UNIVERSITY OF OSLO
Department of Informatics

Majorization, polyhedra and statistical testing problems.

Geir Dahl

Report 240,
ISBN 82-7368-155-6

February 1997

Majorization, polyhedra and statistical testing problems.

Geir Dahl *

February 1997

Abstract

There are important connections between majorization and convex polyhedra. Both weak majorization and majorization are preorders related to certain simple convex cones. We investigate the facial structure of a polyhedral cone C associated with a layered directed graph. A generalization of weak majorization based on C is introduced. It defines a preorder of matrices. An application in statistical testing theory is discussed in some detail.

Keywords: Majorization, polyhedra, cone ordering, statistical testing.

1 Introduction

In this paper we study some problems related to majorization. For $\mathbf{z} \in \mathbb{R}^n$ we let $z_{[j]}$ denote the jth largest number among the components of $\mathbf{z}$. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ one says that $\mathbf{x}$ is *weakly majorized* by $\mathbf{y}$, denoted by $\mathbf{x} \prec_w \mathbf{y}$, provided that

$$\sum_{j=1}^{k} x_{[j]} \leq \sum_{j=1}^{k} y_{[j]} \quad \text{for } k = 1, \ldots, n.$$

If, in addition, equality holds for $k = n$, then $\mathbf{x}$ is majorized by $\mathbf{y}$ and we write $\mathbf{x} \prec \mathbf{y}$. Both $\prec$ and $\prec_w$ are preorderings that reflect how "spread out" the components of the vectors are. These concepts play an important role in different areas of mathematics and statistics, see [10], [1] and other papers

*Institute of Informatics, University of Oslo, P.O.Box 1080, Blindern, 0316 Oslo, Norway (Email:geird@ifi.uio.no)

in the special issue of *Linear Algebra and Its Appl.* (Volume 199, 1994) in honor of I. Olkin.

There are interesting (convex) polyhedra that are related to (weak) majorization. This applies to the polytope Ω_n of $n \times n$ doubly-stochastic matrices as one has the well-known characterization of Hardy-Littlewood-Polya: $\mathbf{x} \prec \mathbf{y}$ if and only if there exists an $\mathbf{S} \in \Omega_n$ with $\mathbf{x} = \mathbf{S}\mathbf{y}$. For a given majorization, say $\mathbf{x} \prec \mathbf{y}$, the polytope $\Omega(\mathbf{x} \prec \mathbf{y})$ consisting of all $\mathbf{S} \in \Omega_n$ satisfying $\mathbf{x} = \mathbf{S}\mathbf{y}$ was studied in [5]. Several combinatorial properties of this polytope were established. As another example, [6] and [7] contains a study of the concept of weak k-majorization from a polyhedral point of view. Both $\prec_w$ and $\prec$ are *cone orderings* corresponding to certain simple polyhedral cones (see [10]). For instance, consider the convex cone $\mathcal{D}_+^n$ of nonincreasing and nonnegative vectors

$$\mathcal{D}_+^n = \{\mathbf{x} \in \mathbb{R}^n : x_1 \geq \ldots \geq x_n \geq 0\}.$$

Assume that both the vectors $\mathbf{x}$ and $\mathbf{y}$ are nonincreasing. Then $\mathbf{x}$ is weakly majorized by $\mathbf{y}$ if and only if $\mathbf{y} - \mathbf{x}$ lies the polar cone of $\mathcal{D}_+^n$ (consisting of the vectors $\mathbf{z}$ satisfying $\mathbf{w}^T\mathbf{z} \geq 0$ for all $\mathbf{w} \in \mathcal{D}_+^n$). Now, due to the simplicity of the cone $\mathcal{D}_+^n$ one can explicitly determine a finite set of generators (a "frame") so that $\mathcal{D}_+^n$ is the set of nonnegative linear combinations of these generators. From this fact one may derive nice characterizations of Schur-convex functions and also different characterizations of weak majorization.

In this paper we study from a polyhedral point of view a class of convex cones that contains $\mathcal{D}_+^n$ as a special case. The motivation comes from statistical testing theory, and this application is discussed in some detail. The paper is organized as follows. In section 2 we introduce a polyhedral cone C associated with "layered" directed graphs. In a special case C consists of nonnegative $m \times n$ matrices where no entry is smaller than any entry in a preceding row. Thus, for $n = 1$ we have $C = \mathcal{D}_+^m$. The faces of C are studied (in the general case) and characterized by means of certain partitions in the graph. In section 3 we relate C to weak majorization and introduce a new preorder based on C. The application in statistical testing theory is presented in section 4. It concerns optimal tests for certain testing problems in discrete experiments.

We describe our notation. S_n denotes the group of permutations on n elements and K_n is the standard simplex in $\mathbb{R}^n$, i.e., $K_n = \{\mathbf{x} \in \mathbb{R}_+^n : \sum_{j=1}^n x_j = 1\}$. For a finite set V we let $\mathbb{R}^V$ denote the vector space of real valued functions from V to $\mathbb{R}$. $\mathbf{0}$ denotes the vector with all components being zero. If $S \subseteq V$ the vector χ^S is the incidence vector of S (so χ_v^S equals 1 if $v \in S$ and 0 otherwise) and we also define $\mathbf{x}(S) = \sum_{v \in S} x_v$ for $\mathbf{x} \in \mathbb{R}^V$. If

$A \subseteq \mathbb{R}^V$ the convex hull (conical hull) of A is denoted by $\text{conv}(A)$ ($\text{cone}(A)$). The relative interior of a convex set C is denoted by $\text{rint}(C)$. For polyhedral theory, we refer to [4], [12], [14]. Some graph terminology is used, but is is fairly standard.

2 A cone of row-ordered vectors

Let n_i for $i = 1, \ldots, m$ be given positive integers. Let $R_i = \{(i,j) : 1 \leq j \leq n_i\}$ for $i = 1, \ldots, m$ and define the index set (or node set) $V = R_1 \cup \ldots \cup R_m$. Each set R_i is called a *row*. We let $D = (V, E)$ denote the directed graph with node set V and with an arc from each node in the row R_i to each node in the next row R_{i+1} for $i = 1, \ldots, m-1$. Thus, $E = \{(u,v) : u \in R_i,\ v \in R_{i+1} \text{ for some } i \leq m-1\}$. D is a layered digraph. In the special case where $n_1 = \ldots = n_m = n$, the nodes correspond to the entries (or indices) of an $m \times n$-matrix. We shall study certain problems in the vector space $\mathbb{R}^V$ of real valued functions from V to the reals. In the "matrix special case" above this vector space may be identified with $\mathbb{R}^{m,n}$.

We are interested in the polyhedral cone $C \subset \mathbb{R}^V$ consisting of the vectors $\mathbf{x} \in \mathbb{R}^V$ that satisfy the following set of homogeneous linear inequalities

$$\begin{array}{lll} \text{(i)} & x_u \geq x_v & \text{for all } (u,v) \in E; \\ \text{(ii)} & x_v \geq 0 & \text{for all } v \in V. \end{array} \tag{1}$$

Thus a nonnegative vector $\mathbf{x} \geq \mathbf{0}$ lies in C iff no component of $\mathbf{x}$ is smaller than a component in the next row. The cone C is full dimensional. If $\mathbf{x} \in C$ we say that $\mathbf{x}$ is *row-ordered*.

Our main task in this section is to study the facial structure of C. The faces and, in particular, the extreme rays of C are of interest in two different contexts in the subsequent sections; majorization and statistical testing.

The faces of C are related to partitions of V as discussed in the following. Consider a partition $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ of V where the sets $N_1, \ldots, N_p$ are nonempty while N_0 may be empty. (A more concise notation would be $\mathcal{N} = (N_0, \{N_1, \ldots, N_p\})$). We say that partitions $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ and $\mathcal{M} = \{M_0, M_1, \ldots, M_q\}$ are *equal* and write $\mathcal{N} = \mathcal{M}$ if $p = q$, $N_0 = M_0$ and $\{N_1, \ldots, N_p\} = \{M_1, \ldots, M_q\}$, so $N_1, \ldots, N_p$ is just a renumbering of the sets $M_1, \ldots, M_p$. (This is consistent with $(N_0, \{N_1, \ldots, N_p\}) = (M_0, \{M_1, \ldots, M_q\})$). The partition $\mathcal{N}$ induces an equivalence relation $\equiv_{\mathcal{N}}$ on N in the usual way, that is, $i \equiv_{\mathcal{N}} j$ if and only if $i, j \in N_k$ for some $k \leq p$. We write $u \equiv_{\mathcal{N}} 0$ if $i \in N_0$. If no confusion should arise, we may write $\equiv$ in stead of $\equiv_{\mathcal{N}}$.

