Convex Polyhedra I: Estimation

Statistics 710

October 3 & 10, 2006

Polyhedral Cones. Let $\mathcal{H} = I\!\!R^n$ with $\langle x, y\rangle = x'Wy$, where W is symmetric and positive definite. Recall that a convex set Ω is said to be cone if $c\theta \in \Omega$ for all $c > 0$ whenever $\theta \in \Omega$. If Ω is a cone, then

$$\Omega^o = \{\omega \in I\!\!R^n : \langle \theta, \omega\rangle \leq 0 \text{ for all } \theta \in \Omega\}.$$

is again a closed convex cone, called *the polar cone*. A *convex polyhedral cone* is a set of the form

$$\Omega = \{\theta \in I\!\!R^n : \langle \gamma_i, \theta\rangle \geq 0,\ i = 1, \cdots, m\}, \tag{1}$$

where $\gamma_1, \cdots, \gamma_m \in \Re^n$. For example, if $W = I_n$, then the set of non-decreasing sequences in $I\!\!R^n$, $\Omega = \{\theta : -\infty < \theta_1 \leq \cdots \leq \theta_n < \infty\}$ is a polyhedral cone with $\gamma_1 = (-1, 1, 0, \cdots, 0)', \cdots, (0, \cdots, -1, 1)'$.

If $m \leq n$ and $\gamma_1, \cdots, \gamma_m$ are linearly independent, then it is possible to describe Ω in a useful way. Let

$$L = \text{span}\{\gamma_1, \cdots, \gamma_m\}.$$

Then $\Omega \supseteq L^\perp$, the orthogonal complement of L. Let

$$\Gamma = [\gamma_1, \cdots, \gamma_m] \quad (n \times m),$$

and

$$\Delta = \Gamma(\Gamma' W \Gamma)^{-1} = [\delta_1, \cdots, \delta_m],\ \text{say}.$$

Then $\Gamma' W \Delta = I_m = \Delta' W \Gamma$, so that $\gamma_1, \cdots, \gamma_m$ and $\delta_1, \cdots, \delta_m$ are biorthogonal, $\langle \gamma_j, \delta_k\rangle = \gamma_j' W \delta_k = 1$ or 0, accordingly as $j = k$ or $j \neq k$, and

$$L = \text{span}\{\delta_1, \cdots, \delta_m\},$$

since the relation between Γ and Δ is invertible. Let $\delta_{m+1}, \cdots, \delta_n$ be an orthonormal basis for $L^\perp$. *If $\gamma_1, \cdots, \gamma_m$ are linearly independent, then Ω consists of all linear combinations*

$$\theta = a_1\delta_1 + \cdots + a_n\delta_n \tag{2}$$

for which $0 \le a_1, \cdots, a_m < \infty$ and $-\infty < a_{m+1}, \cdots, a_n < \infty$, and the polar cone consists of all

$$\omega = b_1\gamma_1 + \cdots + b_m\gamma_m$$

for which $-\infty < b_1, \cdots, b_m \le 0$. To see this observe that any $\theta \in I\!\!R^n$ can be written in the form (2), since $\delta_1, \cdots, \delta_n$ are a basis for $I\!\!R^n$; and if θ is so written, then $\theta \in \Omega$ iff $a_i = \langle \gamma_i, \theta \rangle > 0$ for $i = 1, \cdots, m$. The assertion about the polar cone may be established similarly, with slightly more detail.

Example 1 *If $W = I_n$ and Ω is the set of non-decreasing sequences, then clearly $L^\perp = \text{span}\{\mathbf{1}\} = \{c\mathbf{1} : c \in I\!\!R\}$. Write $\gamma_k = (\gamma_{k,1}, \cdots, \gamma_{k,n})'$ and $\delta_k = (\delta_{k,1}, \cdots, \delta_{k,n})'$. Then $\gamma_{k,i} = -1$ if $i = k$, 1 if $i = k+1$, and 0 otherwise. In this case $\delta_k = (0, \cdots, 0, 1, \cdots, , 1)' - (n-k)\mathbf{1}$, where the first 1 appears in the $(k+1)^{\text{st}}$ position. For with this definition, $\langle \gamma_j, \delta_k \rangle = 1$ if $j = k$ and 0 otherwise.*

