Convex Polyhedra II: Testing  
Statistics 710  
October 12, 2006

The Testing Problems. Again suppose that $W = I_n$ and consider a polyhedral cone in $\mathbb{R}^n$,
\[ \Omega = \{ \theta \in \mathbb{R}^n : \langle \gamma_i, \theta \rangle \geq 0, \ i = 1, \ldots, m \}, \tag{1} \]
where $\gamma_1, \ldots, \gamma_m \in \mathbb{R}^n$ are linearly independent; let $L = \text{span}\{\gamma_1, \ldots, \gamma_m\}$; and suppose that $y \sim \text{Normal}[\theta, \sigma^2 I_n]$. The following three hypotheses are considered: $H_0 : \theta \in L^\perp$, $H_1 : \theta \in \Omega$, and $H_2 : \theta \in \mathbb{R}^n$. For example, in monotone regression, $H_0$ is the hypothesis that the regression function is constant; in convex regression, it is the hypothesis that the regression function is linear.

First consider $H_0$ vs. $H_1 - H_0$. The log-likelihood function is
\[ \ell(\theta, \sigma^2|y) = -\frac{1}{2\sigma^2}||y - \theta||^2 - \frac{1}{2} n \log(\sigma^2), \tag{2} \]
and the least squares estimators of $\theta$ are the maximum likelihood estimators. So, the maximum likelihood estimator under $H_0$ is $\hat{\theta}_o^\circ = \Pi_{L^\perp} y$ and the unconditional maximum likelihood estimator is $\hat{\theta} = \Pi_{\Omega} y$. If $\sigma^2$ is known, then the log-likelihood ratio statistics is
\[ \Lambda_{01} = 2 \left[ \ell(\hat{\theta}, \sigma^2) - \ell(\hat{\theta}_o^\circ, \sigma^2) \right] = \frac{1}{\sigma^2} \left[ ||y - \hat{\theta}_o^\circ||^2 - ||y - \hat{\theta}||^2 \right]. \]
Here $y - \hat{\theta}_o^\circ = y - \hat{\theta} + \hat{\theta} - \hat{\theta}_o^\circ$, and $||y - \hat{\theta}_o^\circ||^2 = ||y - \hat{\theta}||^2 + 2(y - \hat{\theta}, \hat{\theta} - \hat{\theta}_o^\circ) + ||\hat{\theta} - \hat{\theta}_o^\circ||^2 = ||y - \hat{\theta}||^2 + ||\hat{\theta} - \hat{\theta}_o^\circ||^2$ and, therefore,
\[ \Lambda_{01} = \frac{1}{\sigma^2} ||\hat{\theta} - \hat{\theta}_o^\circ||^2. \]

If $\sigma^2$ is unknown, then the maximum likelihood estimators are
\[ \hat{\sigma}^2 = \frac{||y - \hat{\theta}||^2}{n} \quad \text{and} \quad \hat{\sigma}_o^2 = \frac{||y - \hat{\theta}_o^\circ||^2}{n}, \]
and the likelihood ratio statistics is
\[ \Lambda_{01} = 2 \left[ \ell(\hat{\theta}, \sigma^2) - \ell(\hat{\theta}^o, \sigma^2) \right] = n \log \left[ \frac{||y - \hat{\theta}^o||^2}{||y - \hat{\theta}||^2} \right] = \log \left[ \frac{||\hat{\theta}^o - \hat{\theta}||^2 + ||y - \hat{\theta}||^2}{||y - \hat{\theta}||^2} \right]. \]

Of course, an equivalent test is to reject if
\[ \frac{||\hat{\theta}^o - \hat{\theta}||^2}{||\hat{\theta} - \hat{\theta}^o||^2 + ||y - \hat{\theta}||^2} \]
is large.

Next, consider testing \( H_1 \text{ vs } H_2 \), when \( \sigma^2 \) is known. For \( H_2 \), the maximum likelihood estimator is \( y \), and
\[ \Lambda_{12} = 2 \left[ \ell(y, \sigma^2) - \ell(\hat{\theta}^o, \sigma^2) \right] = \frac{1}{\sigma^2} ||y - \hat{\theta}||^2. \]
If \( \sigma^2 \) unknown, then an independent estimate is required.