Partitions with a certain relation to the rows $R_1, \ldots, R_m$ are of interest below, and to define these some more terminology is useful. For integers l and r with $l \leq r$ we call the set $I = \{l, l+1, \ldots, r-1, r\}$ an *interval* and define $l(I) := l$, $r(I) := r$. A family of intervals I_t, $t = 0, \ldots, p$ is called *cross-free* if there is a permutation $\pi \in S_{p+1}$ with $\pi(0) = 0$ such that $r(I_{\pi(t+1)}) \leq l(I_{\pi(t)})$ for $t = 0, 1, \ldots, p-1$. Let $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ be a partition and define the associated sets ("projections")

$$I(N_t) = \{i \leq m : R_i \cap N_t \neq \emptyset\}$$

for $t = 0, \ldots, p$. We say that $\mathcal{N}$ is *cross-free* if $I(N_0), I(N_1), \ldots, I(N_p)$ is a family of cross-free intervals. It is not difficult to verify that $\mathcal{N}$ is cross-free if and only if the following conditions hold

- CF(i) if $u \equiv v$ where $u \in R_{i_1}$, $v \in R_{i_2}$ and $i_1 < i < i_2$ then $u \equiv w$ for each $w \in R_i$;
- CF(ii) for each $i < m$ there is at most one k such that N_k intersects both row R_i and R_{i+1};
- CF(iii) $R_k \cap N_0 \neq \emptyset$ implies that $R_t \subset N_0$ for all $t > k$.

Roughly, this means, e.g. in the matrix case, that the sets $N_0, N_1, \ldots, N_p$ are stacked on top of each other with N_0 at the bottom of the matrix, see Figure 1. Let $\mathcal{P}$ be the set of all cross-free partitions (with equality as defined above).

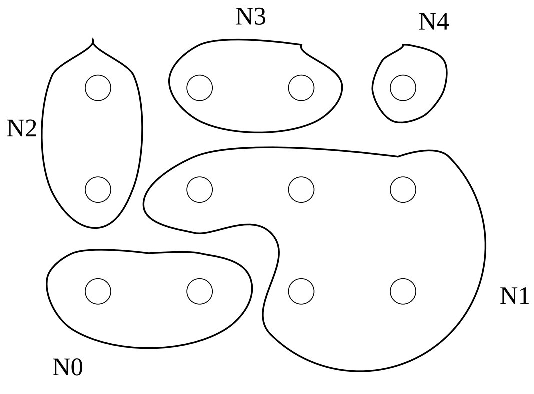

Figure 1: A cross-free partition in the matrix case, $m = 3$ and $n = 4$.

We define a polyhedral cone $C(\mathcal{N})$ associated with the partition $\mathcal{N}$:

$$C(\mathcal{N}) = \{\mathbf{x} \in C : x_u = x_v \text{ when } u \equiv v; \ \ x_v = 0 \text{ when } v \equiv 0\}. \qquad (2)$$

A crucial property of $C(\mathcal{N})$ holds when $\mathcal{N}$ is a cross-free partition.

Lemma 2.1 *Let $\mathcal{N} \in \mathcal{P}$. Then there is a point $\mathbf{x} \in C(\mathcal{N})$ such that for each $u, v \in V$ one has that $x_u = x_v$ if and only if $u \equiv v$.*

Proof. Assume that the partition $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ is cross-free. Then the sets $I(N_0), I(N_1), \ldots, I(N_p)$ are intervals and for some permutation $\pi \in S_{p+1}$ with $\pi(0) = 0$ we have $r(I_{\pi(t+1)}) \leq l(I_{\pi(t)})$ for $t = 0, \ldots, p-1$. Let $\mathbf{x} \in \mathbb{R}^V$ denote the vector with $x_v = t$ when $v \in N_{\pi(t)}$, $t = 0, \ldots, p$. Then, due to the construction, $\mathbf{x} \in C$, and furthermore $x_u = x_v$ if and only if $u \equiv v$. $\square$

There is a bijection between $\mathcal{P}$ and the set of faces of C. This is described in the following proposition.

Proposition 2.2 *(i) For each $\mathcal{N} \in \mathcal{P}$ the cone $C(\mathcal{N})$ is a face of C.*

(ii) If F is a nonempty face of C, there is an $\mathcal{N} \in \mathcal{P}$ with $F = C(\mathcal{N})$.

(iii) Let $\mathcal{N}, \mathcal{M} \in \mathcal{P}$. Then $C(\mathcal{N}) = C(\mathcal{M})$ if and only if $\mathcal{N} = \mathcal{M}$.

(iv) If $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ is a cross-free partition the face $C(\mathcal{N})$ has dimension p.

Proof. (i). Let $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ be a cross-free partition of N. We prove that $C(\mathcal{N})$ (as defined in (2)) is a face of C by finding an equivalent system describing $C(\mathcal{N})$ and consisting of valid inequalities for C set to equality.

From property CF(i) it follows that if $(i_1, j_1) = u \equiv v = (i_2, j_2)$ and $i_1 < i_2$ then there is a path from u to v in D, say $u = u_0, u_1, \ldots, u_t = v$, where each $u_i \equiv u$ for each i. This means that each equality $x_u = x_v$ in (2) is equivalent to the equalities $x_{u_i} = x_{u_{i+1}}$ for $i = 0, \ldots, t-1$. Similarly, by CF(iii), if $u \equiv 0$ there is a path in D consisting of nodes $u = u_0, u_1, \ldots, u_t$ with $u_t \in R_m$ and $u_i \equiv u$ for all i. Thus, $x_u = 0$ is equivalent to $x_u = x_{u_1}, \ldots, x_{u_{t-1}} = x_{u_t}, x_{u_t} = 0$. We may therefore replace all the equalities in (2)) by an equivalent system of valid inequalities for C set to equality, and therefore $C(\mathcal{N})$ is a face of C.

(ii). Observe that all the inequalities $x_v \geq 0$ for $v \in V \setminus R_m$ may be removed in the definition of C; they are implied by the remaining inequalities. Let F be a face of C. Then

$$F = \{\mathbf{x} \in \mathbb{R}^V : x_u = x_v \text{ when } (u,v) \in E_0, \;\; x_v = 0 \text{ when } v \in V_0\}$$

for suitable $E_0 \subseteq E$ and $V_0 \subseteq R_m$. Define for each $v \in V$ the set $N(v) \subseteq V$ by $N(v) = \{u \in V : x_u = x_v \text{ for all } \mathbf{x} \in F\}$. Since $v \in N(v)$ all these sets are nonempty. Moreover, it is easy to see any two sets $N(v_1)$ and $N(v_2)$ are either equal or disjoint. Consider the set $T = \{u \in V : x_u = 0 \text{ for all } \mathbf{x} \in F\}$. This set may be empty, but if it is nonempty, say $v \in T$, then $T = N(v)$. It follows that the sets $N(v)$, $v \in V$ constitute a partition $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ where $N_0 = T$ is possibly empty. As usual, let $\equiv$ denote the equivalence relation induced by this partition. We prove that

$$F = C(\mathcal{N}).$$

Let $\mathbf{x} \in F$. If $u \equiv v$, then $u \in N(v)$ and therefore, due to the definition of $N(v)$, $x_u = x_v$ holds. Also, if $\mathbf{x} \in F$ and $v \in T$, then $x_v = 0$. It follows that $F \subseteq C(\mathcal{N})$. To prove the converse inclusion, let $(u, v) \in E_0$. Then each $\mathbf{x} \in F$ satisfies $x_u = x_v$, so $u \in N(v)$, i.e., $u \equiv v$. Thus each arc in E_0 has both endnodes in the same equivalence class. Consequently all the equalities defining F are also valid equalities for $C(\mathcal{N})$ (because we clearly have $V_0 \subseteq T$). This proves that $F = C(\mathcal{N})$, as desired.

