasticity and the summing up to p_2. Precisely, set $a_{i,2} = 0$ for all $i \leq i_1$ since these rows already have ones in the first column, $a_{i_1+1,2} = \min\{1 - a_{i_1+1,1}, np_2\}$. The first argument of the minimizer expresses that the row sum cannot exceed one, and the second argument that entries in the second column cannot exceed np_2. Finally, set $a_{i,2} = 1$ for all i between $i_1 + 1$ and $i_2 = \lfloor np_2 - a_{i_1+1,2} \rfloor$, and $a_{i_2+1,2} = np_2 - a_{i_1+1,2} - i_2$. Thus, after completing the $i_1 + 1$ row with whatever probability remains after setting $a_{i_1+1,1}$, one sets entries below equal to 1 subject to the column sum being less than np_2, and put whatever fraction remains in the next entry below. Remaining columns are constructed similarly.

By construction, A is stochastically ordered, as is easily checked. Moreover, given any row-stochastic matrix B with the column sums as A, it is intuitive and easy to check that $A \succ_{CCM} B$ since A puts as much weight as possible in the first columns of the first rows and, equivalently, in the last columns of the last row. Precisely, for any column k and row l, the sum of entries in A over all columns with index above k and rows with index above l is maximal, subject to row-stochasticity and column-sum constraints.

Given any $n \times m$ row-stochastic matrix B with column sums $(np_1, \ldots, np_m)$, let Y_i have distribution described by row i of B and let $(Y_1, \ldots, Y_n)$ be independent. From this independent asymmetric distribution, create the symmetric (exchangeable), non-independent distribution, of random variables $(Y_1^s, \ldots, Y_n^s)$, such that for any symmetric supermodular W, $EW(Y_1, \ldots, Y_n) = EW(Y_1^s, \ldots, Y_n^s)$. Then the exchangeable random variables $(Y_1^s, \ldots, Y_n^s)$ each have a marginal distribution given by $(p_1, \ldots, p_m)$, and we will refer to the distribution of $(Y_1^s, \ldots, Y_n^s)$ as the exchangeable distribution generated by the matrix B.

The corollary to Proposition 1 tells us that, for any n, m, and $(p_1, \ldots p_m)$, the exchangeable distribution generated by the matrix $\underline{A}$ is SPM-dominated by the exchangeable distribution generated by any other row-stochastic matrix B with matching column sums.

8 Other Multidimensional Orderings of Interdependence

This section considers other interdependence orderings and relate them to supermodular ordering. We start with defining the orders, and then establish their relations.

In what follows, we distinguish between interdependence concepts and interdependence orderings. For example, the well known notions of affiliation and association are concepts defining a property that any given distribution may have. By contrast, orderings are relations between pairs of distributions. To each concept corresponds an ordering, as will be discussed shortly.

Definition 1 (Weak Association) *An n-dimensional random vector X is weakly associated if for any pair (A, B) of disjoint subsets of $\{1, \ldots, n\}$ and nondecreasing functions r and s of $\mathbb{R}^{|A|}$ and $\mathbb{R}^{|B|}$, respectively,*

$$Cov(r(X_A), s(X_B)) \geq 0.$$

Similarly, X is negatively associated if for all A, B, r, and s,

$$Cov(r(X_A), s(X_B)) \leq 0.$$

The qualification "weak" contrasts with the concept of "association," where the functions r and s can both depend on the entire vector X, rather than on disjoint components thereof.

Definition 2 (Association) *X is associated if for all nondecreasing functions r and s defined on $\mathbb{R}^n$*

$$Cov(r(X), s(X)) \geq 0.$$

As the name suggests, association is a stronger concept than weak association. Negative association is the opposite of *weak* association. The reason is that the opposite of (strong) association defines a trivial order: X can only be negatively associated in a strong sense if it is constant (to see this, consider functions $r = s$). Since this concept is uninteresting, the term "negative association" is thus unambiguously reserved for the opposite of weak association.

The ordering corresponding to weak association is defined as follows.

Definition 3 (Greater Weak Association) *X displays greater weak association than Y if they have identical marginals and for all disjoint subsets A, B of $\{1, \ldots, n\}$ and increasing functions r, s of $\mathbb{R}^{|A|}$ and $\mathbb{R}^{|B|}$ respectively,*

$$Cov(r(X_A), s(X_B)) \geq Cov(r(Y_A), s(Y_B)).$$

In the insurance literature, this ordering has been named the "correlation order" - see Denuit et al. (XXX) and Lu and Zhang (2004). Greater association is defined similarly.

[QN: DO WE NEED TO ASSUME IDENTICAL MARGINALS IN THE DEFINITION OF GREATER WEAK ASSOCIATION? If not, one should get rid of it in the definition and state it as a property.]

In particular, X is weakly associated if and only if it displays greater weak association than the vector Y that has equal marginals and independently distributed components.

Another well known concept of interdependence is that of affiliation. Affiliation is essentially the requirement of association conditional on variables being on any given lattice.

Definition 4 (Affiliation) *X is affiliated if for any sublattice A of $\mathbb{R}^n$ and all nondecreasing functions r and s defined on $\mathbb{R}^n$*

$$Cov[r(X), s(X)|X \in A] \geq 0.$$

IS THE DEFINITION CORRECT? REFERENCES: Paul Milgrom's paper, Meg's paper, etc. Affiliation is stronger than association. The concept also yields an order: greater affiliation.

Definition 5 (Greater Affiliation) *X displays greater affiliation than Y if for any sublattice A of $\mathbb{R}^n$ and all nondecreasing functions r and s defined on $\mathbb{R}^n$*

$$Cov[r(X), s(X)|X \in A] \geq Cov[r(Y), s(Y)|X \in A].$$

IS THAT DEFINITION INTERESTING?

