Model Selection And Sharp Asymptotic Minimaxity

Zheyang Wu
Jointly with Harrison Zhou

Yale School of Medicine
Yale University

May 16, 2009

Problem description

- **Multivariate normal mean problem:**

$$y_i = \theta_i + \sigma_n z_i, \;\; z_i \overset{i.i.d.}{\sim} N(0,1), i = 1, ..., n.$$

- **Sparse parameter space:** $\theta \in \Theta_{n,p}(\eta_n)$, $\Theta_{n,p}(\eta_n)$ is one of

$$\begin{aligned}
l_0 \text{ ball} \quad &: \quad \{\theta \in \mathbb{R}^n : \|\theta\|_0 \leq \eta_n n\} \\
l_p \text{ ball} \quad &: \quad \left\{\theta \in \mathbb{R}^n : \sum_{i=1}^{n} |\theta_i|^p \leq \eta_n^p n, 0 < p < 2\right\} \\
m_p \text{ ball} \quad &: \quad \left\{\theta \in \mathbb{R}^n : |\theta|_{[k]} \leq \eta_n \left(\frac{n}{k}\right)^{1/p}, 0 < p < 2, k = 1, ..., n\right\}
\end{aligned}$$

where $\eta_n \to 0$.

- **Goal:** to estimate $\theta = (\theta_1, ..., \theta_n)$.

Main theorems

- Penalized estimation: $\hat{\theta} = \arg\min_{\mu} \left[\|y-\mu\|_2^2 + \mathrm{Pen}\left(\|\mu\|_0\right) \right]$.
- Let $R_n\left(\Theta_{n,p}\left(\eta_n\right)\right)$ be the minimax risk, $\|\mu\|_0 = k$, $\mathrm{Pen}\left(k\right) = \sum_{i=1}^{k} u_i^2$,

Theorem

For some $0 < \varepsilon < 1$, *and any constant* c', *let*
$2\log\frac{n}{i} - (1-\varepsilon)\log\log\frac{n}{i} \leq u_i^2 \leq 2\log\frac{n}{i} + c'\log\log n$, *then*
$\sup_{\theta\in\Theta_{n,p}(\eta_n)} \mathbb{E}\left\|\hat{\theta}-\theta\right\|_2^2 \sim R_n\left(\Theta_{n,p}\left(\eta_n\right)\right)$.

Theorem

Let $u_i^2 = c_n \log\frac{n}{i}$, $c_n \to c > 2$, *then*
$\sup_{\theta\in\Theta_{n,p}(\eta_n)} \mathbb{E}\left\|\hat{\theta}-\theta\right\|_2^2 \sim c^* R_n\left(\Theta_{n,p}\left(\eta_n\right)\right)$, *where* $c^* = \left(\frac{c}{2}\right)^{1-p/2}$.

Interpretation and significance

Interpretation

- Sharp asymptotic minimaxity if $\mathrm{Pen}(k) = 2k \log \frac{n}{k}$.
- Not sharp asymptotic minimaxity if $\mathrm{Pen}(k) = ck \log \frac{n}{k}$, $c \neq 2$.

Significance

- Sharp asymptotic minimaxity for $\mathrm{Pen}(k)$
  - Foster and Stine (1999): $2 \sum_{i=1}^{k} \log \frac{n}{i}$.
  - George and Foster (2000): $2 \sum_{i=1}^{k} \log \left(\frac{n+1}{i} - 1 \right)$.
  - Berge and Massart (2001): $2k \log \frac{n}{k}$.
  - ABDJ (2006): $\sum_{i=1}^{k} \left(\tilde{\Phi}^{-1} \left(\frac{q_n i}{2n} \right) \right)^2$.
- Not sharp asymptotic minimaxity for $\mathrm{Pen}(k)$
  - Tibshirani and Knight (1999): $4 \sum_{i=1}^{k} \log \frac{n}{i}$.
  - Abramovich, Grinshtein and Pensky (2007): $ck \log \frac{n}{k}$, some $c > 2$.