Lecture 8: Point vs. mixture, Ingster–Suslina method

Lecturer: Yanjun Han

April 21, 2021

Announcements

- HW1 due today
- HW2 released, due two weeks later

Today's plan

Test between a simple hypothesis and a composite hypothesis

- same two-point lower bound
- upper bound TV by Ingster–Suslina χ^2-method
- example: hidden clique detection
- example: uniformity and identity testing
- example: linear functional of sparse Gaussian mean
- example: covariance matrix estimation
- example: quadratic functional estimation

Generalized two-point method

Generalized two-point method

Fix any $\theta_0 \in \Theta$ and $\Theta_1 \subseteq \Theta$. Suppose that the following separation condition holds:

$\{\theta_0\}$, Θ_1 separated apart.

$$\min_{a \in \mathcal{A}} L(\theta_0, a) + L(\theta_1, a) \geq \Delta > 0, \quad \forall \theta_1 \in \Theta_1.$$

Then for any probability distribution π on Θ_1,

$$\inf_T \max_{\theta \in \{\theta_0\} \cup \Theta_1} \mathbb{E}_\theta[L(\theta, T(X))] \geq \frac{\Delta}{2} \left(1 - \|P_{\theta_0} - \mathbb{E}_\pi[P_{\theta_1}]\|_{\mathrm{TV}}\right).$$

$\mathbb{E}_\pi[P_{\theta_1}] \triangleq Q$

$$Q(x) = \mathbb{E}_{\theta_1 \sim \pi}[P_{\theta_1}(x)]$$

$$\begin{aligned}
\text{LHS} &\geq \inf_T \tfrac{1}{2} \mathbb{E}_{\theta_0}[L(\theta_0, T(X))] + \tfrac{1}{2} \mathbb{E}_\pi \mathbb{E}_{\theta_1}[L(\theta_1, T(X))] \\
&= \inf_T \sum_{x \in \mathcal{X}} \left(\tfrac{1}{2} L(\theta_0, T(x)) \cdot P_{\theta_0}(x) + \tfrac{1}{2} \mathbb{E}_\pi[L(\theta_1, T(x)) \cdot P_{\theta_1}(x)] \right) \\
&\geq \sum_{x \in \mathcal{X}} \tfrac{1}{2} \min\{P_{\theta_0}(x), \mathbb{E}_\pi[P_{\theta_1}(x)]\} \cdot \Delta
\end{aligned}$$

Upper bound the TV distance

- naïve upper bound:

$$\|P_{\theta_0} - \mathbb{E}_\pi[P_{\theta_1}]\|_{\mathrm{TV}} \le \mathbb{E}_\pi[\|P_{\theta_0} - P_{\theta_1}\|_{\mathrm{TV}}]$$

- Ingster–Suslina χ^2-method:

$$\chi^2(\mathbb{E}_\pi[P_{\theta_1}], P_{\theta_0}) = \mathbb{E}_{\theta_1,\theta_1'\sim\pi}\left[\sum_{x\in\mathcal{X}} \frac{P_{\theta_1}(x)P_{\theta_1'}(x)}{P_{\theta_0}(x)}\right] - 1$$

- implication on mixture of product models:

$$\chi^2(\mathbb{E}_\pi[P_{\theta_1}^{\otimes n}], P_{\theta_0}^{\otimes n}) = \mathbb{E}_{\theta_1,\theta_1'\sim\pi}\left[\prod_{i=1}^n \sum_{x_i\in\mathcal{X}} \frac{P_{\theta_1}(x_i)P_{\theta_1'}(x_i)}{P_{\theta_0}(x_i)}\right] - 1$$

$\neq (\mathbb{E}_\pi[P_{\theta_1}])^{\otimes n}$

Example I: planted clique problem

- given a graph G, test between

$$H_0 : G \sim \mathcal{G}(n, 1/2) \quad \text{vs.} \quad H_1 : G \sim \mathcal{G}(n, 1/2, k)$$

- claim: statistical threshold $k_n = 2\log_2 n - 2\log_2\log_2 n + \Theta(1)$
- failure of naïve two-point method: knowing where is the clique gives too much power to the learner
- χ^2-method:

$$\begin{aligned}\chi^2(\mathbb{E}_S[P_S], P) &= \mathbb{E}_{S,S'}\left[\sum_G \frac{P_S(G)P_{S'}(G)}{P(G)}\right] - 1\\ &= \mathbb{E}_{S,S'}\left[2^{\binom{|S\cap S'|}{2}}\right] - 1\\ &= \sum_{r=1}^{k} 2^{\binom{r}{2}} \cdot \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}}.\end{aligned}$$

Example II: uniformity testing

- given $X_1, \cdots, X_n \sim P$, test between

  $= (p_1, \cdots, p_k)$

$$H_0 : P = U_k \quad \text{vs.} \quad H_1 : \|P - U_k\|_{\text{TV}} \geq \varepsilon$$

uniform dist. on $[k]$

- claim: sample complexity $n = \Omega(\sqrt{k}/\varepsilon^2)$

  $n = \Theta(\frac{k}{\varepsilon^2})$ - learning complexity

- failure of naïve two-point again

  H_0: $P = U_k$

  H_1: $P = Q$ s.t. $\|Q - U_k\|_{TV} \geq \varepsilon$.

