ANOMALIES OF THE LIKELIHOOD RATIO TEST FOR TESTING RESTRICTED HYPOTHESES

BY J. A. MENÉNDEZ AND B. SALVADOR

Universidad de Valladolid

The first anomaly in the L.R.T. for testing restricted hypotheses was observed by Warrack and Robertson. They found the L.R.T. for testing an order restriction in a normal model to be dominated by a different test.

In this paper we deal with a more general situation in which the L.R.T. for testing a face of an acute cone is dominated by a different test that does not take into account some of the information in the model.

1. Introduction. In a normal model, the likelihood ratio test (L.R.T.) provides a frequently used method for testing means when the hypothesis defines order restrictions on the parameters. The L.R.T. performs well in some testing problems, as can be seen in the book of Barlow, Bartholomew, Bremner and Brunk (1972). The first anomaly of the L.R.T. was observed by Warrack and Robertson (1984). They showed a problem with some order restriction about means in a normal model where the L.R.T. is dominated by another test and they asked for the cause of such an anomaly. We examine this issue in a general context, where answers can be given.

Consider a $k$-dimensional random normal vector $N_k(\theta, \Gamma)$, with unknown mean vector $\theta = (\theta_1, \ldots, \theta_k)'$ and a known covariance. We deal with the L.R.T. for testing the hypotheses:

\begin{equation}
H_0: a'_j \theta = 0, \quad j = 1, \ldots, r; \quad a'_j \theta \geq 0, \quad j = r + 1, \ldots, n.
\end{equation}

\begin{equation}
H_A: a'_j \theta \geq 0, \quad j = 1, \ldots, n.
\end{equation}

given by $k$-dimensional fixed vectors $a_1, \ldots, a_n$ in such a way that $H_A$ defines a polyhedral closed convex cone in $R^k$.

The statistic $T(x) = -2 \ln l(x)$ that defines the L.R.T. ($T > t$) for testing $H_0$ against $H_A - H_0$ is given by

\begin{equation}
T(x) = \|x - x^0\|^2 - \|x - x^A\|^2,
\end{equation}

where $\|x\|^2 = x'\Gamma^{-1}x$ and $x^0, x^A$ are the projections of $x$ on $H_0$, $H_A$, respectively.

Let us consider the hypotheses:

\begin{equation}
H_0^*: a'_j \theta = 0, \quad j = 1, \ldots, r,
\end{equation}

\begin{equation}
H_A^*: a'_j \theta \geq 0, \quad j = 1, \ldots, r.
\end{equation}

Received October 1987; revised February 1990.

$^1$Research partially supported by the DGICYT under grant PB-87-0905-C02-01.


Key words and phrases. Restricted inference, likelihood ratio test, acute cone.
The L.R.T. \( \{ T^* > t \} \) for testing \( H_0^* \) against \( H_A^* - H_0^* \) is given by the statistic
\[
T^*(x) = \| x - x^0 \|^2 - \| x - x^A \|^2 = \| x^A - x^0 \|^2,
\]
the last equality is true since \( H_0^* \) is the least dimension face in the cone \( H_A^* \) and \( x^0 \) and \( x^A \) are the projections of \( x \) on \( H_0^* \) and \( H_A^* \), respectively.

We are concerned with the dominance of the L.R.T. \( \{ T^* > t \} \) when \( H_A \) is an acute cone. The concept of an acute cone was first introduced by Martín and Salvador (1988) who studied the relation between acute cones and the usefulness of the pool adjacent violator algorithm (PAV).

We now give some notation, definitions and a result, which will be useful later. Let \( C \) be a cone, \( C = \{ x \in \mathbb{R}^k : a'_jx \geq 0, \ j = 1, \ldots, n \} \). Any face of \( C \) may be denoted by
\[
K_B = \{ x : a'_jx = 0, \ j \in B; \ a'_jx \geq 0, \ j \in B^c \}
\]
for some subset \( B \) of \( \{ 1, \ldots, n \} \). The subspace associated with \( K_B \) is \( L_B = \{ x : a'_jx = 0, \ j \in B \} \). Denote by \( p(x|C) \) the \( \Gamma^{-1} \)-orthogonal projection onto the cone \( C \), so that
\[
\| x - p(x|C) \|^2 = \inf_{y \in C} \| x - y \|^2.
\]

We will consider the following two definitions which are equivalent to Definition 2.2 in Martín and Salvador (1988).

