36-789: Topics in High Dimensional Statistics II **Fall 2015**

Lecture 7: Nov. 17

Lecturer: Alessandro Rinaldo *Scribes: Lingxue Zhu*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

7.1 Assouad Method

Theorem 7.1 (Assouad) *Suppose* $\exists m \in \mathbb{N}$, *a sub-family* $\{P_v : v \in \{-1,1\}^m\} \subseteq \mathcal{P}$, *and a function* $V : \theta(\mathcal{P}) \to \{-1,1\}^m$, *such that*

$$w\left(d(\theta, \theta(P_v))\right) \geq 2\delta \sum_{j=1}^{m} I_{\{V(\theta)_j \neq v_j\}}\,,\ \forall v \in \{-1,1\}^m\,.$$

That is, for $\forall v \in \{-1,1\}^m$, *there exists* $P_v \in \mathcal{P}$, *such that* $\forall v \neq v'$,

$$w(d(\theta(P_v), \theta(P_{v'}))) \geq 2\delta \sum_{j=1}^{m} I_{\{v_j \neq v'_j\}} = 2\delta\, d_H(v, v')\,.$$

Then we have

$$\inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[w(d(\hat{\theta}, \theta(P))) \right] \geq m\delta \min_{\substack{v,v' \in \{-1,1\}^m \\ d_H(v,v')=1}} \{1 - d_{TV}(P_v, P_{v'})\}\,.$$

Proof: Let $V \sim \text{Unif}(\{-1,1\}^m)$ and $P_{\pm j}$ be the conditional distribution of (X, V) given $V_j = \pm 1$. Notice that

$$P_{\pm j} = \frac{1}{2^{m-1}} \sum_{v \in \{-1,1\}^m} P_{v,\pm j}\,,$$

where $P_{v,\pm j}$ is P_v with $v_j = \pm 1$. Then $\forall \hat{\theta}$,

$$\begin{aligned}
\sup_{P \in \mathcal{P}} \mathbb{E}_P \left[w(d(\hat{\theta}, \theta(P))) \right] &\geq \frac{1}{2^m} \sum_{v \in \{-1,1\}^m} E_{P_v} \left[w(d(\hat{\theta}, \theta(P_v))) \right] \\
&\geq \frac{1}{2^m} \sum_{v \in \{-1,1\}^m} 2\delta \sum_{j=1}^{m} P_v \left(V(\hat{\theta})_j \neq v_j \right) \\
&= 2\delta \sum_{j=1}^{m} \frac{1}{2^m} \left[\sum_{\substack{v \in \{-1,1\}^m \\ v_j = 1}} P_v \left(V(\hat{\theta})_j \neq v_j \right) + \sum_{\substack{v \in \{-1,1\}^m \\ v_j = -1}} P_v \left(V(\hat{\theta})_j \neq v_j \right) \right] \\
&= 2\delta \sum_{j=1}^{m} \frac{1}{2} \left[P_{+j} \left(V(\hat{\theta})_j \neq 1 \right) + P_{-j} \left(V(\hat{\theta})_j \neq -1 \right) \right] \\
&\geq 2\delta \sum_{j=1}^{m} [1 - d_{TV}(P_{+j}, P_{-j})] \geq 2\delta m \min_j [1 - d_{TV}(P_{+j}, P_{-j})]
\end{aligned}$$

Finally, observe that

$$d_{TV}(P_{+j}, P_{-j}) \leq \frac{1}{2^{m-1}} \sum_{v} d_{TV}(P_{v,+j}, P_{v,-j}) \leq \max_{v,j} d_{TV}(P_{v,+j}, P_{v,-j}) = \max_{\substack{v,v' \\ d_H(v,v')=1}} d_{TV}(P_v, P_{v'}) \,.$$

■

As a consequence, if for all v, v' such that $d_H(v, v') = 1$, we have

1. $d_{TV}(P_v, P_{v'}) \leq \alpha$, then the lower bound is $\delta \frac{m}{2}(1-\alpha)$.
2. $H^2(P_v, P_{v'}) \leq \alpha < 2$, then the lower bound is $\delta \frac{m}{2} \left[1 - \sqrt{\alpha(1-\alpha/4)}\right]$.
3. $KL(P_v, P_{v'}) \leq \alpha$ or $\mathcal{X}^2(P_v, P_{v'}) \leq \alpha$, then the lower bound is $\delta \frac{m}{2} \max\left\{\frac{1}{2}e^{-\alpha},\, 1 - \sqrt{\frac{\alpha}{2}}\right\}$.

