Lec 8: Advanced Le Cam's Method

Yanjun Han

General hypothesis testing

$$\begin{cases} H_0: & \theta \in \Theta_0 \\ H_1: & \theta \in \Theta_1 \end{cases}$$

(simple: Θ_0 / Θ_1 is a singleton
composite: Θ_0 / Θ_1 is a set)

In the composite setting, for a test $T \in \{0,1\}$:

$$\text{Type I error} = \sup_{\theta_0 \in \Theta_0} P_{\theta_0}(T=1)$$

$$\text{Type II error} = \sup_{\theta_1 \in \Theta_1} P_{\theta_1}(T=0)$$

Thm.
$$\inf_T \left(\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(T=1) + \sup_{\theta_1 \in \Theta_1} P_{\theta_1}(T=0) \right) = 1 - \inf_{\substack{\pi_0 \in \mathcal{P}(\Theta_0) \\ \pi_1 \in \mathcal{P}(\Theta_1)}} TV\left(\mathbb{E}_{\theta_0 \sim \pi_0}[P_{\theta_0}], \mathbb{E}_{\theta_1 \sim \pi_1}[P_{\theta_1}]\right).$$

Pf. Minimax theorem; details left as exercise.

- In the last lecture, basic Le Cam's two-point method reduces the estimation problem to the hypothesis testing between two simple hypothesis.
- However, it could be helpful to choose one or both hypotheses to be composite, or in other words, to be mixture distributions with a carefully chosen prior π.

Advanced Le Cam I: point vs. mixture

General Thm. Let $\theta_0 \in \Theta$ and $\Theta_1 \subseteq \Theta$. Suppose

$$\inf_{\theta_1 \in \Theta_1} \min_a L(\theta_0, a) + L(\theta_1, a) \geq \Delta,$$

then for any probability distribution π on Θ_1,

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

Pf. Consider the two-point prior $\frac{1}{2}(\delta_{\theta_0} + \pi)$. The Bayes risk lower bound follows from the same two-point proof with

$$P = P_{\theta_0}, \quad Q = \mathbb{E}_{\theta \sim \pi}[P_{\theta_1}].$$

$\square$

How to upper bound $TV(P_{\theta_0}, \mathbb{E}_\pi[P_{\theta_1}])$?

- The "point vs. mixture" structure is only helpful when

$$TV(P_{\theta_0}, \mathbb{E}_\pi[P_{\theta_1}]) \ll \inf_{\theta_1 \in \Theta_1} TV(P_{\theta_0}, P_{\theta_1}).$$

i.e. mixture increases closeness.

- To achieve so, the standard method is Ingster-Suslina χ^2 method or the second-moment method, by upper bounding $\chi^2(\mathbb{E}_\pi[P_{\theta_1}] \| P_{\theta_0})$.

Thm (χ^2-method). $$\chi^2(\mathbb{E}_\pi[P_{\theta_1}] \| P_{\theta_0}) = \mathbb{E}_{\theta_1, \theta_1' \sim \pi}\left[\int \frac{P_{\theta_1} P_{\theta_1'}}{P_{\theta_0}}\right] - 1,$$

where $\theta_1' \sim \pi$ is an independent copy of θ_1.

Pf.
$$\chi^2(\mathbb{E}_\pi[P_{\theta_1}] \| P_{\theta_0}) + 1 = \int \frac{(\mathbb{E}_{\pi_1}[P_{\theta_1}])^2}{P_{\theta_0}}$$
$$= \int \frac{\mathbb{E}_{\theta_1, \theta_1' \sim \pi}[P_{\theta_1} P_{\theta_1'}]}{P_{\theta_0}}$$
$$= \mathbb{E}_{\theta_1, \theta_1' \sim \pi}\left[\int \frac{P_{\theta_1} P_{\theta_1'}}{P_{\theta_0}}\right] \text{ by Fubini.}$$
$\square$

Corollary. For i.i.d. models,

$$\chi^2(\mathbb{E}_\pi[P_{\theta_1}^{\otimes n}] \| P_{\theta_0}^{\otimes n}) = \mathbb{E}_{\theta_1, \theta_1' \sim \pi}\left[\left(\int \frac{P_{\theta_1} P_{\theta_1'}}{P_{\theta_0}}\right)^n\right] - 1.$$

Pf. Check $$\int \frac{P_{\theta_1}^{\otimes n} P_{\theta_1'}^{\otimes n}}{P_{\theta_0}^{\otimes n}} = \left(\int \frac{P_{\theta_1} P_{\theta_1'}}{P_{\theta_0}}\right)^n.$$ $\square$

Example 1.1 (Planted Clique) Given an undirected graph G on n vertices, aim to test between

H_0: $G \sim \mathcal{G}(n, \frac{1}{2})$ (i.e. $\forall i<j$, $\mathbb{P}((i,j) \in E) = \frac{1}{2}$)

H_1: $G \sim \mathcal{G}(n, \frac{1}{2}, k)$ (i.e. there is an unknown $S \subseteq [n]$, $|S| = k$, and $\mathbb{P}((i,j) \in E) = \begin{cases} 1 & \text{if } i,j \in S \\ \frac{1}{2} & \text{o.w.} \end{cases}$)

Target: $\exists$ constant C s.t. if $k < 2\log_2 n - 2\log_2 \log_2 n + C$, then no test can reliably distinguish between H_0 and H_1.

Why mixture structure in H_1 is important? Because for each <u>fixed</u> instance of H_1, the learner knows the set S and can look at if $G[S]$ is a clique or not.

<u>Pf.</u> Let P be the distribution of $G \sim \mathcal{G}(n, \frac{1}{2})$;

P_S be the distribution of G with a clique planted at S;

S be a uniformly random subset of $[n]$ of size k.

