

CHARACTERISATION OF A MULTIVARIATE STOCHASTIC ORDERING

COLM ART O'CONNOR

The multivariate stochastic ordering induced by the convex nondecreasing functions compares a combination of size and variability of random vectors. Closely following methods developed by Strassen, we show that two probability measures are ordered in this way if and only if they are the marginals of some submartingale. The implications of this in majorisation theory are discussed.

1. INTRODUCTION

The purpose of this note is to clarify a somewhat abstruse point in the stochastic ordering and majorisation literature, namely, the equivalence of the "submartingale property" and the ordering of (Borel) probability measure (p.m.s) induced by convex nondecreasing functions. To explain, let μ and ν denote p.m.s on \mathbb{R}^k , $k \geq 1$, whose means are finite (namely, the Euclidean length $|x|$ is both μ - and ν -integrable). Let us write $\mu \leq_{(1)} \nu$ if for every convex function $f: \mathbb{R}^k \rightarrow \mathbb{R}$ we have

$$(1) \quad \int f \, d\mu \leq \int f \, d\nu.$$

We write $\mu \leq_{(2)} \nu$ if (1) holds for every convex *nondecreasing* (that is, nondecreasing in each variable) function f . Note that the integrals in (1) may be $+\infty$, but not $-\infty$. Let us say that the pair (μ, ν) has the *submartingale property* if there exists a pair of random variables (X, Y) with finite means such that X has distribution μ , Y has distribution ν , and (X, Y) is a submartingale:

$$(2) \quad E(Y | X) \geq X \quad a.s.$$

We say that (μ, ν) has the *martingale property* if it has the submartingale property, and equality holds in (2) almost surely; that is, if (X, Y) is a martingale. In this note we prove

Received 5 February 1988

Copyright Clearance Centre, Inc. Serial-fee code: 0004-9729/88 \$A2.00+0.00.

THEOREM 1. *Let μ and ν be p.m.s on \mathbb{R}^k with finite means. Then $\mu \leq_{(2)} \nu$ if and only if (μ, ν) has the submartingale property.*

Strassen [4, Theorem 8] has proved that $\mu \leq_{(1)} \nu$ if and only if (μ, ν) has the martingale property. This result is shown below (Corollary 1) to be an easy consequence of Theorem 1, and in this sense Theorem 1 may be viewed as an extension of Strassen’s result. Strassen’s Theorem 9 also contains the univariate case ($k = 1$) of our Theorem 1. That Theorem 1 may be proved using Strassen’s original techniques is not well recognised in the stochastic ordering or majorisation literature. Our proof below closely parallels the proof of Strassen’s Theorem 8. Stoyan [3, p. 33] remarks that Theorem 1 is known in the one-dimensional case ($k = 1$), as does Whitt [5, p. 1064].

When μ and ν are uniform (that is “equally likely”) p.m.s on finite subsets of \mathbb{R}^k of the same cardinality, the ordering $\leq_{(2)}$ is a multivariate generalisation of the weak majorisation ordering \prec_w of Marshall and Olkin [1]. In this context, Theorem 1 becomes an extension of the multivariate case of Theorem 5.A.9 of their book. Because of the independent interest of majorisation, we provide an alternative proof of the result, in this special case, which does not rely on functional analytic methods.

2. PROOF OF THEOREM 1

The following special case of Strassen’s Theorem 7 is crucial to our proof of Theorem 1. Here π_1 and π_2 denote the two natural projections of $\mathbb{R}^k \times \mathbb{R}^k$ onto \mathbb{R}^k . Let Z^m denote the set of continuous real-valued functions h on \mathbb{R}^m for which

$$\|h\| = \sup\{|h(s)|/(|s| + 1) : s \in \mathbb{R}^m\} < \infty.$$

LEMMA 1. (Strassen). *Let μ and ν be p.m.s on \mathbb{R}^k having finite means. Let Λ be a nonempty convex set of p.m.s on $\mathbb{R}^k \times \mathbb{R}^k$, with finite means, which is closed in the topology generated by the functionals*

$$(3) \quad \lambda \rightarrow \int h \, d\lambda$$

for $h \in Z^{2k}$. Then a necessary and sufficient condition for Λ to contain a measure λ_0 with marginals $\mu = \lambda_0 \circ \pi_1^{-1}$ and $\nu = \lambda_0 \circ \pi_2^{-1}$ is that

$$(4) \quad \int f \, d\mu + \int g \, d\nu \leq \sup_{\lambda \in \Lambda} \int f \circ \pi_1 + g \circ \pi_2 \, d\lambda$$

for all f and g in Z^k .