**Definition 1.1.** The cone \( C_{ij} = \{ x : a'_i x \geq 0, \ a'_j x \geq 0 \} \) is said to be acute (strictly acute) if \( x'\Gamma^{-1}y \geq 0 \ (> 0) \) whenever \( x \in L_i \cap C_{ij} \) and \( y \in L_j \cap C_{ij} \) with \( x \) in the \( \Gamma^{-1} \)-orthogonal subspace to \( L_{ij} \).

**Definition 1.2.** The cone \( C \) is said to be acute (strictly acute) if \( C_{ij} \) is acute (strictly acute) for any couple \( i, j \) in \( \{ 1, \ldots, n \} \).

Note that the acuteness of a cone is preserved by linear transformations of the entire statistical problem and therefore we could use the identity matrix for the covariance of the normal model and the unit metric on \( \mathbb{R}^k \), after performing a linear transformation.

**Proposition 1.1.** \( C \) is an acute cone if and only if \( a'_j p(x|C) = 0 \) for any \( x \) such that \( a'_j x \leq 0 \).

**Proof.** See Martín and Salvador (1988). \( \square \)

In an obvious reference to the PAV algorithm, Martín and Salvador (1988) say that the cone \( C \) is PAV when the sufficient condition in the proposition holds.

We now explain briefly the PAV process for obtaining \( p(x|C) \) when \( C \) is an acute cone [cf. Martín and Salvador (1988), Theorem 2.2].
At the first step, we project \( x \) onto the subspace \( S_1 \) defined by the constraints in \( C \) which are violated by \( x \). If \( p(x|S_1) \in C \), then \( p(x|S_1) = p(x|C) \). If \( p(x|S_1) \notin C \), at the second step we project \( x \) or \( p(x|S_1) \) onto \( S_1 \cap S_2 \), where \( S_2 \) is the subspace associated with the constraints in \( C \) which are violated by \( p(x|S_1) \). If \( p(x|S_1 \cap S_2) \notin C \), we begin a new step and so on. In a finite number of steps we reach \( p(x|C) \). At any rate the solution \( p(x|C) \) is in a subspace of \( S_1 \), as shown by Proposition 1.1.

In Section 2, we find the test \( \{ T > t \} \) to be dominated by \( \{ T^* > t \} \) when testing \( H_0 \) against \( H_A - H_0 \), so that \( T \) and \( T^* \) being equally sized, \( T^* \) becomes more powerful than \( T \). In order to prove that, we give three lemmas, also useful in their own right. The proofs of the lemmas are given in Section 3.

2. Dominance of the L.R.T. The next lemma is a very useful property of an acute cone.

**Lemma 2.1.** Let \( C \) be an acute cone and let \( B \) be a subset of \( \{1, \ldots, n\} \) such that \( K_B \) is not empty. Then, for any \( x \in C \), \( p(x|K_B) = p(x|L_B) \). Moreover, \( p(x|K_B) \neq 0 \), whenever \( x \neq 0 \) and \( C \) is strictly acute.

The next two results generalize to arbitrary acute cones Lemmas 2.1 and 2.2 in Warrack and Robertson (1984).

**Lemma 2.2.** If \( C \) is an acute cone, then for any \( x \in R^k \) and \( \delta \in C \):
\[
a_j'p(x + \delta|C) \geq a_j'p(x|C), \quad j = 1, \ldots, n.
\]

**Lemma 2.3.** Let \( x \) and \( y \) be two elements in a cone \( C \), such that \( a_j'x \leq a_j'y \), \( j = 1, \ldots, n \). If \( C \) is acute, then \( \|x - p(x|K_1, \ldots, n)\|^2 \leq \|y - p(y|K_1, \ldots, n)\|^2 \).

As noted earlier we give proofs of Lemmas 2.1–2.3 in Section 3.

Before presenting the main result (Theorem 2.2), we need to prove the following theorem about the statistic \( T \) defined in (1.2).

**Theorem 2.1.** Let \( \theta \) be an element of \( H_0 \). Then, for any \( x \in R^k \),
\[
T(x + \theta) \geq T(x).
\]

**Proof.** Through this proof we shall denote by \( U_m \), \( m = 1, \ldots, n \), the subspace \( L_{1 \ldots m} \) and we write \( x^0 \) and \( x^A \) instead of \( p(x|H_0) \) and \( p(x|H_A) \), respectively.

Since \( x^0 = p(x|U_m) \) for some \( m \geq r \) and \( x^0 \) is reached projecting \( x \) onto a subspace \( S \) defined by all such constraints which are satisfied with equality by \( x^0 \), the three cases considered below cover all possible situations.