Then
$$\int \frac{P_S P_{S'}}{P} = \sum_G \frac{P_S(G) P_{S'}(G)}{P(G)} = \sum_{(X_{ij}) \in \{0,1\}^{\binom{n}{2}}} \frac{\left(\frac{1}{2}\right)^{\binom{n}{2}-\binom{k}{2}} \prod_{i,j\in S} 1(X_{ij}=1) \cdot \left(\frac{1}{2}\right)^{\binom{n}{2}-\binom{k}{2}} \prod_{i,j\in S'} 1(X_{ij}=1)}{\left(\frac{1}{2}\right)^{\binom{n}{2}}}$$
$$= \frac{\left(\frac{1}{2}\right)^{2\binom{n}{2}-2\binom{k}{2}} \cdot 2^{\binom{n}{2}-2\binom{k}{2}+\binom{|S\cap S'|}{2}}}{\left(\frac{1}{2}\right)^{\binom{n}{2}}} = 2^{\binom{|S\cap S'|}{2}}.$$

$$\Longrightarrow \chi^2(\mathbb{E}[P_S] \,\|\, P) = \mathbb{E}_{S,S'}\left[2^{\binom{|S\cap S'|}{2}}\right] - 1$$
$$= \sum_{r=1}^{k} 2^{\binom{r}{2}} \cdot \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} \underset{\text{by algebra}}{=} o(1) \quad \text{when } k < 2\log_2 n - 2\log_2\log_2 n + C$$

$\square$

<u>Example 1.2 (uniformity testing)</u> Given $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} P = (p_1, \cdots, p_k)$, aim to test

$$H_0: \quad P = \mathrm{Unif}[k]$$
$$\text{vs.} \quad H_1: \quad TV(P, \mathrm{Unif}[k]) \geq \varepsilon.$$

Target: the sample complexity of a reliable uniformity testing is

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

Note: Again, a naive two-point method will not succeed, for if the learner knows the pattern of how P deviates from uniform, then $O(\frac{1}{\varepsilon^2})$ samples will suffice to tell the difference.

Pf of lower bound. WLOG assume k is even.

Under H_0: $P = (\frac{1}{k}, \cdots, \frac{1}{k})$

Under H_1: $P_v = (\frac{1-2\varepsilon v_1}{k}, \frac{1+2\varepsilon v_1}{k}, \cdots, \frac{1-2\varepsilon v_{k/2}}{k}, \frac{1+2\varepsilon v_{k/2}}{k})$, with

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

Note that $TV(P_v, \mathrm{Unif}[k]) = \varepsilon$ for all $v \in \{\pm 1\}^{k/2}$, and

$$\int \frac{P_v P_{v'}}{P} = \sum_{x=1}^{k} \frac{P_v(x) P_{v'}(x)}{P(x)} = \sum_{i=1}^{k/2} \left(\frac{(1-2\varepsilon v_i)(1-2\varepsilon v_i')}{k} + \frac{(1+2\varepsilon v_i)(1+2\varepsilon v_i')}{k} \right)$$

$$= 1 + \frac{8\varepsilon^2}{k} \sum_{i=1}^{k/2} v_i v_i'$$

$$\Longrightarrow \chi^2(\mathbb{E}[P_v^{\otimes n}] \,\|\, P^{\otimes n}) = \mathbb{E}_{v,v'}\left[\left(1 + \frac{8\varepsilon^2}{k} \sum_{i=1}^{k/2} v_i v_i' \right)^n \right] - 1$$

$$\le \mathbb{E}_{v,v'} \exp\left(\frac{8n\varepsilon^2}{k} \underbrace{\sum_{i=1}^{k/2} v_i v_i'}_{\frac{k}{2}\text{-subGaussian}} \right) - 1$$

$$\le \exp\left(\frac{1}{2} \left(\frac{8n\varepsilon^2}{k} \right)^2 \cdot \frac{k}{2} \right) - 1 = \exp\left(\frac{16 n^2 \varepsilon^4}{k} \right) - 1.$$

Therefore, $\chi^2 = O(1)$ when $n = O(\frac{\sqrt{k}}{\varepsilon^2})$. $\square$

Example 1.3 (Linear functional of sparse parameters) $X \sim N(\mu, I_d)$ with $\|\mu\|_0 \le s$.

Target:
$$\inf_T \sup_{\|\mu\|_0 \le s} \mathbb{E}_\mu \left(T - \sum_{i=1}^{d} \mu_i \right)^2 \asymp s^2 \log\left(1 + \frac{d}{s^2}\right).$$

Pf of lower bound. H_0: $\mu = 0$ (call it P)

H_1: $\mu = \rho 1_S$, $S \sim \mathrm{Unif}(\binom{[d]}{s})$ (call it $\mathbb{E}[P_S]$)

Separation condition is satisfied with $\Delta \asymp \rho s^2$.

$$\int \frac{P_S P_{S'}}{P} = \int \frac{\varphi(x - \rho 1_S)\, \varphi(x - \rho 1_{S'})}{\varphi(x)} dx = e^{\rho^2 \langle 1_S, 1_{S'} \rangle} = e^{\rho^2 |S \cap S'|}.$$

To proceed, note that $|S \cap S'| \sim \mathrm{Hypergeometric}(d, s, s)$, so by Hoeffding's lemma,

$$\chi^2(\mathbb{E}[P_S] \,\|\, P_0) + 1 = \mathbb{E}\left[e^{\rho^2 |S \cap S'|} \right] \le \mathbb{E}\left[e^{\rho^2 B(s, \frac{s}{d})} \right] = \left(1 - \frac{s}{d} + \frac{s}{d} e^{\rho^2} \right)^s$$

$$= O(1) \text{ when } \rho \asymp \sqrt{\log\left(1 + \frac{d}{s^2}\right)}.$$

$\square$

Lemma (Hoeffding) Let $C = \{c_1, \cdots, c_N\} \subseteq \mathbb{R}$ be a fixed population, and

$X_1, \cdots, X_n$: n draws from C without replacement

$X_1^*, \cdots, X_n^*$: n draws from C with replacement

Then for convex $f: \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[f(\sum_{i=1}^n X_i)] \leq \mathbb{E}[f(\sum_{i=1}^n X_i^*)].$$

Example 1.4 (Quadratic functional estimation) $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f$, where the density f is supported on $[0,1]^d$, and $\|f^{(s)}\|_\infty = O(1)$ for some integer s.