Remark The Assouad lower bound can also be written in the following form: for $p > 0$,

$$\sup_{v \in \{-1,1\}^m} \mathbb{E}_{P_v} \left[d^p(\hat{\theta}, \theta(P_v))\right] \geq \min_{\substack{v,v' \in \{-1,1\}^m \\ d_H(v,v') \geq 1}} \frac{d^p(\theta(P_v), \theta(P_{v'}))}{d_H(v, v')} \cdot \frac{m}{2} \min_{\substack{v,v' \in \{-1,1\}^m \\ d_H(v,v')=1}} \left[1 - d_{TV}(P_v, P_{v'})\right] \,.$$

Proof: Let $V(\hat{\theta}) \in \{-1,1\}^m$ such that

$$V(\hat{\theta}) = v^* \quad \text{if} \quad d(\hat{\theta}, \theta(P_{v^*})) = \min_{v \in \{-1,1\}^m} d(\hat{\theta}, \theta(P_v)) \,.$$

Then for any $v \in \{-1,1\}^m$, by triangle inequality, we have

$$d\left(\theta\left(P_{V(\hat{\theta})}\right), \theta(P_v)\right) \leq 2 \cdot d(\hat{\theta}, \theta(P_v)) \,.$$

Therefore,

$$2^p \mathbb{E}_{P_v} \left[d^p(\hat{\theta}, \theta(P_v))\right] \geq \mathbb{E}_{P_v} \left[d^p\left(\theta\left(P_{V(\hat{\theta})}\right), \theta(P_v)\right)\right] \geq 2\delta\, \mathbb{E}_{P_v} \left[d_H(V(\hat{\theta}), v)\right]$$

where $2\delta = \min_{v \neq v'} \dfrac{d^p(\theta(P_v), \theta(P_{v'}))}{d_H(v, v')}$. Then proceed as before to reach the desired result. ■

7.2 Minimax Confidence Ball

Suppose $X \sim N_n(\theta, \sigma_n^2 I_n)$, we want to construct a confidence ball $B_n(X)$ for θ, such that

$$\inf_{\theta \in \mathbb{R}^n} \mathbb{P}_\theta(\theta \in B_n) \geq 1 - \alpha \,.$$

Note that $\frac{||X-\theta||^2}{\sigma_n^2} \sim \mathcal{X}_n^2$, the simplest confidence ball is a $\mathcal{X}^2$ ball:

$$B_n = \left\{\theta \in \mathbb{R}^n : ||X - \theta||^2 \leq \sigma_n^2 \mathcal{X}_{n,1-\alpha}^2\right\} \,,$$

where $\mathcal{X}_{n,1-\alpha}^2$ is the $1-\alpha$ quantile of $\mathcal{X}_n^2$. The radius is deterministic, which is in the order of $\sigma_n \sqrt{n}$.

Lepski proposed another way to construct the confidence ball B_n as follows. First, we test the hypothesis

$$H_0 : \theta = 0 \;\; v.s. \;\; H_1 : \theta \neq 0 \,.$$

If we accept H_0, then the ball is centered at 0, with radius $\sigma_n n^{1/4}$. Otherwise, if we reject H_0, then we use the $\mathcal{X}^2$ ball with radius $\sigma_n\sqrt{n}$. This gives a valid $(1-\alpha)$ confidence ball with random radius.

In fact, in general, the rate of $\sigma_n\sqrt{n}$ is optimal; but in some specific scenario, it might be possible to attain $\sigma_n n^{1/4}$.

Claim Let S_n be the random radius of a ball B_n centered at any estimator $\hat{\theta}$ of θ that is a $(1-\alpha)$ confidence ball, then there exists a constant C, such that

(i) $\mathbb{E}_\theta[S_n] \geq C\sigma_n n^{1/4}$, for any $\theta \in \mathbb{R}^n$.

(ii) $\mathbb{E}_\theta[S_n] \geq C\sigma_n n^{1/2}$, for some $\theta \in \mathbb{R}^n$.

Now we prove the first claim.

Theorem 7.2 *Let $\alpha \in (0, \frac{1}{2})$, $B_n = \{\theta \in \mathbb{R}^n : ||\hat{\theta} - \theta|| \leq s_n\}$ for any estimator $\hat{\theta}$, such that B_n is an honest confidence ball:*

$$\inf_{\theta\in\mathbb{R}^n} \mathbb{P}_\theta(\theta \in B_n) \geq 1-\alpha\,.$$

Then $\forall\epsilon \in (0, \frac{1}{2}-\alpha)$,

$$\inf_{\theta\in\mathbb{R}^n} \mathbb{E}_\theta[S_n] \geq \sigma_n n^{1/4}(1-2\alpha-2\epsilon)\left(\log(1+\epsilon^2)\right)^{1/4}\,.$$

