The Low Degree Method

Main Question: Can we decide if a sample was drawn from P or Q?

A test is any function

$$f: S \to \{p, q\}$$

(S: sample space)

and its type I error is

$$\alpha(f) = \mathbb{P}[f(x) = p \mid x \sim Q]$$

and its type II error is

$$\beta(f) = \mathbb{P}[f(x) = q \mid x \sim P]$$

What is the optimal tradeoff btwn α and β?

Def: The likelihood ratio test with threshold η is

$$L_\eta(x) = \begin{cases} p & \text{if } \frac{P(x)}{Q(x)} > \eta \\ q & \text{else} \end{cases}$$

Proposition [Neyman, Pearson] For any $\eta \geq 0$, among all tests f with

$$\alpha(f) \leq \alpha(L_\eta)$$

L_η maximizes the power $1 - \beta(f)$

Proof: Let $R_f = \{x \mid f(x) = p\}$. Then

$$1 - B(f) = \mathbb{P}[f(x) = p \mid x \sim P]$$

$$= \int_{R_f} dP = \int_{R_f} L(x)\, d\varphi$$

Now by assumption

$$\alpha(f) = \varphi[R_f] \leq \varphi[L(x) > \eta] = \alpha(L_\eta)$$

Thus we can write the problem of maximizing the power as

$$\max \int_{R_f} L(x)\, d\varphi$$

$$\text{s.t. } \varphi[R_f] \leq \varphi[L(x) > \eta]$$

Now let's look at the difference of powers

$$(1 - B(L_\eta)) - (1 - B(f))$$

$$= \int_{R_{L_\eta}} L(x)\, d\varphi - \int_{R_f} L(x)\, d\varphi$$

$$= \int L(x)\, d\varphi - \int L(x)\, d\varphi$$

$R_{L_\eta} \setminus R_f$ ($\geq \eta$) $\qquad$ $R_f \setminus R_{L_\eta}$ ($\leq \eta$)

Thus the difference in power

$$\geq n\left(\Phi[R_{L_n} \setminus R_f] - \Phi[R_f \setminus R_{L_n}]\right)$$

$$= n\left(\underbrace{\Phi[R_{L_n}] - \Phi[R_f]}_{\geq 0}\right)$$

□

However computing the likelihood can be hard, b/c it involves summing over all partitions, etc

Main Idea [Hopkins]: Consider the projection onto low degree polynomials

Preliminary ideas appeared in [Barak, Hopkins, Kelner, Kothari, Moitra, Potechin] for SOS lower bounds for planted clique

Consider the following inner-product structure

$$\langle f, g \rangle \triangleq \mathop{\mathbb{E}}_{x \sim \Phi}[f(x)\, g(x)]$$

In particular

$$\|f\|^2 \triangleq \langle f, f \rangle$$

Now consider the following optimization problem:

$$\max \mathop{\mathbb{E}}_{x \sim P}[f(x)]$$

$$\text{s.t.} \mathop{\mathbb{E}}_{x \sim Q}[f(x)^2] = \langle f, f \rangle = 1$$

Claim: The optimal solution is

$$f^* = \frac{L(x)}{\|L\|}, \text{ normalized LRT}$$

Proof: We can rewrite the objective as

$$\mathop{\mathbb{E}}_{x \sim P}[f(x)] = \mathop{\mathbb{E}}_{x \sim Q}[L(x) f(x)] = \langle L, f \rangle$$

And by C-S we have

$$\langle L, f \rangle \leq \|L\| \underbrace{\|f\|}_{=1} = \|L\|$$

$=1$ from constraints

With equality iff $f = cL$ for some constant c. ∎

Intuitively we want f to be large on P while being bounded on Q

Now let's define a low degree variant. Let

$V_n^{\leq D}$ = polynomials on S of degree $\leq D$

e.g. for planted clique, the variables correspond to edges

$$x_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ -1 & \text{else} \end{cases}$$

and we can count subgraphs, e.g.

$$\left(\frac{a+1}{2}\right)\left(\frac{b+1}{2}\right)\left(\frac{c+1}{2}\right) = \begin{cases} 1 & \text{if } a=b=c=1 \\ 0 & \text{if } a=-1 \text{ or } b=-1 \text{ or } c=-1 \end{cases}$$

and so the polynomial

$$\sum_{ijk} \left(\frac{x_{ij}+1}{2}\right)\left(\frac{x_{ik}+1}{2}\right)\left(\frac{x_{jk}+1}{2}\right)$$

counts the # of triangles

Now consider

$$\max \underset{x \sim P_n}{\mathbb{E}} [f(x)]$$

$$\text{s.t. } \underset{x \sim Q_n}{\mathbb{E}} [f(x)^2] = 1$$

$$f(x) \in \mathbb{R}_n^{\leq D}$$

The optimal solution is

$$f^* = \frac{L_n^{\leq D}}{\|L^{\leq D}\|}$$

low-degree projection of LRT

For "nice" distributions

Conjecture [Hopkins] (informal) If for some $\varepsilon > 0$ and $D \geq (\log n)^{1+\varepsilon}$ we have

$$\|L_n^{\leq D}\| = O(1)$$

then there is no polynomial time algorithm B distinguishing P and Q with prob $1-o(1)$