Target:
$$\inf_T \sup_f \mathbb{E}_f \left| T - \int_{[0,1]^d} f(x)^2 dx \right| \asymp n^{-\frac{4s}{4s+d}} + n^{-\frac{1}{2}}.$$

Pf of lower bound. The parametric rate $\Omega(\frac{1}{\sqrt{n}})$ is trivial, by either LAM or a simple two-point argument (try it yourself!). For the $\Omega(n^{-\frac{4s}{4s+d}})$ lower bound:

$H_0: \quad f \equiv 1$

$H_1: \quad f_v(x) = 1 + c\sum_{i=1}^{h^{-d}} v_i h^s g(\frac{x-c_i}{h})$, with $v \sim \text{Unif}(\{\pm 1\}^{h^{-d}})$

- where $g(\cdot)$ is a smooth function on $[0,1]^d$ with $\int g = 0$,

$[0,1]^d$ is partitioned into h^{-d} subcubes with edgelength h

c_i is the lower-left corner of i-th subcube.

For a small absolute constant $c > 0$, can verify $\|f_v^{(s)}\|_\infty = O(1)$ for all v, and

$$\int_{[0,1]^d} f_v(x)^2 dx = 1 + c^2 \sum_{i=1}^{h^{-d}} h^{2s} \int_{[0,1]^d} g^2(\frac{x-c_i}{h}) dx = 1 + c^2 h^{2s} \|g\|_2^2.$$

$\Rightarrow$ Separation condition holds with $\Delta \asymp h^{2s}$.

For indistinguishability,

$$\int \frac{f_v f_{v'}}{f} = 1 + \int_{[0,1]^d} c^2 \sum_{i=1}^{h^{-d}} v_i v_i' h^{2s} g^2(\frac{x-c_i}{h}) dx = 1 + c^2 \|g\|_2^2 h^{2s+d} \sum_{i=1}^{h^{-d}} v_i v_i'$$

$$\Rightarrow \chi^2(\mathbb{E}[f_v^{\otimes n}] \| f^{\otimes n}) + 1 \leq \mathbb{E}\Big[\exp\Big(nc^2 \|g\|_2^2 h^{2s+d} \underbrace{\sum_{i=1}^{h^{-d}} v_i v_i'}_{O(h^{-d})\text{-subGaussian}}\Big)\Big]$$

$$\leq \exp(O(n^2 h^{4s+2d} \cdot h^{-d})) = O(1) \text{ when } h \asymp n^{-\frac{2}{4s+d}}.$$

Advanced Le Cam II: mixture vs. mixture

General Thm. Fix any $\Theta_0 \subseteq \Theta$ and $\Theta_1 \subseteq \Theta$. Suppose that

$$\inf_{\theta_0\in\Theta_0,\,\theta_1\in\Theta_1} \min_a \left(L(\theta_0,a)+L(\theta_1,a)\right) \geq \Delta,$$

then for any probability distributions π_0 and π_1,

$$\inf_T \sup_{\theta\in\Theta_0\cup\Theta_1} \mathbb{E}_\theta[L(\theta,T(X))] \geq \frac{\Delta}{2}\left(1-TV(\mathbb{E}_{\pi_0}[P_\theta],\mathbb{E}_{\pi_1}[P_{\theta_1}])-\pi_0(\Theta_0^c)-\pi_1(\Theta_1^c)\right).$$

Pf. The only new observation is that, if $\tilde{\pi}_0$ is the restriction of π_0 on Θ_0, then

$$TV(\mathbb{E}_{\pi_0}[P_\theta],\mathbb{E}_{\tilde{\pi}_0}[P_\theta]) \leq TV(\pi_0,\tilde{\pi}_0) = \pi_0(\Theta_0^c).$$ $\boxed{E}$

Challenge: What is a good way to upper bound $TV(\mathbb{E}_{\pi_0}[P_{\theta_0}],\mathbb{E}_{\pi_1}[P_{\theta_1}])$, apart from trivial convexity arguments?

Orthogonal functions/polynomials. Suppose $(P_\theta)_{\theta\in[\theta_0-\varepsilon,\,\theta_0+\varepsilon]}$ is a 1-D family of distributions, with likelihood ratio expansion

$$\frac{P_{\theta_0+u}}{P_{\theta_0}}(x) = \sum_{m=0}^{\infty} p_m(x;\theta_0)\frac{u^m}{m!}, \quad \text{for } |u|\leq\varepsilon.$$

Then under some structural condition, $\{p_m(x;\theta_0)\}_{m\geq 0}$ are orthogonal under P_{θ_0}.

Lemma. If $\int \frac{P_{\theta_0+u}P_{\theta_0+v}}{P_{\theta_0}}$ depends only on $(\theta_0, u\cdot v)$, then

$$\mathbb{E}_{X\sim P_{\theta_0}}[p_m(x;\theta_0)p_n(x;\theta_0)] = 0 \quad \forall\, m\neq n.$$

Pf.
$$\int \frac{P_{\theta_0+u}P_{\theta_0+v}}{P_{\theta_0}} = \mathbb{E}_{X\sim P_{\theta_0}}\left[\left(\sum_{m=0}^{\infty} p_m(X;\theta_0)\frac{u^m}{m!}\right)\left(\sum_{n=0}^{\infty} p_n(X;\theta_0)\frac{v^n}{n!}\right)\right]$$

$$= \sum_{m,n=0}^{\infty} \mathbb{E}_{X\sim P_{\theta_0}}[p_m(X;\theta_0)p_n(X;\theta_0)]\frac{u^m v^n}{m!\,n!}.$$

Since this quantity depends on (u,v) through $u\cdot v$, all coefficients on the RHS are 0 for $m\neq n$. $\boxed{E}$

Two important examples.