  $TV(U_k^{\otimes n}, Q^{\otimes n})$ too small

- mixture distribution: Paninski's construction

$$P_v = \left(\frac{1 - \varepsilon v_1}{k}, \frac{1 + \varepsilon v_1}{k}, \cdots, \frac{1 - \varepsilon v_{k/2}}{k}, \frac{1 + \varepsilon v_{k/2}}{k} \right)$$

with $v = (v_1, \cdots, v_{k/2}) \sim \text{Unif}(\{\pm 1\}^{k/2})$

χ^2-method computation

- the central quantity:

$$\sum_{i=1}^{k/2}\left(\frac{\frac{1-\varepsilon v_i}{k}\cdot\frac{1-\varepsilon v_i'}{k}}{\frac{1}{k}}+\frac{\frac{1+\varepsilon v_i}{k}\cdot\frac{1+\varepsilon v_i'}{k}}{\frac{1}{k}}\right), \quad \text{each term} = \frac{2}{k}\varepsilon^2 v_i v_i' + \frac{2}{k}$$

$$\sum_{x\in[k]} \frac{P_v(x)P_{v'}(x)}{P(x)} = 1 + \frac{2\varepsilon^2}{k}\sum_{i=1}^{k/2} v_i v_i'$$

- computation of $\chi^2(\mathbb{E}[P_v], P)$:

$$\chi^2(\mathbb{E}[P_v^{\otimes n}], P^{\otimes n}) = \mathbb{E}_{v,v'}\left[\left(1+\frac{2\varepsilon^2}{k}\sum_i v_i v_i'\right)^n\right] - 1$$

$$\le \mathbb{E}_{v,v'}\left[\exp\left(\frac{2n\varepsilon^2}{k}\sum_i v_i v_i'\right)\right] - 1 \qquad v' \text{ is 1-subGaussian}$$

$$\le \mathbb{E}_v\left[\exp\left(\frac{1}{2}\cdot\left(\frac{2n\varepsilon^2}{k}\right)^2\cdot\sum_i v_i^2\right)\right] - 1, \qquad \sum_i v_i^2 = k/2$$

$$= \exp\left(\frac{n^2\varepsilon^4}{k}\right) - 1 = O(1)$$

$$n = O\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$$

Example III: identity testing

- given $X_1, \cdots, X_n \sim P$, test between

$$H_0 : P = U \quad \text{vs.} \quad H_1 : \|P - U\|_{\mathrm{TV}} \ge \varepsilon$$

- simplifying assumption: $U = (u_1, u_1, \cdots, u_{k/2}, u_{k/2})$ with min. prob. $\ge \varepsilon^3/8$
- claim: sample complexity $n = \Omega(\|U\|_{2/3}/\varepsilon^2)$ $\quad \|U\|_{\frac{2}{3}} = \left(\sum_i u_i^{2/3}\right)^{\frac{3}{2}}$
- mixture distribution: $v = (v_1, \cdots, v_{k/2}) \sim \mathrm{Unif}(\{\pm 1\}^{k/2})$, and

$$P_v = \left(u_1 - \varepsilon_1 v_1, u_1 + \varepsilon_1 v_1, \cdots, u_{k/2} - \varepsilon_{k/2} v_{k/2}, u_{k/2} + \varepsilon_{k/2} v_{k/2}\right)$$

- constraints on ε: $\varepsilon_i \le u_i$ and $\sum_{i=1}^{k/2} \varepsilon_i \ge \varepsilon$
- χ^2-method:

$$\chi^2(\mathbb{E}[P_v^{\otimes n}], P^{\otimes n}) + 1 = \exp\left(2n^2 \sum_{i=1}^{k/2} \frac{\varepsilon_i^4}{u_i^2}\right)$$

