6.S951 Modern Mathematical Statistics Fall 2024

Lecture 8 — October 1, 2024

Prof. Stephen Bates

Scribe: Xander Morgan

1 Outline

Agenda:

1. Le Cam's Method
2. Uniform Location Example
3. Non-parametric Density Estimation Example

Last time:

1. Optimal Hypothesis Testing
2. TV Distance

2 Le Cam's Method

Goal: minimax optimal *rates* (hardness bound)

We will do this for complex estimation poroblems where exactly computing Bayes-optimal risk is hard.

Idea: reduce estimation to an (easier) testing problem. Hardness of testing (which we get from a TV distance calculation) propagates to hardness of estimation.

Setting: $\mathcal{A} = \mathbb{R}^d$ with loss $L(a, \theta)$ creating an implicit distance

$$d(\theta_0, \theta_1) = \inf_{a \in \mathcal{A}} L(a, \theta_0) + L(a, \theta_1).$$

For example, if $L(a, \theta) = (a - \theta)^2$, then $d(\theta_0, \theta_1) = \frac{1}{2}\|\theta_0 - \theta_1\|^2$.

Theorem 1. *Le Cam's Method: Let $P_{\theta_0}, P_{\theta_1}$ be distributions over $\mathcal{X}$ with $d(\theta_0, \theta_1) \geq 2\delta$. Let Q be a uniform prior over θ_0, θ_1. Then,*

$$R_B^*(Q) \geq \frac{\delta}{2}\left(1 - \|P_{\theta_0} - P_{\theta_1}\|_{TV}\right)$$

where $R_B^*(Q)$ *is the Bayes optimal risk,* δ *represents the separation in parameter space, and* $\frac{1}{2}(1 - \|P_{\theta_0} - P_{\theta_1}\|_{TV})$ *is the Bayes risk of testing* P_{θ_0} *versus* P_{θ_1}*. This immediately implies*

$$R_M^* \geq \frac{\delta}{2}(1 - \|P_{\theta_0} - P_{\theta_1}\|_{TV})$$

where

$$R_M^* = \inf_{A:\mathcal{X}\to\mathcal{A}} \sup_{\theta\in\Theta} R(A, \theta)$$

is the minimax optimal risk.

Proof. Consider any estimator $A : \mathcal{X} \to \mathcal{A}$. Define an implied test mapping $\mathcal{X} \to \{0, 1\}$ by

$$A^{\text{test}}(X) = \arg\min_{b\in\{0,1\}} L(A(X), \theta_b).$$

Thus, for any $\theta \in \{\theta_0, \theta_1\}$,

$$L(A(X); \theta) \geq \delta \mathbf{1}\{\theta_{A^{\text{test}}(X)} \neq \theta\}$$

so

$$\begin{aligned}
R_B(A; Q) &= \mathbb{E}[L(A(X), \theta)] \\
&\geq \mathbb{E}[\delta \mathbf{1}\{\theta_{A^{\text{test}}(X)} \neq \theta\}] \\
&= \frac{\delta}{2}\left[P_{\theta_0}(A^{\text{test}}(X) = 1) + P_{\theta_1}(A^{\text{test}}(X) = 0)\right] \\
&\geq \frac{\delta}{2}\left(1 - \|P_{\theta_0} - P_{\theta_1}\|_{\text{TV}}\right).
\end{aligned}$$

The minimax bound follows immediately. □

3 Uniform Location Example

Suppose $X_i \overset{iid}{\sim} \text{Unif}(\theta - 1/2, \theta + 1/2)$ for $i = 1, \ldots, n$ and $\theta \in \mathbb{R}$. We want to estimate θ under loss $L(a, \theta) = (a - \theta)^2$. We will use Le Cam's method to get a minimiax hardness bound. Consider θ_0 versus $0 < \theta_1 < 1$.

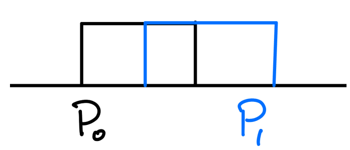

Figure 1: Plot for the distributions we are considering in the uniform location example.

$$\begin{aligned}
\|P_{\theta_0}^n - P_{\theta_1}^n\|_{\text{TV}} &= \sup_{B\subset\mathbb{R}^n} P_{\theta_1}^n(B) - P_{\theta_0}^n(B) \\
&= 1 - (1 - \theta_1)^n \\
&\approx n\theta_1
\end{aligned}$$

where we selected

$$B^{\text{opt}} = \{x | P_{\theta_1}^n(x) \geq P_{\theta_0}^n(x)\} = \{x | \max(x) > 1/2\} \subset \mathbb{R}^n.$$

To get a nontrivial bound, select $\theta_1 = \frac{1}{2n}$ and apply Le Cam's method with $\delta = \frac{1}{8n^2}$, giving

$$R_M^* \geq \frac{\delta}{2}(1 - \|P_{\theta_0} - P_{\theta_1}\|_{\text{TV}}) = \Theta\left(\frac{1}{n^2}\right).$$

so the minimax risk is growing (at least) as $1/n^2$, which is the correct rate for this problem.