Proof: Let $a_n = \dfrac{\sigma_n}{n^{1/4}}\left(\log(1+\epsilon^2)\right)^{1/4}$, and define $\Omega = \{\theta \in \mathbb{R}^n : |\theta_i| = a_n, i = 1, ..., n\}$, hence $|\Omega| = 2^n$. Let f_θ be the density of $N_n(\theta, \sigma_n^2 I_n)$, and $q = \dfrac{1}{2^n}\sum_{\theta\in\Omega} f_\theta$ be the density of a mixture distribution, then

$$\int |q - f_0| \leq \sqrt{\int \frac{q^2}{f_0} - 1}\,.$$

In addition, let $E_1, \cdots, E_n \overset{i.i.d.}{\sim}$ Rademacher, then

$$\begin{aligned}
\int \frac{q^2}{f_0} &= \left(\frac{1}{2^n}\right)^2 \sum_{\theta,\theta'\in\Omega} \int \frac{f_\theta f_{\theta'}}{f_0} \\
&= \left(\frac{1}{2^n}\right)^2 \sum_{\theta,\theta'\in\Omega} \exp\left\{\frac{\langle\theta,\theta'\rangle}{\sigma_n^2}\right\} \\
&= \mathbb{E}\left[\exp\left\{\frac{a_n^2\sum_{i=1}^n E_i}{\sigma_n^2}\right\}\right] \\
&= \prod_{i=1}^n \mathbb{E}\left[\exp\left\{\frac{a_n^2 E_i}{\sigma_n^2}\right\}\right] = \left[\cosh\left(\frac{a_n^2}{\sigma_n^2}\right)\right]^n \\
&\leq \exp\left\{\frac{a_n^4}{\sigma_n^2} n\right\}
\end{aligned}$$

So $\int |f_0 - q| \leq \sqrt{\exp\left\{\frac{a_n^4}{\sigma_n^2} n\right\} - 1} := \epsilon_n$. For any event A, let Q, P_0 be the measure of q, f_0, then

$$P_0(A) \geq Q(A) - \int_A |q - f_0| \geq Q(A) - \epsilon_n\,.$$

Now let $A = \{0 \in B_n\}$, $D = \{\Omega \cap B_n \neq \emptyset\}$, and $c_n = ||\theta|| = a_n\sqrt{n}$ for $\theta \in \Omega$.

Note that $A \cap D \subseteq \{s_n \geq c_n\}$. In addition, since $P_\theta(\theta \in B_n) \geq 1-\alpha$ for $\forall\theta$, we have $P_\theta(D) \geq 1-\alpha$ for $\forall\theta \in \Omega$. Therefore, $Q(D) \geq 1-\alpha$, and

$$\begin{aligned}
P_0(s_n \geq c_n) &\geq P_0(A \cap D) \geq Q(A \cap D) - \epsilon_n \\
&= Q(A) + Q(D) - Q(A \cup D) - \epsilon_n \\
&\geq Q(A) + Q(D) - 1 - \epsilon_n \\
&\geq Q(A) + (1-\alpha) - 1 - \epsilon_n \\
&\geq (P_0(A) - \epsilon_n) + (1-\alpha) - 1 - \epsilon_n \\
&\geq (1-\alpha) - \epsilon_n + (1-\alpha) - 1 - \epsilon_n \\
&= 1 - 2\alpha - 2\epsilon_n
\end{aligned}$$

Finally, the same argument holds for any $\theta \in \mathbb{R}^n$ other than 0. ■

7.3 Equalizer Rule

The risk for $\hat{\theta}$ is $R(\theta, \hat{\theta}) = \mathbb{E}_\theta[d(\hat{\theta}, \theta)]$. Let Π be a distribution over Θ, then the Bayes risk of $\hat{\theta}$ is

$$R(\hat{\theta}, \Pi) = \int_\Theta R(\theta, \hat{\theta})\, d\,\Pi(\theta) = \int_{\mathcal{X}} r(\hat{\theta}|x)\, d\,\mu_x(x)$$

where μ_x is the marginal distribution of X, and $r(\hat{\theta|x})$ is the posterior risk of $\hat{\theta}$ given $X = x$. The Bayes rule $\hat{\theta}(\Pi)$ is the estimator $\hat{\theta}$ that minimizes $R(\hat{\theta}, \Pi)$, or equivalently, minimizes $r(\hat{\theta}|x)$ at every x.

Theorem 7.3 *If a Bayes rule $\hat{\theta}(\Pi)$ has constant risk, that is, $R(\theta, \hat{\theta}(\Pi))$ is constant in θ, then $\hat{\theta}(\Pi)$ is a minimax estimator.*

Proof: Let $\hat{\theta}$ be any estimator, then

$$\sup_\theta R(\theta, \hat{\theta}) \geq \int_\Theta R(\theta, \hat{\theta})\, d\,\Pi(\theta) \geq \int_\Theta R(\theta, \hat{\theta}(\Pi))\, d\,\Pi(\theta) = \sup_\theta R(\theta, \hat{\theta}(\Pi))\,.$$

■