Projections. *The Characterization +.* Now let Ω be as in (1); let $y \in I\!\!R^n$; and let $\hat{\theta} = \Pi_\Omega y$. Thus,

$$\hat{\theta} \in \Omega, \ \langle \hat{\theta}, y - \hat{\theta} \rangle = 0, \ \text{and } \langle y - \hat{\theta}, \xi \rangle \le 0 \tag{3}$$

for all $\xi \in \Omega$. Observe that if $\xi \in I\!\!R^n$ and $\hat{\theta} \pm \alpha\xi \in \Omega$ for all small α, then $\langle y - \hat{\theta}, \xi \rangle = 0$, since $\pm\alpha\langle y - \hat{\theta}, \xi \rangle = \langle y - \hat{\theta}, \hat{\theta} \pm \alpha\xi \rangle \le 0$ for small $\alpha > 0$. In particular, if $\langle \gamma_k, \hat{\theta} \rangle > 0$, then $\langle y - \hat{\theta}, \delta_k \rangle = 0$. For, if $j \ne k$, then $\langle \gamma_j, \hat{\theta} \pm \alpha\delta_k \rangle = \langle \gamma_j, \hat{\theta} \rangle \pm \alpha\langle \gamma_j, \delta_k \rangle = \langle \gamma_j, \hat{\theta} \rangle \ge 0$; and $\langle \gamma_k, (\hat{\theta} \pm \alpha\delta_k \rangle = \hat{a}_k \pm \alpha > 0$ for small α.

A Generalized CSD and GCM. Let $\hat{\Theta} = W\Delta'\hat{\theta}$ and $Y = W\Delta' y$. Then

$$(\Gamma' W \Gamma)\hat{\Theta} \ge 0 \tag{4}$$

and

$$\hat{\Theta}_k \ge Y_k \text{ for all } k \le m \text{ with equality if } \langle \gamma_k, \hat{\theta} \rangle > 0. \tag{5}$$

To see (4) simply observe that $(\Gamma' W \Gamma)\hat{\Theta} = \Gamma' W \theta \ge 0$ by (1), since $[\Gamma' W \theta]_k = \langle \gamma_k, \hat{\theta} \rangle$. For (5), observe that $\hat{\Theta}_k - Y_k = \langle \delta_k, \hat{\theta} - y \rangle \ge 0$ for all k, since $\delta_k \in \Omega$, with equality if $\langle \gamma_k, \hat{\theta} \rangle > 0$, as just explained.

Duality. Clearly condition (3) implies that $y - \hat{\theta} \in \Omega^o$. Moreover, letting $\hat{\xi} = y - \hat{\theta}$, $\langle \hat{\xi}, y - \hat{\xi} \rangle = \langle y - \hat{\theta}, \hat{\theta} \rangle = 0$ and $\langle y - \hat{\xi}, \xi \rangle = \langle \hat{\theta}, \xi \rangle \le 0$ for all $\xi \in \Omega^o$. So, *$\hat{\xi} = y - \hat{\theta}$ is the projection of y onto Ω^o, and*

$$y = \hat{\theta} + \hat{\xi} = \Pi_\Omega y + \Pi_{\Omega^o} y. \tag{6}$$

To go further, write $\hat{\theta} = \hat{a}_1\delta_1 + \cdots + \hat{a}_n\delta_n$, where $\hat{a}_1, \cdots, \hat{a}_m \geq 0$, as in (2), and let

$$\hat{J} = \hat{J}(y) = \{j \leq m : \langle \gamma_j, \hat{\theta} \rangle > 0\}. \tag{7}$$

Then

$$\hat{\theta} = \sum_{j \in \hat{J}} \hat{a}_j \delta_j + \sum_{i=m+1}^{n} \hat{a}_j \delta_j,$$

For $J \subseteq \{1, \cdots, m\}$, let

$$K_J = \text{span}\{\delta_j : j \in J\}$$

and

$$K_J = L_J \oplus L^{\perp} = \text{span}\{\delta_j : j \in J, \text{or } j > m\}.$$