It remains to prove that the partition $\mathcal{N}$ is cross-free. Assume first that $i_1 < i < i_2$ and $u \in R_{i_1}$, $v \in R_{i_2}$, $w \in R_i$ with $u \equiv v$. Then, by definition of $\mathcal{N}$, we have that $x_u = x_v$ for all $\mathbf{x} \in F$ (and F is assumed nonempty). But for $\mathbf{x} \in F \subseteq C$, we have $x_u \geq x_w \geq x_v$, and therefore $x_u = x_w$ for all $\mathbf{x} \in F$ so $u \equiv w$. This proves that CF(i) holds. Next, assume $(u_1, v_1), (u_2, v_2) \in E_0$ where $u_1, u_2 \in R_i$. Let $\mathbf{x} \in F$. Then $x_{u_1} \geq x_{v_2} = x_{u_2} \geq x_{v_1} = x_{u_1}$ so all these numbers are equal and $u_1 \equiv u_2$. This proves CF(ii). Finally, let $u \in R_k \cap N_0$ and let $\mathbf{x} \in F$. Thus, $x_u = 0$ and therefore, if $v \in R_t$ with $t > k$, we get $0 = x_u \geq x_v \geq 0$ and $v \in N_0$. This proves that $\mathcal{N}$ is cross-free.

(iii). Consider two cross-free partitions $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ and $\mathcal{M} = \{M_0, M_1, \ldots, M_q\}$ such that $C(\mathcal{N}) = C(\mathcal{M})$. Assume that there are nodes u and v with $u \equiv_{\mathcal{N}} v$ and $u \not\equiv_{\mathcal{M}} v$. Using Lemma 2.1 we can find an $\mathbf{x} \in C(\mathcal{M})$ with $x_u \neq x_v$. But $C(\mathcal{N}) = C(\mathcal{M})$ so $\mathbf{x} \in C(\mathcal{N})$ which gives (as $u \equiv_{\mathcal{N}} v$) that $x_u = x_v$; a contradiction. It follows that the equivalence classes induced by $\mathcal{N}$ and those induced by $\mathcal{M}$ coincide. Similar arguments give that $N_0 = M_0$. Therefore $\mathcal{N} = \mathcal{M}$. (The converse, that $C(\mathcal{N}) = C(\mathcal{M})$ when $\mathcal{N} = \mathcal{M}$ is trivial).

(iv). Let $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ be a cross-free partition and consider the face $C(\mathcal{N})$. It follows from Lemma 2.1 that no inequality $x_u \geq x_v$ where $u \not\equiv v$ is an implicit equality for $C(\mathcal{N})$. Thus the dimension of $C(\mathcal{N})$ may be found from the rank r of the equalities in (2). It is easy to check that $r = |N_0| + \sum_{t=1}^{p}(|N_t| - 1)$ and from the dimension formula for polyhedra (see [12]) we get $\dim(C(\mathcal{N})) = |V| - r = p$. □

The cone $C(\mathcal{N})$ for the cross-free partition $\mathcal{N}$ shown in Figure 1 has dimension 4.

Note that there are many subsystems of (1) that induce the same face of C. In fact two such subsystems induce the same face if and only if they define the same cross-free partition according to the procedure described in the proof of property (ii) of Proposition 2.2.

We present some further relations between cross-free partitions and the faces of C. Let $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ and $\mathcal{M} = \{M_0, M_1, \ldots, M_q\}$ be two cross-free partitions. We say that $\mathcal{N}$ is *finer* than $\mathcal{M}$, and write $\mathcal{N} \subseteq \mathcal{M}$, if $N_0 \subseteq M_0$ and each set N_t is contained in some set M_s. It is easy to see $(\mathcal{P}, \subseteq)$ is a partially ordered set.

Remark 2.3 It is useful to see what this partial ordering corresponds to in the digraph D. For $\mathcal{N} \subseteq \mathcal{P}$ define the arc set

$$E(\mathcal{N}) = \{(u,v) \in E : u \equiv_{\mathcal{N}} v\}.$$

consisting of arcs with both endnodes in the same equivalence class. Conversely, for any subset E' of E the connected components of the subgraph (V, E') (ignoring arc directions) gives rise to a partition of V, although it may not be cross-free. We observe that $\mathcal{N} \subseteq \mathcal{M}$ if and only if $N_0 \subseteq M_0$ and $E(\mathcal{N}) \subseteq E(\mathcal{M})$.

The following result is a strengthening of Proposition 2.2 (iii).

Proposition 2.4 *Let $\mathcal{N}$ and $\mathcal{M}$ be cross-free partitions. Then $\mathcal{N} \subseteq \mathcal{M}$ if and only if $C(\mathcal{N}) \supseteq C(\mathcal{M})$.*

Proof. Assume that $\mathcal{N} \subseteq \mathcal{M}$ and let $\mathbf{x} \in C(\mathcal{M})$. If $u \equiv_{\mathcal{N}} v$, then $u \equiv_{\mathcal{M}} v$ and therefore $x_u = x_v$. If $u \equiv_{\mathcal{N}} 0$, then $u \equiv_{\mathcal{M}} 0$ and $x_u = 0$. This proves that $\mathbf{x} \in C(\mathcal{N})$, and we conclude that $C(\mathcal{N}) \supseteq C(\mathcal{M})$.

Conversely, assume that $C(\mathcal{N}) \supseteq C(\mathcal{M})$. Assume that there are nodes u and v with $u \equiv_{\mathcal{N}} v$ and $u \not\equiv_{\mathcal{M}} v$. Due to Lemma 2.1) we may find an $\mathbf{x} \in C(\mathcal{M})$ such that $x_u \neq x_v$. This implies that $\mathbf{x} \notin C(\mathcal{N})$ (for $u \equiv_{\mathcal{N}} v$) which contradicts that $C(\mathcal{N}) \supseteq C(\mathcal{M})$. It follows that whenever $u \equiv_{\mathcal{N}} v$ we also have $u \not\equiv_{\mathcal{M}} v$. Furthermore, if $u \equiv_{\mathcal{N}} 0$, the each $\mathbf{x} \in C(\mathcal{N})$ satisfies $x_u = 0$ and, in particular, this holds for each $\mathbf{x} \in C(\mathcal{M})$ (as $C(\mathcal{N}) \supseteq C(\mathcal{M})$). This proves that $\mathcal{N} \subseteq \mathcal{M}$. □

Let $\mathcal{F}_C$ denote the set of all faces of the cone C. It is well known that $(\mathcal{F}_C, \subseteq)$ is a lattice, called the face-lattice of C (the partial ordering is setwise containment), see [4], [14]. We let, in any lattice, $F \vee G$ ($F \wedge G$) denote the smallest upper bound or join (greatest lower bound or meet) of the elements F and G.

Corollary 2.5 *$(\mathcal{P}, \subseteq)$ is a lattice which is anti-isomorphic to the face lattice $(\mathcal{F}_C, \subseteq)$.*

Proof. This may be proved directly, but it also follows from Proposition 2.5 as follows. Due to Proposition 2.2 the function $f : \mathcal{P} \to \mathcal{F}_C$ given by $f(\mathcal{N}) = C(\mathcal{N})$ is a bijection. From Proposition 2.5 it follows that $\mathcal{N} \vee \mathcal{U} = f^{-1}(C(\mathcal{N}) \wedge C(\mathcal{M}))$ and $\mathcal{N} \wedge \mathcal{U} = f^{-1}(C(\mathcal{N}) \vee C(\mathcal{M}))$. This proves that both meet and join exist in $\mathcal{P}$ so this is a lattice and, in addition, f is a lattice anti-isomorphism from $\mathcal{P}$ into $\mathcal{F}_C$. $\square$