In this paper, we are interested in orderings of distributions, rather than properties of a given distribution. Greater weak association plays an important role in comparative statics. Another concept, closer to the supermodular ordering, is the convex-modular ordering. It is well know that $\phi(x+y)$ is supermodular in x, y whenever ϕ is convex, and in fact the same is true for $\phi(r^1(x_1)+\ldots+r^n(x_n))$ whenever ϕ is convex and r^i's are increasing. A function of the above form is called "convex modular". A natural question is whether the cone generated by convex modular functions coincides with supermodular functions. Equivalently, can all supermodular functions be represented as positive sums of convex modular functions? Such functions are used in the insurance literature and called the stop loss order. This concept captures the payoff profile of a reinsurer that must cover all losses above a certain threshold, over a given pool of risky assets. In that case, the function $\phi(x) = \max\{a-x, 0\}$, where x is the threshold, is convex and described the reinsurer's liability. OTHER MOTIVATION FOR CONVEX MODULAR ORDERING? RELATION TO STOP LOSS ORDER.

DEFINITION 6 (CONVEX-MODULAR ORDERING) *X dominates Y according to the convex-modular ordering if and only if $E[w(X)] \geq E[w(Y)]$ for all convex-modular functions w.*

That definition is equivalent to the requirement that $E[w(X)] \geq E[w(Y)]$ for all w's that are positive sums of convex-modular functions.

Another order frequently mentioned is the concordance ordering. That order is a natural generalization of first order stochastic dominance to multivariate settings. Like first-order stochastic dominance, it ranks probability distributions by comparing their cumulative distribution functions.

DEFINITION 7 *X dominates Y according to the concordance ordering if and only if for all vectors a of $\mathbb{R}^n$,*

$$Pr(X \geq a) \geq Pr(Y \geq a)$$

and

$$Pr(X \leq a) \geq Pr(X \leq a).$$

Intuitively, components of X are more likely to be either all high or all low, relative to Y's.

DOES CONCORDANCE IMPLY IDENTICAL MARGINALS?

How are the orders described above related?

THEOREM 7 *[Orderings: Two Dimensions] When $n = 2$, affiliation is strictly stronger than association, which is strictly stronger than weak association. The following orders* *are equivalent: weak association, supermodular ordering, convex-modular ordering, concordance ordering.*

The equivalence between weak association, the supermodular ordering and the concordance ordering is well known. See Meyer (1990) and Muller and Stoyan (2002). The equivalence between the convex-modular ordering and other concepts is shown in Meyer (1990). That association is strictly stronger than weak association, even for two dimensions, has been shown by Hu, Muller and Scarsini (XXXX). Appendix B shows an (THEIR?) example for completeness.

THEOREM 8 *[Orderings: Three or More Dimensions] For $n \geq 3$, weak association is strictly stronger than the supermodular ordering, which is strictly stronger than the convex-modular ordering, which is strictly stronger than the concordance ordering.*

A consequence of Theorem 8 is that supermodular functions cannot be generated by convex modular functions. *Proof.* To prove that greater weak association implies the supermodular ordering, one adapts the proof of Cristofides and Vaggelatou's [SPELLING?] to show that weak association implies positive supermodular dependence (that is, the relation between orderings is proved similarly to the way the relation between concepts is shown). That proof appears in Rüschendorf (2004). [SHOULDN'T WE PUT THE UMLAUT ON MULLER's U AS WELL?] The remaining nontrivial part is to show that the convex-modular ordering implies the concordance ordering. Setting $\phi(z) = 0$ for $z \leq n-1$ and $\phi(z) = z - (n-1)$ for $z \geq n-1$, and letting $r^i(x_i) = 1_{x_i \geq a_i}$ for all i shows that, if X dominates Y according to the convex-modular ordering, then $Pr(X \geq a) \geq Pr(Y \geq a)$. That $Pr(X \leq a) \geq Pr(Y \leq a)$ is shown similarly. To show that greater weak association is not equivalent to the supermodular ordering, we will provide a separate example showing that greater weak association is not a "linear ordering," cannot be characterized by duality. This is a key difference, which is discussed in detail in Section 9. To show that the supermodular ordering is strictly stronger than the convex-modular ordering we construct an extreme supermodular function that cannot be replicated by any convex-modular one (see Appendix B) SHOW CONVEX MODULAR STRICTLY STRONGER CONCORDANCE??.

ALSO CITE LUCIANO DI CASTRO, AND SEND OUR PAPER TO HIM WHEN WE HAVE A DRAFT.

9 Characterization of Orderings

This section provides characterizations of orderings other than the supermodular ordering, which we call "linear orderings," as they have a particular linear structure, which allows us, for example, the use of duality theorems. We use the general duality approach to characterize orders combining supermodularity and componentwise convexity, or full convexity. Since convexity on lattices is a nontrivial concept, we also show how to characterize it in terms of elementary transformations, which is an interesting result in itself.

Recall from (1) that any class $\mathcal{C}$ of functions on L define an order by $f \prec_{\mathcal{C}} g \Leftrightarrow E[w|f] \leq E[w|g] \quad \forall w \in \mathcal{C}$. To clarify the analysis to follow, we formally state the intuitive fact that larger classes of distributions make it harder to compare functions, hence result in a coarser order.

THEOREM 9 (ORDER MONOTONICITY) *If* $\mathcal{C} \subset \mathcal{D}$ *and* $f \prec_{\mathcal{D}} g$*, then* $f \prec_{\mathcal{C}} g$

Proof. Trivial and omitted.

Theorem 9 implies that any property of the order generated from a class of objective functions must be inherited by the order generated from any larger class of objective functions. This property is illustrated in the next two results, which imply that if g dominates f according to the stochastic supermodular ordering, then i) g and f must has the same marginal distributions, and ii) $Cov(Y_i, Y_j) \geq Cov(X_i, X_j)$ for any $i \neq j$ and random vectors X and Y respectively distributed according to f and g.

ONE KEY ADVANTAGE OF LINEAR ORDERINGS IS THAT IT IS CLEAR HOW TO INCREASE INTERDEPENDENCE ACCORDING TO THEM, in contrast to ASSOCIATION and AFFILIATION...

The Modular Ordering A function w on L is *modular* if there exist one dimensional functions w_i, $i \in N$ such that $w(x) = \sum_i w_i(x_i)$. Modular functions form a linear subspace included in the cone of supermodular functions. Let $\mathcal{M}$ denote the set of modular functions on L. The first result states that the order generated by M consists of equivalence classes of distributions, where each equivalence class if characterized by marginal distributions for each component $i \in N$.