Note that \( H_0 \) is an acute cone in \( U_r \).

The restrictions that define \( U_r \) are always verified with equality by \( x^0 \), so that \( S \) can be defined by \( L_B \), with \( \{1, 2, \ldots, r\} \subset B \). Without loss of generality we can take \( B = \{1, \ldots, m\} \), \( m \geq r \), for each \( x \) under consideration.
CASE 1. Let \( x \in \mathbb{R}^k \) be such that \( x^0 = p(x|U_m) \). If \( \theta \in H_0 \), then 
\[
(x + \theta)^0 = x^0 + \theta,
\]
so that
\[
\|x - x^0\|^2 = \|(x + \theta) - (x + \theta)^0\|^2.
\]

By Theorem 2.1 in Robertson and Wegman (1978), \( \|x - x^A\|^2 \geq \|(x + \theta) - (x + \theta)^A\|^2 \), and therefore \( T(x + \theta) \geq T(x) \).

CASE 2. Let \( x \in \mathbb{R}^k \) and \( \theta \in H_0 \) be such that \( x^0 = p(x|U_m) \) and 
\[
(x + \theta)^0 = p(x + \theta|U_m),
\]
for some \( m > r \). Consider \( T' \), the L.R.T. for testing \( H_0' \); \( a'_j \theta = 0, j = 1, \ldots, m; a'_j \theta \geq 0, j = m + 1, \ldots, n \) against \( H_A' - H_0' \).

If \( \theta \in U_m \), then by Case 1, \( T'(x + \theta) \geq T'(x) \) and it is easy to prove that \( T'(x) = T(x) \) and \( T'(x + \theta) = T(x + \theta) \).

If \( \theta \notin U_m \), consider \( \theta^m = p(\theta|U_m) \), then as before, \( T(x + \theta^m) \geq T(x) \).

Let us consider the cone \( C^{(m)} = \{ x: a'_j x \geq 0, j = 1, \ldots, m \} \) and \( y^{(m)} = p(x + \theta|C^{(m)}) \) and \( z^{(m)} = p(x + \theta^m|C^{(m)}) \).

It is obvious that \( y^0 = p(x + \theta|U_m) = p(x + \theta^m|U_m) = z^0 \). Note that \( y^{(m)} = y^{(m),0} \) and \( z^{(m)} = z^{(m),0} \).

Decomposing \( x + \theta = x + \theta^m + \theta - \theta^m \) and since \( U_m \) is the least dimension face in \( C^{(m)} \), the Lemma 2.2 guarantees that \( a'_j y^{(m)} \geq a'_j z^{(m)} \geq 0, j = 1, \ldots, m \), and by Lemma 2.3, \( \|y^{(m)} - y^0\|^2 \geq \|z^{(m)} - z^0\|^2 \) and therefore \( \|y^{(m)}\|^2 \geq \|z^{(m)}\|^2 \).

Also we note that \( y^{(m)} - z^{(m)} \in C^{(m)} \cap U_m \), therefore \( p(y^{(m)} - z^{(m)}) \in C^{(m+1)} \cap U_m = 0 \), being \( C^{(m+1)} = \{ x: a'_j x \geq 0, j = 1, \ldots, m + 1 \} \), which implies, by Lemma 2.1, that \( y^{(m)} - z^{(m)} \notin C^{(m+1)} \) or \( y^{(m)} = z^{(m)} \). In any case, 
\[
a^{m+1}_j z^{(m)} \leq a^{m+1}_j y^{(m)}.
\]
In the same way, \( a'_j y^{(m)} \leq a'_j z^{(m)} \), \( j = m + 1, \ldots, n \). Now suppose \( y^{(m)} \in H_A \), then \( z^{(m)} \in H_A \); and both of them coincide respectively with \( (x + \theta)^A \) and \( (x + \theta^m)^A \) and the result follows. In the other case, \( a'_j y^{(m)} < 0 \) for some \( j \). Without loss of generality we can assume \( j = m + 1 \). Consider \( y^{(m+1)} = p(x + \theta|C^{(m+1)}) \) and \( z^{(m+1)} = p(x + \theta^m|C^{(m+1)}) \). We deal with two situations:

(a) \( a^{m+1}_j z^{(m)} \leq 0 \).

Let us consider the affine hyperplanes \( H_j^z = \{ x: a'_j x = a'_j z^{(m)} \}, j = 1, \ldots, m + 1 \).