We may also give an explicit description of how the lattice operations act on $\mathcal{P}$. Let $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$ and $\mathcal{M} = \{M_0, M_1, \ldots, M_q\}$ be cross-free partitions. We first determine $\mathcal{N} \wedge \mathcal{M}$. The sets $N_i \cap M_j$ for $1 \leq i \leq p$, $i \leq j \leq q$ define a partition $\mathcal{U} = \{U_0, U_1, \ldots, U_r\}$ where $U_0 = N_0 \cap M_0$ and $U_1, \ldots, U_r$ are the remaining nonempty sets $N_i \cap M_j$. It is easy to see that $\mathcal{U}$ is cross-free (all the properties CF(i)–(iii) hold) and that it must be the greatest lower bound of $\mathcal{N}$ and $\mathcal{M}$, i.e., $\mathcal{U} = \mathcal{N} \wedge \mathcal{M}$. It is somewhat more complicated to determine $\mathcal{N} \vee \mathcal{M}$. Let $\mathcal{U} = \{U_0, U_1, \ldots, U_r\}$ be an upper bound (in $\mathcal{P}$) for two cross-free partitions $\mathcal{N}$ and $\mathcal{M}$. Thus, due to Remark 2.3, $N_0 \cup M_0 \subseteq U_0$ and $E(\mathcal{N}) \cup E(\mathcal{M}) \subseteq E(\mathcal{U})$. Let $\mathcal{W} = \{W_0, W_1, \ldots, W_t\}$ denote the partition induced by the connected components in the subgraph $(V, E(\mathcal{N}) \cup E(\mathcal{U}))$ of D, and let W_0 be the union of the (one or two) components intersecting either N_0 or M_0. Observe that $\mathcal{W} \subseteq \mathcal{U}$. $\mathcal{W}$ may not be cross-free, but we can modify it into a cross-free partition as follows. If we can find arcs (u_1, v_1) and (u_2, v_2) with $u_1 \equiv_{\mathcal{W}} v_1$, $u_2 \equiv_{\mathcal{W}} v_2$ and $u_1 \not\equiv_{\mathcal{W}} u_2$, then we replace these two equivalence classes (containing u_1 resp. u_2) by their union. We repeat this procedure until there is no arc pair left with the mentioned properties. Next, if there are two equivalence classes, say W_1 and W_2, with all the sets $W_1 \cap R_{i-1}$, $W_1 \cap R_{i+1}$ and $W_2 \cap R_i$ nonempty, then we replace W_1 and W_2 by their union $W_1 \cup W_2$. This is repeated until there is no pair of equivalence classes with these properties. Let $\mathcal{W}'$ denote the new partition obtained by this procedure. It is not difficult to check that $\mathcal{W}'$ is cross-free and that $\mathcal{W}' \subseteq \mathcal{U}$. Thus, we must have $\mathcal{W}' = \mathcal{N} \vee \mathcal{M}$.

Observe that the construction of the meet and join in the lattice $\mathcal{P}$ also translates (via the bijection f) to finding the meet and join for a pair of faces of the cone C.

Let $k \in \{1, \ldots, m-1\}$. We call a subset S of V a *k-block* in D if $S = R_1 \cup \cdots \cup R_k \cup S'$ for some nonempty subset S' of R_{k+1}. A *block* is a k-block for some k.

Corollary 2.6 *The extreme rays of C are the faces $C(\mathcal{N})$ constructed from cross-free partitions $\mathcal{N} = \{N_0, N_1\}$, i.e. $p = 1$. Moreover, C is generated by the incidence vectors χ^S where S is either a block in D or S consist of a single node in R_1, i.e., each vector in C may be written as a nonnegative linear combination of the mentioned vectors.*

Proof. The extreme rays have dimension 1 so the first part follows from Proposition 2.2. The second form follows from the cross-free property of the partition $\mathcal{N}$. □

Each of the inequalities $x_v \geq 0$ for $v \in R_m$ and $x_u \geq x_v$ for $(u,v) \in E$ defines a facet of C. This is easy to prove directly, and it also follows from Proposition 2.2. Let e and f denote the number of extreme rays and facets of C, respectively. We see that

$$e = n_1 + \sum_{i=2}^{m}(2^{n_i} - 1), \quad f = \sum_{i=1}^{m-1} n_i n_{i+1} + n_m.$$

In the matrix case with $n_i = n$ for all i, we obtain $e = n + (m-1)(2^n - 1)$ and $f = (m-1)n^2 + n$. In particular, the number of extreme rays grows exponentially in n except when $n = 1$.

The generators of C determined in Corollary 2.6 may be used to give a parametric form of each face of C. Consider a face F of C, so (by Proposition 2.2) $F = C(\mathcal{P})$ for some cross-free partition $\mathcal{N} = \{N_0, N_1, \ldots, N_p\}$. The sets $I(N_0), I(N_1), \ldots, I(N_p)$ are intervals and there is a permutation $\pi \in S_{p+1}$ with $\pi(0) = 0$ and $r(I_{\pi(t+1)}) \leq l(I_{\pi(t)})$ for $t = 0, \ldots, p-1$. Define, for $t = 0, \ldots, p$ the node set $G_t = \cup_{k=t}^{p} N_{\pi(k)}$. Then each G_t is either a block or consist of a single node in R_1. Moreover, χ^{G_t}, $t = 0, \ldots, p$ are affinely independent and they generate the face F, i.e.,

$$F = \text{cone}(\{\chi^{G_t} : t = 1, \ldots, p\}).$$

We omit the proof of these facts.

Finally, let us consider the two-dimensional faces of C. Each such face is spanned by two generators of C. We say that that two distinct generators $\mathbf{z}$ and $\mathbf{w}$ are *adjacent* if $F = \text{cone}(\{\mathbf{z}, \mathbf{w}\})$ is a face of C (and then $\dim(F) = 2$). One may derive from Corollary 2.6 that (i) χ^v for each $v \in R_1$ is adjacent to all other generators, and (ii) for distinct blocks S and T the generators χ^S and χ^T are adjacent if and only if $S \subset T$ or $T \subset S$. This implies that the "diameter" of C is two: any two generators are either adjacent or they are both adjacent to some other generator.

As an application of the results above we consider a polytope obtained by intersecting C with a certain hyperplane. Let $\mathbf{p} \in \mathbb{R}^V$ be a vector satisfying $p_v > 0$ for $v \in V$ and $\sum_{v \in V} p_v = 1$. Later, $\mathbf{p}$ will be viewed as a discrete probability distribution on the node set V. Consider the polyhedron

$$C(\mathbf{p}) = \{\mathbf{x} \in C : \sum_{v \in V} p_v x_v = 1\} \tag{3}$$

Observe that $C(\mathbf{p})$ is bounded as $C \subset \mathbb{R}^n_+$ and each p_v is positive. Therefore, $C(\mathbf{p})$ is a polytope. The faces of $C(\mathbf{p})$ are the intersection between the faces of C and the hyperplane $\{\mathbf{x} \in \mathbb{R}^V : \sum_{v\in V} p_v x_v = 1\}$. In particular, we may determine the vertices of $C(\mathbf{p})$ from Corollary 2.6. Recall the notation $\mathbf{p}(S) = \sum_{v\in S} p_v$ for each subset S of V.

Corollary 2.7 *The vertices of $C(\mathbf{p})$ are the points $(1/\mathbf{p}(S))\chi^S$ where S is either a block in D or S consist of a single node in R_1.*

3 Row-ordered majorization

In this section we introduce a vector ordering which may be seen as a generalization of weak majorization. Some properties of the new ordering are given.

We consider again the digraph D introduced in section 2, but restrict the attention to the matrix case where $|R_i| = n$ for $i = 1, \ldots, m$. Thus each $\mathbf{x} \in \mathbb{R}^V$ may be viewed as a real $m \times n$-matrix with (i, j)th entry $x_{i,j}$, and this is done throughout the present section. The results below also hold for a general node set V, but the matrix case is of special interest. We identify the vector spaces $\mathbb{R}^V$ and $\mathbb{R}^{m,n}$. This space is equipped with the usual inner product for vectors which in matrix form is $\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{i,j} x_{i,j} y_{i,j} = \mathrm{Trace}(\mathbf{x}^T \mathbf{y})$. We let $\mathbf{0}$ denote the matrix (suitably dimensioned) with all zeros.

Let K denote the the set of generators for the cone C, that is (see Corollary 2.6) K contains χ^v for $v \in R_1$ and the incidence vectors of blocks in D. Thus we have

$$C = \mathrm{cone}(K).$$

The *polar cone* (sometimes called the dual cone) of C is the convex cone $C^\circ = \{\mathbf{x} \in \mathbb{R}^{m,n} : \langle \mathbf{y}, \mathbf{x}\rangle \geq 0 \text{ for all } \mathbf{y} \in C\}$.