This result is a straightforward consequence of Strassen’s Theorem 7. The hypotheses of Lemma 1 apply to the set Λ of all p.m.s λ which are the distribution of

some \mathbb{R}^k -valued submartingale (X, Y) with finite means. This may be seen as follows. Strassen points out that the topology generated by the functionals (3) is metrisable. Thus it suffices to show that if $\{\lambda_n\} \subseteq \Lambda$ is a sequence converging to a p.m. λ in the sense that

$$(5) \quad \int h \, d\lambda_n \rightarrow \int h \, d\lambda \quad \text{as } n \rightarrow \infty$$

for all $h \in Z^{2k}$, then $\lambda \in \Lambda$. Towards this goal, let g be a bounded nonnegative continuous function on \mathbb{R}^k . The submartingale property of λ_n implies that

$$\int (g \circ \pi_1)\pi_2 \, d\lambda_n \geq \int (g \circ \pi_1)\pi_1 \, d\lambda_n.$$

But the functions $(g \circ \pi_1)\pi_2 : \mathbb{R}^{2k} \rightarrow \mathbb{R}$ and $(g \circ \pi_1)\pi_1 : \mathbb{R}^{2k} \rightarrow \mathbb{R}$ are in Z^{2k} , so that (5) implies that this inequality persists in the limit:

$$\int (g \circ \pi_1)\pi_2 \, d\lambda \geq \int (g \circ \pi_1)\pi_1 \, d\lambda.$$

It follows that this inequality holds also for bounded nonnegative measurable g , and thus that $\lambda \in \Lambda$ as required. Note that Λ is not closed under weak convergence, as we see from the example (for $k = 1$) $(1/n)\delta_{(0,n)} + (1 - 1/n)\delta_{(0,-n/(n-1))} \in \Lambda, n = 2, 3, \dots$, where δ_x denotes the point mass at x . The topology on p.m.s of Lemma 1 ensures that not only mass (namely, “ $\lambda(dx)$ ”), but also moment (namely “ $x\lambda(dx)$ ”) cannot vanish “in the limit.”

By the concave nonincreasing majorant of a given Borel function g with (Borel) domain $T \subseteq \mathbb{R}^k$ we mean the least (extended-real-valued) concave nonincreasing function on \mathbb{R}^k which is greater than or equal to g on T . The following lemma gives a simple characterisation of the concave nonincreasing majorant. In this section it is enough to have this result for $T = \mathbb{R}^k$; in Section 3 we shall find it useful to take T to be a finite set.

LEMMA 2. *The concave nonincreasing majorant of a real-valued Borel function g with domain $T \subseteq \mathbb{R}^k$ is given by \hat{g} where*

$$(6) \quad \hat{g}(s) = \sup_{\mu \in M_s(T)} \int g \, d\mu, \quad s \in \mathbb{R}^k,$$

$M_s(T)$ being the set of p.m.s concentrated on T with mean componentwise greater than or equal to s .

Note:. Although g may not be defined on all of \mathbb{R}^k , \hat{g} is (being $-\infty$ whenever $M_s(T)$ is empty).