Gaussian. For $P_\theta = N(\theta,1)$, then $\int \frac{P_u P_v}{P_0} = \exp(uv)$.

The corresponding $p_m(x;\theta_0=0)$ is called Hermite polynomials $H_m(x)$, with

$$\mathbb{E}_{X\sim N(0,1)}[H_m(X)\,H_n(X)] = n!\cdot 1(m=n).$$

Poisson. For $P_\theta = \mathrm{Poi}(\theta)$, then $\int \frac{P_{\lambda+u}P_{\lambda+v}}{P_\lambda} = \sum_{k=0}^{\infty} \frac{e^{-\lambda-u-v}}{k!}\left(\frac{(\lambda+u)(\lambda+v)}{\lambda}\right)^k = \exp\left(\frac{uv}{\lambda}\right)$.

The corresponding $p_m(x;\theta_0=\lambda)$ is called Poisson-Charlier polynomial $c_m(x;\lambda)$, with

$$\mathbb{E}_{X\sim \mathrm{Poi}(\lambda)}\left[c_m(X;\lambda)\,c_n(X;\lambda)\right] = \frac{n!}{\lambda^n}\,1(m=n).$$

Bounding TV and χ^2: methods of moments

Thm (Gaussian mixture) For $\mu\in\mathbb{R}$ and RVs U, V:

$$\mathrm{TV}\left(\mathbb{E}[N(\mu+U,1)],\ \mathbb{E}[N(\mu+V,1)]\right) \le \frac{1}{2}\left(\sum_{m=0}^{\infty}\frac{(\mathbb{E}[U^m]-\mathbb{E}[V^m])^2}{m!}\right)^{1/2}.$$

If in addition $\mathbb{E}[V]=0$, $\mathbb{E}[V^2]\le M^2$, then

$$\chi^2\left(\mathbb{E}[N(\mu+U,1)]\ \|\ \mathbb{E}[N(\mu+V,1)]\right) \le e^{M^2/2}\cdot\sum_{m=0}^{\infty}\frac{(\mathbb{E}[U^m]-\mathbb{E}[V^m])^2}{m!}.$$

Pf. WLOG assume $\mu=0$, and let $\Delta_m := \mathbb{E}[U^m]-\mathbb{E}[V^m]$. Then

$$\begin{aligned}
\mathrm{TV}(\mathbb{E}[N(U,1)],\ \mathbb{E}[N(V,1)]) &= \frac{1}{2}\int_{\mathbb{R}}\left|\mathbb{E}_{\theta\sim U}[\varphi(x-\theta)] - \mathbb{E}_{\theta\sim V}[\varphi(x-\theta)]\right|dx \\
&= \frac{1}{2}\int_{\mathbb{R}}\varphi(x)\left|\mathbb{E}_{\theta\sim U}\left[\sum_{m=0}^{\infty}H_m(x)\frac{\theta^m}{m!}\right] - \mathbb{E}_{\theta\sim V}\left[\sum_{m=0}^{\infty}H_m(x)\frac{\theta^m}{m!}\right]\right|dx \\
&= \frac{1}{2}\mathbb{E}_{X\sim N(0,1)}\left|\sum_{m=0}^{\infty}H_m(X)\frac{\Delta_m}{m!}\right| \\
&\le \frac{1}{2}\left(\mathbb{E}_{X\sim N(0,1)}\left|\sum_{m=0}^{\infty}H_m(X)\frac{\Delta_m}{m!}\right|^2\right)^{1/2} \\
&= \frac{1}{2}\left(\sum_{m=0}^{\infty}\frac{\Delta_m^2}{m!}\right)^{1/2}.
\end{aligned}$$

For χ^2 upper bound, we lower bound the denominator as

$$\mathbb{E}_{\theta\sim V}[\varphi(x-\theta)] = \varphi(x)\cdot\mathbb{E}_{\theta\sim V}\left[\exp\left(\theta x - \frac{\theta^2}{2}\right)\right] \geq \varphi(x)\cdot\exp\left(\mathbb{E}_{\theta\sim V}\left[\theta x - \frac{\theta^2}{2}\right]\right) \geq \varphi(x) e^{-M^2/2}.$$

The rest is the same. $\square$

Similarly, we have the following result for Poisson mixtures.

Thm (Poisson mixture) For $\lambda > 0$ and RVs U, V supported on $[-\lambda, \infty)$:

$$\mathrm{TV}\left(\mathbb{E}\mathrm{Poi}(\lambda+U), \mathbb{E}\mathrm{Poi}(\lambda+V)\right) \leq \frac{1}{2}\left(\sum_{m=0}^{\infty}\frac{\Delta_m^2}{m!\,\lambda^m}\right)^{1/2}, \quad \Delta_m := \mathbb{E}[U^m] - \mathbb{E}[V^m].$$

If in addition $\mathbb{E}[V]=0$ and $|V| \leq M$, then

$$\chi^2\left(\mathbb{E}[\mathrm{Poi}(\lambda+U)] \,\|\, \mathbb{E}[\mathrm{Poi}(\lambda+V)]\right) \leq e^M \cdot \sum_{m=0}^{\infty}\frac{\Delta_m^2}{m!\,\lambda^m}.$$

Pf. Exercise.

Example 2.1 (Generalized uniformity testing) $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} P = (p_1, \cdots, p_k)$. Aim to test:

$$H_0: \quad P = \mathrm{Unif}(S) \text{ for some } S \subseteq [k]$$

$$\text{vs.} \quad H_1: \quad \min_{S\subseteq[k]} \mathrm{TV}(P, \mathrm{Unif}(S)) \geq \frac{\varepsilon}{2}.$$

Target: sample complexity for a reliable test is $\Theta\left(\frac{\sqrt{k}}{\varepsilon^2} + \frac{k^{2/3}}{\varepsilon^{4/3}}\right)$.