Let \( A \) be the set of the indices \( j \) for which \( H_j^z \) separates \( y^{(m)} \) and \( y^{(m+1)} \). \( A \) is not empty, since \( m + 1 \in A \). For each \( j \in A \), there exists \( \lambda_j \) such that 
\[
\lambda_j y^{(m)} + (1 - \lambda_j) y^{(m+1)} = H_j^z.
\]

Consider \( \lambda_0 = \max_{j\in A} \lambda_j \) and \( y' = \lambda_0 y^{(m)} + (1 - \lambda_0) y^{(m+1)} \).

Then, we have \( y' - z^{(m)} \in C^{(m+1)} \) and \( y^{(m+1)} = y^{(m+1)} \), so that 
\[
p(y^{(m+1)}|U_{m+1}) = p(y^{(m+1)}|U_{m+1}) = p(z^{(m+1)}|U_{m+1}),
\]
the last equality because \( U_{m+1} \) is the least dimension face of \( C^{(m+1)} \) and \( y^{(m+1)} = z^{(m+1)} \).

If we decompose \( y' = z^{(m)} + (y' - z^{(m)}) \), then by Lemma 2.2, \( a'_j y^{(m+1)} \geq a'_j z^{(m+1)} \geq 0, j = 1, \ldots, m + 1 \), and by Lemma 2.3, \( \|y^{(m+1)} - p(y^{(m+1)}|U_{m+1})\|^2 \geq \|z^{(m+1)} - p(z^{(m+1)}|U_{m+1})\|^2 \). Therefore \( \|y^{(m+1)}\|^2 \geq \|z^{(m+1)}\|^2 \).

(b) \( a^{m+1}_j z^{(m)} > 0 \).
The hyperplane \( \{a_{m+1}'x = 0\} \) separates \( y^{(m)} \) and \( z^{(m)} \), so there is \( \lambda \in (0,1) \) such that \( y' = \lambda y^{(m)} + (1-\lambda)z^{(m)} \in C^{(m+1)} \) and \( \|y'\|^2 \geq \|z^{(m)}\|^2 = \|y^{(m+1)}\|^2 \).

The points \( y^{(m+1)} \) and \( y' \) are in case (a) and therefore \( \|y^{(m+1)}\|^2 \geq \|y'\|^2 \). In both (a) and (b) situations, the points \( y^{(m+1)} \) and \( z^{(m+1)} \) verify the conditions that \( y^{(m)} \) and \( z^{(m)} \) verified at the previous step, so that we could repeat the same procedure with the cone \( C^{(m+2)} \) and so on.

The PAV algorithm for acute cones guarantees [Martín and Salvador (1988)], that in a finite number of steps, \( (x + \theta)^A \) and \( (x + \theta^m)^A \) are reached. For these points, \( \|(x + \theta)^A\|^2 \geq \|(x + \theta^m)^A\|^2 \). Therefore \( T(x + \theta) \geq T(x + \theta^m) \), since \( (x + \theta)^0 = (x + \theta^m)^0 \).

**CASE 3.** Consider \( x \) such that \( x^0 = p(x|U_m) \) and let \( \theta \) be in \( H_0 \) such that \( (x + \theta)^0 = p(x + \theta|U_s), \ r \leq s \leq m \leq n \). Denote \( x^{(s)} = p(x|U_s) \) and \( \theta^{(s)} = p(\theta|U_s) \). Note that \( x^{(s)} \notin H_A \), whenever \( s < m \). Without loss of generality, we can suppose \( a_{s+1}'x^{(s)} < 0 \).

\( (x + \theta)^0 \in H_0 \subset H_A \), so that \( a_{s+1}'(x + \theta)^0 = a_{s+1}'(x^{(s)} + \theta^{(s)}) \geq 0 \).

Therefore \( a_{s+1}'(x^{(s)} + \theta^{(s)}) > 0 \).

Consider \( \lambda_1 = -(a_{s+1}'x^{(s)})/(a_{s+1}'\theta^{(s)}), \ 0 \leq \lambda_1 \leq 1 \).

\( a_{s+1}'(x^{(s)} + \lambda_1\theta^{(s)}) = 0 \) and we can write \( p(x + \lambda_1\theta|U_s) = p(x + \lambda_1\theta|U_{s+1}) \).

On decomposing \( x + \theta = x + \lambda_1\theta + \theta - \lambda_1\theta \), we have \( \theta - \lambda_1\theta \in H_0 \) and \( x + \theta \) and \( x + \lambda_1\theta \) are in Case 2 hence \( T(x + \theta) \geq T(x + \lambda_1\theta) \). Repeating the procedure, we obtain \( (x + \lambda_1\theta)^0 = p(x + \lambda_1\theta|U_{s+1}) \).