Optimizing over ε

- target: subject to the constraint $\sum_{i=1}^{k/2} \varepsilon_i \geq \varepsilon$, minimize $\sum_{i=1}^{k/2} \frac{\varepsilon_i^4}{u_i^2}$ $\sim \frac{\varepsilon^4}{\|U\|_{2/3}^2}$
- generalized Hölder's inequality: for non-negative reals $(a_{i,j})_{i\in[m],j\in[n]}$

$$\prod_{i=1}^{m}\left(\sum_{j=1}^{n} a_{i,j}\right)^{1/m} \geq \sum_{j=1}^{n}\left(\prod_{i=1}^{m} a_{i,j}\right)^{1/m}$$

$\sum_j a_{i,j} = 1$ in LHS RHS $\leq \sum_j \frac{1}{m} \sum_i a_{ij} = 1 =$ LHS

- implication:

we construct

$$\left(\sum_{i=1}^{k/2} \frac{\varepsilon_i^4}{u_i^2}\right)\left(\sum_{i=1}^{k/2} u_i^{2/3}\right)^3 \geq \left(\sum_{i=1}^{k/2} \varepsilon_i\right)^4 = \frac{\varepsilon^4}{16},$$

with identity $\varepsilon_i \propto u_i^{2/3}$

Example IV: linear functional of sparse Gaussian mean

- model: $X \sim \mathcal{N}(\mu, \sigma^2 I_p)$ with $\|\mu\|_0 \le s$
- loss function: $L(\mu, T) = (T - L(\mu))^2$ with $L(\mu) = \sum_{i=1}^p \mu_i$
- claim:

$$R^\star_{n,s,p,\sigma} = \Omega\left(\sigma^2 s^2 \log\left(1 + \frac{p}{s^2}\right)\right) = \begin{cases} \Omega(\sigma^2 s^2 \log p) & \text{if } s \ll \sqrt{p} \\ \Omega(\sigma^2 p) & \text{if } s \ge \sqrt{p} \end{cases}$$

- reduce to hypothesis testing:

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_1 : \mu = \rho 1_S, S \sim \text{Unif}(\text{size-}s \text{ subset of } [p])$$

- separation condition: $\Delta \asymp \rho s^2$
- indistinguishability condition:

$$\chi^2(\mathbb{E}_S[P_S], P) + 1 = \mathbb{E}_{S,S'}\left[\exp\left(\frac{\rho^2}{\sigma^2}|S \cap S'|\right)\right]$$

Convex ordering: $\mathrm{HG}(N, K, n) \preceq_{\mathrm{c}} \mathrm{B}(n, K/N)$

Theorem (Hoeffding, 1963)

Let the population be $\mathcal{C} = \{c_1, \cdots, c_N\}$. Let $X_1, \cdots, X_n$ denote random samples without replacement from $\mathcal{C}$, and $X_1^\star, \cdots, X_n^\star$ denote random samples with replacement. Let S and $S^\star$ be the respective sums. Then for any convex function f, it holds that

$$\mathbb{E}[f(S)] \le \mathbb{E}[f(S^\star)].$$

Lemma (Aldous, 1985)

There exists a coupling between $(S, S^\star)$ such that $\mathbb{E}[S^\star \mid S] = S$.

Example of $n = 2$:

$$\mathbb{P}(x_1^* = a, x_2^* = b) = \underbrace{\mathbb{P}(x_1 = a, x_2 = b)}_{= \frac{1}{N(N-1)}} \cdot \mathbb{P}(x_1^*, x_2^* \mid x_1, x_2) = \frac{1}{N(N-1)} \cdot \left(1 - \frac{1}{N}\right) = \frac{1}{N^2}, \quad a \neq b$$

$$\mathbb{P}(x_1^\star, x_2^\star \mid x_1, x_2) = \begin{cases} 1/(2N) & \text{if } (x_1^\star, x_2^\star) = (x_1, x_1) \\ 1/(2N) & \text{if } (x_1^\star, x_2^\star) = (x_2, x_2) \\ 1 - 1/N & \text{if } (x_1^\star, x_2^\star) = (x_1, x_2) \end{cases}.$$

Applying convex ordering

$$
\begin{aligned}
\chi^2(\mathbb{E}_S[P_S], P) + 1 &= \mathbb{E}_{S,S'}\left[\exp\left(\frac{\rho^2}{\sigma^2}|S \cap S'|\right)\right] \\
&\le \mathbb{E}\left[\exp\left(\frac{\rho^2}{\sigma^2} \cdot \mathrm{B}(s, s/p)\right)\right] \\
&= \left(1 - \frac{s}{p} + \frac{s}{p}\exp(\rho^2/\sigma^2)\right)^s
\end{aligned}
$$

