S&DS 411 (Fall 2019) Selected Topics in Statistical Decision Theory **Lecture: 6**

Sparse Vector Estimation

Lecturer: Harrison Zhou *Scribe: Cynthia Yue, Oct 2, 2019*

1 Sparse Vector Estimation

We will look at the lower bound for sparse vector estimation and Bayesian contraction.

Let us consider the model, where $Z_i \sim \mathcal{N}(0,1)$ and $i = 1, \dots, p$:

$$Y_i = \theta_i + \frac{1}{\sqrt{n}} Z_i.$$

We will work under the assumption for the parameter space that

$$\Theta = \{\theta : \|\theta\|_0 \leq s\}.$$

From the previous lecture, we found the risk upper bound to be:

$$\sup_{\theta \in \Theta} \mathbb{E}\|\hat{\theta} - \theta\|^2 \leq c \frac{s \log \frac{ep}{s}}{n}.$$

2 Lower Bound

We have the following goal:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}\|\hat{\theta} - \theta\|^2 \geq c \frac{s \log \frac{ep}{s}}{n}$$

under the assumption $p^\epsilon \leq s \leq p^{1-\epsilon}$ where ϵ small > 0.

Let us review. For $i = 1, \dots, s$,

$$\theta_i = \begin{cases} \frac{1}{\sqrt{n}} & \frac{1}{2} \\ 0 & \frac{1}{2}. \end{cases}$$

We showed that

$$\sup_{\theta \in \Theta_{sub}} \mathbb{E}\|\hat{\theta} - \theta\|^2 \geq c \frac{s}{n}.$$

For the lower bound, we consider some first ideas. We could change the support to $s \log \frac{ep}{s}$ and change $\theta_i = \begin{cases} \frac{1}{\sqrt{n}} \sqrt{\log \frac{ep}{s}} \\ 0. \end{cases}$ However, this is not in the parameter set. Instead, another way is to extend the sub-parameter space.

Let us consider the sub-parameter space Θ_{sub}

$$\theta_i = \begin{cases} \frac{1}{\sqrt{n}}\sqrt{\log\frac{p}{s}} & i \in T \\ 0 & i \notin T. \end{cases}$$

We have $T \subset \{1,\ldots,p\}$ and $\mathrm{Card}(T) = s$. There are $\binom{p}{s}$ ways to get T.

Let us consider the following facts:

1.

$$\max_{1\le i\le p}\{\frac{1}{\sqrt{n}}Z_i\} = (1+O(1))\frac{1}{\sqrt{n}}\sqrt{2\log p}.$$

2.

$$|Z|_{[1]} \ge |Z|_{[2]} \ge \ldots \ge |Z|_{[n]}$$
$$\frac{1}{\sqrt{n}}Z_{[s]} = (1+O(1))\frac{1}{\sqrt{n}}\sqrt{2\log\frac{p}{s}}.$$

It is impossible to identify because the signal is too weak, smaller than the noise. The signal strength is smaller than T, so we can't identify T.

With Bayesian contraction, we can consider the prior π:

$$\theta_i = \begin{cases} \frac{1}{\sqrt{n}}\sqrt{\log\frac{p}{s}} & q = \frac{s}{2p} \\ 0 & 1-q. \end{cases}$$

constrained on Θ i.i.d.

$$\begin{aligned}\mathbb{P}\left\{\sqrt{\mathrm{Bin}(p,\frac{s}{2p})} - \sqrt{s} \ge 0\right\} &= \mathbb{P}\left\{\sqrt{\mathrm{Bin}(p,\frac{s}{2p})} - \sqrt{\frac{s}{2}} \ge \sqrt{s} - \frac{\sqrt{s}}{2}\right\} \\ &\le \mathrm{e}^{(-c's)} \\ &\le \mathrm{e}^{(-cp^{\epsilon})} \\ &\le \mathrm{e}^{(-cn^{\epsilon'})} \\ &\le n^{-100}.\end{aligned}$$

Let us prove the lower bound:

Proof.

$$\begin{aligned}
\sup_\theta \mathbb{E}\|\hat{\theta}_B - \theta\|^2 &\geq \int_\Theta \mathbb{E}\|\hat{\theta}_B - \theta\|^2 d\pi / \pi(\Theta) \quad \text{where } \pi(\Theta) \leq 1 \\
&\geq \int_\Theta \mathbb{E}\|\hat{\theta}_B - \theta\|^2 d\pi \\
&= \underbrace{\int \mathbb{E}\|\hat{\theta}_B - \theta\|^2 d\pi}_{\text{dominating term}} - \underbrace{\int_{\Theta^c} \mathbb{E}\|\hat{\theta}_B - \theta\|^2 d\pi}_{\text{low-order term}} .
\end{aligned}$$

$$\int_{\Theta^c} \mathbb{E}\|\hat{\theta}_B - \theta\|^2 d\pi \leq 4 \log \frac{p}{s} \pi(\Theta^c) = O(n^{-2})$$

$$\text{where } \|\theta\|^2 \leq p \left(\frac{1}{\sqrt{12}n} \sqrt{\log \frac{p}{s}} \right)^2 = \frac{p}{n} \log \frac{p}{s}.$$

