Week 7
Spring 2009

Lecture 13. Le Cam's method – Two-point argument

We will introduce Le Cam's method to derive minimax lower bounds. The essential of this approach is the Neyman-Pearson Lemma.

Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures with densities p and q w.r.t. a measure μ. The affinity between $\mathbb{P}$ and $\mathbb{Q}$ is defined as

$$\alpha_1(\mathbb{P}, \mathbb{Q}) = \int p \wedge q d\mu = 1 - \frac{1}{2} \int |p - q| \, d\mu.$$

Lemma 1

$$\inf_f \mathbb{P}_0 f + \mathbb{P}_1 (1 - f) = \alpha_1 (\mathbb{P}_0, \mathbb{P}_1), \; 0 \le f \le 1.$$

Proof. Let p_0 and p_1 be probability densities of $\mathbb{P}_0$ and $\mathbb{P}_1$ respectively w.r.t. a measure μ. The result follows form the following equation

$$\int (p_0 - p_1)(f - I(p_0 < p_1)) \, d\mu \ge 0$$

i.e.,

$$\int [p_0 f + p_1 (1 - f)] \, d\mu \ge \int [p_0 I (p_0 < p_1) + p_1 I (p_0 \ge p_1)] \, d\mu.$$

The equality holds when $f = I(p_0 < p_1)$. ■

Remark 2 *Neyman-Pearson test. Let f be any rejection region such that $\int p_0 f d\mu \le \alpha$. Find c such that $\int p_0 I(p_0 < cp_1) \, d\mu = \alpha$, then*

$$\int (p_0 - cp_1)(f - I(p_0 < cp_1)) \, d\mu \ge 0$$

which implies

$$0 \ge \int p_0 [f - I(p_0 < cp_1)] \, d\mu \ge c \int p_1 [f - I(p_0 < cp_1)] \, d\mu$$

so $\int p_1 f d\mu \le \int p_1 I(p_0 < cp_1) \, d\mu$.

Corollary 3

$$\inf_{f \ge 0, g \ge 0, f + g \ge 1} \mathbb{P}_0 f + \mathbb{P}_1 g \ge \alpha_1 (\mathbb{P}_0, \mathbb{P}_1)$$

The Hellinger affinity is defined as

$$\alpha_2(\mathbb{P}, \mathbb{Q}) = \int \sqrt{pq} d\mu.$$

It is easy to see

$$pq \le (p \vee q)(p \wedge q) \le (p + q)(p \wedge q)$$

then the Cauchy-Schwarz inequality implies

$$\begin{aligned}
\left(\int \sqrt{pq} d\mu\right)^2 &= \left(\int \sqrt{(p \vee q)(p \wedge q)} d\mu\right)^2 \\
&\leq \left(\int \sqrt{(p + q)(p \wedge q)} d\mu\right)^2 \leq \int (p+q)\, d\mu \int (p \wedge q)\, d\mu = 2\int (p \wedge q)\, d\mu
\end{aligned}$$

i.e.,

$$\frac{1}{2}\left[\alpha_2(\mathbb{P}_0, \mathbb{P}_1)\right]^2 \leq \alpha_1(\mathbb{P}_0, \mathbb{P}_1).$$

Corollary 4

$$\inf_{f \geq 0, g \geq 0, f+g \geq 1} \mathbb{P}_0 f + \mathbb{P}_1 g \geq \frac{1}{2}\left[\alpha_2(\mathbb{P}_0, \mathbb{P}_1)\right]^2$$

Le Cam's method.

Example 5 *Show that the minimax rate in estimating θ for i.i.d.* $\mathrm{U}(0, \theta)$ *is $1/n^2$ for the squared error loss, where $\theta \in [a, b]$ with $a < b$. Let $\widehat{\theta}$ an estimator of θ. We need to show for some $c > 0$*

$$\sup_\theta \mathbb{E}\left(\widehat{\theta} - \theta\right)^2 \geq c\frac{1}{n^2}$$

We know

$$\begin{aligned}
\sup_\theta \mathbb{E}\left(\widehat{\theta} - \theta\right)^2 &\geq \sup_{\theta \in \{\theta_1, \theta_2\}} \mathbb{E}\left(\widehat{\theta} - \theta\right)^2 \\
&\geq \frac{1}{2}\mathbb{E}_{\underline{Y}|\theta_1}\left(\widehat{\theta} - \theta_1\right)^2 + \frac{1}{2}\mathbb{E}_{\underline{Y}|\theta_2}\left(\widehat{\theta} - \theta_2\right)^2.
\end{aligned}$$