THEOREM 10 (MODULAR ORDERING) $f \prec_{\mathcal{M}} g$ *if and only if* f *and* g *have equal marginal distributions. In particular* $f \prec_{\mathcal{M}} g \Leftrightarrow g \prec_{\mathcal{M}} f$.

Proof. Fix $i \in N$. Consider modular functions w such that $w_j = 0$ for $j \neq i$. Since $-w$ is also modular, $f \prec_{\mathcal{M}} g$ implies that $(g_i - f_i) \cdot w_i = 0$ for all functions w_i on L_i, where g_i and f_i are the marginal distributions of g and f respectively, hence that these marginal distributions are equal. ■

The affine ordering, where functions are restricted to be of the form $w(x) = w_0 + \sum_i w_i x_i$ for some real coefficients w_i, is strictly finer than the modular ordering. Affine functions are modular, but do not generate the set of modular functions (for an obvious proof of this, note that set of affine functions is closed under convex combinations, so cannot generate non affine functions). For example, if $n = 1$ and $l_1 = 3$, any distribution g obtained from another distribution f by a mean preserving spread is equivalent to f according to the affine ordering. More generally, we have the following result. Let $\mathcal{A}$ denote the class of affine functions.

THEOREM 11 (AFFINE ORDERING) *$f \prec_{\mathcal{A}} g$ if and only if f and g have identical first moments for each dimension $i \in N$. In particular, $f \prec_{\mathcal{A}} g \Leftrightarrow g \prec_{\mathcal{A}} f$.*

Proof. Taking $w(x) = x_i$, $f \prec_{\mathcal{A}} g$ implies that $\sum_{x_i \in L_i} x_i f_i \leq \sum_{x_i \in L_i} x_i g_i$, where f_i is f's marginal distribution along the i^{th} component, whereas taking $w(x) = -x_i$ yields the opposite inequality. Therefore, f_i and g_i have the same mean. Multiplying the previous equality by w_i and summing over i shows that $E[w|f] = E[w|g]$ for all affine functions. ■

Equal means is a weak requirement than equal marginals, which is not surprising in light of the fact that $\mathcal{A} \subset \mathcal{M}$ and of Theorem 9.

The Quadratic Ordering We now consider the subset $\mathcal{Q}$ of supermodular functions that are quadratic, i.e. generated by affine functions and by the pairwise[3] products $x_i x_j$ for $i \neq j$. Such functions are supermodular as is easily checked. Let X and Y denote random vectors distributed according to f and g, respectively.

THEOREM 12 (QUADRATIC ORDERING) *$f \prec_{\mathcal{Q}} g$ if and only if $E[X_i] = E[Y_i]$ for all i and $Cov(X_i, X_j) \leq Cov(Y_i, Y_j)$ for all $i \neq j$.*

Proof. Since $\mathcal{A} \subset \mathcal{Q}$, $f \prec_{\mathcal{Q}} g$ implies that f and g have equal means, from Theorems 11 and 9. It thus suffices to show that $E[X_i X_j] \leq E[Y_i Y_j]$ for $i \neq j$. This is exactly the

[3]We rule out functions x_i^2 in order to get an equivalence in the next theorem. For the entire class of supermodular quadratic functions, necessity of covariance relations is implied by combining Theorems 12 and 9.

inequality obtained by taking $w(x) = x_i x_j$. For the reverse direction, observe that equal means and ordered covariances implies that $E[w|f] \leq E[w|g]$ for all w that is affine or a product $x_i x_j$, $i \neq j$. Since these functions precisely generate $\mathcal{Q}$, this shows that $f \prec_{\mathcal{Q}} g$. ■

Componentwise Convex/Concave Objective Functions In several applications, objective functions may have other properties than supermodularity. For example, if the objective is a welfare function and each variable entering the multivariate distribution represents the random income of an individual, componentwise concavity may express the social planner's preference for reducing risk faced by each individual. We now show how the duality approach in the case of the stochastic supermodular ordering can be extended to such situations. In what follows, we consider the case of objective functions that are supermodular and componentwise convex, but the case of supermodular, componentwise concave objective functions can be obtained similarly (see discussion below Theorem **??**).

In Section 4, we used the fact that supermodular functions are characterized by a list of inequalities which correspond to nonnegativity of their scalar product with all elementary transformations of the type defined earlier. To accommodate the introduction of other types of elementary transformations, let $\mathcal{T}(\mathcal{S})$ denote the set of elementary transformations characterizing $\mathcal{S}$.

A function w is componentwise convex if for any i in N and x, y in L such that $x_j = y_j$ for all $j \neq i$ and any $\lambda \in [0,1]$ such that $\lambda x + (1-\lambda)y$ belongs to L, $w(\lambda x + (1-\lambda)y) \leq \lambda w(x) + (1-\lambda)w(y)$. Let $\mathcal{X}$ denote the set of componentwise convex functions on L.

For any x and i let t_i^x denote the function on L that vanishes everywhere except at three consecutive nodes, such that $t_i^x(x) = t_i^x(x + 2e_i) = 1$ and $t_i^x(x + e_i) = -2$, and let $\mathcal{T}(\mathcal{X})$ denote the set all such functions. As is easily checked, these functions entirely characterize componentwise convex functions, that is:

$$w \in \mathcal{X} \Leftrightarrow w \cdot t \geq 0 \quad \forall t \in \mathcal{T}(\mathcal{X}).$$

Proceeding as in Section 4, we can characterize the set of distributions ordered according to $\mathcal{X}$ as follows.

THEOREM 13 *$f \prec_{\mathcal{X}} g$ if and only if there exist nonnegative coefficients α_t, $t \in \mathcal{T}(\mathcal{X})$, such that*

$$g = f + \sum_{t \in \mathcal{T}(\mathcal{X}))} \alpha_t t.$$

The proof is identical to the proof of Theorem 1 and therefore omitted.