4 Density Estimation Example

Suppose $X_i \stackrel{iid}{\sim} P_\theta$ where

$P_\theta \in \mathcal{P} = \{$distributions on $[0,1]$ with density f_θ such that $0 < c_1 < f_\theta(x) < c_2$ and $|f''(x)| < c_3\}$.

We wish to estimate $\psi(f_\theta) = \int_0^1 (f'(x))^2\, dx$ and want to know how statistically hard this problem is.

Theorem 2. *Consider $d(\theta_0, \theta_1) = \frac{1}{2}(\psi(\theta_0) - \psi(\theta_1))^2 \in \mathbb{R}$ with $L(a, \theta) = (a - \psi(\theta))^2$. Then, the minimax rate is at least $1/n$. (It is actually slightly higher, but we just show the $1/n$ bound.)*

Proof. We will apply Le Cam's Method with $P_{\theta_0} = \text{Unif}(0,1)$, and P_{θ_1} will be constructed according to a more complex density. Let P_{θ_0} have density f_0 and $\mathbb{P}_{\theta_1}$ have density f_1.

First, we construct f_1. Let $g : [0,1] \to \mathbb{R}$ be twice differentiable with $\int_0^1 g(x)\, dx = 0$, $\int_0^1 (g(x))^2\, dx = a > 0$, and $\int_0^1 (g'(x))^2\, dx = b > 0$. We split $(0,1)$ into m intervals of size $1/m$ with centers x_m.

Next, let $g_j(x) = C\frac{1}{m^2} g(mx - x_j)$ for $j = 1, \ldots, m$ and $f_1(x) = 1 + \sum_{j=1}^m g_j(x)$. To compute $\|P_{\theta_0}^n - P_{\theta_1}^n\|_{\text{TV}}$, we relate TV distance to Hellinger distance. We first define Hellinger distance.

Definition 3. *Hellinger distance. For P_0, P_1 represented as densities,*

$$H^2(P_0, P_1) = \frac{1}{2}\int (\sqrt{P_0} - \sqrt{P_1})^2\, dx = 1 - \int \sqrt{P_0 P_1}\, dx.$$

We have the following bound relating TV distance to Hellinger distance:

$$H^2(P_0, P_1) \leq \|P_0 - P_1\|_{\text{TV}} \leq \sqrt{2H^2(P_0, P_1)} = \sqrt{2}H(P_0, P_1).$$

Continuing our proof, by Taylor expansion,

$$H^2(P_{\theta_0}, P_{\theta_1}) = O\left(\sum_{j=1}^m \int_0^1 (g_j(x))^2\, dx\right) = O(mC^2 a m^{-5}) \propto m^{-4}.$$

Therefore,

$$\begin{aligned}
\|P^n_{\theta_0} - P^n_{\theta_1}\|_{\mathrm{TV}} &\le \sqrt{2} H(P^n_{\theta_0}, P^n_{\theta_1}) \\
&= \sqrt{2}\left(1 - \int_{[0,1]^n} \prod_{i=1}^n \sqrt{f_0(x_i) f_1(x_i)}\, dx\right)^{1/2} \\
&= \sqrt{2}\left(1 - (1 - H^2(P_{\theta_0}, P_{\theta_1}))^n\right)^{1/2} \\
&\to c \in (0,1) \text{ if } m = n^{1/4}.
\end{aligned}$$

Also note that, up to higher order terms,

$$\psi(f_1) - \psi(f_0) = \int_0^1 (f_1'(x))^2\, dx = \sum_{j=1}^m \int_0^1 (g_j'(x))^2\, dx = Cm^{-2}$$

so we get the scaling

$$\delta \propto m^{-4} = n^{-1}.$$

Finally, we apply Le Cam's Method to get

$$R^*_M \gtrsim \delta(1 - \|P^n_{\theta_0} - P^n_{\theta_1}\|_{\mathrm{TV}}) \gtrsim 1/n.$$

Note: this rate is not quite tight, but it is close. □

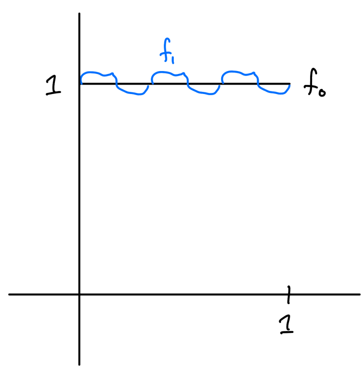

Figure 2: Example of what the construction for f_1 might look like. In this example, $m = 3$.