Since

$$\left(\widehat{\theta} - \theta_1\right)^2 + \left(\widehat{\theta} - \theta_2\right)^2 \geq \frac{1}{2}(\theta_1 - \theta_2)^2$$

i.e.,

$$\frac{\left(\widehat{\theta} - \theta_1\right)^2}{\frac{1}{2}(\theta_1 - \theta_2)^2} + \frac{\left(\widehat{\theta} - \theta_2\right)^2}{\frac{1}{2}(\theta_1 - \theta_2)^2} \geq 1,$$

we have

$$\sup_\theta \mathbb{E}\left(\widehat{\theta} - \theta\right)^2 \geq \frac{(\theta_1 - \theta_2)^2}{4} \cdot \frac{1}{2}\left[\alpha_2(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2})\right]^2.$$

Let $\theta_1 = 1$ and $\theta_2 = \theta_1 + \frac{1}{n}$. It is easy to show

$$\alpha_2(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) \geq \prod_{i=1}^n \int I_{[0,1]}(x_i)\frac{1}{1+1/n} I_{[0,1+\frac{1}{n}]}(x_i)\, dx_i = \left(1 + \frac{1}{n}\right)^{-n} \to e^{-1}.$$

Consider the general problem of finding a lower bound for the minimax risk

$$\sup_{\theta} \mathbb{E} L\left(\hat{\theta}, \theta\right) \geq ?$$

Let

$$d\left(\theta_0, \theta_1\right) = \inf_{t}\left[L\left(t, \theta_0\right) + L\left(t, \theta_1\right)\right].$$

Lemma 6

$$\sup_{\theta} \mathbb{E} L\left(\hat{\theta}, \theta\right) \geq \frac{1}{2} d\left(\theta_0, \theta_1\right) \cdot \alpha_1\left(\mathbb{P}_0, \mathbb{P}_1\right)$$

where $f_i = L\left(\hat{\theta}, \theta_i\right) / d\left(\theta_0, \theta_1\right)$.

Lecture 14. Le Cam's method (cont.) –Multiple comparisons

Let

$$d(\theta_0,\theta_1)=\inf_t\left[L(t,\theta_0)+L(t,\theta_1)\right] \text{ and } d_{\min}=\inf_{1\le i\le n} d(\theta_0,\theta_i)$$

and

$$\begin{aligned} f_0 &= L\left(\hat\theta,\theta_0\right)/d_{\min} \\ f_a &= \inf_i L\left(\hat\theta,\theta_i\right)/d_{\min}. \end{aligned}$$

Note that there is an i such that

$$f_0+f_a=\frac{L\left(\hat\theta,\theta_0\right)+L\left(\hat\theta,\theta_i\right)}{d_{\min}}\ge 1.$$

Then

$$\begin{aligned} \sup_\theta \mathbb{E}L\left(\hat\theta,\theta\right) &\ge \frac{1}{2n}\sum_{i=1}^n\left[\mathbb{P}_0L\left(\hat\theta,\theta_0\right)+\mathbb{P}_iL\left(\hat\theta,\theta_i\right)\right] \\ &\ge \frac{d_{\min}}{2}\cdot\frac{1}{n}\sum_{i=1}^n\left[\mathbb{P}_0\frac{L\left(\hat\theta,\theta_0\right)}{d_{\min}}+\mathbb{P}_i\frac{L\left(\hat\theta,\theta_i\right)}{d_{\min}}\right] \\ &\ge \frac{1}{2}d_{\min}\left[\mathbb{P}_0f_0+\left(\frac{1}{n}\sum_{i=1}^n\mathbb{P}_i\right)f_a\right]\ge\frac{1}{2}d_{\min}\cdot\alpha_1\left(\mathbb{P}_0,\frac{1}{n}\sum_{i=1}^n\mathbb{P}_i\right) \end{aligned}$$

Lemma 7

$$\sup_\theta \mathbb{E}L\left(\hat\theta,\theta\right)\ge\frac{1}{2}d_{\min}\cdot\alpha_1\left(\mathbb{P}_0,\frac{1}{n}\sum_{i=1}^n\mathbb{P}_i\right).$$

