Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 228–235

DUAL CONVEX CONES OF ORDER RESTRCITIONS WITH APPLICATIONS[1]

BY RICHARD L. DYKSTRA
University of Iowa

The concept of closed convex cones in finite dimensional Euclidian space and their duals has proven to be a useful construct. Here dual cones are exhibited for specific closed, convex cones including those pertaining to starshaped orderings and concave (convex) functions.

Applications include finding projections involving starshaped orderings, generalizations of Chebyshev's (Kimball's) inequality, an inequality for concave (convex) functions and a characterization of certain kinds of positive dependence.

1. Introduction. Several authors have made extensive use of the concept of convex cones and their duals in $\mathcal{R}^n$. Among these are Rockafellar (1970), Robertson and Wright (1981), and Barlow and Brunk (1972). Here we wish to specifically exhibit certain convex cones and their duals and discuss the implications.

To be precise, we call $K \subset R^n$ a convex cone if (a) $\mathbf{x},\mathbf{y} \in K \Rightarrow \mathbf{x} + \mathbf{Y} \in K$, and (b) $\mathbf{x} \in K$, $a \geqslant 0 \Rightarrow a\,\mathbf{x} \in K$. Of course if K is a convex cone, so is $-K = \{\mathbf{x}: -\mathbf{x} \in K\}$ which we will call the "negative" of K.

Another important convex cone induced by K is the "dual" of K. For a fixed positive vector $\mathbf{w}$, the dual of K is given by

$$K^{\mathbf{w}*} = \{\mathbf{y}:(\mathbf{x},\mathbf{y}) = \Sigma_{i=1}^n x_i y_i w_i \leqslant 0 \text{ for all } \mathbf{x} \in K\}.$$

(Some authors prefer the term "polar" to "dual." Some also define the dual as the negative of our dual.) Of course if K is closed, then $(K^{\mathbf{w}*})^{\mathbf{w}*} = K$. It is evident that if $K_1 \subset K_2$, then $K_1^{\mathbf{w}*} \supset K_2^{\mathbf{w}*}$.

New convex cones can be formed from existing cones in several ways. Two important methods are through intersections and direct sums.

If the closed, convex cones $K_1, \ldots, K_n$ are sufficiently nice (say finitely generated), the direct sum $\Sigma_{i=1}^n K_i = \{\Sigma_{i=1}^n \mathbf{x}_i | \mathbf{x}_i \epsilon K_i,\ i=1, \ldots, n\}$ is also a closed, convex cone. However, in general the closure property is not guaranteed (see Hestenes (1975), pp. 196–198). Nevertheless, intersections and direct sums of closed, convex cones are closely related because it is always true that $(\Sigma_{i=1}^n K_i)^{\mathbf{w}*} = \bigcap_{i=1}^n K_i^{\mathbf{w}*}$ and

$$(\bigcap_{i=1}^n K_i)^{\mathbf{w}*} = \Sigma_{i=1}^n K_i^{\mathbf{w}*} \tag{1.1}$$

if the latter cone is closed. This is guaranteed if the relative interiors of the K_i have a point in common (see Rockafellar (1970), p. 146) or, as we said, if the $K_i^{\mathbf{w}*}$ are finitely generated. (1.1) is equivalent to the well-known Farkas' Lemma if the K_i are generated by a single vector.

An important cone, especially in the area of isotone regression, is the cone of vectors which are nondecreasing, i.e.

$$K_I = \{\mathbf{x} | x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n\}. \tag{1.2}$$

The dual cone here, as discussed in Barlow and Brunk (1972), is

---

[1] This work was partially supported by ONR Contract N00017–78–C–0655.

AMS 1980 subject classifications. Primary 52A40, secondary 52A20.

Key words and phrases: Convex cones, dual cones, polars, projections, starshaped orderings, concave (convex) functions, Chebyshev's (Kimball's) Inequality, positively dependent random variables, order restrictions.