Consider an $m \times n$-matrix $\mathbf{x}$. We say that $\mathbf{x}$ *C-majorizes* $\mathbf{0}$ and write $\mathbf{x} \succ_C \mathbf{0}$, or $\mathbf{0} \prec_C \mathbf{x}$, provided that $\mathbf{x} \in C^\circ$. Since K generates C, $\mathbf{x} \succ_C \mathbf{0}$ holds if and only if

$$\langle \mathbf{g}, \mathbf{x}\rangle \geq 0 \quad \text{for all } \mathbf{g} \in K. \tag{4}$$

If $\mathbf{y} \in \mathbb{R}^{m,n}$ we say that $\mathbf{x}$ *C-majorizes* by $\mathbf{y}$ and write $\mathbf{x} \succ_C \mathbf{y}$ or $\mathbf{y} \prec_C \mathbf{x}$ if $\mathbf{x} - \mathbf{y} \succ_C \mathbf{0}$.

We may write the concept of C-majorization in a more transparent form. For a real number a we write $a^- := \max\{-a, 0\}$.

Proposition 3.1 *Let* $\mathbf{z} \in \mathrm{I\!R}^{m,n}$. *Then* $\mathbf{z} \succ_C \mathbf{0}$ *if and only if*

$$\begin{array}{lll} \text{(i)} & z_{1,j} \geq 0 & \text{for } j = 1, \ldots, n; \\ \text{(ii)} & \sum_{i=1}^k \sum_{j=1}^n z_{i,j} \geq \sum_{j=1}^n z_{k+1,j}^- & \text{for } k = 1, \ldots, m. \end{array} \tag{5}$$

where we define $z_{m+1,j} = 0$ *for* $j \leq n$. *Thus,* $\mathbf{x} \succ_C \mathbf{y}$ *if and only if* $\mathbf{z} = \mathbf{x} - \mathbf{y}$ *satisfies (5).*

Proof. The result follows from Corollary 2.6 by observing that the minimum value of $\langle \mathbf{g}, \mathbf{z} \rangle$ taken over all k-blocks is equal to $\sum_{i=1}^k \sum_{j=1}^n z_{i,j} - \sum_{j=1}^n z_{k+1,j}^-$ (and removing some redundant inequalities). $\square$

There is a relation between C-majorization and weak majorization. Let $\mathbf{x} \in \mathrm{I\!R}^{m,n}$ and define $\mathbf{s} = (s_1, \ldots, s_m)$ to be the vector of row sums in $\mathbf{x}$, so $s_i = \sum_{j=1}^n x_{i,j}$ for $i = 1, \ldots, m$. Assume that $s_1 \geq s_2 \geq \ldots \geq s_m$. Since $\sum_{i=1}^k s_k \geq \sum_{i=1}^k \sum_{j=1}^n x_{i,j} - \sum_{j=1}^n x_{k+1,j}^-$, we see that $\mathbf{x} \succ_C \mathbf{0}$ implies that $\sum_{i=1}^k s_k \geq 0$ for $k = 1, \ldots, m$, that is, $\mathbf{s} \succ_w \mathbf{0}$.

One may check that $\prec_C$ is a partial ordering on $\mathrm{I\!R}^{m,n}$. Furthermore, $\prec_C$ is a *vector ordering* in the sense that it is compatible with the vector space operations: $\mathbf{x} \prec_C \mathbf{y}$ implies that $\mathbf{x} + \mathbf{z} \prec_C \mathbf{x} + \mathbf{z}$ and $\lambda \mathbf{x} \prec_C \lambda \mathbf{y}$ for all $\mathbf{z}$ and $\lambda \geq 0$. All these properties are due to the fact C° is a convex cone. In fact, $\prec_C$ is a *cone ordering*, see [8] for a discussion of cone orderings on vector spaces. We remark that $\prec_C$ may be extended by introducing symmetries as is the case for (weak) majorization (where $\prec$ and $\prec_w$ are permutation invariant), but we do not discuss this here.

Example 3.2 *(componentwise ordering).* If $m = 1$, the matrices are n-dimen-sional row vectors and (omitting the redundant index) $\mathbf{y} \prec_C \mathbf{x}$ means that $y_j \leq x_j$ for $j = 1, \ldots, n$, i.e., componentwise ordering.

Example 3.3 *(weak majorization).* Let $n = 1$ so our matrices reduce to m-dimensional column vectors (where we again omit an index). Let $\mathbf{z} \in \mathrm{I\!R}^m$ be nonincreasing, so $z_1 \geq \ldots \geq z_m$ (and define $z_{m+1} = 0$). Observe that $\sum_{i=1}^k z_i - z_{k+1}^-$ equals either $\sum_{i=1}^k z_i$ or $\sum_{i=1}^{k+1} z_i$, depending on the sign of z_{k+1}. From this it follows that $\mathbf{z} \succ_C \mathbf{0}$ if and only if $\mathbf{z}$ weakly majorizes $\mathbf{0}$. Thus, C-majorization reduces to weak majorization in the case of $n = 1$.

We give a possible economic interpretation of the ordering $\mathbf{z} \succ_C \mathbf{0}$. Consider n different economic activities over a span of m time periods (say years). Let $z_{i,j}$ denote the expected (discounted) payoff of activity j in time period i, so we obtain a "payoff matrix" $\mathbf{z} \in \mathrm{I\!R}^{m,n}$. What does $\mathbf{z} \succ_C \mathbf{0}$ mean in this setting? Conditions (5) says that (i) no payoff during the first year is negative, and (ii) the *loss* (negative payoff) during one single year does not exceed the accumulated payoff at the beginning of that year. If $\mathbf{x} \succ_C \mathbf{y}$ the payoff matrix $\mathbf{x}$ is "better" than $\mathbf{y}$ in this the sense that $\mathbf{x} - \mathbf{y}$ has the properties (i) and (ii) just given.

Let $f : A \to \mathbb{R}$ be a function defined on some subset A of $\mathbb{R}^{m,n}$. It is of interest to consider those functions that are $\prec_C$*-isotone* in the sense that $f(\mathbf{x}) \geq f(\mathbf{y})$ whenever $\mathbf{x}, \mathbf{y} \in A$ and $\mathbf{x} \succ_C \mathbf{y}$. In [3] order preserving functions for cone ordering are discussed, and a general characterization of isotone functions in terms of generators is presented (see also [10]). Note that $(C^\circ)^\circ = C$ because C is a closed convex cone. The generators of C° may be seen from (1). This set consists of (i) the matrices with a 1 in some row and a -1 in the next row and all other elements being zero, and (ii) the matrices with a 1 in the last row and all other elements zero. By applying the general result of [3] we obtain that f is $\succ_C$-isotone if and only if

$$\frac{\partial f}{\partial x_{i,j}}(\mathbf{x}) \geq \frac{\partial f}{\partial x_{i+1,k}}(\mathbf{x}) \geq 0$$

holds for $i < m$, $j \leq n$, $k \leq n$ and for each $\mathbf{x}$ in the interior of A. This is equivalent to the condition that the $m \times n$-matrix of partial derivatives $\frac{\partial f}{\partial x_{i,j}}(\mathbf{x})$ lies in C for each $\mathbf{x}$ in the interior of A.

4 Applications to statistical hypothesis testing

In this section we study some mathematical problems arising in statistical testing theory, and show how some of the results concerning the cone C are useful.

First we give the relevant statistical background (see [9] for the theory of testing statistical hypothesis). We consider a (discrete) statistical experiment where a random variable Z is observed. The sample space is finite, say $\{1, \ldots, n\}$ and we assume that $n \geq 2$. The distribution of Z may be described by a vector $\mathbf{r}$ where $Pr(Z = j) = r_j$ for $j = 1, \ldots, n$ (here $Pr(\cdot)$ denotes probability). We consider the situation where it is known that $\mathbf{r} \in \{\mathbf{p}, \mathbf{q}\}$, but it is unknown whether $\mathbf{r} = \mathbf{p}$ or $\mathbf{r} = \mathbf{q}$. Here $\mathbf{p}, \mathbf{q} \in K_n$ are given vectors. The *testing problem* is to test, based on the observed value of Z, the null hypothesis H_0: $\mathbf{r} = \mathbf{p}$ against the alternative H_1: $\mathbf{r} = \mathbf{q}$. A test is a rule which specifies whether H_0 should be rejected (and thereby claiming that $\mathbf{q}$ is the true distribution). More precisely, a *test* is simply a function $\delta : \{1, \ldots, n\} \to [0, 1]$ where δ_j is the probability of rejection when $Z = j$ is observed. We also view δ as a vector in $\mathbb{R}^n$, so $\delta = (\delta_1, \ldots, \delta_n)$. The *level* of a test δ is defined as $\sum_{j=1}^n \delta_j p_j$ and the *power* of δ is $\sum_{j=1}^n \delta_j q_j$. The level is equal to the probability of rejection when $\mathbf{r} = \mathbf{p}$ (an "error of the first kind"), and the power equals the probability of rejection when $\mathbf{r} = \mathbf{q}$ (a correct decision).