Consider \( \lambda_2\theta \) in such a way that \( \lambda_1\theta - \lambda_2\theta \in H_0 \) and \( p(x + \lambda_2\theta|U_{s+1}) = p(x + \lambda_2\theta|U_{s+2}) \) and therefore \( T(x + \lambda_1\theta) \geq T(x + \lambda_2\theta) \).

In this way, after \( m - s - 1 \) steps we obtain \( \lambda_{m-s-1}\theta \) such that \( (x + \lambda_{m-s-1}\theta)^0 = p(x + \lambda_{m-s-1}\theta|U_m) \). Therefore \( T(x + \lambda_{m-s-1}\theta) \geq T(x) \). The chain of inequalities obtained proves that \( T(x + \theta) \geq T(x) \). □

**Theorem 2.2.** The L.R.T., \( \{T > t\} \) for testing \( H_0 \) against \( H_A - H_0 \) is dominated by \( (T^* > t) \).

**Proof.** (a) Fix a point \( \theta_0 \) in \( RI(H_0) = \{a_j'\theta = 0, \ j = 1, \ldots, r; \ a_j'\theta > 0, \ j = r + 1, \ldots, n\} \). For each \( x \) in \( R^k \), there is a \( \lambda \), depending on \( x \), such that \( T(x + \lambda\theta_0) = T^*(x + \lambda\theta_0) \).

Let \( x \) be a point of \( R^k \) and \( z = p(x|H_0^*) \). Consider

\[ \delta_j' = a_j'z, \quad j = 1, \ldots, r, \]

\[ \delta_j' = -\frac{a_j'z}{a_j'\theta_0}, \quad j = r + 1, \ldots, n \quad \text{and} \quad \lambda = \max\{\delta_1, \ldots, \delta_n\} \].

Then \( z + \lambda\theta_0 \in H_A \).

\( \theta_0 \in H_0 \subset H_0^* \), with \( H_0^* \) the least dimension face in \( H_A^* \), so that

\[ p(x + \lambda\theta_0|H_A^*) = z + \lambda\theta_0 = p(x + \lambda\theta_0|H_A) \]
and
\[ p(x + \lambda \theta_0 | H_0^* ) = p(x + \lambda \theta_0 | H_0). \]

As a consequence \( T(x + \lambda \theta_0 ) = T^*(x + \lambda \theta_0) \).

(b) \[ P_\theta(T > t) \leq P_\theta(T^* > t) \quad \forall \, \theta, \quad \forall \, t. \]

We shall prove that \( T(x) \leq T^*(x) \quad \forall \, x \), so that (b) will become an obvious consequence. Let \( x \) be a point in \( R^k \) and \( z = p(x | H_A^* ) \). \( H_A \) is an acute cone, so that \( z^A \) and \( x^A \) can be reached by projecting \( x \) on the same subspace. The same is true for \( z^0 \) and \( x^0 \). Therefore \( z^A = x^A \) and \( z^0 = x^0 \) and \( T(x) = T(z) \).

By Theorem 2.1, \( T(z) \leq T(z + \theta) \quad \forall \, \theta \in H_0 \). According to (a), we can choose \( \theta \in H_0 \) in such a way that \( z + \theta \in H_A \). Using then the Lemma 2.1, we conclude that \( T(z + \theta) = T^*(z + \theta) \).

\( T^*(z + \theta) = T^*(z) \) follows, since \( \theta \in H_0 \subset H_0^* \) and \( H_0^* \) is the least dimension face in \( H_A^* \).

Finally, from the definition of \( T^* \), \( T^*(z) = T^*(x) \) and we can assure that, \( \forall \, x \), \( T(x) \leq T^*(x) \).

Inequality (b) proves the L.R.T. to be less powerful than the test \( \{ T^* > t \} \).

Now, we only need to prove that the same significance level is reached by both tests.

(c) \[ \forall \, t, \, \sup_{\theta \in H_0} P_\theta(T > t) = P_0(T^* > t). \]

Let \( t \) be a real number with \( P_0(T^*(X) > t) = \alpha \), where \( X \sim N_0(k, \Gamma) \).