(1.3) $$K_I^{\mathbf{w}*} = \{\mathbf{y} | \Sigma_{j=1}^{i} y_j w_j \geq 0, i = 1, \ldots, n-1, \Sigma_{j=1}^{n} y_j w_j = 0\}.$$

We note in passing that the important concept of majorization as discussed extensively in Marshall and Olkin (1979) is closely connected with the cone in (1.3). If the vectors $\mathbf{x}$ and $\mathbf{y}$ are each ordered from largest to smallest to form $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, $\mathbf{x}$ majorizes $\mathbf{y}$ iff, $\bar{\mathbf{x}} - \bar{\mathbf{y}} \in K_I^{\mathbf{l}*}$. (We let $\mathbf{l}$ denote a vector containing all 1's.) Further discussion of such cone orderings is given in Marshall, Walkup, and Wets (1967).

If the cone specified in (1.2) is modified to require that it contain only nonnegative vectors, i.e., $K = \{\mathbf{x} | 0 \leq x_1 \leq x_2 \leq \ldots \leq x_n\}$, the dual is equivalent to that given in (1.3) with a modification of the last equality. In this case,

$$K^{\mathbf{w}*} = \{\mathbf{y} | \Sigma_{i=1}^{n} y_j w_j \geq 0, i = 1, \ldots, n\}.$$

Much of our interest in dual cones hinges on a duality result discussed in Barlow and Brunk (1972). In particular if g^* solves the problem

(1.4) $$\underset{\mathbf{x} \in K}{\text{Minimize}} \; \Sigma_{j=1}^{n} (g_i - x_i)^2 w_i$$

where K is a closed convex cone, then $\mathbf{g} - \tilde{\mathbf{g}}^*$ solves

(1.5) $$\underset{\mathbf{x} \in K^{\mathbf{w}*}}{\text{Minimize}} \; \Sigma_{i=1}^{n} (g_i - x_i)^2 w_i.$$

Robertson and Wright (1980) make extensive use of this duality in dealing with stochastic ordering restrictions for multinomial parameters. This duality is also important in deriving distributional theory, i.e., see Robertson and Wegman (1978).

2. The Starshaped Ordering. An interesting order restriction is that a vector be starshaped. Shaked (1979) defines a vector $\mathbf{x}$ to be lower (upper) starshaped with respect to the positive weights $\mathbf{w}$ if $\bar{x}_1 \geq \bar{x}_2 \geq \ldots \geq \bar{x}_n \geq 0$ $(0 \leq \bar{x}_1 \leq x_2 \ldots \leq \bar{x}_n)$ where

(2.1) $$\bar{x}_i = \Sigma_{j=1}^{i} x_j w_j / \Sigma_{j=1}^{i} w_j.$$

Shaked is concerned with finding maximum likelihood estimates of Poisson and normal means which must satisfy starshaped restrictions.

Dykstra and Robertson (1982) use the term "decreasing (increasing) on the average" when the nonnegativity restrictions in (2.1) are omitted, and are concerned with such restrictions when testing for trend.

Surprisingly the dual cone of "increasing on the average" vectors is closely associated with the cone of "decreasing on the average" vectors.

THEOREM 2.1. *If* $K_{IA} = \{\mathbf{x} | \bar{x}_1 \leq \bar{x}_2 \leq \ldots \leq \bar{x}_n\}$, *then* $K_{IA}^{\mathbf{w}*} = \{\mathbf{y} | \bar{y}_1 \geq \bar{y}_2 \geq \ldots \geq \bar{y}_n = 0\}$.

Proof. Note that we can write

$$K_{IA} = \{\mathbf{x} | \bar{x}_i - \bar{x}_{i+1} \leq 0, i = 1, \ldots, n-1\} = \bigcap_1^{n-1} K_i$$

where

(2.2) $$K_i = \{\mathbf{x} | \bar{x}_1 - \bar{x}_{i+1} \leq 0\}.$$

Now we claim that

(2.3) $$H_i = \{\mathbf{y} | 0 \leq y_1 = y_2 = \ldots = y_i, \Sigma_1^{i+1} y_j w_j = 0, y_j = 0, j > i+1\}$$

is actually $K_i^{\mathbf{w}*}$. If $\mathbf{y} \in H_i$, then $y_{i+1} = -y_1 W_i w_{i+1}^{-1}$ where $W_i = \Sigma_1^i w_j$.