Pf of lower bound. ① $n = \Omega\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$ follows from uniformity testing (Example 1.2)

② For $n = \Omega\left(\frac{k^{2/3}}{\varepsilon^{4/3}}\right)$, assume Poissonization, where the observations are $\bigotimes_{i=1}^{k}\mathrm{Poi}(np_i)$.

Construct two product priors: under H_0, $p_1, \cdots, p_k \overset{\text{i.i.d.}}{\sim} \mathrm{Law}(U)$

under H_1: $p_1, \cdots, p_k \overset{\text{i.i.d.}}{\sim} \mathrm{Law}(V)$

where

$$U = \begin{cases} 0 & \text{w.p. } \frac{\varepsilon^2}{1+\varepsilon^2} \\ \frac{1+\varepsilon^2}{k} & \text{w.p. } \frac{1}{1+\varepsilon^2} \end{cases}, \qquad V = \begin{cases} \frac{1-\varepsilon}{k} & \text{w.p. } \frac{1}{2} \\ \frac{1+\varepsilon}{k} & \text{w.p. } \frac{1}{2} \end{cases}$$

Note that: ① Under H_0, $p_i \in \{0, \frac{1+\varepsilon^2}{k}\}$, so $(p_1, \cdots, p_k)$ is generalized uniform.

② Under H_1, $(p_1, \cdots, p_k)$ is $\Omega(\varepsilon)$-far from generalized uniform w.h.p.

③ $\mathbb{E}[U] = \mathbb{E}[V] = \frac{1}{k}$, so under both H_0 and H_1, $(p_1, \cdots, p_k)$ is a pmf <u>in expectation</u>. Additional arguments need to be made to ensure that it suffices to consider "approximate pmfs", but we're omitting them here.

④ $\mathbb{E}[U^2] = \mathbb{E}[V^2] = \frac{1+\varepsilon^2}{k^2}$, and

$$\left| \mathbb{E}[(U-\tfrac{1}{k})^m] - \mathbb{E}[(V-\tfrac{1}{k})^m] \right| \le \frac{2\varepsilon^2}{k^m}, \quad m \ge 3.$$

Now by the Poisson mixture result,

$$\chi^2(\mathbb{E}[\mathrm{Poi}(nU)] \,\|\, \mathbb{E}[\mathrm{Poi}(nV)]) \le e^{\frac{n}{k}} \sum_{m=3}^{\infty} \frac{4\varepsilon^4 (\frac{n}{k})^{2m}}{m! \, (\frac{n}{k})^m} \underset{\uparrow}{=} O\left(\frac{n^3 \varepsilon^4}{k^3}\right)$$

$\frac{k^{2/3}}{\varepsilon^{4/3}} \ge \frac{\sqrt{k}}{\varepsilon^2} \iff k \ge \frac{1}{\varepsilon^4}$, so

$n \le \frac{k^{2/3}}{\varepsilon^{4/3}} \Rightarrow n \le k \Rightarrow \frac{n}{k} \le 1$.

$$\Longrightarrow \chi^2\left(\mathbb{E}_U\left[\bigotimes_{i=1}^{k} \mathrm{Poi}(np_i)\right] \,\Big\|\, \mathbb{E}_V\left[\bigotimes_{i=1}^{k} \mathrm{Poi}(np_i)\right]\right) + 1 \underset{\uparrow}{\le} \left(1 + O\left(\frac{n^3\varepsilon^4}{k^3}\right)\right)^k$$

tensorization of χ^2

$$\le \exp\left(O\left(\frac{n^3 \varepsilon^4}{k^2}\right)\right) = O(1)$$

if $n = O\left(\frac{k^{2/3}}{\varepsilon^{4/3}}\right)$. $\square$

Remark: This construction matches the first two moments of (U, V).

Can we match more? No!

<u>Lemma</u>. Let μ be a prob. measure supported on $\{0, x_1, \cdots, x_{k-1}\} \subseteq [0, \infty)$.

Let ν be another prob. measure supported on $[0, \infty)$ s.t.

$$\mathbb{E}_\mu[X^m] = \mathbb{E}_\nu[X^m] \text{ for all } m = 0, 1, \cdots, 2k-1.$$

Then $\mu = \nu$.

<u>Pf</u> $0 = \mathbb{E}_\mu[X(X-x_1)^2 \cdots (X-x_{k-1})^2] = \mathbb{E}_\nu[X(X-x_1)^2 \cdots (X-x_{k-1})^2] \ge 0 \Rightarrow \mathrm{supp}(\nu) \subseteq \{0, x_1, \cdots, x_{k-1}\}$

$\Rightarrow \nu = \mu$. $\square$

Example 2.2 (ℓ_1-norm estimation) $X \sim N(\theta, I_n)$ with $\|\theta\|_\infty \le 1$.

Target: $$\inf_T \sup_{\|\theta\|_\infty \le 1} \mathbb{E}_\theta |T - \|\theta\|_1| \asymp n \cdot \frac{\log\log n}{\log n}.$$

Pf of lower bound. Idea: test between $H_0: \|\theta\|_1 \le \rho_0$ (assign $\theta \sim \mu_0^{\otimes n}$) vs. $H_1: \|\theta\|_1 \ge \rho_1$ (assign $\theta \sim \mu_1^{\otimes n}$)

Desired properties: ① $\chi^2(\mu_0 * N(0,1) \,\|\, \mu_1 * N(0,1)) = O(\frac{1}{n})$.

② $\mu_0^{\otimes n}(H_0^c) + \mu_1^{\otimes n}(H_1^c) = o(1)$.

③ $\rho_1 - \rho_0 = \Omega(n \cdot \frac{\log\log n}{\log n})$.

We design $(\rho_0, \rho_1, \mu_0, \mu_1)$ for these properties separately.