We assume hereafter (for simplicity) that $p_j > 0$ for each j. Let $0 \leq \alpha \leq 1$. The problem of finding a test with maximum power among all tests with level at most α may be formulated as the linear programming problem

$$\max\{\sum_{j=1}^{n} q_j \delta_j : \sum_{j=1}^{n} p_j \delta_j \leq \alpha; \;\; 0 \leq \delta_j \leq 1 \text{ for } j = 1, \ldots, n\}. \tag{6}$$

A basic result in statistical testing theory is the *Neyman-Pearson lemma* (see [9]) which describes the solution of problem (6). An optimal solution δ^* of (6), called a *Neyman-Pearson test*, is found as follows. Determine a permutation $\pi \in S_n$ such that

$$q_{\pi(1)}/p_{\pi(1)} \geq \ldots \geq q_{\pi(n)}/p_{\pi(n)}, \tag{7}$$

that is, the fractions q_j/p_j are ordered nonincreasingly. Let t be maximal with $\sum_{j<t} p_{\pi(j)} \leq \alpha$ and define $\gamma = (\alpha - \sum_{j<t} p_{\pi(j)})/p_{\pi(t)}$. The Neyman-Pearson test δ^* is then given by

$$\delta^*_{\pi(j)} = \begin{cases} 1 & \text{for } j < t; \\ \gamma & \text{for } j = t; \\ 0 & \text{for } j > t. \end{cases} \tag{8}$$

The fact that δ^* is optimal in (6) may also be derived directly using linear programming duality (see e.g. [2]). The problem (6), where $\mathbf{p}$ and $\mathbf{q}$ are arbitrary vectors, is known as the continuous knapsack problem. A discussion of this problem and other algorithms for solving it may be found in [11].

The test δ^* is a function of the permutation π. There may be several permutations satisfying (7) as one may reorder elements i and j for which $q_i/p_i = q_j/p_j$. In fact, these permutations determine *all* the optimal solutions of (6): the set of optimal solutions is the convex hull of Neyman–Pearson tests based on permutations satisfying (7). Thus the ordering in (7) characterizes the solutions of the testing problem.

Consider the testing problem where $\mathbf{p}$ is fixed (and $p_j > 0$ for all j) but $\mathbf{q}$ may vary and it will be replaced by the variable $\mathbf{y} = (y_1, \ldots, y_n)$. Let $i, j \leq n$ be distinct and assume that $\mathbf{y} \in K_n$ satisfies the ordering inequality

$$y_i/p_i \geq y_j/p_j. \tag{9}$$

Then, for each $\alpha \in [0, 1]$, there is a Neyman–Pearson test δ satisfying $\delta_i \geq \delta_j$, i.e., the rejection probability at i is no smaller than the rejection probability at j. Moreover, if the inequality in (9) is strict, then *every* Neyman–Pearson test must satisfy $\delta_i \geq \delta_j$. Therefore, inequalities of the form (9) lead to specific properties of Neyman–Pearson tests. We shall consider two different sets of ordering inequalities of type (9). In each case the relation to the cone C studied in section 2 is explained and statistical interpretations of the results are given.

Let $1 \leq k < n$ and consider the polyhedron $N_k(\mathbf{p}) \in \mathbb{R}^n$ defined by

$$N_k(\mathbf{p}) = \{\mathbf{y} \in K_n : y_i/p_i \geq y_j/p_j \quad \text{for all } i \leq k < j\}. \tag{10}$$

We see that $\mathbf{y} \in \mathrm{rint}(N_k(\mathbf{p}))$ if and only if the *unique* Neyman–Pearson test δ with level $\alpha_k = \sum_{j=1}^{k} p_j$ for testing $\mathbf{p}$ against $\mathbf{y}$ is given by $\delta_j = 1$ for $j \leq k$ and $\delta_j = 0$ for $j > k$.

Consider the node set V being the union of the two rows $R_1 = \{1, \ldots, k\}$ and $R_2 = \{k+1, \ldots, n\}$. Recall the polytope $C(\mathbf{p})$ defined in (3). Consider the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^V$ given by $T(\mathbf{y}) = \mathbf{x}$ where $x_{1,j} = y_j/p_j$ for $1 \leq j \leq k$ and $x_{2,j-k} = y_j/p_j$ for $k+1 \leq j \leq n$. Thus, the jth variable is scaled by $1/p_j$ and the variables are placed consecutively in the two rows. We see that

$$C(\mathbf{p}) = T(N_k(\mathbf{p}))$$

and T is a bijection from $N_k(\mathbf{p})$ to $C(\mathbf{p})$. This implies that all the faces of $N_k(\mathbf{p})$ are found from the faces of $C(\mathbf{p})$. Thus, F is a face of $N_k(\mathbf{p})$ if and only if $T(F)$ is a face of $C(\mathbf{p})$. Moreover, F and $T(F)$ have the same dimension. In particular, we may determine the vertices of $N_k(\mathbf{p})$.

Let $S \subseteq \{1, \ldots, n\}$ be such that either $S = \{i\}$ for some $i \leq k$, or $\{1, \ldots, k\} \subset S$ (with strict inclusion). Let $\mathcal{S}$ denote the set of all such subsets S. For each $S \in \mathcal{S}$ we define $\mathbf{p}^S \in K_n$ by

$$\mathbf{p}^S_j = \begin{cases} p_j/\mathbf{p}(S) & \text{for } j \in S; \\ 0 & \text{for } j \notin S. \end{cases}$$

Note that $\mathbf{p}^S = \mathbf{e}_i$ when $S = \{i\}$, $i \leq k$. Let Z be a random variable on $\{1, \ldots, n\}$ with distribution given by $\mathbf{p}$. Then the conditional probability of the event $Z = j$ given that $Z \in S$ is equal to $p_j/\mathbf{p}(S)$ if $j \in S$ and 0 otherwise. Therefore $\mathbf{p}^S$ may be interpreted as the conditional distribution of Z given that $Z \in S$. From Corollary 2.7 we now obtain

Corollary 4.1 *The vertices of $N_k(\mathbf{p})$ are the points $\mathbf{p}^S$ for $S \in \mathcal{S}$.*

The polytope $N_k(\mathbf{p})$ has therefore $2^{n-k} + k - 1$ vertices.

We next study a second set of ordering constraints of type (9). Consider the polytope $M(\mathbf{p}) \in \mathbb{R}^n$ given by

$$M(\mathbf{p}) = \{\mathbf{y} \in K_n : y_1/p_1 \geq \ldots \geq y_n/p_n\}. \tag{11}$$

The relative interior of $M(\mathbf{p})$ consists of all those vectors $\mathbf{y}$ such that for each level $\alpha \in [0,1]$ the Neyman–Pearson test δ is unique and has the form $\delta_j = 1$ for $j < t$, $\delta_j = \gamma$ for $j = t$ and $\delta_j = 0$ for $j > t$.

This polytope $M(\mathbf{p})$ is also (affinely) isomorphic to a polytope $C(\mathbf{p})$ for suitable node set V. In fact, let $|R_i| = 1$ for $i = 1, \ldots, n$ and consider the linear transformation T (from $\mathbb{R}^n$ to $\mathbb{R}^V$) given by $x_{i,1} = y_i/p_i$ for $i = 1, \ldots, m$. Then we have that $C(\mathbf{p}) = T(M(\mathbf{p}))$ so again all faces of $M(\mathbf{p})$ are given via those of $C(\mathbf{p})$.