Consider \( \delta > 0 \) and \( E \) a sphere centered at the origin, such that
\[ P_0(\{ T^*(X) > t \} \cap E) \geq \alpha - \delta. \]

This is always possible by considering \( E \) with \( P_0(X \in E) \geq 1 - \delta \). For all \( \lambda \) and \( \theta_0 \in RI(H_0) \), we have
\[ P_0(\{ T^*(X) > t \} \cap E) = P_{\lambda \theta_0}(\{ T^*(X + \lambda \theta_0) > t \} \cap \{ E + \lambda \theta_0 \}). \]

From (a) and the boundedness of \( E \), there exists \( \lambda_0 \) such that \( T^*(x + \lambda_0 \theta_0) = T(x + \lambda_0 \theta_0) \quad \forall \, x \in E \).

Therefore \( P_{\lambda \theta_0}(\{ T(X + \lambda \theta_0) > t \} \cap \{ E + \lambda \theta_0 \}) \geq \alpha - \delta. \)

This inequality beside (b) proves (c) and the theorem follows. \( \square \)

Figure 1 sketches the results in the proof of Theorem 2.2. \( H_A \) is given by \( a_1 = (0, 1) \) and \( a_2 = (\frac{1}{2}, -1) \) and we can see \( \{ T > t \} \), the striped region, to be contained in \( \{ T^* > t \} \), the dotted and striped region which shows the critical region for testing \( H_0^* \) (defined by \( a_1 \)) against \( H_A^* - H_0^* \).

REMARC 1. If we consider part (c) in the proof of Theorem 2.2 and Figure 1, we may obtain an intuitive idea for getting the significance level of the test \( \{ T > t \} \).

The significance level is reached as we consider \( \lambda_0 \to \infty \) since for \( \theta_0 \in RI(H_0) \), \( P_{\lambda \theta_0}(T(X) > t) \) is an increasing function of \( \lambda \). This also implies the test \( \{ T > t \} \) is biased. When \( \theta_0 \in RI(H_0) \) and \( \lambda_0 \to \infty \), the only sensible
constraints defining $H_A$ are those defining $H^*_A$, so that at infinity $T^*$ and $T$ become equivalents.

**Remark 2.** The region $\{T^* > t\}$ in Figure 1 yields the uniformly most powerful level $\alpha$ test of $H_0$ against $H_A - H_0$. Although in general that is not true, possibly $\{T^* > t\}$ will always be admissible. (We are in debt to a referee for this remark).

Figure 2 shows how the Theorem 2.2 fails when $H_A$ is not acute. $H_A$ is defined by $a_1 = (0,1)$ and $a_2 = (1,1)$. By considering $t$ in such a way that $P_\theta(T^* > t) = \alpha \forall \theta \in H^*_A$, it can be seen $\{T^* > t\} \subset \{T > t\}$, so that $P_\theta(T > t) > \alpha \forall \theta \in H_0$, and the L.R.T. for testing $H_0$ against $H_A - H_0$ is not

**Fig. 2.** Critical regions given by $T$ and $T^*$ for $H_A$ nonacute.
dominated by \( \{T^* > t\} \) (\( \{T > t\} \) is the dotted region added to the striped region \( \{T^* > t\} \)).

### 3. Proof of the lemmas

We give proofs of the Lemmas used in Section 2.

**Proof of Lemma 2.1.** Let \( x \) be in \( C \). We use induction on the number of elements in \( B \). Consider \( B = \{i\} \) and \( x^i = p(x|L_i) \). Suppose \( x^i \notin C \). Then there is a \( j \neq i \) such that \( a'_j x^i < 0 \).

Consider \( x_{ij} = p(x|L_{ij}) \). We have \( x_{ij} - x^i \in L_i \cap C_{ij} \), since \( a'_j(x_{ij} - x^i) = 0 \) and \( a'_j(x_{ij} - x^i) = -a'_j x^i > 0 \). Also, \( x_{ij} - x^i \) is orthogonal to \( L_{ij} \) because \( x_{ij} = p(x_i|L_{ij}) \).

Let \( z = x - x_{ij} \) and \( y = x^i - x_{ij} \). They satisfy \( a'_j z \geq 0 \), \( a'_j z \geq 0 \), \( a'_j y = 0 \) and \( a'_j y < 0 \), so that \( L_j \) separates \( y \) and \( z \). There is an \( \alpha \in (0, 1) \) such that \( t = \alpha z + (1 - \alpha) y \in L_j \). Consequently \( a'_j(t - x_{ij}) = 0 \) and \( a'_j(t - x_{ij}) > 0 \), that is, \( t - x_{ij} \in L_j \cap C_{ij} \).

