

Convexification of Stochastic Ordering*

Darinka Dentcheva[†]

Andrzej Ruszczyński[‡]

Abstract

We consider sets defined by the usual stochastic ordering relation and by the second order stochastic dominance relation. Under fairly general assumptions we prove that in the space of integrable random variables the closed convex hull of the first set is equal to the second set.

Keywords: Stochastic Dominance, Convexity.

1 Stochastic Ordering Relations

The notion of stochastic ordering (or *stochastic dominance of first order*) has been introduced in statistics in [11, 10] and further applied and developed in economics [15, 6]. It is defined as follows. For a random variable $X \in \mathcal{L}_1$ we consider its distribution function, $F(X; \eta) = P[X \leq \eta]$, $\eta \in \mathbb{R}$. We say that a random variable X *dominates in the first order* a random variable Y if

$$F(X; \eta) \leq F(Y; \eta) \quad \text{for all } \eta \in \mathbb{R}.$$

We denote this relation $X \succeq_{(1)} Y$.

For two integrable random variables X and Y , we say that X *dominates Y in the second order* if

$$\int_{-\infty}^{\eta} F(X; \alpha) d\alpha \leq \int_{-\infty}^{\eta} F(Y; \alpha) d\alpha \quad \text{for all } \eta \in \mathbb{R}.$$

We denote this relation $X \succeq_{(2)} Y$. The second order dominance has been introduced in [8]. We refer the reader to [13, 14, 17] for a modern perspective on stochastic ordering and to [1, 7] for recent applications in statistics.

In recent publications [3, 4], we have introduced a new stochastic optimization model involving stochastic dominance relations of second (or higher) order as constraints. These constraints allow us to use random reference outcomes, instead of fixed thresholds. We have discovered the role of concave utility functions as Lagrange multipliers associated with dominance constraints of second order and higher orders.

In this paper we analyze sets defined by first order stochastic dominance constraints.

Let us introduce some notation used throughout the paper. An abstract probability space is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$. The expected value operator is denoted by \mathbb{E} . The standard symbol $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^n)$ (shortly \mathcal{L}_1^n) denotes the space of all integrable mappings X from Ω to \mathbb{R}^n .

*appeared (under this title) in *Comptes Rendus Acad. Bulgare Sci.* 57 (2004), No. 3, 5–10, and also as part of “Semi-Infinite Probabilistic Optimization: First Order Stochastic Dominance Constraints,” *Optimization* 53 (2004) 583–601.

[†]Stevens Institute of Technology, Department of Mathematical Sciences, Hoboken, NJ 07030, USA.

[‡]Rutgers University, Department of Management Science, 94 Rockafeller Rd, Piscataway, NJ 08854, USA.

2 Sets Defined by Dominance Constraints

Define for a given random variable $Y \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ two sets

$$\begin{aligned} A_{(1)}(Y) &= \{X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) : X \succeq_{(1)} Y\}, \\ A_{(2)}(Y) &= \{X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) : X \succeq_{(2)} Y\}. \end{aligned}$$

The random variable $Y \in \mathcal{L}_1$ plays the role of a fixed reference outcome.

Lemma 1 *The set $A_{(1)}(Y)$ is closed in $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$.*

Proof. The result follows from the fact that convergence in \mathcal{L}_1 implies convergence in probability. \square

Lemma 2 ([4]) *The set $A_{(2)}(Y)$ is convex and closed in $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$.*

The set $A_{(1)}(Y)$ is not convex in general.

Example 1 Suppose that $\Omega = \{\omega_1, \omega_2\}$, $\mathbb{P}[\omega_1] = \mathbb{P}[\omega_2] = 1/2$ and $Y(\omega_1) = -1$, $Y(\omega_2) = 1$. Then $X_1 = Y$ and $X_2 = -Y$ both dominate Y in the first order. However, $X = (X_1 + X_2)/2 = 0$ is not an element of $A_{(1)}(Y)$ and thus the set $A_{(1)}(Y)$ is not convex.

We may notice, though, that X dominates Y in the second order, and this is the starting point of our analysis. Our main result here is that under fairly general assumptions the set $A_{(2)}(Y)$ is a convexification of $A_{(1)}(Y)$.