If $\mathbf{x} \in K_i$ and $\mathbf{y} \in H_i$, then

$$(\mathbf{x},\mathbf{y}) = \Sigma_1^n x_j y_j w_j = y_1[\Sigma_1^n x_j w_j - x_{i+1} W_i] \leqslant 0$$

by (2.2) and (2.3). Since $H_i^{\mathbf{w}*}$ is clearly K_i, we have that $H_i = K_i^{\mathbf{w}*}$.

Since from (1.1)

$$(\bigcap_{i=1}^{n-1} K_i)^{\mathbf{w}*} = \Sigma_{i=1}^{n-1} K_i^{\mathbf{w}*},$$

we need to show that $\Sigma_{i=1}^{n-1} K_i^{\mathbf{w}*} = \{\mathbf{y}|\bar{y}_1 \geqslant \bar{y}_2 \geqslant \ldots \geqslant \bar{y}_n = 0\}$.

First assume $\mathbf{x}_i \in K_i^{\mathbf{w}*}$, $i = 1, \ldots, n-1$. Then we may write

$$\mathbf{x}_1 = (x_1, -x_1 W_1 w_2^{-1}, 0, \ldots, 0) \qquad (x_1 \geqslant 0)$$
$$\mathbf{x}_2 = (x_2, x_2, -x_2 W_2 w_3^{-1}, 0, \ldots, 0) \qquad (x_2 \geqslant 0)$$
$$\ldots$$
$$\mathbf{x}_{n-1} = (x_{n-1}, x_{n-1}, \ldots, x_{n-1}, -x_{n-1} W_{n-1} w_n^{-1}) \qquad (x_{n-1} \geqslant 0)$$

After adding coordinates we see that

$$(\Sigma_{i=1}^{n-1} \mathbf{x}_j)_i - (\Sigma_{i=1}^{n-1} \mathbf{x}_j)_{i+1} = x_i \geqslant 0, i = 1, \ldots, n-1$$

and $(\Sigma_{i=1}^{n-1} \mathbf{x}_j)_n = 0$. Thus $\Sigma_{i=1}^{n-1} K_i^{\mathbf{w}} \subset \{\mathbf{y}: \bar{y}_1 \geqslant \bar{y}_2 \geqslant \ldots \geqslant 0\}$.

Conversely, consider $\mathbf{y} = (y_1, \ldots, y_n)$ such that $\bar{y}_1 \geqslant \bar{y}_2 \geqslant \ldots \geqslant \bar{y}_n = 0$. Recalling that $w_i = \Sigma_1^i W_j$, we partition $\mathbf{y}$ as follows:

$$\mathbf{x}_1 = (-w_2 W_1^{-1} z_1, z_1, 0, \ldots, 0)$$
$$\mathbf{x}_2 = (-w_3 W_2^{-1} z_2, -w_3 W_2^{-1} z_2, z_1, 0, \ldots, 0)$$
$$\ldots$$
$$\mathbf{x}_{n-1} = (-w_n W_{n-1}^{-1} z_{n-1}, \ldots, -w_n W_{n-1}^{-1} z_{n-1}, z_{n-1})$$

where

$$\mathbf{z}_{i-1} = y_i + W_i^{-1} \Sigma_{i+1}^n y_j w_j.$$

It can be verified that the i-th column of the above array sums to y_i and that each row is such that $\Sigma_1^n x_{ij} w_j = 0$.

Finally we note that

$$\bar{y}_{i-1} \geqslant \bar{y}_i \Longleftrightarrow W_i \Sigma_1^{i-1} y_j w_j \geqslant w_{i-1} \Sigma_1^i y_j w_j \Longleftrightarrow \Sigma_1^{i-1} y_j w_j \geqslant W_{i-1} y_i.$$

Therefore

$$0 = W_i^{-1}[\Sigma_1^{i-1} y_j w_j + \Sigma_i^n y_j w_j] \geqslant W_i^{-1}[W_{i-1} y_i + \Sigma_i^n y_j w_j]$$
$$= y_i + W_i^{-1} \Sigma_{i+1}^n y_j w_j = z_{i-1},$$

so that $-\mathrm{w}_{i+1} W_i^{-1} z_i \geqslant 0$, and hence $\mathbf{x}_i \in K_i^{\mathbf{w}*}$. Thus we have that

$$\{\mathbf{y}|\bar{y}_1 \geqslant \bar{y}_2 \geqslant \ldots \geqslant \bar{y}_n = 0\} \subset K_1^{\mathbf{w}*} + \ldots + K_{n-1}^{\mathbf{w}*}$$

so that equality holds. □

The dual cones of lower and upper starshaped vectors discussed by Shaked (1979) can also be found. First we handle the lower starshaped vector.