$$\int \mathbb{E}\|\hat{\theta}_B - \theta\|^2 d\pi \geq \int \mathbb{E}\|\mathbb{E}(\theta|Y) - \theta\|^2 d\pi.$$

$$\theta_i = \begin{cases} \underbrace{\frac{1}{\sqrt{n}} \sqrt{\log \frac{p}{s}}}_{\text{a}} & q = \frac{s}{2p} \\ 0 & 1 - q. \end{cases}$$

$$\begin{aligned}
\mathbb{P}(\theta_i = 0|Y_i) &= \frac{(1-q)\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(Y_i - 0)^2}{2(\frac{1}{\sqrt{n}})^2} n}}{(1-q)\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(Y_i-0)^2}{2} n} + q \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(Y_i - a)^2}{2} n}} \\
&= \frac{1-q}{(1-q) + q \mathrm{e}^{Y_i a n} \mathrm{e}^{-\frac{na^2}{2}}}. \\
\mathbb{P}(\theta_i = a|Y_i) &= 1 - \mathbb{P}(\theta_i = 0|Y_i).
\end{aligned}$$

Therefore, we have

$$\int \mathbb{E}\|\hat{\theta}_B - \theta\|^2 d\pi = \sum_{i=1}^p \left[q \Big(\mathbb{E}(\theta_i|Y) - a \Big)^2 + (1-q) \Big(\mathbb{E}(\theta_i|Y) - 0 \Big)^2 \right].$$

For the posterior norm, we have

$$\begin{aligned}
\mathbb{E}(\theta_i|Y_i) &= \mathbb{P}(\theta_i = a|Y_i)a + \mathbb{P}(\theta_i = 0|Y_i)0 \\
&= \mathbb{P}(\theta_i = a|Y_i)a.
\end{aligned}$$

If we plug this in, we get the desired lower bound:

$$\begin{aligned}
\int \mathbb{E}\|\hat{\theta}_B - \theta\|^2 d\pi &= \sum_{i=1}^{p} \left[q\Big(\mathbb{E}(\theta_i|Y) - a\Big)^2 + (1-q)\Big(\mathbb{E}(\theta_i|Y) - 0\Big)^2 \right] \\
&\geq q \sum_{i=1}^{p} \mathbb{E}_{Y_i|\theta_i=a}\Big(\mathbb{E}(\theta_i|Y_i) - a\Big)^2 \\
&= q \sum_{i=1}^{p} \mathbb{E}_{Y_i|\theta_i=a}\Big(\mathbb{P}(\theta_i = a|Y_i)a - a\Big)^2 \\
&= qa^2 \sum_{i=1}^{p} \underbrace{\mathbb{E}_{Y_i|\theta_i=a}\Big(\mathbb{P}(\theta_i = 0|Y_i)\Big)^2}_{\leq \text{C}} \\
&\geq qa^2 pC.
\end{aligned}$$

$\square$

Now,

$$\begin{aligned}
Y_i &= a + \frac{1}{\sqrt{n}} Z_i. \\
\mathbb{P}(\theta_i = 0|Y_i) &= \frac{1-q}{1-q+\frac{s}{2p}\mathrm{e}^{(a+\frac{1}{\sqrt{n}}Z_i)an - \frac{1}{2}na^2}} \\
&= \frac{1-q}{1-q+\frac{1}{2}\mathrm{e}^{\sqrt{\log\frac{p}{s}}Z_i - \frac{1}{2}\log\frac{p}{s}}} \\
&\to 1 \text{ as } p \to +\infty
\end{aligned}$$

so we can find constant C_1.

Another way to do this is to use Fano's lemma.

For Fano's lemma, if we have the joint distribution $Y = (Y_1, \ldots, Y_p)$, we can use the lemma to show that if we have $\{\theta_0, \theta_1, \ldots, \theta^{M^*}\} \subset \Theta$ and if $d(\theta^i, \theta^j) \geq 2\epsilon$ where we choose $\epsilon = \sqrt{\frac{s\log\frac{p}{s}}{n}}\gamma$, and $\frac{1}{M^*}\sum_{k=1}^{M^*} k(P_{\theta^k}|P_{\theta^0}) \leq c\log M*$ for some $0 < c < \frac{1}{8}$, then for any $\hat{\theta}$,

$$\sup_{d\leq k\leq M*} \mathbb{E}d^2(\hat{\theta}, \theta_k) \geq \epsilon^2 \underbrace{\frac{\sqrt{M*}}{1+\sqrt{M*}}(1-2c-\sqrt{\frac{2c}{\log M*}})}_{\geq\frac{1}{2}}$$

For Θ_{sub},

$$\begin{aligned}
&\theta^0 = 0 \\
&\theta_i = \begin{cases} \sqrt{\log\frac{p}{s}\gamma}/\sqrt{n} & i \in T \\ 0 & i \notin T \end{cases} \\
&T : \binom{p}{s} \text{ and } \mathrm{Card}(T) = s
\end{aligned}$$

Claim. By Varshomov-Gilbert Lemma, we can find $\theta^1, \ldots, \theta^{M*}$ with $\log M^* \geq c_1 s\log\frac{p}{s}$ for some c_1 s.t.

$$d(\theta^i, \theta^j) \geq 2\sqrt{\frac{s\log\frac{p}{s}}{n}\frac{1}{16}}.$$