3 Convexification

Directly from the definition we see that first order dominance implies second order dominance, and thus $A_{(1)}(Y) \subseteq A_{(2)}(Y)$. Since the set $A_{(2)}(Y)$ is convex we also have

$$\text{conv } A_{(1)}(Y) \subseteq A_{(2)}(Y). \quad (1)$$

Here the symbol $\text{conv}A$ denotes the convex hull of the set A . We shall use $\overline{\text{conv}}A$ for the closure of the convex hull. We are interested in establishing the inverse inclusion.

We start from the special case of discrete distributions.

Theorem 1 *Assume that $\Omega = \{1, \dots, N\}$, \mathcal{F} is the set of all subsets of Ω and $\mathbb{P}[k] = 1/N$, $k = 1, \dots, N$. If Y is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ then*

$$\text{conv } A_{(1)}(Y) = A_{(2)}(Y).$$

Proof. To prove the inverse inclusion to (1), suppose that $X \in A_{(2)}(Y)$. Define $x_i = X(i)$ and $y_i = Y(i)$, $i = 1, \dots, N$. We can identify X and Y with vectors $x = (x_1, \dots, x_N)$ and $y = (y_1, \dots, y_N)$. Since the probabilities of all elementary events are equal, the second order stochastic dominance relation coincides with the concept of *weak majorization*:

$$\left[X \succeq_{(2)} Y \right] \Leftrightarrow \left[\sum_{k=1}^l x_{[k]} \geq \sum_{k=1}^l y_{[k]}, \quad l = 1, \dots, N \right],$$

where $x_{[k]}$ denotes the k th smallest component of x (see [12]).

It follows from the theorem by Hardy, Littlewood and Polya [9] (see also [12, Proposition D.2.b]) that weak majorization is equivalent to the existence of a doubly stochastic matrix Π such that $x \geq \Pi y$. By Birkhoff's Theorem [2], we can find permutation matrices Q^1, \dots, Q^M such that $\Pi \in \text{conv}\{Q^1, \dots, Q^M\}$. Defining $z^j = Q^j y$ we conclude that there exist nonnegative reals $\alpha_1, \dots, \alpha_M$ totaling 1, such that $x \geq \sum_{j=1}^M \alpha_j z^j$. Identifying with the vectors z^j random variables Z^j on $(\Omega, \mathcal{F}, \mathbb{P})$ we also see that

$$X(\omega) \geq \sum_{j=1}^M \alpha_j Z^j(\omega),$$

for all $\omega \in \Omega$. Since each vector z^j is a permutation of y and the probabilities are equal, the distribution of Z^j is identical with the distribution of Y . Thus $Z^j \succeq_{(1)} Y$, $j = 1, \dots, M$. Let us define

$$\hat{Z}^j(\omega) = Z^j(\omega) + \left(X(\omega) - \sum_{k=1}^M \alpha_k Z^k(\omega) \right), \quad \omega \in \Omega, \quad j = 1, \dots, M.$$

Then the last two inequalities render $\hat{Z}^j \in A_{(1)}(Y)$, $j = 1, \dots, M$, and $X(\omega) = \sum_{j=1}^M \alpha_j \hat{Z}^j(\omega)$, as required. \square

The result of the above theorem is not true for general probability spaces, as the following example illustrates.

Example 2 Suppose that $\Omega = \{\omega_1, \omega_2\}$, $\mathbb{P}[\omega_1] = 1/3$, $\mathbb{P}[\omega_2] = 2/3$ and $Y(\omega_1) = -1$, $Y(\omega_2) = 1$. Then it is easy to see that $X \succeq_{(1)} Y$ if and only if $X(\omega_1) \geq -1$ and $X(\omega_2) \geq 1$. Thus $A_{(1)}(Y)$ is convex.

Consider the random variable $Z = \mathbb{E}[Y] = 1/3$. It dominates Y in the second order, but it does not belong to $\text{conv } A_{(1)}(Y) = A_{(1)}(Y)$.