COROLLARY 2.2. *If* $K_{LS} = \{\mathbf{x}|\bar{x}_1 \geqslant \bar{x}_2 \geqslant \ldots \geqslant \bar{x}_n \geqslant 0\}$, *then*

$$K_{LS}^{\mathbf{w}*} = \{\mathbf{y}|\bar{y}_1 \leqslant \bar{y}_2 \leqslant \ldots \leqslant \bar{y}_n \leqslant 0\}.$$

(Note that this dual also has the property that $K_{LS}^{\mathbf{w}*} = -K_{LS}$).

Proof. Note that

$$K_{LS} = K_{DA} \quad \{\mathbf{x}|\Sigma_{i=1}^n x_j w_j \geqslant 0\}.$$

Since the dual of this last cone is

$$\{\mathbf{y}|y_1 = y_2 = \ldots = y_n \leqslant 0\}, \tag{2.4}$$

the identity in (1.1) implies that $K^{\mathbf{w}*}_{LS}$ is the direct sum of $K^{\mathbf{w}*}_{DA}$ and the cone in (2.4). This can be shown to be the desired cone. □

The dual cone of the upper starshaped vectors is not quite as elegant.

COROLLARY 2.3. *If* $K_{US} = \{\mathbf{x} | 0 \leqslant \bar{x}_1 \leqslant \bar{x}_2 \leqslant \ldots \leqslant \bar{x}_n\}$, *then* $K^{\mathbf{w}*}_{US} = \{\mathbf{y} | y_{i+1} - \bar{y}_i \leqslant (\Sigma_1^i w_j)^{-1} \Sigma_{j=1}^n y_j w_j \leqslant 0$ *for* $i = 1, \ldots, n-1\}$.

Proof. The proof follows by writing

$$K_{US} = K_{IA} \bigcap \{\mathbf{x}: x_1 \geqslant 0\},$$

recognizing that

$$\{\mathbf{x} | x_1 \geqslant 0\}^{\mathbf{w}*} = \{\mathbf{y} | y_1 \leqslant 0, y_2 = y_3 = \ldots = y_n = 0\}$$

and using (1.1) and Theorem 2.1. □

3. The Concave Ordering. A frequently occurring closed convex cone in $\mathcal{R}^n$ is the class of concave (convex) functions $K_{CC}(K_{CV})$ defined on the set of real numbers $\{x_1, \ldots, x_n\}$. Thus a point $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{R}^n$ is interpreted as the function whose image of x_i is y_i. If we let $\Delta y_i = y_{i+1} - y_i$ and $\Delta x_i = x_{i+1} - x_i$, we can write $K_{CC} = \bigcap_{i=1}^{n-2} H_i$ where $H_i = \{\mathbf{y} | \Delta y_i / \Delta x_i \geqslant \Delta y_{i+1} / \Delta x_{i+1}\}$. The dual cone of $K_{CC}(K_{CV})$ is surprisingly tractable.

THEOREM 3.1. *The dual cone of the set of concave functions on* $\{x_1, \ldots, x_n\}$ *is given by*

$$K^{\mathbf{w}*}_{CC} = \left\{\mathbf{z} \middle| \Sigma_{i=1}^n z_i w_i = 0, \Sigma_{i=1}^{n-l-1} (x_{n-l} - x_i) z_i w_i \begin{Bmatrix} =0, l=0 \\ \geqslant 0, l=1,2,\cdots,n-2 \end{Bmatrix} \right\}.$$

Our proof proceeds similarly to Theorem 2.1 and is not given. The theorem is closely related to a result of Brunk (1956).

4. Applications. Of course by their very definitions, a convex cone K and its dual $K^{\mathbf{w}*}$ give rise to natural inequalities. In particular, if $\mathbf{x} \in K$ and $\mathbf{y} - \mathbf{z} \in K^{\mathbf{w}*}$, then

$$\Sigma_1^n x_j (y_j - z_j) w_j \leqslant 0. \tag{4.1}$$

This has some straightforward implications in terms of sample covariances by taking $\mathbf{w} = \mathbf{1}$.