On the other hand \( (x_{ij} - x^i)(t - x_{ij}) = (x_{ij} - x^i)(x_{ij} - x_{ij} + \alpha(x - x_{ij}) - \alpha x^i) = -\|x^i - x_{ij}\|^2 < 0 \), which is in contradiction with the acuteness of \( C_{ij} \). Now, let us consider \( x \neq 0 \) and \( C \) strictly acute and suppose \( x^i = 0 \). Then we can write \( x = x - x^i \in L_i^+ \) so that \( x = \lambda a_i \), where \( \lambda > 0 \) since \( x \neq 0 \) and \( x \in C \). \( C \) strictly acute implies \( a'_j a_i < 0 \), \( j \neq i \); therefore \( a'_j x = \lambda a'_j a_i \lambda x < 0 \), in contradiction with \( x \in C \), so that we conclude \( x^i = p(x|L_i) \neq 0 \).

Now, we suppose the result is right when \( B \) contains \( r \) elements. Let \( B \) be some subset with \( r + 1 \) elements in \( \{1, \ldots, n\} \). Consider \( x_{B^{-i}} = p(x|L_{B^{-i}}) \) for \( i \in B \).

By the induction hypothesis, \( x_{B^{-i}} \in L_{B^{-i}} \cap C \).

Now \( L_B = L_i \cap L_{B^{-i}} \), so that \( x_B = p(x|L_B) = p(x_{B^{-i}}|L_B) \).

In the subspace \( L_{B^{-i}} \), the cone \( L_{B^{-i}} \cap C \) is acute and we can use the preceding arguments in order to obtain \( x_B \in C \) and also \( x_B \neq 0 \), whenever \( x \neq 0 \) and \( C \) strictly acute. \( \Box \)

**Proof of Lemma 2.2.** If \( x \in C \) the result is obvious.

Given \( x \) in \( R^\delta \), consider \( B = \{i: a'_j x^c = 0\} \) and \( B^\delta = \{i: a'_j(x + \delta)^c = 0\} \), where \( x^c = p(x|C) \) and \( (x + \delta)^c = p(x + \delta|C) \).

Let \( \delta \) be in \( C \). We begin by showing \( B^\delta \subset B \).

Consider \( B_r = \{i: a'_j x^r \leq 0\} \) with \( x^{r+1} = p(x|L_{B_r}) \) and

\[
B^\delta = \{i: a'_j(x + \delta)^r \leq 0\} \quad \text{with} \quad (x + \delta)^{r+1} = p(x + \delta|L_{B^\delta})
\]

for \( r = 0, 1, \ldots, [x^0 = x \text{ and (}x + \delta)^0 = x + \delta] \). It is obvious that \( B_0 \subset B_1 \subset \cdots \subset B \) and \( B_0^\delta \subset B_1^\delta \subset \cdots \subset B^\delta \).

\( B^\delta_0 \subset B_0 \), since \( a'_j x \leq a'_j x + a'_j \delta \leq 0 \) for \( i \in B^\delta_0 \).

We shall now prove that \( B^\delta_1 \subset B_1 \).

\[
(x + \delta)^1 = p(x + \delta|L_{B^\delta_1}) = p(x|L_{B^\delta_1}) + p(\delta|L_{B^\delta_1})
\]

Let \( i \in B^\delta_1 \), then \( a'_i(x + \delta)^1 \leq 0 \).
By Lemma 2.1, \( p(\delta|L_{B_0}) \subseteq C \), so that \( a'_i p(\delta|L_{B_0}) \geq 0 \), therefore \( a'_i p(x|L_{B_0}) \leq 0 \).

But \( L_{B_0} \subseteq L_{B_0} \), since \( B_0 \subseteq B_0 \), so that \( x^1 = p(x|L_{B_0}) = p(p(x|L_{B_0})|L_{B_0}) \) and we can assume that \( x^1 \) has been obtained by projecting \( x \) on \( L_{B_0} \) after projecting \( x \) on \( L_{B_0} \) in a step of the PAV process applied to \( x \). In this way, either \( x^1 \) has the restriction given by \( a_i \) as an active constraint or does not, as it happens with \( p(x|L_{B_0}) \). Therefore \( a'_i x^1 \leq 0 \) and \( i \in B_1 \).

In the same way, it can be proved that \( \forall r, B_r \subseteq B_r \).

If \( B^r = B^r \) and \( B_0 = B \), then \( r \leq s \), because if \( p(x|L_{B_0}) \subseteq C \), then \( p(x + \delta|L_{B_0}) \subseteq C \) by Lemma 2.1. Therefore, \( B^r \subseteq B \).