**Least Favorable Configurations.** Since both null hypotheses are composite, the dependence of the test statistics on parameters, under the hypotheses must be assessed. For \( H_0 \text{ vs } H_1 \) this is simple. *The distributions of \( ||\hat{\theta}^o - \hat{\theta}||^2 \text{ and } ||y - \hat{\theta}||^2 \text{ are the same for all } \theta \in L^+ \).* This is a simple consequence of the following: if \( z \in \mathbb{R}^n \) and \( \theta \in L^+ \), then
\[ \hat{\theta}(z + \theta) = \hat{\theta}(z) + \theta \quad \text{and} \quad \hat{\theta}^o(z + \theta) = \hat{\theta}^o(z) + \theta. \] (3)
To establish the first of these assertions, it suffices to show that \( \hat{\theta}(z) + \theta \) satisfies the necessary and sufficient conditions for \( \hat{\theta}(z + \theta) \). Clearly, \( \hat{\theta}(z) + \theta \in \Omega \) and
\[ \langle z + \theta - [\hat{\theta}(z) + \theta], \xi \rangle = \langle z - \hat{\theta}(z), \xi \rangle \leq 0 \]
for all \( \xi \in \Omega \). Also,
\[ \langle z + \theta - [\hat{\theta}(z) + \theta], \hat{\theta}(z) + \theta \rangle = \langle z - \hat{\theta}(z), \hat{\theta}(z) + \theta \rangle = 0, \]
since \( \hat{\theta}(z) \pm \theta \in \Omega \). The second assertion in (3) may be established similarly (and more easily). To complete the argument, observe that if \( y \sim \text{Normal}(\theta, I_n) \), where \( \theta \in L^+ \), then \( y \) has the same distribution as \( z + \theta \), where \( z \sim \Phi^n \). It follows that
\[ \|\hat{\theta}(y) - \hat{\theta}^o(y)\|^2, ||y - \hat{\theta}(y)||^2 \overset{d}{=} [||\hat{\theta}(z) - \hat{\theta}^o(z)||^2, ||z - \hat{\theta}(z)||^2]. \]

The situation is slightly more complicated for testing \( H_1 \text{ vs } H_2 \), since the distribution of \( ||y - \hat{\theta}(y)||^2 \) does depend on \( \theta \in \Omega \), but a bound can be derived. If \( y = z + \theta \), where \( z \in \mathbb{R}^n \) and \( \theta \in \Omega \), then \( \hat{\theta}(z) + \theta \in \Omega \), so that
\[ ||y - \hat{\theta}(y)||^2 = \inf_{\xi \in \Omega} ||y - \xi||^2 \leq ||z + \theta - [\hat{\theta}(z) + \theta]||^2 = ||z - \hat{\theta}(z)||^2. \]
where

\[ \max_{\theta \in \theta} P_\theta[\|y - \hat{\theta}(y)\|^2 > u] \leq P[\|z - \hat{\theta}(z)\|^2 > u] = P_0[\|y - \hat{\theta}(y)\|^2 > u]. \]

**The Null Distribution.** The main result is that if \( \theta \in L \), then

\[
P_\theta\left[ \frac{1}{\sigma^2} \| \hat{\theta} - \hat{\theta}^* \|^2 \leq u, \frac{1}{\sigma^2} \| y - \hat{\theta} \|^2 \leq v \right] = \sum_{k=m}^{n} P[\chi_{k-m}^2 \leq u] P[\chi_{n-k}^2 \leq v] q(n, k),
\]

where

\[ q(n, k) = P_0[D = k]. \]

Two preliminary results are need to establish this. First, recall the relation \( \hat{\theta} = \Pi_{L,j} y + \Pi_{L,j} y \), where \( J = \{ j \leq m : \langle \gamma_j, \hat{\theta} \rangle > 0 \} \). Recall too the definitions of \( \Gamma_j \) and \( \Delta_j \) and observe that

\[ \Gamma'_{j} \Pi_{L,j} = (\Delta'_{j} \Delta_{j})^{-1} \Delta'_{j} \] and \( \Delta'_{j} \Pi_{K,j} = (\Gamma'_{j} \Gamma_{j})^{-1} \Gamma_{j} \). It follows easily that

\[ \{ y : \hat{\theta}(y) = J \} = \{ y \in \mathbb{R}^n : \Gamma'_{j} \Pi_{L,j} y > 0 \text{ and } \Delta'_{j} \Pi_{K,j} y \leq 0 \}. \]

Next, recall that if \( z \sim \Phi^0 \), then \( \|z\| \) and \( z/\|z\| \) are independent. In fact, if \( Q \neq 0 \) is any projection matrix, then \( \|Qz\| \) and \( Qz/\|Qz\| \) are independent. To see this recall that the eigen values of a projection matrix are either 0 or 1, so that \( Q \) may be written as \( Q = C \text{diag}(I_k, 0)C' \), where \( 1 \leq k \leq n \) and \( C \) is orthogonal. Then \( Cz \sim \text{Normal}[0, \text{diag}(I_k, 0)] \), so that \( Cz = [w', 0, \ldots, 0]' \), where \( w \sim \Phi^k \). The independence of \( \|z\| \) and \( z/\|z\| \) now follows easily from that of \( \|w\| \) and \( w/\|w\| \).