COROLLARY 4.1. *Suppose* $\mathbf{x}, \mathbf{y}$ *and* $\mathbf{z}$ *are vectors in* $\mathcal{R}^n$. *If*

$$i^{-1} \Sigma_{j=1}^i x_j \geqslant (i+1)^{-1} \Sigma_{j=1}^{i+1} x_j, \; i = 1, \ldots, n-1, \tag{4.2}$$

and

$$i^{-1} \Sigma_{j=1}^i (y_j - z_j) \geqslant (i+1)^{-1} \Sigma_{j=1}^{i+1} y_j - z_j, \; i = 1, \ldots, n-1, \tag{4.3}$$

then the sample covariance of $(\mathbf{x}, \mathbf{y})$ *is at least as large as the sample covariance of* $(\mathbf{x}, \mathbf{z})$.

Proof. Condition (4.2) states that $\mathbf{x} \in K_{DA}$. Condition (4.3) implies that $\mathbf{z} - \mathbf{y} \in K_{IA}$ which is equivalent to saying $(\mathbf{z} - \mathbf{y}) - (\bar{\mathbf{z}} - \bar{\mathbf{y}}) \in K^{\mathbf{1}*}_{DA}$ (where $\bar{\mathbf{a}} = (\bar{a}, \bar{a}, \ldots, \bar{a})$). Thus

$$\Sigma_1^n (x_i - \bar{x})(z_i - \bar{z}) = \Sigma_1^n x_i (z_i - \bar{z}) \leqslant \Sigma_1^n x_i (y_i - \bar{y}) = \Sigma_1^n (x_i - \bar{x})(y_i - \bar{y}). \quad \square$$

Of course if $\mathbf{z} = \mathbf{0}$, this result is equivalent to saying that if $\mathbf{x}, \mathbf{y} \in K_{DA}(K_{IA})$ then $(\mathbf{x}, \mathbf{y}) \geqslant n\bar{x}\bar{y}$. Of course since $K_{DA} = -K_{IA}$, if $x \in K_{IA}$ and $y \in K_{DA}$ (or vice versa) $(\mathbf{x}, \mathbf{y}) \leqslant n\bar{x}\bar{y}$. These inequalities are as strong as possible in the sense if $\mathbf{x} \notin K_{DA}(K_{IA})$, one can find a $\mathbf{y} \in K_{DA}(K_{IA})$ such that $(\mathbf{x}, \mathbf{y}) < (>) n\bar{x}\bar{y}$. Note that the above inequality generalizes the well known result for nondecreasing (nonincreasing) vectors.

Another application concerns Shaked's paper (1979). In this paper Shaked wants to find a weighted least squares projection of say $\mathbf{g}$ onto the cone K_{LS}. However Shaked actually finds the projection, say $\mathbf{g}^*$, onto the cone K_{DA} and hopes that $\mathbf{g}^*$ is in K_{LS} (in which case $\mathbf{g}^*$ is also the projection onto K_{LS}). However, if $\mathbf{g}^*$ is *not* in K_{LS}, i.e., $\Sigma_1^n g_j^* w_j < 0$, one can say that the true projection $\hat{\mathbf{g}}$ has the property that $\Sigma_1^n \hat{g}_j w_j = 0$ (see page 89, Barlow, Bartholomew, Bremner and Brunk (1972)). In this case, we know that $\hat{\mathbf{g}}$ must be the projection onto the dual of K_{IA}.