Combined Properties of Objective Functions As mentioned earlier, one may be interested in more restrictive classes of objective functions that supermodular ones. Such restrictions are important as they may refine the resulting order on distributions (from Theorem 9), i.e. allow one to compare distributions that were not comparable under the stochastic supermodular order. The following result, based on duality, provides a general method to characterize the order based on objective functions that combine several properties. Let $\mathcal{C}$ and $\mathcal{D}$ denote two classes of functions that are each stable under positive combinations (i.e. $\mathcal{C}$ and $\mathcal{D}$ are convex cones seen as subsets of $\mathbb{R}^d$. Also let $\mathcal{T}$ and $\mathcal{U}$ denote their respective set of elementary transformations. In this generalized setting, elementary transformations are the extreme rays of the dual cones of $\mathcal{C}$ and $\mathcal{D}$.

THEOREM 14 *$f \prec_{\mathcal{C}\cap\mathcal{D}} g$ if and only if there exist nonnegative coefficients α_t and β_u such that*

$$g = f + \sum_{t\in\mathcal{T}} \alpha_t t + \sum_{u\in\mathcal{U}} \beta_u u.$$

Proof. The dual cone of the intersection of two polyhedral cones is equal to the (Minkowski) sum of the dual cones (see Goldman and Tucker, 1956). Therefore, $f \prec_{\mathcal{C}\cap\mathcal{D}} g$ if and only if $g - f$ belongs to $\mathcal{C}^* + \mathcal{D}^*$, where $\mathcal{C}^*$ and $\mathcal{D}^*$ are respectively the dual (or polar) cones of $\mathcal{C}$ and $\mathcal{D}$. Since these dual cones are the convex hulls of $\mathcal{T}$ and $\mathcal{U}$, the result obtains.

Theorem 14 extends to any set of properties that can be described by polyhedral cones.

COROLLARY 2 *Let $\mathcal{SX}$ denote the set of objective functions that are both supermodular and componentwise convex. Then $f \prec_{\mathcal{SX}} g$ if and only if there exists a sequence of elementary transformations of either type t_i^x or $t_{i,j}^x$ that, added to f, yield g.*

Convexity Discrete convexity is harder to characterize than componentwise convexity. The very notion of convexity in discrete settings has received several definitions, several of which are compared in Murota and Shioura (2001). We focus here on a notion, natural to economists, of convex-extensibility. A function $w : L \to \mathbb{R}$ is *convex extensible* if there exists a convex function $\bar{w} : \mathbb{R}^n \to \mathbb{R}$ such that $w(x) = \bar{w}(x)$ for all $x \in L$. Concavity is defined similarly. This definition is natural in economic settings: it characterizes usual convexity or concavity properties of an objective function defined on all possible outcomes in a situation where only discrete outcomes are available.[4] To apply the duality technique

[4] Although natural in economics, this definition of discrete convexity is criticized by Murota (1998),

used so far in this section, we need to characterize convexity as a set of inequalities, each of which corresponds to an elementary transformation. For example, suppose that $L = L_{3,3}$, i.e. the square lattice of $\mathbb{R}^2$ with 3 points along each dimension. In that case, convexity is clearly a stronger requirement than componentwise convexity: the two diagonals of the square each imply a convexity relation that involves both dimensions. As a first guess, then, could it be that discrete convexity on L is characterized by the componentwise convexity relations plus the inequalities $w(0,0) + w(2,2) \geq 2w(1,1)$ and $w(0,2) + w(2,0) \geq 2w(1,1)$. As it turns out, this requirement is not enough to guarantee convexity. For example consider the function w on L with the following values:

w	$x_1 = 0$	$x_1 = 1$	$x_1 = 2$
$x_2 = 0$	0	1	2
$x_2 = 1$	1	1	1
$x_2 = 2$	2	1	2

Component-wise convexity as well as convexity along the main diagonals is satisfied. However, even though $(1,1)$ is the barycenter of $(0,0)$, $(1,2)$ and $(2,1)$ with equal weights, we have $w(1,1) > (w(0,0)+w(1,2)+w(2,1))/3$, which precludes the existence of a convex function $\bar{w}$ extending w. For real variables, the following relations are equivalent for any convex set $\mathcal{X}$ of $\mathbb{R}^n$ and $w : \mathcal{X} \to \mathbb{R}$.

$$w(\alpha x + (1-\alpha)y) \leq \alpha w(x) + (1-\alpha)w(y) \qquad \forall (x, y, \alpha) \in \mathcal{X}^2 \times [0,1]$$

$$w\left(\sum_{i=1}^{p} \alpha_i x_i\right) \leq \sum_{i=1}^{p} \alpha_i w(x_i) \quad \forall (x, \alpha) \in \mathcal{X}^p \times \Delta_n$$

However, this equivalence fails for discrete variables, as the above example illustrates. In that example, all convexity conditions involving convex combinations of two variables are satisfied, but convexity is violated by a convex combination of three variables. The reason is that the usual induction argument to reduce a p-variable convex relation to a 2-variable one fails, as the intermediate convex combinations it involves typically do not belong to the lattice.

In that case, how to characterize convex-extendibility? Can a function w satisfying all convexity inequalities on a lattice be extended to convex function of continuous variables? The answer is yes, and in fact one only needs to consider convex combinations of at most $(n+1)$ variables, where n is the dimension of the space. The following characterization is new to our knowledge, although a similar statement based on epigraph comparisons for

who shows by means of examples that the notion of differentials and natural extensions of gradient methods do not a slightly different class of functions appears in Kiselman (2005), and a method of proof using LP duality for local convex extensions is given in Murota (2003).

THEOREM 15 (DISCRETE CONVEXITY) *Let L denote any finite lattice of $\mathbb{R}^n$. The following two statements are equivalent.*

- *(i) w is convex extendible.*
- *(ii) For all $(x_0, \ldots, x_n) \in L$ and $\alpha \in \Delta_n$,*

$$w\left(\sum_{i=0}^{n} \alpha_i x_i\right) \leq \sum_{i=0}^{n} \alpha_i w(x_i).$$

Proof. Clearly (i) implies (ii). We now show the reverse. We follow the approach of Murota (2003). For all $x \in \mathbb{R}^n$, Let

$$\bar{w}(x) = \sup_{(p,\gamma)\in\mathbb{R}^n\times\mathbb{R}} \{p \cdot x + \gamma | p \cdot y + \gamma \leq w(y) \quad \forall y \in L\}. \tag{14}$$