①: If μ_0, μ_1 match the first K moments, then

$$\chi^2(\mu_0 * N(0,1) \,\|\, \mu_1 * N(0,1)) \le O(1) \cdot \sum_{m=K+1}^{\infty} \frac{2^{m+1}}{m!} \le \left(\frac{O(1)}{K}\right)^K.$$

by shifting μ_1 if necessary

To make it $O(\frac{1}{n})$, choose $K \asymp \frac{\log n}{\log\log n}$.

② Choose $$\rho_0 = n \cdot \mathbb{E}_{\mu_0}|\theta_1| + \omega(\sqrt{n}),$$ $$\rho_1 = n \cdot \mathbb{E}_{\mu_1}|\theta_1| - \omega(\sqrt{n}).$$

Since under $\mu_0^{\otimes n}$, $\|\theta\|_1$ concentrates around $n \cdot \mathbb{E}_{\mu_0}|\theta_1|$ with fluctuation $O(\sqrt{n})$, Chebyshev's inequality gives $\mu_0^{\otimes n}(H_0^c), \mu_1^{\otimes n}(H_1^c) = o(1)$.

③ It remains to solve the following optimization program:

$$\begin{cases} \max & \mathbb{E}_{\mu_1}|\theta_1| - \mathbb{E}_{\mu_0}|\theta_1| \\ \text{s.t.} & \mu_0, \mu_1 \text{ supported on } [-1,1],\ \mathbb{E}_{\mu_1}[\theta_1^m] = \mathbb{E}_{\mu_0}[\theta_1^m] \text{ for } 0 \le m \le K. \end{cases}$$

There is a duality result between moment matching & best polynomial approximation.

Thm. Let $I \subseteq \mathbb{R}$ be a compact set, and f be continuous on I.

Let
$$V^* = \begin{cases} \max & \mathbb{E}_{\mu_1}[f(X)] - \mathbb{E}_{\mu_0}[f(X)] \\ \text{s.t.} & \mathrm{supp}(\mu_0), \mathrm{supp}(\mu_1) \subseteq I, \text{ with matching } K \text{ first moments} \end{cases}$$

$$E^* = \inf_{P: \deg P \le K} \sup_{x \in I} |f(x) - P(x)|$$

Then
$$V^* = 2 \cdot E^*.$$

Pf. The direction $V^* \le 2E^*$ is easy (exercise).

For the hard direction $V^* \ge 2E^*$, consider $F = \mathrm{span}\{1, x, \cdots, x^K, f(x)\}$.

Define a linear functional L on F with $L(x^m) = 0 \quad \forall m = 0, 1, \cdots, K$

$L(f) = E^*$

Then $\|L\| := \sup_{\substack{h \in F \\ \|h\|_{L^\infty(I)} \le 1}} |Lh|$ is 1. To see it, let $P^*(x)$ be the best approximating polynomial, with $\|f - P^*\|_{L^\infty(I)} = E^*$.

Then any $h \in F$ can be written as $h = c(f - P^*) + P$, with $\|h\|_{L^\infty(I)} \ge |c| E^*$ by definition of P^*.

$$\Longrightarrow \frac{|Lh|}{\|h\|_{L^\infty(I)}} \le \frac{|c|E^*}{|c|E^*} = 1, \text{ with equality if } P = 0.$$

Now by Hahn–Banach, L can be extended to $C(I)$ with $\|L\| = 1$.

by Riesz representation, $Lh = \int_I h \, d\mu$ for a signed measure μ.

Apply Jordan decomposition $\mu = \mu_+ - \mu_-$, then $L1 = 0 \Rightarrow \mu_+(I) = \mu_-(I)$
$\|L\| = 1 \Rightarrow \mu_+(I) + \mu_-(I) = 1$ $\Bigg\} \Rightarrow \mu_+(I) = \mu_-(I) = \frac{1}{2}$.

Also, $L(x^m) = 0 \Rightarrow \int x^m d\mu_+ = \int x^m d\mu_-$ for all $m = 0, \cdots, K$.

Finally, choose $\mu_1 = 2\mu_+$, $\mu_0 = 2\mu_-$, we get

$$E^* = Lf = \mathbb{E}_{\mu_+}[f] - \mathbb{E}_{\mu_-}[f] = \frac{1}{2}\left(\mathbb{E}_{\mu_1}[f] - \mathbb{E}_{\mu_0}[f]\right). \quad \boxed{\Xi}$$

By approximation theory, the uniform approximation error of $|\theta|$ by $\mathrm{span}\{1, \cdots, \theta^K\}$ is $\Theta(\frac{1}{K})$, so we get
$$p_1 - p_0 = \Omega\left(\frac{n}{K}\right) = \Omega\left(n \frac{\log \log n}{\log n}\right).$$

Combining ①–③ gives the target lower bound. $\boxed{\Xi}$

Special topic: dualizing Le Cam (Polyanskiy & Wu '2019)

Setting: $\theta_1, \cdots, \theta_n \overset{\text{i.i.d.}}{\sim} \pi$, $X_i \mid \theta_i \overset{\text{ind.}}{\sim} P_{\theta_i}$. observations: $(X_1, \cdots; X_n)$

(in other words, $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathbb{E}_{\theta\sim\pi}[P_\theta] =: \pi P$)

Target: estimate a linear functional $T(\pi)$, and characterize

$$r^* = \inf_{\hat{T}} \sup_{\pi\in\Pi} \mathbb{E}_\pi[(\hat{T}(X_1,\cdots,X_n) - T(\pi))^2]$$

(A related setting: $X_i \mid \theta_i \overset{\text{ind.}}{\sim} P_{\theta_i}$, where $(\theta_1, \cdots, \theta_n)$ is an individual sequence. The target is to estimate $T(\pi_\theta) = \frac{1}{n}\sum_{i=1}^n h(\theta_i)$, which is linear in $\pi_\theta = \frac{1}{n}\sum_i \delta_{\theta_i}$. Note that this covers the setting of functional estimation such as l_1-norm estimation in Example 2.2. A similar result holds in this setting; see paper)