$\sum_{i=1}^{s} Y_i$

$\frac{s}{p}\left(e^{\rho^2/\sigma^2} - 1\right) = O\left(\frac{1}{s}\right)$

- choice of ρ: $\rho^2 \asymp \sigma^2 \log(1 + p/s^2)$

Example V: covariance estimation

- model: $X_1, \cdots, X_n \sim \mathcal{N}(0, \Sigma)$ with $\Sigma \in \mathbb{R}^{p \times p}$ and $\|\Sigma\|_{\mathrm{op}} \le 1$
- target: estimate Σ under loss $L(\Sigma, T) = \|T - \Sigma\|_{\mathrm{op}}$
- claim: $R_n^\star = \Omega(1 \wedge \sqrt{p/n})$
- not a low-dimensional target, but could lower bound L by another loss $L_0(\Sigma, T) = |T - \lambda_{\max}(\Sigma)|$, for

$$|\lambda_{\max}(T) - \lambda_{\max}(\Sigma)| \le \|T - \Sigma\|_{\mathrm{op}}$$

- hypothesis testing:

$$H_0 : \Sigma = aI_p \quad \text{vs.} \quad H_1 : \Sigma = aI_p + bvv^\top, v \sim \mathrm{Unif}(\mathbb{S}^{p-1})$$

- condition on (a, b): $a, b > 0, a + b \le 1$

Lower bound analysis

- separation condition: $\Delta = b$
- indistinguishability condition:

$$\chi^2(\mathbb{E}[P_v], P) + 1 = \mathbb{E}_{v,v'}\left[\left(1 - \frac{b^2}{a^2}(v^\top v')^2\right)^{-\frac{n}{2}}\right]$$

$$\leq \mathbb{E}_{v,v'}\left[\exp\left(\frac{nb^2}{a^2}(v^\top v')^2\right)\right]$$

as long as $2a \geq b$

$v' = (1, 0, 0, \dots, 0)$

- choice of parameters: since $|v^\top v'| = O(1/\sqrt{p})$ typically, we choose

$$a = \frac{1}{2}, \qquad b = \frac{1}{4} \wedge \sqrt{\frac{p}{n}}.$$

Example VI: quadratic functional estimation

- model: $X_1, \cdots, X_n \sim f$ with Hölder smoothness $s > 0$
- target: estimate $Q(f) = \int_0^1 f(x)^2 dx$, with $L(f, T) = |T - Q(f)|$
- claim: elbow effect

$t \mapsto t^2$

dominate if $s \le \frac{1}{4}$

$$R^\star_{n,s} = \Omega\left(n^{-\frac{4s}{4s+1}} + n^{-\frac{1}{2}}\right)$$

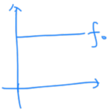
- hypothesis testing:

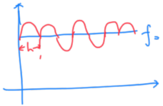
$$H_0 : f \equiv 1 \quad \text{vs.} \quad H_1 : f_v(x) = 1 + c_0 \cdot \sum_{i=1}^{1/h} v_i h^s g\left(\frac{x-(i-1)h}{h}\right)$$

with $v_1, \cdots, v_{1/h} \sim \mathrm{Unif}(\{\pm 1\})$

g: smooth, supported on $[0,1)$

$\int g = 0$





Remaining details

$Q(f_0) = 1$

$Q(f_v) = 1 + \Theta(h^{2s})$

- separation condition: $\Delta \asymp h^{2s}$
- indistinguishability condition:

$$
\begin{aligned}
1 + \chi^2(\mathbb{E}[f_v^{\otimes n}], f^{\otimes n}) &= \mathbb{E}_{v,v'}[(1 + (c_0\|g\|_2)^2 h^{2s+1} \cdot v^\top v')^n] \\
&\leq \mathbb{E}_{v,v'}[\exp(nc_0^2\|g\|_2^2 h^{2s+1} \cdot (v^\top v')^2)] \\
&= \mathbb{E}_{v,v'}\left[\exp\left(O\left(n^2 h^{4s+2} \cdot \frac{1}{h}\right)\right)\right]
\end{aligned}
$$

- choice of h: $h_n \asymp n^{-\frac{2}{4s+1}}$
- same lower bound holds for any locally quadratic functional

References

- Yuri Ingster and Irina A. Suslina. "Nonparametric goodness-of-fit testing under Gaussian models." Vol. 169. Springer Science & Business Media, 2012.
- T. Tony Cai and Harrison H. Zhou. "Optimal rates of convergence for sparse covariance matrix estimation." *The Annals of Statistics* 40, no. 5 (2012): 2389-2420.
- Gregory Valiant and Paul Valiant. "An automatic inequality prover and instance optimal identity testing." *SIAM Journal on Computing* 46, no. 1 (2017): 429-455.
- Sivaraman Balakrishnan and Larry Wasserman. "Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates." *Annals of Statistics* 47, no. 4 (2019): 1893-1927.

Next lecture: two mixtures, orthogonal polynomial, moment matching