In this event (see (1.5)), $\hat{\mathbf{g}} = \mathbf{g} - \tilde{\mathbf{g}}$ where $\tilde{\mathbf{g}}$ is the projection of g onto K_{IA} which is a problem that Shaked also solves. From Shaked's solution we can verify that $\mathbf{g} - \tilde{\mathbf{g}} = \mathbf{g}^* \cdot \bar{\mathbf{g}}$. Thus the projection onto K_{LS} is given by

$$\hat{\mathbf{g}} = \begin{cases} \mathbf{g}^*, \text{ if } \Sigma_1^n g_j^* w_j \geq 0 \\ \mathbf{g}^* - \bar{\mathbf{g}}, \text{ if } \Sigma_1^n g_j^* w_j < 0. \end{cases}$$

A useful inequality attributed to Chebyshev (see Hardy, Littlewood and Pólya (1959, p. 43)) and discussed and generalized in various places such as Horn (1979), Kimball (1951), and Dykstra, Hewett and Thompson (1973) concerns the expected value of a product of monotone functions of a random variable. Thus, for example, if f, g are nondecreasing (nonincreasing) functions,

$$Ef(X) \cdot g(X) \geq Ef(X) \cdot Eg(X) \tag{4.4}$$

assuming the expectations are defined. We can develop similar types of inequalities based upon closed convex cones and their duals.

COROLLARY 4.2. *If f, g are real valued functions in the class*

$$A_X = \{f: f(X) \textit{ is integrable}, E[f(X)I_{[X \leq x]}]/P(X \leq x) \\ \textit{is nondecreasing over } \{x: P(X \leq x) > 0\}\},$$

then

$$Ef(X)g(X) \geq Ef(X) \cdot Eg(X).$$

Proof. Suppose first that X is finitely discrete on the set $\{x_1, \ldots, x_n\}$. If we let $\mathbf{w} = (w_1, \ldots, w_n)$ where $w_i = P(X = x_i)$, then the condition that $f \in A_X$ is equivalent to saying

$$(f(x_1), f(x_2), \ldots, f(x_n)) \in K_{IA}.$$

If $g \in A_X$, $Eg(X) - g$ must belong to $K^{\mathbf{w}*}_{IA}$ and the result follows.

In the general case, we let $x_{n,j}$, $j = 0, \ldots, k(n)$ be a series of nested partitions covering the support of X which generate the Borel sets in the support of X. We define

$$f_n(X) = \Sigma_{i=1}^{k(n)} E(f(X)I_{A_{n,i}}(X)) \cdot I_{A_{n,i}}(x)/w_{n,i}$$
$$g_n(X) = \Sigma_{i=1}^{k(n)} E(g(X)I_{A_{n,i}}(x)) I_{A_{n,i}}(X)/w_{n,i}$$

where $A_{n,i} = (x_{n,i-1}, x_{i,i}]$ and $w_{n,i} = P(X \in A_{n,i})$. (We take $x_{n,0} = -\infty$.)

Viewing $f_n(X)$ and $g_n(X)$ as conditional expectations, we can use Theorem 5.21 of Breiman (1968) to argue that $f_n(X) \xrightarrow[\text{a.s.}]{L_1} f(X)$ and $g_n(X) \xrightarrow[\text{a.s.}]{L_1} g(X)$.

We have from the first part of the proof that

$$Ef(X)\,Eg(X) = Ef_n(X)\,\mathrm{E}g_n(X) \leq Ef_n(X)g_n(X) \quad \text{for all } n.$$

Therefore if f is bounded above, by Fatou's lemma,

$$\begin{aligned} Ef(X)Eg(X) &\leq \limsup Ef_n(X) \cdot g_n(X) \\ &\leq E \limsup f_n(X)g_n(X) = Ef(X)g(X). \end{aligned} \tag{4.5}$$

Finally, noting that if $h \in A_X$, so does $\min\{h, c\}$ for any positive constant c, we have the desired result for $\min\{f, c\}$ and $\min\{g, c\}$. Note that (4.5) guarantees that $E[f(X)g(X)^-] < \infty$. If $E[f(X)g(X)^+] = \infty$, the desired result clearly holds, so we may assume that $f(X)g(X)$

is integrable. Finally, letting $c \to \infty$ and using the Dominated Convergence Theorem on each side concludes the proof. □

We can obtain similar type inequalities by working with other cones and their duals. For example, we can establish the following corollary which is closely related to the basic lemma of Marshall and Proschan (1970).

COROLLARY 4.3. *If f is a real-valued nondecreasing function with $f(X)$ integrable and g is a real-valued function in the class*

$$B_x = \{g: g(X) \text{ is integrable}, E[g(X)I_{(X \leq x)}] \leq Eg(X) \quad \text{for all } x\},$$

then

$$Ef(X)g(X) \geq Ef(X)g(X).$$

The proof follows the lines of Corollary 4.2 and is not given.