It follows from this example that the probability space must be sufficiently rich to observe our phenomenon. If we could define a new probability space $\Omega' = \{\omega_1, \omega_{21}, \omega_{22}\}$, in which the event ω_2 is split in two equally likely events ω_{21}, ω_{22} , then we could use Theorem 1 to obtain the equality $\text{conv } A_{(1)}(Y) = A_{(2)}(Y)$. The theorems of Strassen follow this line (see [16, 14]). In our problem, and in the optimization context in general, the probability space has to be fixed at the outset and we are interested in sets of random variables as elements of functional spaces $\mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^n)$, rather than sets of their distributions. Therefore, the Strassen theory cannot be directly applied here.

Theorem 2 Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is nonatomic and $Y \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$. Then

$$A_{(2)}(Y) = \overline{\text{conv}} A_{(1)}(Y).$$

Proof. Assume that Y has a continuous probability distribution. Suppose that $X \succeq_{(2)} Y$. By the integrability of X , for every $\varepsilon > 0$ we can find a finite σ -subalgebra \mathcal{B} associated with some finite partition $\{\Omega_1, \dots, \Omega_N\}$ of Ω such that the discrete random vector $\hat{X} = \mathbb{E}[X|\mathcal{B}]$ satisfies the inequality $\|\hat{X} - X\| \leq \varepsilon$. Here and everywhere else in this proof the symbol $\|X\|$ refers to the \mathcal{L}_1 -norm of X . For each Ω_i , $i = 1, \dots, N$, the conditional distribution of Y is continuous. Therefore, for any $\delta > 0$, we can find $\Omega'_i \subseteq \Omega_i$ such that the probability of Ω'_i is a rational number and $\mathbb{P}[\Omega'_i] \geq \mathbb{P}[\Omega_i] - \delta$, $i = 1, \dots, N$. Define

$$\tilde{X}(\omega) = \mathbb{E}[X|\Omega'_i] \quad \text{for } \omega \in \Omega'_i, \quad i = 1, \dots, N+1, \quad (2)$$

where $\Omega'_{N+1} = \Omega \setminus \bigcup_{i=1}^N \Omega'_i$. Since X is integrable, we can always choose δ small enough so that $\|\tilde{X} - \hat{X}\| \leq \varepsilon$. Since the probabilities of Ω'_i , $i = 1, \dots, N+1$, are rational, we can use the conditional

distributions of Y to split these events in such a way that all elements of the resulting partition have equal probabilities. Thus we obtain a partition $\{B_1, \dots, B_K\}$ of Ω such that $\mathbb{P}[B_k] = 1/K, k = 1, \dots, K$. The random variable \tilde{X} is constant on each B_k and $\|\tilde{X} - X\| \leq 2\varepsilon$.

Jensen's inequality and (2) imply that for any concave function $u(\cdot)$ $\mathbb{E}[u(\tilde{X})] \geq \mathbb{E}[u(X)]$, whenever $\mathbb{E}[u(X)]$ exists [5, 10.2.7]. Thus \tilde{X} dominates X in the second order. The stochastic dominance relations are transitive and therefore

$$\tilde{X} \in A_{(2)}(Y).$$

For each B_k the conditional distribution of Y is continuous. Denote by $q_k(\alpha)$ the α -quantile of this conditional distribution with the convention $q_k(0) = -\infty$ and $q_k(1) = +\infty$. We partition each B_k into equally probable disjoint subsets $B_{kl}, l = 1, \dots, L$, such that

$$Y(B_{kl}) \subset \left(q_k\left(\frac{l-1}{L}\right), q_k\left(\frac{l}{L}\right) \right], \quad l = 1, \dots, L.$$

In this way we define a certain subalgebra \mathcal{B}' of \mathcal{F} .

Since Y is integrable, we can choose L big enough so that the discrete random variable $\tilde{Y} = \mathbb{E}[Y|\mathcal{B}']$ satisfies the inequality

$$\|\tilde{Y} - Y\| \leq \frac{\varepsilon}{K}. \quad (3)$$

Consider the function

$$F^{(2)}(Y, \eta) = \int_{-\infty}^{\eta} \mathbb{P}[Y \leq \alpha] d\alpha = \mathbb{E}[(\eta - Y)_+].$$