By construction, $\bar{w}$ is convex and such that $\bar{w}(x) \leq w(x)$ for all $x \in L$. We will show that $\bar{w}(x) \geq w(x)$ for all $x \in L$, which will conclude the proof. Since L is finite, the number d of constraints defining (14) is finite, and the objective is well defined and finite. By strong LP duality (see e.g. Bertsimas and Tsitsiklis, 1997, Theorem 4.4), this implies that for all $x \in \mathbb{R}^n$,

$$\bar{w}(x) = \inf_{\lambda\in\mathbb{R}^d} \left\{\sum_{y\in L} \lambda_y w(y) | \sum_{y\in L} \lambda_y y = x, \sum_{y\in L} \lambda_y = 1, \lambda_y \geq 0\right\}.$$

Moreover, there exists a basic feasible solution $\lambda^* \in \mathbb{R}^d$ to this dual program, i.e. such that λ^* vanishes except for a set $Y(x)$ of at most $n+1$ components (see Bertsimas and Tsitsiklis, Theorem 2.4). That is,

$$\bar{w}(x) = \sum_{y\in Y(x)} \lambda_y^* w(y).$$

From (ii), this implies that $\bar{w}(x) \geq w(x)$, which concludes the proof.[5] ■

Theorem 15 allows us to characterize the convex order. For each subset $\chi = \{x_0, \ldots, x_n\} \subset L$ of $n+1$ elements and weights $\alpha \in \Delta_{n+1}$ such that $y = \sum_{i=0}^{n} \alpha_i x_i \in L \setminus \chi$, let $t(\chi, \alpha)$

[5]The result can also be proved by adapting the approach of Kiselman (2005), by showing that the epigraph of w in $\mathbb{Z}^n \times \mathbb{R}$ is $\mathbb{Z}^n \times \mathbb{R}$ convex. With this approach, Carathéodory's theorem is used to reduce the number of convex combinations entering the characterization.

denote the function on L such that $t(x_i) = \alpha_i$ for $0 \le i \le n$, $t(y) = -1$, and $t(x) = 0$ for $x \in L \setminus (\chi \cup \{y\})$, and let $\mathcal{T}_x$ denote the set of all such transformations, and let $\mathcal{C}_x$ denote the set of convex-extendible functions on L. Proceeding as for Theorem 1 and using Theorem 15, we get the following result.

Theorem 16 (Convex Ordering) *$f \prec_{\mathcal{C}_x} g$ if and only if*

$$g = f + \sum_{t \in \mathcal{T}^x} \alpha_t t$$

for nonnegative coefficients $\{\alpha_t\}_{t \in \mathcal{T}_x}$.

Theorem 3 has shown that elementary transformations corresponding to the supermodular ordering are non-redundant, i.e. cannot be reduced in one's search of a positive sequence bridging f to g. This result does not hold for the convex order. For example, consider in $L_{3,3}$ the 3-point convex combination where $(0,0)$ and $(2,0)$ receive weight $1/4$ and $(1,2)$ receives weight $1/2$. The resulting barycenter is $(1,1)$. In this case however, the convex combination can be decomposed into two simpler ones, one putting weights $1/2$ on $(1,2)$ and $(1,0)$, and the other putting weights $1/2$ on $(0,0)$ and $(2,0)$. In terms of elementary transformations, we have

$$t(\{(0,0),(2,0),(1,2)\},(.25,.25,.5)) = t(\{(1,2),(1,0)\},(.5,.5)) + \frac{1}{2} t(\{(0,0),(2,0)\},(.5,.5)).$$

Therefore, some "elementary transformations" in $\mathcal{T}_x$ are redundant, although some 3-point transformations are not, as illustrated earlier.

For the class of supermodular and convex objective functions, Theorem 14 implies that $f \prec_{\mathcal{S} \cap \mathcal{C}_x} g$ if and only if g can be obtained by adding to g a positive sum of elementary transformations from $\mathcal{T}(\mathcal{S})$ and $\mathcal{T}_x$. In that case, redundancy is even more severe. In fact, preliminary investigation suggests, for the case of two dimensions, that one can dispense of all convex elementary transformations based on 3-point convex combinations.

10 General Coarsening

In many applications, there is some arbitrariness in the way variables are constructed. For example, empirical income distributions may be formed by lumping together close income levels into categories. When comparing such distributions, it is desirable that the resulting ranking be robust with respect to the particular chosen categories. Most importantly, one should not “lose” important properties or comparisons of distributions by coarsening them through categories.

In Section 4, we have shown that the stochastic supermodular ordering is stable under coarsening. The technique used in the proof relied on the linear structure of the order, and particularly on its conic representation. However, some important orders do not have this structure. For example, association and affiliation, which involve covariances or conditional distributions, does not have this structure. It is important to determine whether such widely used orders is invariant to coarsening. At the same time, the convex ordering (Section **??**) has a conic structure but is not invariant to coarsening. Intuitively, convexity is a property that depends on the evenness with which data is split. To make this intuition precise, we need a more general approach clarifying what structural property of the order guarantees coarsening-invariance.

In order to apply our coarsening theorem to the aforementioned orderings as well as to expose its underlying argument, we use the following flexible setting. Given a lattice L with d nodes, let $\Delta_L \subset \mathbb{R}^d$ denote the set of probability distributions defined on L.

An *expectations-based* order $\prec$ on Δ_L is given by

- A class $\mathcal{C}(L)$ of $\mathbb{R}^k$-valued functions defined on L,
- A criterion function $\Theta : \mathbb{R}^k \to \mathbb{R}$,

such that

$$f \prec g \Leftrightarrow \Theta(E[w_1|f], \ldots, E[w_k|f]) \leq \Theta(E[w_1|g], \ldots, E[w_k|g])$$

for all $w \in \mathcal{C}(L)$. The supermodular (convex) ordering corresponds to the case in which $k = 1$, $\Theta(z) = z$, and $\mathcal{C}(L)$ denotes the class of supermodular (convex) functions. “Higher association” is also an expectations-based order: a random vector Y is “more associated” than a random vector X if $Cov(m(Y), n(Y)) \geq Cov(m(X), n(X))$ for all increasing functions m and n. This corresponds to the case $k = 3$, $\Theta(z_1, z_2, z_3) = z_1 - z_2 z_3$, and $\mathcal{C}(L)$ consists of 3-tuple of functions (w_1, w_2, w_3) such that i) w_2 and w_3 are increasing, and ii) $w_1 = w_2 w_3$. Orders involving conditional expectations are also expectations-based orders. For example, expressions like $E[m(X)|n(X) \geq z]$ with m, n increasing can be rewritten as

