Week 10
April 3 – April 5

Lecture 19. **Linear Minimaxity**
We consider a general setting

$$y_i = \theta_i + \epsilon z_i,\ z_i \overset{i.i.d.}{\sim} N(0,1),\ \theta \in \Theta$$

where Θ is an ellipsoid in $l_2(\mathbb{N})$:

$$\Theta = \left\{ \theta : \sum_i a_i^2 \theta_i^2 \le M \right\}.$$

The set Θ is orthosymmetric and convex. Without loss of generality, assume that $\epsilon = 1$. (**John Hartigan's calculation of avoiding minimax theorem**) Recall that

$$R_L(\Theta) = \inf_{(c_i)} \sup_{\sum_i a_i^2\theta_i^2 \le M} \sum_{i=1}^{\infty} c_i^2 + \theta_i^2 (1-c_i)^2$$

which is equal to

$$\begin{aligned} & \inf_{(c_i)} \left[\sup_i (1-c_i)^2 \frac{M}{a_i^2} + \sum_{i=1}^{\infty} c_i^2 \right] \\ = & \inf_{\lambda} \inf_{\left\{ \sup_i (1-c_i)^2 \frac{1}{a_i^2} = \lambda \right\}} \left[\sup_i (1-c_i)^2 \frac{M}{a_i^2} + \sum_{i=1}^{\infty} c_i^2 \right] \end{aligned}$$

When $\sup_i (1-c_i)^2 \frac{1}{a_i^2} = \lambda^{-2}$, then $c_i \ge (1 - a_i/\lambda)_+$ for all i and equality holds for at least one i, which implies

$$\inf_{\left\{ \sup_i (1-c_i)^2 \frac{1}{a_i^2} = \lambda \right\}} \left[\sup_i (1-c_i)^2 \frac{M}{a_i^2} + \sum_{i=1}^{\infty} c_i^2 \right] = M\lambda^{-2} + \sum (1 - a_i/\lambda)_+^2$$

Thus

$$R_L(\Theta) = \inf_{\lambda} \left[M\lambda^{-2} + \sum (1 - a_i/\lambda)_+^2 \right]$$

then solve the quadratic minimization for λ. The solution is

$$\sum_{i=1}^{\infty} a_i (\lambda - a_i)_+ = M$$

For general ϵ, we have

$$\epsilon^2 \sum a_i (\lambda - a_i)_+ = M$$

and

$$c_i = (1 - a_i/\lambda)_+$$

which corresponds to the prior

$$\theta_i \sim N\left(0, \tau_i^2\right), \tau_i^2 = (\lambda/a_i - 1)_+$$

The corresponding maximum is

$$R_L(\Theta, \epsilon) = \epsilon^2 \sum (1 - a_i/\lambda)_+ .$$

If a_i is a monotone sequence, let $I = \max\{i, 1 \geq a_i/\lambda\}$. Then

$$R_L(\Theta) = \inf_{\lambda} \left[\frac{M}{\epsilon^2} \lambda^{-2} + \sum_{i=1}^{I} (1 - a_i/\lambda)^2 \right]$$

The inf is achieved at

$$\lambda = \frac{\sum_{i=1}^{I} a_i^2 + \frac{M}{\epsilon^2}}{\sum_{i=1}^{I} a_i},$$

and

$$\begin{aligned} R_L(\Theta, \epsilon) &= \epsilon^2 \sum_{i=1}^{I} (1 - a_i/\lambda) \\ &= \epsilon^2 \left(I - \frac{\sum_{i=1}^{I} a_i}{\lambda} \right) \end{aligned}$$

Homework problem

Now consider the application of this result to Sobolev ellipsoid.

$$\mathcal{F} = \left\{ f : \sum_{k=1}^{\infty} (2\pi k)^{2m} \left(\theta_{2k}^2 + \theta_{2k+1}^2 \right) \leq M \right\}.$$