The last equation can be obtained by changing the order of integration. It follows from the last two relations that

$$F^{(2)}(\tilde{Y}, \eta) = \mathbb{E}[(\eta - \tilde{Y})_+] \geq \mathbb{E}[(\eta - Y)_+] - \mathbb{E}[|Y - \tilde{Y}|] \geq F^{(2)}(Y, \eta) - \frac{\varepsilon}{K}.$$

Since \tilde{X} dominates Y in the second order, for every $\eta \in \mathbb{R}$ we can continue the last chain of inequalities as follows

$$F^{(2)}(\tilde{Y}, \eta) \geq F^{(2)}(Y, \eta) - \frac{\varepsilon}{K} \geq F^{(2)}(\tilde{X}, \eta) - \frac{\varepsilon}{K}.$$

Observe that $F^{(2)}(\tilde{X}, \cdot)$ is convex and piecewise linear and its smallest nonzero slope equals at least $1/K$. Moreover, $F^{(2)}(\tilde{Y}, \cdot)$ is nonnegative. Shifting $F^{(2)}(\tilde{X}, \eta)$ by ε to the right yields a lower bound for $F^{(2)}(\tilde{Y}, \eta)$. Thus the last inequality implies that

$$F^{(2)}(\tilde{Y}, \eta) \geq F^{(2)}(\tilde{X}, \eta - \varepsilon) \quad \text{for all } \eta \in \mathbb{R}.$$

Equivalently,

$$\tilde{X} + \varepsilon \succeq_{(2)} \tilde{Y}.$$

We can consider the random variables \tilde{X} and \tilde{Y} as defined on a finite probability space $\tilde{\Omega} = \{B_{kl} : k = 1, \dots, K, l = 1, \dots, L\}$ with equally likely elementary events.

As in the proof of Theorem 1, we associate with \tilde{X} and \tilde{Y} vectors x and y in \mathbb{R}^{KL} . We conclude that there exist nonnegative numbers $\alpha_m, m = 1, \dots, M$, totaling 1, and permutations Q^m such that

$$x + \varepsilon \mathbb{1} \geq \sum_{m=1}^M \alpha_m Q^m y, \quad (4)$$

where $\mathbb{1} = (1, \dots, 1) \in \mathbb{R}^{KL}$. Each Q^m is a permutation matrix of dimension $KL \times KL$. Consider the vector $z^m = Q^m y$. It can be interpreted as the vector of realizations of a discrete random variable Z^m . Inequality (4) can be rewritten as

$$\tilde{X}(\omega) + \varepsilon \geq \sum_{m=1}^M \alpha_m \tilde{Z}^m(\omega), \quad \omega \in \Omega. \quad (5)$$

By construction, the distribution of \tilde{Z}^m is the same as that of \tilde{Y} and therefore $Z^m \in A_{(1)}(\tilde{Y})$.

Suppose that Q^m maps the event B_{kl} to the event B_{st} , that is, the value of Z^m at all $\omega \in B_{st}$ is equal to the value of \tilde{Y} at all $\omega \in B_{kl}$. Our aim is to define an analogous ‘‘permutation’’ V^m of the original reference outcome Y . We want to have the distribution of V^m the same as that of Y , and thus $V^m \succeq_{(1)} Y$. Moreover, V^m will be close to Z^m .

Consider an arbitrary $\omega \in \Omega$. Suppose that $\omega \in B_{st}$ for some (s, t) . Let $F_s(\cdot)$ be the conditional distribution function of Y , given $\tilde{X} = x_s$. Clearly, $F_s(Y(\omega))$, given $\omega \in B_{st}$, is uniform in $[(t-1)/L, t/L]$. We now find (k, l) such that Q^m maps the event B_{kl} to the event B_{st} . This allows us to translate the interval $[(t-1)/L, t/L]$ to the interval $[(l-1)/L, l/L]$. Finally, we define

$$V^m(\omega) = F_k^{-1}\left(F_s(Y(\omega)) + \frac{l-t}{L}\right).$$

If the equation $F_k(v) = \eta$ has many solutions, we can choose any of them as $F_k^{-1}(\eta)$. The probability of such a situation is zero.