Note that if we define the class of real-valued functions

$$C_X = \{g: g(X) \quad \text{is integrable and } g \text{ is nondecreasing}\},$$

then $C_X \subset A_X \subset B_X$. Thus both Corollary 4.2 and Corollary 4.3 generalize the Chebyshev inequality (4.4). The results of this section enable us to obtain some insight into certain types of positive dependence as discussed in Lehmann (1966) and elsewhere.

Let us say that the random variables (X, Y) satisfy the following kinds of positive dependence: (1) Type I if $P[X \leq x, Y \leq y] \geq P[X \leq x]P[Y \leq y]$ for all x, y, (2) Type II if $P[Y \geq y|X \leq x]$ is nondecreasing in x for all y, and (3) Type III if $P[Y \geq y|X = x]$ is nondecreasing in x for all y. Assuming that all quantities are defined, each of the above types of dependence can be characterized by the inequality

$$Ef(X) \cdot g(X) \geq Ef(X) \cdot Eg(X) \tag{4.6}$$

as shown in the following Theorem.

THEOREM 4.1. *Assume $g \in C_y$. Then (X, Y) exhibits Type I, II, or III dependence iff (4.6) holds for all $f \in C_X$, A_X, or B_X respectively.*

Proof. The result for Type I dependence is handled in Lehmann (1966). For Type II, let $h(t) = P[Y \geq y|X = t]$. Then $h \in A_X$ iff

$$E[P\{Y \geq y|X\}I_{(X \leq x)}]/P[X \leq x] = P[Y \geq y|X \leq x]$$

is nondecreasing in x. Thus if f also belongs to A_X, we have by Corollary 4.2

$$\begin{aligned} E[f(X)h(X)] &= Ef(X) \cdot I_{(Y \geq y)} \\ &\geq Ef(X) \cdot P(Y \geq y), \quad \text{for all } y. \end{aligned} \tag{4.7}$$

Thus

$$Ef(X)\Sigma a_i I_{(Y \geq y)} \geq Ef(X)\Sigma a_i P(Y \geq y_i)$$

for all nonnegative a_i. A passage to the limit will imply the desired result for a nonnegative, nondecreasing g in C_Y from which the result follows. If $P[Y \geq y|X \leq x]$ is *not* nondecreasing in x, then $h \notin A_X$ which implies there is an $f \in A_X$ such that (4.7) does not hold.

The case of Type III dependence is handled similarly. □

We note that while Type I dependence is symmetric in X and Y, Types II and III are not as is evident from our characterizations. In some sense, the size of the sets C_X, A_X, and B_X is a measure of the relative strengths of the dependence relations.

We can use the dual cones derived in section 3 to obtain inequalities for concave (convex) functions somewhat similar to those given in Corollary 4.2. To set some notation, we note that if the random variables X and $f(X)$ are square integrable, then the linear function of X which is closest to $f(X)$ in the sense of minimizing $E(f(X)-(aX+b))^2$ is given by $l_f(X) = a_fX+b_f$ where

$$(4.8) \qquad a_f = E(Xf(X)) - E(X)Ef(X)/\sigma_X^2,\ b_f = Ef(X) - a_fE(X).$$

It is well known that $Ef(X) = E\ l_f(X)$ and $EXf(X) = EX\ l_f(X)$. Interestingly, if f and g are both concave (convex) functions such that $f(X)$ and $g(X)$ are integrable, then replacing $f(X)$ and/or $g(X)$ by their linear approximations can only decrease the expected value of the product. We begin with a more general result for discrete random variables.

COROLLARY 4.4. *If the random variable X is finitely discrete (on the values $x_1 < x_2 < \ldots < x_n$), f is concave on the range of X and g is such that* (1) $Eg(X) = 0$, (2) $EXg(X) = 0$, (3) $E(x-X)g(X)I_{(X<x)} \geq 0$ *for all x in the support of X, then $Ef(X)g(X) \leq 0$.*

Proof. The proof follows directly from Theorem 3.1 by letting $\mathbf{w_i} = P(X = x_i)$. □

An important class of functions which satisfies the above conditions is given in the following theorem.