$$E[m(X)1_{n(X)\geq z}]/E[1_{n(X)\geq z}],$$

which corresponds to the case $k = 2$, $\Theta(z_1, z_2) = z_1/z_2$, and $\mathcal{C}(L)$ consists of all pairs (w_1, w_2) where $w_2 = 1_A$ is a nondecreasing indicator function (i.e. corresponding to a so-called "increasing set", A) and w_1 is the product of any increasing function and of w_2. In what follows, we fix Θ, so that the expectations-based order will simply be characterized by the class $\mathcal{C}(L)$ and denoted by $\prec_{\mathcal{C}(L)}$.

Consider a coarsening M of L along with the surjective map $\phi : L \to M$ defined in Section 4. Let $\mathcal{C}(L)$ and $\mathcal{C}(M)$ denote classes of $\mathbb{R}^k$-valued functions respectively defined on L and M. Although not required for the analysis to follow, one should think of these classes as being characterized by a common property, such as supermodularity or convexity, or a combination thereof.

For any $w \in \mathcal{C}(M)$, the L-extension of w is defined by

$$w^\phi(x) = w(\phi(x))$$

for all $x \in L$. Say that $\mathcal{C}(M)$ is *embedded* in $\mathcal{C}(L)$ if for all w in $\mathcal{C}(M)$, the L-extension of w belongs to $\mathcal{C}(L)$. Finally, recall that for any distribution $f \in \Delta_L$, the M-coarsening of f is given by

$$f^\phi(y) = \sum_{x \in L : \phi(x) = y} f(x)$$

for all $y \in M$.

THEOREM 17 *Suppose that $f \prec_{\mathcal{C}(L)} g$ and that $\mathcal{C}(M)$ is embedded in $\mathcal{C}(L)$. Then, $f^\phi \prec_{\mathcal{C}(M)} g^\phi$.*

Proof. It suffices to show that for any $w \in \mathcal{C}(M)$, there exists $\tilde{w} \in \mathcal{C}(L)$ such that $E[w_i|h^\phi] = E[w_i|h]$ for all $h \in \Delta_L$ and $i \in \{1, \ldots, k\}$. Taking $\tilde{w} = w^\phi$ yields the result since, by construction of the coarsening,

$$\sum_{y \in M} h^\phi(y) g(w_I(y)) = \sum_{x \in L} h(x) g(w_I^\phi(x)). \blacksquare$$

As a corollary, we can recover the coarsening result for the stochastic supermodular ordering. Indeed, suppose that $w : M \to \mathbb{R}$ is supermodular. For any $x \in L$ and components $i, j \in N$, there are two cases. First, it could be that $\phi(x) = \phi(x + e_i)$ or $\phi(x) = \phi(x + e_j)$. In that case, we necessarily have $\phi(x + e_j) = \phi(x + e_i + e_j)$ or, respectively, $\phi(x + e_i) = \phi(x + e_i + e_j)$. Either way, this implies that $w^\phi(x) + w^\phi(x + e_i + e_j) = w(\phi(x)) + w(\phi(x + e_i + e_j)) = w^\phi(x + e_i) + w^\phi(x + e_j)$. Second, it could be that the four elements x, $x + e_i$, $x + e_j$, $x + e_i + e_j$ of L have distinct images under the coarsening ϕ. In that case, $w^\phi(x) + w^\phi(x + e_i + e_j) - w^\phi(x + e_i) + w^\phi(x + e_j) = w(\phi(x)) + w(\phi(x) + e_i + e_j) - w(\phi(x) + e_i) + w(\phi(x) + e_j)$, which is nonnegative by supermodularity of w. This shows that the class of supermodular functions on M is embedded in the class of supermodular functions on L, and hence, by Theorem 17, that the stochastic supermodular ordering is coarsening invariant.

Similarly, one can show that higher association is coarsening invariant. For this, it suffices to show that the L-extension of an nondecreasing function on M is also nondecreasing, which is a straightforward exercise.

11 Relation to Copulas

An increasingly popular way to think about interdependence across random variables is the concept of copulas. A common view is that copulas capture interdependence by separating marginal distributions from joint distributions. This view is based on Sklar's seminal theorem, which we recall here. For simplicity let us say that C is a *copula* if it is the joint distribution of n uniform random variables.

Theorem 18 (Sklar, 1959) *Let F be any distribution function of n variables, with marginals $F_1, \ldots, F_n$. There exists a copula C such that*

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

Suppose that copulas indeed contain all interesting information about interdependence. There still remains to compare copulas for different distributions. If the comparison of two joint distributions only depends on their copulas, how should one compare these copulas? A natural idea, followed by Decancq (2007) is to apply the stochastic supermodular ordering to copulas rather than to the distribution themselves.

Our analysis challenges the use of copulas for comparing interdependence. Firstly, Theorem 10 provides a sharp observation: for two distributions to be comparable according to the modular ordering (and, therefore the more restrictive supermodular ordering), they *must have identical marginals.* Therefore, the apparent gain provided by copulas to abstract from differing marginal distributions disappears when interdependence is based on the supermodular ordering.