Let $i \in \{1, \ldots, n\}$. We define $s_i = \sum_{k=1}^{i} p_k$ and $\mathbf{p}^{(i)} = (p_1^{(i)}, \ldots, p_n^{(i)}) \in K_n$ by

$$p_j^{(i)} = \begin{cases} p_j/s_i & \text{for } j \leq i; \\ 0 & \text{for } j > i \end{cases} \tag{12}$$

In particular, $\mathbf{p}^{(1)} = \mathbf{e}_1$ and $\mathbf{p}^{(n)} = \mathbf{p}$. Consider again a stochastic variable Z with sample space $\{1, \ldots, n\}$ and distribution given by $\mathbf{p}$. We may interprete $\mathbf{p}^{(i)}$ as the conditional distribution of Z given the event that $Z \leq i$.

Proposition 4.2 *$M(\mathbf{p})$ is an $(n-1)$-simplex with vertices $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(n)}$.*

Proof. This form of the vertices follows from Corollary 2.7 and the affine independence is easy to check. Therefore $M(\mathbf{p})$ is a simplex of dimension $n-1$. □

Statistically, this says that the set of distributions $\mathbf{q}$ for which the Neyman–Pearson test is unique and nonincreasing coincides with the set of convex combinations ("mixings" of the conditional distributions for Z given that $Z \leq i$. In Figure 2 the polytope $M(\mathbf{p})$ is shown for $\mathbf{p} = (0.1, 0.4, 0.5)$.

So far we have assumed that $p_j > 0$ for all j. We now consider the general case where some components in $\mathbf{p}$ may be zero. We therefore assume that, for some k, $p_j = 0$ for $j \leq k$ and $p_j > 0$ for $j > k$. Observe that, for $\alpha > 0$, each Neyman–Pearson test δ must satisfy $\delta_j = 1$ for $j \leq k$. Thus we are lead to consider the polyhedron

$$M_k(\mathbf{p}) = \{\mathbf{y} \in K_n : y_{k+1}/p_{k+1} \geq \ldots \geq y_n/p_n\}. \tag{13}$$

Note that $M_0(\mathbf{p}) = M(\mathbf{p})$.

Proposition 4.3 *Let $\mathbf{p}$ be as above. Then $M_k(\mathbf{p})$ is an $(n-1)$-simplex with vertices $\mathbf{e}_1, \ldots, \mathbf{e}_k, \mathbf{p}^{(k+1)}, \ldots, \mathbf{p}^{(n)}$.*

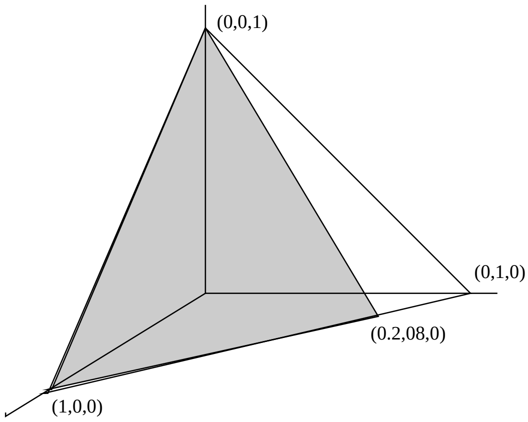

Figure 2: The simplex $M(\mathbf{p})$ for $\mathbf{p} = (0.1, 0.4, 0.5)$.

Proof. Observe first that all the inequalities $y_j \geq 0$ for $k+1 \leq j < n$ are redundant. Let $\mathbf{y}$ be a vertex of $M_k(\mathbf{p})$. Then $\mathbf{y}$ is determined by n linearly independent equations consisting of $\sum_{j=1}^n y_j = 1$ plus $n-1$ equations among the n equations (i) $y_i = 0$ for $i \leq k$ and (ii) $y_n = 0$ and $y_i/p_i = y_{i+1}/p_{i+1}$ for $k \in \{i, \ldots, n-1\}$. If we leave out an equation from group (i), say $y_i = 0$ for some $i \leq k$, then $\mathbf{y}$ satisfies $y_j = 0$ for $j \neq i$ and $j \leq k$, $y_k/p_k = \ldots = y_n/p_n$, $y_n = 0$ and finally $\sum_{j=1}^n y_j = 1$. A little calculation gives that $\mathbf{y} = \mathbf{e}_i$. If we leave out the equation $y_n = 0$, we obtain $\mathbf{y} = \mathbf{p}$. Finally, if we leave out the equation $y_i/p_i = \mathbf{y}_{i+1}/p_{i+1}$ for some $i \in \{k+1, \ldots, n-1\}$, then $y_j = 0$ for $j \in \{1, \ldots, k, i+1, \ldots, n\}$ and we get $\mathbf{y} = \mathbf{p}^{(i)}$. This proves the result. $\square$

Similar results to those above may also be derived for more complicated sets of ordering constraints (9) by transforming the problem to questions concerning faces of C. We do not pursue this here, but rather we mention some geometrical properties of $M(\mathbf{p})$ (similar results hold for $M_k(\mathbf{p})$). Let $v_k(C)$ denote the k-dimensional volume of a convex set C of dimension k.

Let $S = \text{conv}(\{\mathbf{0}, \mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(n)}\})$ which (due to Proposition 4.2) is the convex hull of $M(\mathbf{p}) \cup \{\mathbf{0}\}$. Since $\mathbf{0}, \mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(n)}$ are affinely independent, S is an n-simplex in $\mathbb{R}^n$. The relation between the volumes of $M(\mathbf{p})$ and S is given by

$$v_n(S) = \frac{1}{n\sqrt{n}} v_{n-1}(M(\mathbf{p})). \tag{14}$$

This follows from a well-known volume relation (see [13]) using the facts that $M(\mathbf{p})$ lies in the hyperplane defined by $\sum_{j=1}^n x_j = 1$ and the distance from $\mathbf{0}$ to this hyperplane equals $1/\sqrt{n}$. The volume of S can be determined as follows

$$v_n(S) = (1/n!)|\det(\mathbf{A})|$$

where $\mathbf{A} \in \mathbb{R}^{n,n}$ is given by

$$\mathbf{A} = \left[\ \mathbf{p}^{(1)} - \mathbf{0}, \ldots, \mathbf{p}^{(n)} - \mathbf{0}\ \right] = \begin{bmatrix} 1 & \frac{p_1}{s_2} & \frac{p_1}{s_3} & \cdots & \frac{p_1}{s_{n-1}} & \frac{p_1}{s_n} \\ 0 & \frac{p_2}{s_2} & \frac{p_2}{s_3} & \cdots & \frac{p_2}{s_{n-1}} & \frac{p_2}{s_n} \\ 0 & 0 & \frac{p_3}{s_3} & \cdots & \frac{p_3}{s_{n-1}} & \frac{p_3}{s_n} \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \frac{p_{n-1}}{s_{n-1}} & \frac{p_{n-1}}{s_n} \\ 0 & 0 & 0 & 0 & 0 & \frac{p_n}{s_n} \end{bmatrix}.$$

Therefore we have

$$\det(\mathbf{A}) = \prod_{i=1}^{n} p_i/s_i.$$

Putting these results together shows that the volume of $M(\mathbf{p})$ is given by

$$v_{n-1}(M(\mathbf{p})) = \frac{\sqrt{n}}{(n-1)!} \cdot \prod_{i=1}^{n} \frac{p_i}{s_i}. \tag{15}$$

For instance, for $n = 2$ the volume (length) is $p_2\sqrt{2}$ and for $n = 3$ it is $(1/2)\sqrt{3}; p_2p_3/(1-p_3)$.

The $(n-1)$-volume of the standard simplex K_n is $\sqrt{n}/((n-1)!)$, so we have

$$v(\mathbf{p}) := v_{n-1}(M(\mathbf{p}))/v_{n-1}(K_n) = \prod_{i=1}^{n} \frac{p_i}{s_i}. \tag{16}$$

This number $v(\mathbf{p})$, the *relative volume*, lies strictly between 0 and 1 and may be given a probabilistic interpretation. Assume that a vector $\mathbf{q}$ is drawn at random from K_n according to the uniform distribution, and recall the remark given after (11). It follows that $v(\mathbf{p})$ is the probability that $\mathbf{q}$ lies in $M(\mathbf{p})$) or, equivalently, that for each level $\alpha \in [0, 1]$ there is a *unique* nonincreasing Neyman–Pearson test δ.