THEOREM 4.2. *If $g(x)$ is convex then $g(x) - (a_gx + b_g)$ (as defined in 4.8) satisfies conditions 1), 2) and 3) of Corollary 4.4.*

Proof. The proof is trivial if g is linear so assume that it is not. It is easily shown that conditions 1) and 2) hold so we consider condition 3). Now by the convexity assumption, $g(x) - (a_gx + b_g)$ must be positive, negative and positive again. Thus $\Sigma_{j=1}^{i} g(x_j) - (a_gx_j+b_g)$ must first be nonnegative and then nonpositive as i increases from 1 to n. Thus $g(x) - (a_gx + b_g)$ is in the cone K_I^{w*} (see 1.3) for the weights $w_i = P(X = x_i)$. Since for each $i, h(x_j) = \sup\{x_i - x_j, 0\}$ is in $-K_I$ (see (1.2)), condition 3) must hold by the definition of dual convex cones. □

This leads to the following corollary which also holds for the continuous case. Note that b) is similar to the Chebyshev inequality (4.4) with monotonicity replaced by concavity (convexity).

COROLLARY 4.5. *If f and g are both concave (convex) functions such that $X, f(X)$ and* $g(X)$ are all square integrable, then (a)

$$Ef(X)g(X) \geq Ef(X)(a_gX+b_g) = E(a_fX+b_f)(a_gX+b_g).$$

Moreover, if $EXf(X) - EXEf(X)$ and $EXg(X) - EXEg(X)$ have the same sign, then (b) $Ef(X)g(X) \geq Ef(X)\,Eg(X)$.

Proof. The first inequality follows by considering finer and finer partitions of the support of X, noting that f and g are concave on the partition points, and employing Theorem 4.2 and Corollary 4.4 together with limiting arguments. The equality in (a) follows from $a_gx + b_g$ being both concave and convex. Inequality (b) then follows from Chebyshev's inequality on the last part of a). □

Acknowledgement. The author would like to thank Tim Robertson for helpful discussions on the subject of convex cones.

REFERENCES

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M., and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.

BARLOW, R. E. and BRUNK, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc.* *67* 140–147.

BREIMAN, LEO (1968). *Probability*. Addison Wesley, Reading, Mass.

BRUNK, H. D. (1956). On an inequality for convex functions. *Proc. Amer. Math. Soc.* *7* 817–824.

DYKSTRA, RICHARD L. and ROBERTSON, TIM (1982). Order restricted statistical tests on multinomial and Poisson Parameters: the starshaped restriction. *Ann. Statist.* *10* 1246–1252.

HARDY, G. H., LITTLEWOOD, J. E., and PÓLYA, G. (1959). *Inequalities,* 2nd ed. Cambridge Univ. Press, London and New York.

HESTENES, MAGNUS R. (1975). *Optimization Theory,* Wiley, New York.

HORN, M. (1979). Some inequalities for the expectation of a product of functions of a random variable and for the multivariate distribution function at a random point. *Biometrical J.* *21* 243–245.

KIMBALL, A. W. (1951). On dependent tests of significance in the analysis of variance. *Ann. Math. Statist.* *22* 600–602.

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* *37* 1137–1153.

MARSHALL, ALBERT W. and OLKIN, INGRAM (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

MARSHALL, ALBERT W. and PROSCHAN, FRANK (1970). Mean life of series and parallel systems. *J. Appl. Probab.* *7* 165–174.

MARSHALL, A. W., WALKUP, D. W., and WETS, R. J. B. (1967). Order preserving functions: Applications to majorization and order statistics. *Pacific J. Math.* *23* 569–584.

ROBERTSON, TIM and WEGMAN, E. J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* *6* 485–505.

ROBERTSON, TIM and WRIGHT, F. T. (1981). Likelihood ratio tests for and against a stochastic ordering between multinomial populations. *Ann. Statist.* *9* 1248–1257.

ROCKAFELLAR, A. TYRELL (1970). *Convex Analysis*. Princeton University Press, Princeton, New Jersey.

SHAKED, MOSHE (1979). Estimation of starshaped sequences of Poisson and normal means. *Ann. Statist.* *7* 729–741.