For the proof of (*) we may suppose that \( \theta = 0 \) and \( \sigma = 1 \). Then

\[
P_0\left[ \|\hat{\theta} - \hat{\theta}^*\|^2 \leq u, \|y - \hat{\theta}\|^2 \leq v \right] = \sum_{J} P[J(y) = J, \|\Pi_{L,j} y\|^2 \leq u, \|\Pi_{K,j} y\|^2 \leq v]
\]

Here \( \Pi_{L,j} y \) and \( \Pi_{K,j} y \) are independent. So,

\[
P[J(y) = J, \|\Pi_{L,j} y\|^2 \leq u, \|\Pi_{K,j} y\|^2 \leq v]
= P[\Gamma'_{J} \Pi_{L,j} y > 0, \Delta_{J} \Pi_{K,j} \leq 0, \|\Pi_{L,j} y\|^2 \leq u, \|\Pi_{K,j} y\|^2 \leq v]
= P[\Gamma'_{J} \Pi_{L,j} y > 0, \|\Pi_{L,j} y\|^2 \leq u] \times P[\Delta_{J} \Pi_{K,j} \leq 0, \|\Pi_{K,j} y\|^2 \leq v]
\]

Next, using the independence of norms and angles

\[
P[\Gamma'_{J} \Pi_{L,j} y > 0, \|\Pi_{L,j} y\|^2 \leq u] = P[\Gamma'_{J} \Pi_{L,j} y > 0] P[\|\Pi_{L,j} y\|^2 \leq u]
\]

and

\[
P[\Delta_{J} \Pi_{K,j} \leq 0, \|\Pi_{K,j} y\|^2 \leq v].
\]
So, letting $k = \# J$

\[
P[\hat{J}(y) = J, \|\Pi_{L_J}y\|^2 \leq u, \|\Pi_{K_{\hat{J}}}y\|^2 \leq v] = P[\Gamma'_J\Pi_{L_J}y > 0] \times P[\|\Pi_{L_J}y\|^2 \leq u] \times P[\Delta_{\hat{J}}\Pi_{K_{\hat{J}}} \leq 0] \times P[\|\Pi_{K_{\hat{J}}}y\|^2 \leq v]
\]

\[
= P[\chi^2_{k-m} \leq u]P[\chi^2_{n-k} \leq v]P[\hat{J} = J]
\]
in which the independence of $\Pi_{L_J}y$ and $\Pi_{K_{\hat{J}}}$ has been used again. Relation (*) then follows by writing

\[
\sum_{J} = \sum_{k=0}^{n-m} \sum_{\# J = k} \cdot \sum_{J}.
\]

So, for the case of known $\sigma^2$,

\[
P_{\theta} [\Lambda_{01} > c] = P_0 \left[ \frac{1}{\sigma^2} \|\hat{\theta} - \theta^o\|^2 > c \right] = \sum_{k=m}^{n} P[\chi^2_{k-m} > c]q(n, k)
\]

for all $\theta \in L^\perp$, and this may set equal to any given $\alpha$, by appropriate choice of $c$. For unknown $\sigma^2$, recall that if $U$ and $V$ are independent chi-squared variables with $r$ and $s$ degrees of freedom, then

\[
\frac{U}{U+V} \sim \beta\left(\frac{r}{2}, \frac{s}{2}\right).
\]

So,

\[
P_{\theta} \left[ \frac{\|\hat{\theta}^o - \hat{\theta}\|^2}{\|\hat{\theta}^o - \hat{\theta}\|^2 + \|y - \hat{\theta}\|^2} > c \right] = \sum_{k=m}^{n} P \left[ \beta\left(\frac{k-m}{2}, \frac{n-k}{2}\right) > c \right]q(n, k)
\]

for all $\theta \in L^\perp$.

**Remark.** This material is adapted from [1].

**References**