By construction, the distribution of $V^m(\omega)$ for $\omega \in B_{st}$ is exactly the same as the distribution of $Y(\omega)$ for $\omega \in B_{kl}$. All events B_{ij} are equally likely and simply permuted, and therefore the distribution of V^m is identical with the distribution of Y . Consequently, $V^m \in A_{(1)}(Y)$. Thus,

$$\sum_{m=1}^M \alpha_m V^m(\omega) \in \text{conv } A_{(1)}(Y).$$

Furthermore, by the construction of V^m and inequality (3), we have $\|V^m - Z^m\| = \|Y - \tilde{Y}\| \leq \varepsilon$. Observe that

$$A_{(1)}(Y) = A_{(1)}(Y) + \mathcal{L}_1^+,$$

where $\mathcal{L}_1^+ = \{X \in \mathcal{L}_1 : X \geq 0 \text{ a.s.}\}$. Using (5) and the last three relations we obtain the following estimate of the \mathcal{L}_1 -distance of X to the set $\text{conv } A_{(1)}(Y)$:

$$\begin{aligned} d(X, \text{conv } A_{(1)}(Y)) &= d(X, \text{conv } A_{(1)}(Y) + \mathcal{L}_1^+) \\ &\leq \|X - \tilde{X} - \varepsilon\| + d(\tilde{X} + \varepsilon, \text{conv } A_{(1)}(Y) + \mathcal{L}_1^+) \\ &\leq 3\varepsilon + d(\tilde{X} + \varepsilon, \sum_{m=1}^M \alpha_m V^m + \mathcal{L}_1^+) \\ &\leq 3\varepsilon + d(\tilde{X} + \varepsilon, \sum_{m=1}^M \alpha_m Z^m + \mathcal{L}_1^+) + \sum_{m=1}^M \alpha_m \|Z^m - V^m\| \\ &\leq 3\varepsilon + 0 + \sum_{m=1}^M \alpha_m \varepsilon = 4\varepsilon. \end{aligned}$$

Since ε was arbitrary, we conclude that $d(X, \text{conv } A_{(1)}(Y)) = 0$, which proves the statement.

For a general distribution of Y we approximate it by continuous random variables, and the result follows by passing to the limit. \square

References

- [1] E. Arjas, D. Gasbarra, Bayesian inference of survival probabilities, under stochastic ordering constraints, *J. Amer. Statist. Assoc.* 91 (1996) 1101–1109.
- [2] G. Birkhoff, Tres observaciones sobre el algebra lineal, *Univ. Nac. Tucumán Rev. Ser. A* 5 (1946) 147–151.
- [3] D. Dentcheva and A. Ruszczyński, Optimization under linear stochastic dominance, *Comptes Rendus de l'Academie Bulgare des Sciences* 56 (2003), No. 6, pp. 5–10.
- [4] D. Dentcheva and A. Ruszczyński, Optimization with stochastic dominance constraints, *SIAM Journal on Optimization*, accepted for publication.
- [5] R.M. Dudley, *Real Analysis and Probability*, Cambridge University Press, Cambridge, 2002.
- [6] P.C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons, New York, 1970.
- [7] R.E. Gangnon and W.N. King, Minimum distance estimation of the distribution functions of stochastically ordered random variables, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 51 (2002) Issue 4 485–498.
- [8] J. Hadar and W. Russell, Rules for ordering uncertain prospects, *The American Economic Review* 59 (1969) 25–34.
- [9] G.H. Hardy, J.E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, MA, 1934.
- [10] E. Lehmann, Ordered families of distributions, *Ann. Math. Statistics* 26 (1955) 399–419.
- [11] H.B. Mann and D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Statistics* 18 (1947) 50–60.
- [12] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, San Diego, 1979.
- [13] K. Mosler and M. Scarsini (Eds.), *Stochastic Orders and Decision Under Risk*, Institute of Mathematical Statistics, Hayward, California, 1991.
- [14] A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risks*, John Wiley & Sons, Chichester, 2002.
- [15] J.P. Quirk and R. Saposnik, Admissibility and measurable utility functions, *Review of Economic Studies* 29 (1962) 140–146.
- [16] V. Strassen, The existence of probability measures with given marginals, *Ann. Math. Statistics* 38 (1965) 423–439.
- [17] R. Szekli, *Stochastic Ordering and Dependence in Applied Probability* Springer, New York, 1995.