Show that

$$R_L(\Theta, \epsilon) \sim \epsilon^2 \pi^{-1} \frac{m}{m+1} \left[\frac{(m+1)(2m+1)}{m} \frac{\pi M}{\epsilon^2} \right]^{1/(2m+1)} \pi^{-2r} = P_r M^{(1-2r)} \pi^{-2r} (\epsilon)^{4r},$$

where $r = m/(2m+1)$ and *Pinsker constant* P_r is

$$P_r = r(1-r)^{r-1} (2-r)^{-r}.$$

Lecture 20. Adaptive Minimaxity
Model:

$$y_i = \theta_i + \epsilon z_i,\ z_i \overset{i.i.d.}{\sim} N(0,1),\ \theta \in \Theta_M$$

where Θ_M is an ellipsoid in $l_2(\mathbb{N})$

$$\Theta(m, M) = \left\{ f : \sum_{i=1}^{\infty} a_i^2 \theta_i^2 \le M,\ a_{2k} = a_{2k+1} = (2\pi k)^m \right\}$$

In this lecture, we apply blockwise James-Stein estimator to achieve the sharp linear adaptive minimaxity.

Lemma: $X \sim N_d(\mu, \epsilon I)$

$$r\left(\widehat{\mu}^{JS}, \mu\right) \le 2\epsilon^2 + \frac{(d-2)\epsilon^2 \|\mu\|^2}{(d-2)\epsilon^2 + \|\mu\|^2} \le 2\epsilon^2 + \frac{d\epsilon^2 \|\mu\|^2}{d\epsilon^2 + \|\mu\|^2}$$

Proof: Without loss of generality, we assume that $\epsilon = 1$. Since $\|X\|^2$ can be seen as a mixture of χ^2_{d+2N} and $N \sim$Poisson$\left(\|\mu\|^2/2\right)$, and

$$r\left(\widehat{\mu}^{JS}, \mu\right) = d - (d-2)^2 E_\mu \|X\|^{-2},$$

then

$$r\left(\widehat{\mu}^{JS}, \mu\right) = d - (d-2)^2 E\frac{1}{d-2+2N} \underset{\text{Jensen}}{\le} d - (d-2)^2 \frac{1}{d-2+\|\mu\|^2}.$$

Remark: For $\epsilon = 1$, this Lemma implies

$$r\left(\widehat{\mu}^{JS}, \mu\right) \le 2 + r\left(\widehat{\mu}^{IS}, \mu\right).$$

where $r\left(\widehat{\mu}^{IS}, \mu\right) = \inf_c r(\widehat{\mu}_c, \mu)$.

Proof of adaptive minimaxity: Define

$$B_b = \left\{ i : \left[a^{b-1}\right] \le i < \left[a^b\right],\ a = 1 + 1/\log n \right\}.$$

Set L such that $a^{L-1} > 3\log n + 1$, for instance, $L = 4$. Starting from the L-th block, we apply James-Stein estimator to the observations in each block. Then we have

$$r_\epsilon\left(\widehat{\theta}^{BJS}, \theta\right) \le \left(2\log_a n - 2L + a^L\right)\frac{1}{n} + \sum_{b=L}^{\log_a n} r\left(\widehat{\theta}^{IS}_{(b)}, \theta_{(b)}\right) + \sum_{i \ge n} \theta_i^2$$

Let L be finite. It is easy to see

$$\left(2\log_a n - 2L + a^L\right)\frac{1}{n} + \sum_{i \ge n} \theta_l^2 = O\left(n^{-2m}\right) = o\left(n^{-2m/(2m+1)}\right).$$

Note that

$$\begin{aligned}\sup_{\Theta(m,M)} \sum_{b=L}^{\log_a n} r\left(\widehat{\theta}_{(b)}^{IS}, \theta_{(b)}\right) &\leq \sup_{\sum_b (\pi a)^{2(b-1)m} \|\theta\|_{(b)}^2 \leq M} \sum_{b=L}^{\log_a n} r\left(\widehat{\theta}_{(b)}^{IS}, \theta_{(b)}\right) \\ &\sim R_L\left(\Theta\left(m, M\right), \epsilon\right)\end{aligned}$$

by using the Lagrangian multiplier.