Example 8 *Consider the multivariate normal mean problem:*

$$Y_i=\mu_i+\frac{1}{\sqrt{n}}Z_i, Z_i\overset{i.i.d.}{\sim}N(0,1),\ i=1,2,\dots,p.$$

under the assumption that $\Theta_n=\{\mu\in\mathbb{R}^n:\|\mu\|_0\le 1\}$. *Show that*

$$\sup_{\Theta_n}\mathbb{E}\sup_{1\le i\le n}\|\hat\mu_i-\mu_i\|^2\ge c\frac{\log p}{n},\ \textit{for some } c>0.$$

In this lecture notes, c is a generic constant whose value may vary from place to place. Let $\mathbb{P}_0$ be $\mathrm{N}\left(0,\frac{1}{n}I_p\right)$, and $\mathbb{P}_i$ be the joint distribution of $\underline{Y}$ with $\mu_i=\sqrt{a\frac{\log p}{n}}$ and $\mu_j=0$ for $j\ne i$. Let $a<1$. Let f_i be the density $\mathbb{P}_i$ w.r.t. the Lebesgue measure. Since

$$\begin{aligned} \int\frac{\left(\frac{1}{p}\sum_{i=1}^p f_i-f_0\right)^2}{f_0}d\mu &= \int\left(\frac{1}{p}\sum_{i=1}^p f_i\right)^2/f_0d\mu-1=\frac{1}{p^2}\sum_{i,j}\left(\int\frac{f_if_j}{f_0}d\mu-1\right) \\ &= \frac{1}{p^2}\sum_{i=1}^p\left(\int\frac{f_i^2}{f_0}d\mu-1\right)=\frac{1}{p}\exp(a\log p)-\frac{1}{p}\to 0 \end{aligned}$$

which implies

$$\alpha_1\left(\mathbb{P}_0, \frac{1}{n}\sum_{i=1}^{n}\mathbb{P}_i\right) \geq c, \textit{ for some } c > 0$$

by the well known fact that the square of L_1 *distance is bounded by* χ^2 *distance. Then the desired lower bound follows immediately from* $d_{\min} = a^2\frac{\log p}{n}$. *The upper bound* $O\left(\frac{\log p}{n}\right)$ *can be obtained by Bonferroni correction.*

Example 9 *Sparse signals estimation. Consider the multivariate normal mean problem:*

$$Y_i = \mu_i + \frac{1}{\sqrt{n}}Z_i, Z_i \overset{i.i.d.}{\sim} N(0,1), \ i = 1, 2, \ldots, p.$$

under the assumption that

$$\Theta_{n,p} = \left\{\mu \in \mathbb{R}^n : \frac{1}{n}\sum_{i=1}^{n}|\mu_i|^p \leq \eta_n^p\right\}, \eta_n^p = \ n^{-\delta}, 0 < \delta < 1.$$

Show that

$$\sup_{\Theta_n}\mathbb{E}\left\|\hat{\mu} - \mu\right\|^2 \geq cn\eta_n^p\left(2\log\eta_n^{-p}\right)^{(2-p)/2}, \textit{ for some } c > 0.$$

Why is this true? Here is an intuitive argument which can be made to be rigorous. We divide n *into* k *nonoverlapping blocks with block size* $\sim n/k$ *for each block, and in each block there is only one nonzero signal with magnitude* $\sqrt{a\log(n/k)}$. *The signal in each block is weak to be detected. That suggests you have to just estimate* μ *by* 0. *For the example above we expect*

$$\sup_{\Theta_n}\mathbb{E}\left\|\hat{\mu} - \mu\right\|^2 \geq ck\left(a\log(n/k)\right), \textit{ for some } c > 0,$$

where k *is chosen such that* $\frac{k}{n}\left(a\log(n/k)\right)^p = \eta_n^p$. *Then the desired lower bound follows immediately.*

Example 10 *Covariance matrix estimation. Observe* $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ *i.i.d. from a p-variate Gaussian distribution,* $N(\boldsymbol{\mu}, \Sigma_{p\times p})$. *Bickel and Levina (2008) considered covariance matrix estimation with the parameter space as follows*

$$\mathcal{F}(\alpha, M) = \left\{\begin{array}{c}\Sigma : |\sigma_{ij}| \leq Mk^{-(\alpha+1)} \textit{ for all } |i-j| = k \\ 0 < \varepsilon \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq 1/\varepsilon\end{array}\right\}. \tag{1}$$