More generally:

degree D polynomials $\approx$ $n^{\tilde{\Theta}(D)}$-time algorithms

---

Further Discussion

① sum-of-squares

[Parrilo] [Lasserre] Powerful semidefinite programming hierarchy

Goal: Find an operator that behaves like the expectation over a distribution on solutions

$$\tilde{\mathbb{E}} : \mathcal{V}_n^{\leq D} \to \mathbb{R}$$

Pseudoexpectation

Constraints

1. $\tilde{\mathbb{E}}$ is linear
2. $\tilde{\mathbb{E}}[1]$ is 1
3. $\tilde{\mathbb{E}}[P^2] \geq 0$ for all poly. P of $\deg(P) \leq \frac{D}{2}$

Specialization to P.C.

4. $\tilde{\mathbb{E}}[x_i^2 P] = \tilde{\mathbb{E}}[x_i P]$ booleanity
5. $\tilde{\mathbb{E}}[\sum x_i] = k$ clique size
6. $\tilde{\mathbb{E}}[x_i x_j P] = 0$ $\forall (i,j) \notin E$, $\deg(P) \leq D-2$

Given a clique, can define a valid $\tilde{\mathbb{E}}$ as

$$\tilde{\mathbb{E}}[P(x_1, \ldots x_n)] = P(a_1, a_2 \ldots a_n)$$

where $a_1, \ldots, a_n$ is indicator vector of a k-clique

Claim: This $\tilde{\mathbb{E}}$ satisfies 1.–6.

Proposition: There is an $n^{O(D)}$ time algorithm for finding a valid $\tilde{\mathbb{E}}$, if it exists

Main idea: Constaint 3. is a PSD-ness condition

Thm [Barak et al] For any $D = o(\log n)$ SOS fails to detect cliques of size $n^{1/2 - o(1)}$

Many other problems where we believe there is a computational vs statistical

tradeoff are open

Does SOS fail? Probably, but very difficult to show

---

(2) overlap gap property

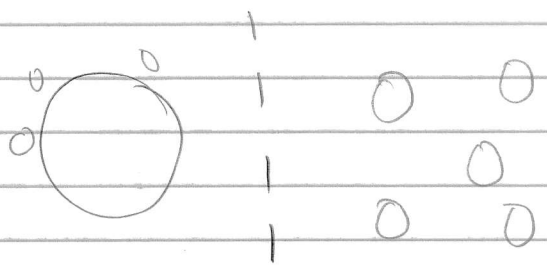
Given $G \sim G(n, 1/2)$, we know that whp the largest clique has size $(2 \pm o(1)) \log n$

Best known efficient algorithm only finds a clique of size $\log n$.

Main Idea [Gamarnik, Sudan], many others
Does the landscape exhibit a gap?

* pair of near optimal solutions has distance $\leq \alpha$ or $\geq \beta$

Let's look at space of k cliques



giant component $k = \log n$ shatters

---

③ statistical query lower bounds

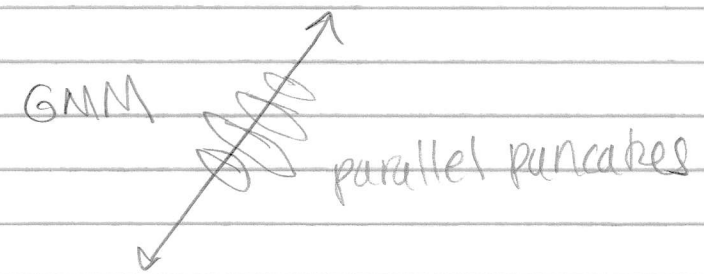
[Kearns] Don't directly see samples but can query functions $f: \mathbb{R}^n \to [-1, 1]$

[Feldman et al] VSTAT(t): Get response

$$v \in \left[\mathop{E}_{x \sim D}[f(x)] \pm \tau \right]$$

where $\tau = \max\left\{ \frac{1}{t}, \sqrt{\frac{\mathrm{var}_{x \sim D}[f(x)]}{t}} \right\}$

Example of computational vs statistical tradeoff in distribution learning



[Diakonikolas, Kane, Stewart] Any SQ algorithm for learning k-GMMs need $n^{\Omega(k)}$ queries or tolerance $n^{-O(k)}$

Intuition: It's hard to find the interesting direction, and o.w. it looks like a single Gaussian

④ search vs. refutation, e.g. for the Sherrington-Kirkpatrick model

$$\max_{\sigma \in \{\pm 1\}^n} \sigma^T A \sigma = P_* n$$

A: random Gaussian entries

P_*: Parisi constant

Thm [Montanari] There is a poly-time algorithm to find a near maximizer

Thm [Bandeira, Kunisky, Wein] It is hard to certify a better upper bound than $\|A\| n$