Secondly, the use of copulas can only increase complexity of the comparison. With discrete support, there is an uncountable infinity of copula representations for any distribution F. The only constraint (other than usual conditions for any function to be a copula) is that copulas must coincide on the range of values of the marginal distributions. This point, on which we will come back, can be illustrated by the simplest example: suppose that $L = L_2$, i.e. consists of a one-dimensional two-point support, and that $F(0) = 1/4)$. Then, any nondecreasing function $C : [0,1] \rightarrow [0,1]$ such that $C(1/4) = 1/4$, $C(0) = 0$ and $C(1) = 1$, provides a representation of F in Sklar's theorem. It is generally impossible to reconstruct a distribution from its copula. To illustrate, suppose in the previous example that $C(x) = 0$ for $x < 1/4$, $C(x) = 1/4$ for $1/4 \leq x < 1/2$, $C(x) = 1/2$ for $1/2 < x < 1$ and $C(1) = 1$. One could mistakenly infer that there are three points in the support of F, since the copula has three jumps. Or, if one already knows the initial distribution has a two-point support, how to determine which value of $1/4$ or $1/2$ corresponds to $F(0)$? One could impose the rule of picking a particular copula that is constant between any two values in the range of F, but then the copula coincides with F, except that the domain is scaled by the values of marginal distributions. Therefore, even with this rule, copulas do not offer any advantage compared to working with the initial distribution. In conclusion, the use of copulas should be rejected because i) distributions can only be compared if they have identical marginals, so that advantage of copulas disappears, and ii) copulas are only well defined on the range of values of marginal distributions, and contain no other useful information. To compare copulas according to the stochastic supermodular ordering, one has to essentially reconstruct the initial joint distribution.

12 Discussion

The Quasi-Supermodular Ordering

We have argued that the supermodular ordering was a natural notion to compare interdependence in multivariate distributions. We also considered the more restrictive class of quadratic objective functions. However, there exist other classes of functions that capture some notion of interdependence. A larger class consists of quasisupermodular functions. Recall that a function w defined on a lattice L is quasisupermodular if for all z, z' in L, $w(z \wedge z') \leq (<) w(z) \Rightarrow w(z') \leq (<) w(z \vee z')$. While potentially interesting, this class of functions has two problems. First, if one requires that the objective also be increasing, then quasisupermodularity always holds, as is easily checked. That is, the class of increasing quasisupermodular functions coincides with the class of increasing functions. This is true because any increasing function satisfies both the premise and the conclusion of any quasisupermodularity condition. Similarly, the class of decreasing quasisupermodular functions coincides with the class of decreasing functions, because the premise of any quasisupermodular condition is never satisfied for $x \neq y$. Another problem with the quasisupermodular ordering is that the class of quasisupermodular functions is not convex, i.e. the sum of two quasisupermodular functions is not necessarily quasisupermodular. Therefore the dual cone approach undertaken in this paper cannot be extended to this ordering.

For the case of two variables, say x and y, a function w is quasisupermodular if and only it satisfies the single crossing property in both (x, y) and (y, x) (recall that variables entering the definition of the single crossing property play an asymmetric role). In the context of decision making under uncertainty, the single crossing property is associated with the notion of affiliation. Let $f(x, y)$ denote the joint probability. Affiliation means (among other possible definitions) that the ratio $f(x', y)/f(x, y)$ is nondecreasing in y whenever $x' \geq x$. Stated differently, the function f is log-supermodular, i.e. $f(x', y')f(x, y) \geq f(x', y)f(x, y')$ for all $x \leq x'$ and $y \leq y'$. Affiliation is a symmetric property and a stronger notion than first order stochastic dominance, which is closely related to the stochastic supermodular ordering. Given the close tie between single-crossing and affiliation, hence between quasi-supermodularity and affiliation, it seems that, for the class of quasisupermodular functions, the corresponding order on distributions should correspond to variables being "more affiliated". A related notion has been introduced by Lehmann (1988) to compare signal informativeness. Interpreting x as a signal and y as the state of the world, x is more informative with distribution G than with distribution F if $G^{-1}(F(x|y)|y)$ is increasing in y for all x, where $F(\cdot|y)$ and $G(\cdot|y)$ are the conditional distributions of x given y and $G^{-1}(\cdot|y)$ is the (generalized) inverse of $G(\cdot|y)$. Lehmann informativeness is related to the single crossing property through several comparative statics theorem (see Jewitt (2006) and Quah and Strulovici (2008)). Higher informativeness captures a strong notion of higher interdependence between x and y, and we may conjecture a strong notion, similar to higher informativeness, should correspond to the interdependence order generated by the class of quasisupermodular objective functions. This question would deserve a separate inquiry.

13 Conclusion

14 References

Adler, M. D., and Sanchirico, C. W., "Inequality and Uncertainty: Theory and Legal Applications", University of Pennsylvania Law Review, 2006, 155, 279.

Athey, S., "Characterizing Properties of Stochastic Objective Functions", unpublished paper, Sept. 2000.

Atkinson, A. B., and Bourguignon, F., "The Comparison of Multi-Dimensioned Distributions of Economic Status", Review of Economic Studies, 1982, 49, 183-201.

Avis, D., and Fukuda, K., "A Pivoting Algorithm for Convex Hulls and Vertex Enumeration of Arrangements and Polyhedra", Discrete and Computational Geometry, 1992, 8, 295-313.

Avis, D., and Bremner, D., "How Good Are Convex Hull Algorithms?", 11^{th} Annual ACM Symposium on Computational Geometry, Vancouver, B.C., 1995, pp. 20–28.

Ben-Porath, E., Gilboa, I., and Schmeidler, D., "On the Measurement of Inequality under Uncertainty", Journal of Economic Theory, 1997, 75, 194-204.

Bertsimas, D., and Tsitsiklis, J. *Introduction to Linear Optimization*, Athena Scientific, Nashua, N.H., 1997.

Boland, P. J., and Proschan, F., "Multivariate Arrangement Increasing Functions with Applications in Probability and Statistics", Journal of Multivariate Analysis, 1988, 25, 286-298.

Bond, P., and Gomes, A., "Multitask Principal-Agent Problems: Optimal Contracts, Fragility, and Effort Misallocation", Journal of Economic Theory, 2009, 144, 175-211.

Bourguignon, F., and Chakravarty, S.R., "Multi-dimensional poverty orderings", DELTA W.P. 2002-22.

Christofides, T.C., and Vaggelatou, E., "A Connection between Supermodular Ordering and Positive/Negative Association", Journal of Multivariate Analysis, 2004, 88, 138-151.

Denuit, M., Dhaene, J., Goovaerts, M.J., and Kaas, R., Actuarial Theory for Dependent Risks: Measures, Orders, and Models, 2005, Wiley.

Dhaene, J., and Goovaerts, M.J., "Dependency of Risks and Stop-Loss Order", ASTIN Bulletin, 1996, 26, 201-212.