Then

$$K_J^{\perp} = \text{span}\{\gamma_j : j \in J^c\},$$

where $J^c = \{1, \cdots, m\}$. Denote the right side of the last line by M. Then, clearly $M \subseteq K_J^{\perp}$; and if $z \in K_J^{\perp}$, then $z = \sum_{i=1}^{m} c_i\gamma_i + \sum_{i=m+1}^{n} c_i\delta_i$, where $c_i = \langle \delta_i, z \rangle = 0$ if $i \in J$ or $i > m$, so that $z \in M$. It follows easily that

$$\hat{\theta} = \Pi_{K_{\hat{J}}} y \qquad \text{and} \qquad y - \hat{\theta} = \Pi_{K_{\hat{J}}^{\perp}} y. \tag{8}$$

To see that $\hat{\theta} = \Pi_{K_{\hat{J}}} y$, it suffices to show that $\hat{\theta} \in K_{\hat{J}}$ and that $\langle y - \hat{\theta}, \xi \rangle = 0$ or all $\xi \in K_{\hat{J}}$. That $\hat{\theta} \in K_{\hat{J}}$ is clear. If $\xi \in K_{\hat{J}}$, then $\hat{\theta} \pm \alpha\xi \in \Omega$ for all sufficiently small α. For if $j \in \hat{J}$, then

$$\langle \gamma_j, \hat{\theta} \pm \alpha\xi \rangle = \langle \gamma_j, \hat{\theta} \rangle \pm \alpha \langle \gamma_j, \xi \rangle = \hat{a}_j \pm \alpha \langle \gamma_j, \xi \rangle$$

which is positive for all small α; and if $j \notin \hat{J}$, then $\xi = \sum_{j \notin \hat{J}} c_j\delta_j + \sum_{j=m+1}^{n} c_j\delta_j$, so that $\langle \gamma_j, \xi \rangle = 0$. So, from (3), $\pm\alpha\langle y - \hat{\theta}, \xi \rangle = \langle y - \hat{\theta}, \hat{\theta} \pm \alpha\xi \rangle \leq 0$ and, therefore, $\langle y - \hat{\theta}, \xi \rangle = 0$.

Problem 1 *Show that* $y - \hat{\theta} = \Pi_{\Omega^o} = \Pi_{K_{\hat{J}}^{\perp}}$.

Continuity and Differentiability. First $\hat{\theta}$ is Lipschitz continuous; that is $\|\hat{\theta}(y) - \hat{\theta}(z)\| \leq \|z - y\|$. To see this observe that

$$\langle y - \hat{\theta}(y), \hat{\theta}(z) - \hat{\theta}(y) \rangle \leq 0 \qquad \text{and} \qquad \langle z - \hat{\theta}(z), \hat{\theta}(z) - \hat{\theta}(y) \rangle \geq 0,$$

by (3). Subtracting,

$$\langle y - z, \hat{\theta}(z) - \hat{\theta}(y) \rangle + \|z - y\|^2 \leq 0$$

and, therefore, $\|z - y\|^2 \leq \langle z - y, \hat{\theta}(z) - \hat{\theta}(y) \rangle \leq \|z - y\| \times \|z - y\|$, from which the assertion follows.

For $J \subseteq \{1, \cdots, m\}$, let

$$B_J = \{y \in I\!R^n : \hat{J}(y) = J\}, \tag{9}$$

so that $\hat{\theta} = \Pi_{K_J} y$ for $y \in B_J$. Let Π_{K_J} denote (also) the projection matrix onto K_J. It follows that

$$\left[\frac{\partial \hat{\theta}_j(y)}{\partial y_i}\right] = \Pi_{K_J} \tag{10}$$

on the interior of each B_J. In particular, the divergence of $\hat{\theta}$ is just the dimension of K_J,

$$D(y) = \sum_{j=1}^{n} \frac{\partial \hat{\theta}_j(y)}{\partial y_j} = \mathrm{tr}\,[\Pi_{K_J}] = \dim(K_J).$$

I now claim that $\bar{B}_I \cap \bar{B}_J$ is of Lebesgue measure 0 for any two different subsets of $\{1, \cdots, m\}$, so that (10) holds almost everywhere. To see this, simply observe that if $y \in \bar{B}_I \cap \bar{B}_J$, then $\Pi_{K_J} y = \hat{\theta} = \Pi_{K_I} y$, so that $(\Pi_{K_J} - \Pi_{K_I})y = 0$. Thus, $\bar{B}_I \cap \bar{B}_J$ is contained in a linear subspace of dimension less than n and the assertion follows.