How does $v(\mathbf{p})$ depend on $\mathbf{p}$? A first observation is that if $\mathbf{p} \in K_n$ is nondecreasing, i.e., $p_1 \leq \ldots \leq p_n$, then $v(\mathbf{p}) \geq v(\pi(\mathbf{p}))$ for all permutations $\pi \in S_n$. Next, $v(\mathbf{p})$ can be made arbitrarily close to 0 or 1 by suitable choices of $\mathbf{p}$. In fact, let $\mathbf{p}^{(m)} = (1-(n-1)/m, 1/m, \ldots, 1/m)$ for $m = 1, 2, \ldots$. Then $M(\mathbf{p}^{(m)})$ converges to $\{\mathbf{0}\}$ in the Hausdorff norm. It is also possible to find another sequence $\mathbf{w}^{(m)}$, $m = 1, 2, \ldots$ such that $M(\mathbf{w}^{(m)})$ converges to K_n (in the Hausdorff norm), say $\mathbf{w}^{(m)} = (1/m, 1/m, \ldots, 1-(n-1)/m)$. Thus, by the continuity of the volume function w.r.t. the Hausdorff distance we get $\lim_{m\to\infty} v(\mathbf{p}^{(m)}) = 0$ and $\lim_{m\to\infty} v(\mathbf{w}^{(m)}) = 1$. This can be checked by direct calculation using (16). Both these observations are quite intuitive due to the definition of $M(\mathbf{p})$ or the probabilistic interpretation of $M(\mathbf{p})$ in terms of Neyman–Pearson tests.

The function $v(\cdot)$ is neither convex nor concave, but it has another interesting property.

Proposition 4.4 *$v(\cdot)$ is Schur-concave on $\mathcal{D}_+^n \cap \mathrm{rint}(K_n)$.*

Proof. We use the well known fact that the function $\mathbf{x} \to \sum_{j=1}^n g(x_j)$ is Schur-convex when $g : \mathbb{R} \to \mathbb{R}$ is convex, see [10]. Therefore the function $\mathbf{x} \to \sum_{j=1}^n \ln(x_j)$ is Scur-concave. Since the exponential function is increasing it follows that $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(\mathbf{x}) = \exp(\sum_{j=1}^n \ln(x_j)) = \prod_{j=1}^n x_j$ is Schur-concave.

Assume that $\mathbf{x}, \mathbf{y} \in \mathcal{D}_+^n \cap \mathrm{rint}(K_n)$ and that $\mathbf{x} \prec \mathbf{y}$. Since f above is Schur-concave we have

$$x_1 \cdots x_n \geq y_1 \cdots y_n$$

and combining this with the fact that $\sum_{j=1}^k x_j \leq \sum_{j=1}^k y_j$ for $k = 1, \ldots, n-1$ (here we use the assumption that $\mathbf{x}, \mathbf{y} \in \mathcal{D}_+^n$), we see that

$$v(\mathbf{x}) \geq v(\mathbf{y})$$

as desired. □

We remark that the function $v(\cdot)$ is not Schur-concave on $\mathrm{rint}(K_n)$ since it is not symmetric.

There is another property of the polytope $M(\mathbf{p})$ that relates to majorization.

Proposition 4.5 *Assume that $\mathbf{p} \in \mathcal{D}_+^n \cap \mathrm{rint}(K_n)$. Then the vertices of $M(\mathbf{p})$ constitute a chain in the partial ordering given by $\prec$:*

$$\mathbf{p} = \mathbf{p}^{(n)} \prec \mathbf{p}^{(n-1)} \prec \ldots \prec \mathbf{p}^{(1)} = \mathbf{e}_1.$$

Proof. First we observe that $\mathbf{p} \in \mathcal{D}_+^n$ implies that $\mathbf{p}^{(i)} \in \mathcal{D}_+^n$ for $i = 1, \ldots, n$ and therefore $p_{[j]}^{(i)} = p_j^{(i)}$ for all i and j. Let $i \in \{1, \ldots, n-1\}$. As $p_j^{(i)} = 0$ for $j > i$ we get for each $k \leq n$ that

$$\Delta_k^{(i)} := \sum_{j=1}^k p_j^{(i)} - \sum_{j=1}^k p_j^{(i+1)} = \frac{1}{s_i} \sum_{j=1}^{k \wedge i} p_j - \frac{1}{s_{i+1}} \sum_{j=1}^{k \wedge (i+1)} p_j$$

where $a \wedge b$ denotes $\min\{a, b\}$. If $k \leq i$ we get $\Delta_k^{(i)} = (\frac{1}{s_i} - \frac{1}{s_{i+1}}) \sum_{j=1}^k p_j > 0$ because $p_{i+1} > 0$. If $k \geq i+1$ then $\Delta_k^{(i)} = \frac{1}{s_i} \sum_{j=1}^i p_j - \frac{1}{s_{i+1}} \sum_{j=1}^{i+1} p_j = 1 - 1 = 0$. This shows that $\Delta_k^{(i)} \geq 0$ and $\mathbf{p}^{(i+1)} \prec \mathbf{p}^i$. □

A consequence of Proposition 4.5 is that for every symmetric vector norm $\|\cdot\|$ (so $\|\pi(\mathbf{x})\| = \|\mathbf{x}\|$ for every $\mathbf{x} \in \mathbb{R}^n$ and $\pi \in S_n$) the norms of the vertices $\mathbf{p}^{(i)}$ when $\mathbf{p} \in \mathcal{D}_+^n \cap \mathrm{rint}(K_n)$ are ordered by

$$\|\mathbf{p}^{(n)}\| \leq \|\mathbf{p}^{(n-1)}\| \leq \ldots \leq \|\mathbf{p}^{(1)}\|.$$

This is due the fact that such a norm is Schur-convex, see [10].

A final geometrical property we mention concerns the angles between the vectors $\mathbf{p}^{(i)}$. If we let $\alpha_{i,j}$ denote the angle between $\mathbf{p}^{(i)}$ and $\mathbf{p}^{(j)}$ when $1 \leq i < j \leq n$ (angles are calculated using the Euclidean norm), then we obtain that

$$\cos(\alpha_{i,j}) = \sqrt{\sum_{t=1}^{i} p_j^2 / \sum_{t=1}^{j} p_j^2}.$$

In particular, for $\mathbf{p} = (1/n, \ldots, 1/n)$ we have $\cos(\alpha_{i,j}) = \sqrt{i/j}$.

References

[1] T. Ando. Majorization and inequalities in matrix theory. *Linear algebra and its Applications*, 199:17–67, 1994.

[2] T.S. Arthanari and Y. Dodge. *Mathematical programming in statistics.* Wiley, 1981.

[3] D.W. Walkup A.W. Marshall and R.J.B. Wets. Order preserving functions: applications to majorization and order statistics. *Pacific J. Math.*, 23:569–584, 1967.

[4] A. Brøndsted. *An introduction to convex polytopes.* Springer, New York, 1983.

[5] R. A. Brualdi. The doubly stochastic matrices of a vector majorization. *Linear algebra and its Applications*, 61:141–154, 1984.

[6] G. Dahl. Polyhedra and optimization in connection with a weak majorization ordering. *Springer Lecture Notes in Computer Science*, 920:426–437, 1995.

[7] G. Dahl and F. Margot. Weak k-majorization and polyhedra. *Mathematical Programming (to appear).*

[8] M.L. Eaton. On group induced orderings, monotone functions, and convolutions theorems. In R.C. Bose and T.A. Dowling, editors, *Inequalities in statistics and probability*, volume 5, pages 13–25. IMS Lecture notes - Monograph Series, 1984.

[9] E.L. Lehmann. *Testing statistical hypothesis.* Wiley, 1986.

[10] A.W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York, 1979.

[11] G. Nemhauser and L.A. Wolsey. *Integer and combinatorial optimization.* Wiley, 1988.

[12] A. Schrijver. *Theory of linear and integer programming.* Wiley, Chichester, 1986.

[13] R. Webster. *Convexity.* Oxford University Press, Oxford, 1994.

[14] G. Ziegler. *Lectures on polytopes.* Springer, 1994.