Thm. If T is linear and Π is convex, under regularity conditions,

$$\frac{1}{7}\delta_{\chi^2}\left(\frac{1}{\sqrt{n}}\right)^2 \le r^* \le \delta_{\chi^2}\left(\frac{1}{\sqrt{n}}\right)^2,$$

where $\delta_{\chi^2}(t)$ is the χ^2 modulus of continuity:

$$\delta_{\chi^2}(t) = \sup\{|T(\pi') - T(\pi)| : \chi^2(\pi' P \| \pi P) \le t^2, \pi, \pi' \in \Pi\}.$$

Remark. ① δ_{χ^2} is the best separation constant subject to the χ^2 indistinguishability constraint, and $r^* \ge \frac{1}{7}\delta_{\chi^2}(\frac{1}{\sqrt{n}})^2$ trivially follows from Le Cam's two-point method

② The upper bound shows that for linear T, Le Cam's method can be "dualized" to get statistical upper bounds.

Pf of upper bound. Try $\hat{T}(X_1,\cdots,X_n) = \frac{1}{n}\sum_{i=1}^n g(X_i)$ for some function $g: \mathcal{X} \to \mathbb{R}$.

By bias-variance analysis,

$$\sup_\pi \mathbb{E}_\pi[(\hat{T} - T(\pi))^2] = |T(\pi) - \pi P g|^2 + \frac{1}{n}\mathrm{Var}_{\pi P}(g).$$

Therefore, it suffices to show that

$$\inf_g \sup_\pi \underbrace{|T(\pi) - \pi P g| + \frac{1}{\sqrt{n}}\sqrt{\mathrm{Var}_{\pi P}(g)}}_{L(\pi, g)} \le \delta_{\chi^2}\left(\frac{1}{\sqrt{n}}\right).$$

Ideally we'd like to apply minimax duality to $L(\pi, g)$. Note that

- $L(\pi, g)$ is convex in g;
- $\sqrt{\mathrm{Var}_{\pi P}(g)}$ is concave in π; but
- $|T(\pi) - \pi P g|$ is convex in π.

To mitigate the last issue, write

$$L(\pi, g) \le \sup_{\pi'} \sup_{0 \le \xi \le 2} (T(\pi) - \pi P g) - \xi(T(\pi') - \pi' P g) + \frac{1}{\sqrt{n}}\sqrt{\mathrm{Var}_{\pi P}(g)}$$

$$= \sup_{\pi_2 \in \underbrace{\{\xi\pi : \pi \in \Pi,\ 0 \le \xi \le 2\}}_{=: \Pi_2}} \underbrace{(T(\pi) - \pi P g) - (T(\pi_2) - \pi_2 P g) + \frac{1}{\sqrt{n}}\sqrt{\mathrm{Var}_{\pi P}(g)}}_{\text{concave in } (\pi, \pi_2) \text{ thanks to linearity of } T}$$

Therefore, we may apply minimax theorem to get

$$\inf_g \sup_\pi L(\pi, g) \le \sup_{\substack{\pi \in \Pi \\ \pi_2 \in \Pi_2}} \underbrace{\inf_g (T(\pi) - \pi P g) - (T(\pi_2) - \pi_2 P g) + \frac{1}{\sqrt{n}}\sqrt{\mathrm{Var}_{\pi P}(g)}}_{= -\infty \text{ if } \pi_2 \notin \Pi, \text{ by choosing } g \equiv c \to \pm\infty}$$

$$= \sup_{\pi, \pi' \in \Pi} \inf_g T(\pi) - T(\pi') + (\pi' - \pi) P g + \frac{1}{\sqrt{n}}\sqrt{\mathrm{Var}_{\pi P}(g)}.$$

Recall that $\chi^2(\pi' P \| \pi P) = \sup\{ |(\pi' - \pi) P g|^2 : \mathrm{Var}_{\pi P}(g) \le 1\}$.

So if $\chi^2(\pi' P \| \pi P) > \frac{1}{n}$, then $\exists\, g_0$ with $\mathrm{Var}_{\pi P}(g_0) \le 1$ and $(\pi' - \pi) P g_0 < -\frac{1}{\sqrt{n}}$.

Now choosing $g = c g_0$ with $c \to \infty$ gives that $\inf_g (\pi' - \pi) P g + \frac{1}{\sqrt{n}}\sqrt{\mathrm{Var}_{\pi P}(g)} = -\infty$.

On the other hand, if $\chi^2(\pi' P \| \pi P) \le \frac{1}{n}$, then $\inf_g (\pi' - \pi) P g + \frac{1}{\sqrt{n}}\sqrt{\mathrm{Var}_{\pi P}(g)} = 0$.

Therefore,

$$\inf_g \sup_\pi L(\pi, g) \le \sup_{\pi, \pi' \in \Pi} \left\{ T(\pi) - T(\pi') : \chi^2(\pi' P \| \pi P) \le \frac{1}{n} \right\}$$

$$\le \delta_{\chi^2}\left(\frac{1}{\sqrt{n}}\right).$$

<u>Example (Fisher's species problem)</u> Let $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim}$ pmf p supported on $\mathbb{N}$.

Let $m = nr$, and hypothetically draw $X_1', \cdots, X_m' \overset{\text{i.i.d.}}{\sim} p$, aim to estimate

$$U = \left| \{X_1', \cdots, X_m'\} \setminus \{X_1, \cdots, X_n\} \right| \quad (\text{\# of "new" species}).$$

Q: Characterize $\quad r^* = \inf_{\hat{U}} \sup_p \mathbb{E}_p\left[\frac{1}{n^2}(\hat{U} - U)^2\right]$

A:

$$r^* = \begin{cases} \Theta\left(\frac{1}{n}\right) & \text{if } r \le 1, \\ \tilde{\Theta}\left(n^{-\frac{2}{r+1}}\right) & \text{if } r > 1. \end{cases}$$

<u>Pf of upper bound</u>. First we make some simplifications:

① Poissonization: the histograms $N_x = \sum_{i=1}^{n} \mathbb{1}(X_i = x) \sim \text{Poi}(np_x)$, and $N_x' \sim \text{Poi}(mp_x)$ are independent Poisson RVs.