PROOF: $\hat{g} \geq g$ on T since $\delta_s \in M_s(T)$ if $s \in T$. \hat{g} is nonincreasing since if $t \geq s$ (componentwise) then $M_t(T) \subseteq M_s(T)$. \hat{g} is concave since if $\mu_1 \in M_s(T)$ and $\mu_2 \in M_t(T)$ then $p\mu_1 + q\mu_2 \in M_{ps+qt}(T)$, whenever p and q are nonnegative and sum to 1. Thus \hat{g} is concave and nonincreasing. It remains to show that if h is another concave nonincreasing function satisfying $g \leq h$ on T , then $\hat{g} \leq h$ also. To show this, note that for such an h , and any $\mu \in M_s(T)$, we have

$$h(s) \geq \int h \, d\mu.$$

Thus $\hat{h}_T \leq h$, where h_T denotes the restriction of h to T . Now $g \leq h_T$ on T , so $\hat{g} \leq \hat{h}_T \leq h$, by monotonicity of $\hat{\cdot}$ as defined by (6), and so $\hat{g} \leq h$ as required.

PROOF OF THEOREM 1: If (μ, ν) has the submartingale property, (1) may be deduced from (2) for convex nondecreasing functions through an application of Jensen’s inequality. Conversely, suppose that $\mu \leq_{(2)} \nu$. The result will follow from Lemma 1 once we prove (4) for the set Λ of distributions of \mathbb{R}^k -valued submartingales with finite means. Thus let f and g be elements of Z^k . We have

$$\begin{aligned} (7) \quad \int f \, d\mu + \int g \, d\nu &\leq \int f \, d\mu + \int \hat{g} \, d\nu, \text{ since } \hat{g} \geq g, \\ &\leq \int f \, d\mu + \int \hat{g} \, d\mu, \text{ since } -\hat{g} \text{ is convex nondecreasing, and } \mu \leq_{(2)} \nu, \\ &= \int f + \hat{g} \, d\mu \leq \sup_{x \in \mathbb{R}^k} [f(x) + \hat{g}(x)]. \end{aligned}$$

Now let a be a real number for which

$$(8) \quad a < \sup_{x \in \mathbb{R}^k} [f(x) + \hat{g}(x)].$$

Let us choose a point $s \in \mathbb{R}^k$ for which $a < f(s) + \hat{g}(s)$. By Lemma 2 there exists a measure $\theta \in M_s(\mathbb{R}^k)$ such that

$$a < f(s) + \int g \, d\theta.$$

Then the p.m. λ on $\mathbb{R}^k \times \mathbb{R}^k$ defined by $\lambda(\{s\} \times E) = \theta(E)$ for Borel $E \subset \mathbb{R}^k$ (that is the p.m. giving all mass to the affine set $\{s\} \times \mathbb{R}^k$, with θ determining the distribution of mass on the second component) is in Λ , and further

$$\int f \circ \pi_1 + g \circ \pi_2 \, d\lambda = f(s) + \int g \, d\theta > a.$$

As a was arbitrary subject to (8) we have

$$\sup_{\lambda \in \Lambda} \int f \circ \pi_1 + g \circ \pi_2 \, d\lambda \geq \sup_{x \in \mathbb{R}^k} [f(x) + \hat{g}(x)] \geq \int f \, d\mu + \int g \, d\nu,$$

using (7) in the second inequality. This demonstrates (4), and we are done. ■

Remark. To relate the submartingale (X, Y) whose existence is guaranteed by Theorem 1 to Marshall and Olkin's Theorem 5.A.9, note that their vector u corresponds to our $E(Y | X)$, and their v corresponds to our $Y + X - E(Y | X)$.

It is easy to deduce the corresponding representation for the ordering induced by all convex functions. We have

COROLLARY 1. (Strassen). *Let μ and ν be p.m.s on \mathbb{R}^k with finite means. Then $\mu \leq_{(1)} \nu$ if and only if the pair (μ, ν) has the martingale property.*

PROOF: Sufficiency of the martingale property follows from Jensen's inequality. Conversely, if $\mu \leq_{(1)} \nu$ then clearly $\mu \leq_{(2)} \nu$ also. Thus by Theorem 1 we can find (X, Y) with the correct marginals for which $E(Y | X) \geq X$. But $\mu \leq_{(1)} \nu$ also implies that $E(X) = E(Y)$, since (1) holds for all affine functions f , and this forces $E(Y | X) = X$ a.s. ■