Let \( x \not\in C \), then for any \( j \in B \), \( a'_j(x + \delta)^c \geq a'_j x^c \geq 0 \) and

\[
(x + \delta)^c = p(x + \delta|L_{B^r}) = p(x|L_{B^r}) + p(\delta|L_{B^r}).
\]

Denote by \( y, z \) the first and second terms on the right-hand side. By Lemma 2.1, \( z \in C \). Moreover \( L_B \subseteq L_{B^r} \), since \( B^r \subseteq B \) and so \( x^c = p(x|L_B) = p(y|L_B) \).

Now, we prove, for \( y \in C \),

\[
(3.1) \quad a'_j y \geq a'_j x^c \quad \forall j \in B.
\]

Assume that there is \( j \in B \) such that \( a'_j y < a'_j x^c \). Consider \( x^c_j = p(x|L_B \cap L_j) \) and \( y_j = x^c_j + (y - x^c) \). Then \( a'_j y_j < 0 \) and \( a'_j x^c = a'_j x^c_j \).

Consider \( t = \lambda(y - y_j) \). There is a \( t \) such that \( a'_j t = 0 \). For this \( t \), set \( t_B = t - (y - x^c) \). Then \( t_B = p(t_l|L_B) \) and \( a'_j t_B > 0 \) which is not possible since \( C \) is acute. Therefore (3.1) holds and we can say \( \forall j \not\in B, a'_j x^c \leq a'_j y + a'_j z = a'_j(x + \delta)^c \).

Now, when \( y \in C \), we have \( B = B^r \) and \( x^c = y \) and therefore,

\[
\forall j, a'_j x^c = a'_j y \leq a'_j y + a'_j z = a'_j(x + \delta)^c. \quad \Box
\]

**Proof of Lemma 2.3.** Let \( x, y \) be elements in \( C \) with \( a'_j x \leq a'_j y \), \( j = 1, \ldots, n \). If \( a'_j x = a'_j y \) for all \( j \), then both \( x \) and \( y \) are in \( K_1, \ldots, n + x \) and the result follows.

Consider \( D = \{ j : a'_j x < a'_j y \} \), \( x^D = p(x|L_D) \) and \( y^D = p(y|L_D) \) and denote \( x^\phi = p(x|K_1, \ldots, n) \) and \( y^\phi = p(y|K_1, \ldots, n) \). We have

\[
\begin{align*}
\|x - x^\phi\|^2 &= \|x - x^D\|^2 + \|x^D - x^\phi\|^2, \\
\|y - y^\phi\|^2 &= \|y - y^D\|^2 + \|y^D - y^\phi\|^2, \\
\|y - y^D\|^2 &\geq \|y^D - y^D\|^2 = \|x - x^D\|^2,
\end{align*}
\]

where \( y^D = p(y|L_D + x) \) and \( L_D + x = \{ z : z = y + x, y \in L_D \} \) so that the result will be proved if we prove that \( \|y^D - y^\phi\|^2 \geq \|x^D - x^\phi\|^2 \).

\( C \) is an acute cone and \( y - x \in C \), therefore \( y^D - x^D \in C \) by Lemma 2.1, so that \( a'_j y^D \geq a'_j x^D \), \( j = 1, \ldots, n \).

Also \( y^D, x^D \) are in \( C \cap L_D \), which is an acute cone in the subspace \( L_D \). In this way, \( x^D \) and \( y^D \) are, with respect to \( C \cap L_D \), in the same situation as \( x \) and \( y \) were respect to \( C \). If \( \forall j, a'_j x^D = a'_j y^D \), then \( \|y^D - y^\phi\|^2 = \|x^D - x^\phi\|^2 \).
In the other case, we can apply to $x^D$ and $y^D$ the procedure applied to $x$ and $y$ and so on. This iterative procedure gives pairs $x^F$, $y^F$ satisfying for all $j$, $a'_j x^F \leq a'_j y^F$. If we have at least one strict inequality for $F \subset \{1, \ldots, n\}$ in every step, then we shall obtain $x^\phi$ and $y^\phi$ that verify the result. □

Acknowledgments. The authors are grateful to the Associate Editor, Professor R. Dykstra and the referees for their careful reading of the manuscript and for their valuable comments and suggestions that resulted in a much improved version of the paper.

REFERENCES


Departamento de Estadística e I.O.
Facultad de Ciencias
Universidad de Valladolid
47071 Valladolid
Spain