② Replace by expectation: can show $U \approx \mathbb{E}U$ w.h.p., so it's equivalent to estimate

$$\mathbb{E}[U] = \mathbb{E}\left[\sum_x \mathbb{1}(N_x = 0, N_x' > 0)\right] = \sum_x e^{-np_x}(1 - e^{-mp_x}).$$

③ The support size of p is at most $O(n)$. In this case, let $\theta_x = np_x$ and $\pi \sim \text{Unif}(\{\theta_x\})$, then $\frac{1}{n}\mathbb{E}[U]$ is equivalent to

$$\mathbb{E}_\pi[h(\theta)] = \mathbb{E}_{\theta\sim\pi}\left[e^{-\theta} - e^{-(1+r)\theta}\right].$$

By the previous result, it suffices to show that

$$\delta_{\chi^2}\left(\frac{1}{\sqrt{n}}\right) = \sup\left\{ \left|\mathbb{E}_{\pi' - \pi}[h(\theta)]\right| : \chi^2(\pi' P \,\|\, \pi P) \le \frac{1}{n} \right\} \lesssim n^{-\min\{\frac{1}{2}, \frac{1}{1+r}\}}. \quad P = \text{Poi}$$

Let $t = \frac{1}{\sqrt{n}}$. Since $\chi^2 \le t^2$ implies $TV \le t$, we have

$$\delta_{\chi^2}(t) \le \sup\left\{ \left|\int h \, d\Delta\right| : \|\Delta\|_{TV} \le 1, \ \|\Delta P\|_{TV} \le t \right\}.$$

where $\Delta = \pi' - \pi$ is a signed measure. To upper bound this quantity, we use complex analysis. Let

$$f_\Delta(z) = \int_{\mathbb{R}_+} e^{az} \Delta(da) \quad (\text{Laplace transform})$$

$$f_{\Delta P}(z) = \sum_{m=0}^{\infty} z^m \Delta P(m) \quad (z\text{-transform})$$

then

$$\int h \, d\Delta = \int_{\mathbb{R}_+} \left(e^{-\theta} - e^{-(r+1)\theta}\right) \Delta(d\theta) = f_\Delta(-1) - f_\Delta(-1-r).$$

In addition,

$$f_{\Delta P}(z) = \sum_{m=0}^{\infty} z^m \int e^{-a} \frac{a^m}{m!} \Delta(da) = \int e^{a(z-1)} \Delta(da) = f_\Delta(z-1),$$

Finally,

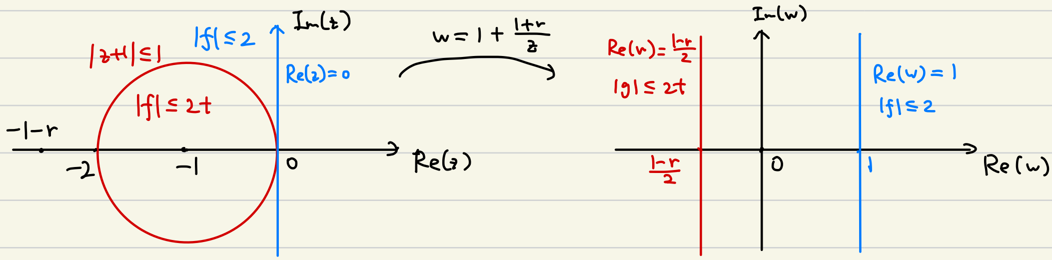
$$|f_\Delta(z)| \le \int_{\mathbb{R}_+} |\Delta|(da) \le 2 \qquad \text{for} \quad \mathrm{Re}(z) \le 0,$$

$$|f_{\Delta P}(z)| \le \sum_{m=0}^{\infty} |\Delta P(m)| \le 2t \qquad \text{for} \quad |z| \le 1.$$

Consequently,

$$\delta_{\chi^2}(t) \le \sup_\Delta \Big\{ |f_\Delta(-1) - f_\Delta(-1-r)| : \|f_\Delta\|_{H^\infty(\mathrm{Re}\, z \le 0)} \le 2,\ \|f_\Delta\|_{H^\infty(\underbrace{D-1}_{\text{disk } \{|z+1| \le 1\}})} \le 2t \Big\}$$

$$\le \sup_f \Big\{ |f(-1) - f(-1-r)| : \|f\|_{H^\infty(\mathrm{Re}\, z \le 0)} \le 2,\ \|f\|_{H^\infty(D-1)} \le 2t,\ f \text{ holomorphic on } \{z : \mathrm{Re}(z) \le 0\} \Big\}.$$



1) If $r \le 1$, then $|f(-1) - f(-1-r)| \le 4t$ as $-1-r$ lies in the disk $D-1$.

2) If $r > 1$, use the conformal map $f(z) = g(1 + \frac{1+r}{z})$. By Hadamard three-line theorem,

$$|f(-1-r)| = |g(0)| \le \|g\|_{H^\infty(\mathrm{Re}\, w = \frac{1-r}{2})}^{\frac{1}{1-\frac{1-r}{2}}} \|g\|_{H^\infty(\mathrm{Re}\, w = 1)}^{\frac{\frac{r-1}{2}}{1-\frac{1-r}{2}}} = O(t^{\frac{2}{1+r}})$$

$$\Rightarrow |f(-1-r) - f(-1)| = O(t^{\frac{2}{1+r}} + t) = O(t^{\frac{2}{1+r}}) \quad \text{as } t = \tfrac{1}{\sqrt{n}} \le 1.$$