To understand the projection matrices in more detail, recall that $K_J = L_J \oplus L^\perp$ so that $\Pi_{K_J} y = \Pi_{L_J} y + \Pi_{L^\perp} y$ in (8); and if $J = \{j_1, \cdots, j_k\}$, where $1 \le j_1 < \cdots < j_k \le m$, let $\Delta_J = [\delta_{j_1}, \cdots, \delta_{j_k}]$ and $\Gamma_J = [\gamma_{j_1}, \cdots, \gamma_{j_k}]$, so that

$$\Pi_{L_J} = \Delta_J (\Delta_J' \Delta_J)^{-1} \Delta_J' \qquad \text{and} \qquad \Pi_{K_J^\perp} = \Gamma_J (\Gamma_J' \Gamma_J)^{-1} \Gamma_J'.$$

Properties of the Estimator. Suppose now that $W = I_n$ and that y is normally distributed with mean $\theta \in \Omega$ and covariance matrix $\sigma^2 I_n$, where $\sigma^2 > 0$. Then there are both an unbiased estimator and a bound on the mean squared error in terms of D:

$$E_\theta \|\hat{\theta} - \theta\|^2 = E_\theta(U), \tag{11}$$

where

$$U = \|y - \hat{\theta}\|^2 + 2\sigma^2 D - n\sigma^2,$$

and

$$E_\theta \|\hat{\theta} - \theta\|^2 \le \sigma^2 E_\theta(D) \tag{12}$$

for all $\theta \in \Omega$. For (11), write $y - \hat{\theta} = y - \theta - (\hat{\theta} - \theta)$, $\|y - \hat{\theta}\|^2 = \|y - \theta\|^2 - 2\langle y - \theta, \hat{\theta} - \theta\rangle + \|\hat{\theta} - \theta\|^2$, and

$$\begin{aligned} E_\theta \|y - \hat{\theta}\|^2 &= E\|y - \theta\|^2 - 2E_\theta \langle y - \theta, \hat{\theta} - \theta\rangle + E_\theta \|\hat{\theta} - \theta\|^2 \\ &= n\sigma^2 - 2\sigma^2 E_\theta(D) + E_\theta \|\hat{\theta} - \theta\|^2 \end{aligned}$$

where the last step uses Stein's Identity. Equation (11) follows by rewriting the expression. To see (12) observe that by (3)

$$0 \le \langle y - \hat{\theta}, \hat{\theta} - \theta\rangle = \langle y - \theta, \hat{\theta} - \theta\rangle - \|\hat{\theta} - \theta\|^2$$

so that

$$E_\theta\|\hat{\theta}-\theta\|^2 \le E_\theta\left[\langle y-\theta, \hat{\theta}-\theta\rangle\right] = \sigma^2 E_\theta(D),$$

where the equality follows from Stein's Identity.

Estimating σ^2. As a corollary

$$n\sigma^2 - 2\sigma^2 E_\theta(D) \le E\|y-\hat{\theta}\|^2 \le n\sigma^2 - \sigma^2 E_\theta(D). \tag{13}$$

Under regularity conditions, Meyer and Woodroofe [3] showed that $E\|y-\hat{\theta}\|^2 \approx n\sigma^2 - \kappa\sigma^2 E_\theta(D)$, where $\kappa \approx 1.6$ and suggested an estimator of the form

$$\hat{\sigma}^2 = \frac{\|y-\hat{\theta}\|^2}{n-\kappa D}$$

for the case of unkown σ^2.

Remarks. This material is adapted from [2] and [3]. Shrinkage estimation is consider in [4] and [1].

References

[1] Amirdjanova, A. and Michael Woodroofe (2004). Shrinkage estimation for convex polyhedral cones. *Stat. Prob. Lttrs.*, **70**, 87-94.

[2] Meyer, Mary (1996). Shape restricted estimation with applications to ·. Ph. D. Thesis (Statistics), The University of Michigan.

[3] Meyer, Mary and Michael Woodroofe (2000). On the degrees of freedom in shape restricted estimation. *Ann. Statist.*, **28**, 1083-1104.

[4] Sengupta, D. and P.K. Sen (1991). Shrinkage estimation in a restricted parameter space. *Sankhya, A*, **53**, 389-411.