Detailed calculations using the Lagrangian multiplier:

Recall that

$$R_L\left(\Theta\right) = \sup_{\theta \in \Theta} \left\{ \sum_i \frac{\epsilon^2 \theta_i^2}{\epsilon^2 + \theta_i^2} : \sum a_i^2 \theta_i^2 \leq M \right\}$$

The maximum is attained at

$$\theta_i^2 = \epsilon^2 \left(\lambda / a_i - 1\right)_+ .$$

where the Lagrangian multiplier parameter $\lambda = \lambda_\epsilon$ is uniquely determined by the equation

$$\sum a_i^2 \theta_i^2 = \epsilon^2 \sum a_i \left(\lambda - a_i\right)_+ = M.$$

and maximum is

$$R_L\left(\Theta, \epsilon\right) = \sum \frac{\epsilon^2 \theta_i^2}{\epsilon^2 + \theta_i^2} = \epsilon^2 \sum \left(1 - a_i/\lambda\right)_+$$

For the Sobolev ball with $a_{2k} = a_{2k+1} = (2\pi k)^m$, we have

$$\begin{aligned}R_L\left(\Theta, \epsilon\right) &= \epsilon^2 \sum_{i=1}^{I} \left(1 - a_i/\lambda\right) \\ &\approx \epsilon^2 \left(I - \frac{\sum_{i=1}^{I} \left(\pi i\right)^m}{\lambda}\right)\end{aligned}$$

where

$$\begin{aligned}\left(\pi I\right)^m &\sim \lambda, \text{ i.e., } I \sim \lambda^{1/m}/\pi \\ \lambda^{(2m+1)/m} &\sim \frac{\left(m+1\right)\left(2m+1\right)}{m} \pi M/\epsilon^2\end{aligned}$$

Key observation: Let $a^{b-1} = i$, i.e., $a^b = ai$, then

$$\sum_i^{ai} i^m \leq \sum_i^{ai} j^m \leq \sum_i^{ai} \left(ai\right)^m = a^m \sum_i^{ai} i^m$$

since $a^m = 1 + o\left(1\right)$.

In our setting, we need to consider

$$\sup_{\theta\in\Theta}\left\{\sum_b \frac{|B_b|\,\epsilon^2\,\|\theta\|_{(b)}^2}{|B_b|\,\epsilon^2+\|\theta\|_{(b)}^2} : \sum_b (\pi a)^{2(b-1)m}\,\|\theta\|_{(b)}^2 \le M\right\}.$$

Simple calculus shows that the maximum is attained at

$$\|\theta\|_{(b)}^2 = |B_b|\,\epsilon^2\left(\lambda/\left(\pi a\right)^{(b-1)m}-1\right)_+.$$

the Lagrangian multiplier parameter $\lambda=\lambda_\epsilon$ is uniquely determined by the equation

$$|B_b|\,\epsilon^2\sum_{b=L}^{\log_a n}(\pi a)^{(b-1)m}\left(\lambda-(\pi a)^{(b-1)m}\right)_+ = M.$$

Then

$$R_{BL}\left(\Theta,\epsilon\right)=\epsilon^2\,|B_b|\sum_{b=L}^{B}\left(1-\frac{(\pi a)^{(b-1)m}}{\lambda}\right)$$

Roughly we have

$$(\pi a)^{(B-1)m}\sim\lambda\sim(\pi I)^m$$

i.e.,

$$I\sim a^{(B-1)}$$

because

$$|B_b|\,(\pi a)^{(b-1)m}\le\sum_{i=a^{b-1}}^{a^b-1}(\pi i)^m\le|B_b|\left(\pi a^b\right)^m=\left(1+o\left(1\right)\right)|B_b|\,(\pi a)^{(b-1)m}.$$

An application of Leung and Barron (2004). (more details in class).

Homework problem.
Observe $y\sim N\left(\theta,\epsilon\right)$. Consider a set of priors

$$\mathcal{P}=\left\{G:\int\theta^2G\left(d\theta\right)\le\tau^2\right\}.$$

For any π, define

$$B\left(G\right)=\inf_{\widehat{\theta}}\int E_{y|\theta}\left(\theta-\widehat{\theta}\right)^2G\left(d\theta\right)$$

Show that

$$\sup_{\mathcal{P}}\left\{B\left(G\right)\right\}=\epsilon^2\tau^2/\left(\epsilon^2+\tau^2\right)$$

equality holds if and only if G is Gaussian.