Epstein. L., and Tanny, S., "Increasing Generalized Correlation: A Definition and Some Economic Consequences", Canadian Journal of Economics, 1980, 13, 16-34.

Esary, J.D., Proschan, F., and Walkup, D.W., "Association of Random Variables, with Applications", Annals of Mathematical Statistics, 1967, 44, 1466-1474.

Fernandez, R., and Gali, J., "To Each According to? Markets, Tournaments, and the Matching Problem with Borrowing Constraints", Review of Economic Studies, 1999, 66, 799-824.

Fishburn, P. C., and Lavalle, I. H., "Stochastic Dominance on Unidimensional Grids", Mathematics of Operations Research, 1995, 20, 513-525.

Fukuda, K., and Prodon, A., "Double description method revisited." In Combinatorics and Computer Science, volume 1120 of Lecture Notes in Computer Science, pages 91–111. Springer, 1996.

Gajdos, T., and Maurin, E., "Unequal Uncertainties and Uncertain Inequalities: An Axiomatic Approach", Journal of Economic Theory, 2004, 116, 93-118.

Goldman, A. J. and Tucker, A. W. "Polyhedral convex cones." In Linear inequalities and related systems, Annals of Mathematics Studies 38, eds. H. W. Kuhn and A. W. Tucker. Princeton University Press, 1956, pp. 19-39.

Hardy, G.H., Littlewood, J.E., and Polya, G., *Inequalities*, Cambridge: Cambridge University Press, 2nd edition, 1952.

Hennessy, D. A., and Lapan, H. E., "A Definition of 'More Systematic Risk' with Some Welfare Implications", Economica, 2003, 70, 493-507.

Hu, T., Muller, A., and Scarsini, M., "Some Counterexamples in Positive Dependence", Journal of Statistical Planning and Inference, 2004, 153-158.

Hu, T., and Wu, Z., "On Dependence of Risks and Stop-Loss Premiums", Insurance:

Mathematics and Economics, 1999, 24, 323-332.

Hu, T., Xie, C., and Ruan, L, "Dependence Structures of Multivariate Bernoulli Random Vectors", Journal of Multivariate Analysis, 2005, 94, 172-195.

Joe, H., "Multivariate Concordance", Journal of Multivariate Analysis, 1990, 35, 12-30.

Karlin, S., and Novikoff, A., "Generalized Convex Inequalities", Pacific Journal of Mathematics, 1963, 13, 1251-1279.

Kiselman, C., "Convex Functions on Discrete Sets", Combinatorial Image Analysis, 2005, Springer.

Kroll, Y., and Davidovitz, L., "Inequality Aversion versus Risk Aversion", Economica, 2003, 70, 19-29.

Levy, H., and Paroush, J., "Toward Multivariate Efficiency Criteria", Journal of Economic Theory, 1974, 7, 129-42.

Lorentz, G.G., "An Inequality for Rearrangements", American Mathematical Monthly, 1953, vol. 60, 176-179.

Luenberger, D., *Vector Space Optimization*, Wiley, New York, 1969.

Lu, T.-Y., and Zhang, Y., "Generalized Correlation Order and Stop-Loss Order", Insurance: Mathematics and Economics, 2004, 35, 69-76.

Marinacci, M., and Montrucchio, L., "Ultramodular Functions", Mathematics of Operations Research, 2005, 30, 311-332.

McAfee, R. P., "Coarse Matching", Econometrica, 2002, 70, 2025-34.

Meyer, M.A., "Interdependence in Multivariate Distributions: Stochastic Dominance Theorems and an Application to the Measurement of Ex Post Inequality under Uncertainty", Nuffield College Disc. Paper No. 49, April 1990.

Meyer, M. A., and Mookherjee, D., "Incentives, Compensation and Social Welfare", Review of Economic Studies, 1987, 45, 209-226.

Meyer, M. A., and Rothschild, M., "(In)efficiency in Sorting and Matching", mimeo, Oxford, 2004.

Motzkin, T. S., Raiffa, H., Thompson, G. L. , and Thrall, R. M.. "The Double Description Method". In H. W. Kuhn and A. W. Tucker Ed., Contributions to the Theory of Games II, 1953, 51-73.

Muller, A., and Scarsini, M., "Some Remarks on the Supermodular Order", Journal of Multivariate Analysis, 2000, 73, 107-119.

Muller, A., and Stoyan, D., Comparison Methods for Stochastic Models and Risks, 2002, Wiley.

Murota, K., "Discrete Convex Analysis", Mathematical Programming, 1998, 83, 313-371.

Murota, K., *Discrete Convex Analysis"*, SIAM Monographs on Discrete Mathematics and Applications, Philadelphia, PA, 2003.

Ostrovsky, M., "Stability in Supply Chain Networks", American Economic Review, 2008, 98, 897-923.

Prat, A., "Should a Team Be Homogenous?", European Economic Review, 2002, 46, 1187-1207.

Rothschild, M. and Stiglitz, J. E. "Increasing risk: I. A definition", Journal of Economic Theory, 1970, 2, 225-243.

Ruschendorf, L., "Comparison of Multivariate Risks and Positive Dependence", Journal of Applied Probability, 41, 2004, 391-406.

Shaked, M., and Shanthikumar, J.G., "Supermodular Stochastic Orders and Positive Dependence of Random Vectors", Journal of Multivariate Analysis, 1997, 61, 86-101.

Sklar, A. "Fonctions de répartition à n dimensions et leurs marges", Publ. Inst. Statist. Univ. Paris, 1959, 8, 229-231.

Tchen, A.H., "Inequalities for Distributions with Given Marginals", Annals of Probability, 1980, 8, 814-827.

Topkis, D.M., *Ordered Optimal Solution*, Doctoral Dissertation, Stanford University, 1968.

Topkis, D.M., "Minimizing a Submodular function on a Lattice", Operations Research, 1978, 26, 305-331.

Yi, Z., and Weng, C., "On the Correlation Order", Statistics and Probability Letters, 2006, 76, 1410-1416.

Appendices

A Appendix: Implementation of the Double Description Method

B Appendix: Relation between orderings

Greater Weak Association is not a linear order

Every linear order satisfies Theorem 5, as is easily checked. However, association does not.

Supermodular Function that cannot be replicated by convex-modular ones.