We need $\frac{\log p}{n} \to 0$ *in the covariance matrix estimation similar to Example 8. For* $1 \leq m \leq p_1$, *let* Σ_m *be a diagonal covariance matrix with* $\sigma_{mm} = 1 + \sqrt{\tau\frac{\log p_1}{n}}$, $\sigma_{ii} = 1$ *for* $i \neq m$, *and let* Σ_0 *be the identity matrix. Let* $\mathbf{X}_l = \left(X_1^l, X_2^l, \ldots, X_p^l\right)^T \sim N(0, \Sigma_m)$, *and denote the joint density of* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *by* f_m, $1 \leq m \leq p_1$ *with* $p_1 \leq \max\{p, \exp(n/2)\}$, *which can be written as follows*

$$f_m = \prod_{1\leq i\leq n, 1\leq j\leq p, j\neq m}\phi_1\left(x_j^i\right) \cdot \prod_{1\leq i\leq n}\phi_{\sigma_{mm}}\left(x_m^i\right)$$

*where ϕ_σ, $\sigma = 1$ or σ_{mm}, is the density of $N\left(0, \sigma^2\right)$. Let $\theta_m = \Sigma_m$ for $0 \le m \le p_1$ and the loss function L be the squared operator norm. It is easy to see $d\left(\theta_0, \theta_m\right) = \frac{1}{2}\tau \frac{\log p_1}{n}$ for all $1 \le m \le p_1$. Then the lower bound (**??**) follows immediately if there is a constant $c > 0$ such that*

$$\left\|\mathbb{P}_{\theta_0} \wedge \bar{\mathbb{P}}\right\| \ge c. \tag{2}$$

Since $\int q_0 \wedge q_1 d\mu = 1 - \frac{1}{2}\int |q_0 - q_1|\, d\mu$ for any two densities q_0 and q_1, and the Jensen's inequality implies

$$\left[\int |q_0 - q_1|\, d\mu\right]^2 = \left(\int \left|\frac{q_0 - q_1}{q_1}\right| q_1 d\mu\right)^2 \le \int \frac{(q_0 - q_1)^2}{q_1} d\mu = \int \frac{q_0^2}{q_1} d\mu - 1.$$

Hence $\int q_0 \wedge q_1 d\mu \ge 1 - \frac{1}{2}\left(\int \frac{q_0^2}{q_1} d\mu - 1\right)^{1/2}$. To establish equation (2), it thus suffices to show that $\int \left(\frac{1}{p_1}\sum_{m=1}^{p_1} f_m\right)^2 / f_0 d\mu - 1 \to 0$, i.e.,

$$\int \frac{1}{p_1^2}\sum_{m=1}^{p_1} \frac{f_m^2}{f_0} d\mu + \frac{1}{p_1^2}\sum_{m \ne j} \frac{f_m f_j}{f_0} d\mu - 1 \to 0. \tag{3}$$

We now calculate $\int \frac{f_m f_j}{f_0} d\mu$. For $m \ne j$ it is easy to see

$$\int \frac{f_m f_j}{f_0} d\mu - 1 = 0.$$

When $m = j$, we have

$$\begin{aligned} \int \frac{f_m^2}{f_0} d\mu &= \frac{\left(\sqrt{2\pi\sigma_{mm}}\right)^{-2n}}{\left(\sqrt{2\pi}\right)^{-n}} \prod_{1 \le i \le n} \int \exp\left[\left(x_m^i\right)^2 \left(-\frac{1}{\sigma_{mm}} + \frac{1}{2}\right)\right] dx_m^i \\ &= \left[1 - (1 - \sigma_{mm})^2\right]^{-n/2} = \left(1 - \tau\frac{\log p_1}{n}\right)^{-n/2}. \end{aligned}$$

Thus

$$\begin{aligned} \int \left(\frac{1}{p_1}\sum_{m=1}^{p_1} f_m\right)^2 / f_0 d\mu - 1 &= \frac{1}{p_1^2}\sum_{m=1}^{p_1}\left(\int \frac{f_m^2}{f_0} d\mu - 1\right) \le \frac{1}{p_1}\left(1 - \tau\frac{\log p_1}{n}\right)^{-n/2} - \frac{1}{p_1} \\ &= \exp\left[-\log p_1 - \frac{n}{2}\log\left(1 - \tau\frac{\log p_1}{n}\right)\right] - \frac{1}{p_1} \to 0 \end{aligned} \tag{4}$$

for $0 < \tau < 1$, where the last step follows from the inequality $\log(1 - x) \ge -2x$ for $0 < x < 1/2$.

Remark 11 *In literature, people only considered rate optimality by Le Cam's method. Can we obtain the optimal constant too? This is a battleground unexplored.*