3. MAJORISATION

Let $S = \{x_1, x_2, \dots, x_n\}$ and $T = \{y_1, y_2, \dots, y_n\}$ be subsets of \mathbb{R}^k of the same cardinality. Following Marshall and Olkin let us write $S \prec_w T$ if for every convex nondecreasing function $f: \mathbb{R}^k \rightarrow \mathbb{R}$ we have

$$\sum_{i=1}^n f(x_i) \leq \sum_{i=1}^n f(y_i).$$

Let us write $S \prec T$ if this holds for all convex f . Then we have

COROLLARY 2. *$S \prec_w T$ if and only if there exists a doubly stochastic matrix $P = (p_{ij})$ such that*

$$\sum_{j=1}^n p_{ij}y_j \geq x_i, i = 1, 2, \dots, n.$$

Furthermore, $S \prec T$ if and only if such a P may be found giving equality here.

PROOF: Let μ be the p.m. giving mass $1/n$ to each point of S and let ν be the p.m. giving mass $1/n$ to each point of T . Then the hypothesis $S \prec_w T$ is equivalent to $\mu \leq_{(2)} \nu$. The result is a restatement of the conclusions of Theorem 1 and Corollary 1 in this situation. ■

Since the idea of majorisation is of interest in its own right, and is a finite-dimensional notion, it is somewhat unsatisfactory that in proving Corollary 2 we implicitly used the functional analytic ideas of Lemma 1. We now provide a proof of Lemma 1 in the finite-dimensional case which uses only finite-dimensional methods. It is appropriate to use the language of vectors and matrices now, rather than that of functions and measures. In the following, all vectors are taken to be columns, and T denotes transpose. We denote by e_j the j -vector all of those entries are ones.

THEOREM 2. Let $\mu \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^n$ be nonnegative vectors whose entries sum to one. Let Λ be a closed convex set of nonnegative $m \times n$ matrices whose entries sum to one. Then Λ contains a matrix λ_0 with row sums $\mu = \lambda_0 e_n$ and column sums $\nu^T = e_m^T \lambda_0$ if and only if

$$(9) \quad f^T \mu + g^T \nu \leq \sup_{\lambda \in \Lambda} [f^T \lambda e_n + e_m^T \lambda g]$$

for all $f \in \mathbb{R}^m$ and $g \in \mathbb{R}^n$.

PROOF: Necessity being obvious, we only prove sufficiency. Let Δ be the set of all real $m \times n$ matrices with row and column sums μ and ν^T , respectively, and assume, seeking to contradict (9), that Δ has no element in common with Λ . Since both Λ and Δ are closed and convex, and Λ is compact, there exists, by the (finite-dimensional) separating hyperplane theorem, a linear functional φ for which

$$(10) \quad \inf_{\lambda \in \Delta} \varphi(\lambda) > \sup_{\lambda \in \Lambda} \varphi(\lambda).$$

As Λ is nonempty, such a φ is bounded below on Δ . But since Δ is an affine set, the functional φ can only be bounded below there if it is constant there, and so φ is a function of the row and column sums of its argument only. Thus there exist f and g with

$$\varphi(\lambda) = f^T \lambda e_n + e_m^T \lambda g.$$

But this with (10) contradicts (9), since $\varphi(\lambda) = f^T \mu + g^T \nu$ for $\lambda \in \Delta$, and we are done. ■

This result, along with a proof of Lemma 2 in the case of T a finite set, engenders a proof of Corollary 2 using only finite-dimensional methods.

REFERENCES

- [1] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications* (Academic Press, New York, 1979).
- [2] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, New Jersey, 1970).
- [3] D. Stoyan, *Comparison Methods for Queues and other Stochastic Models* (John Wiley, New York, 1983).
- [4] V. Strassen, 'The existence of probability measures with given marginals', *Ann. Math. Statist.* **36** (1965), 423–439.
- [5] W. Whitt, 'The effect of variability in the GI/G/s Queue', *J. Appl. Probab.* **17** (1980), 1062–1071.

Statistics Division
 Department of Mathematical Sciences
 University of Arkansas
 Fayetteville, Arkansas 72701
 United States of America