Introduction to “Low-Degree Method” for Computational Hardness

Tim Kunisky

(based on a survey with Afonso Bandeira and Alex Wein, **which is based on deep ideas not original to us!**)

Courant Institute of Mathematical Sciences

February 26, 2020

Question:

How to predict when **statistical inference** will be **computationally hard**?

What is statistical inference?

For this talk, **statistical inference** = **hypothesis testing**.

What is statistical inference?

For this talk, **statistical inference** = **hypothesis testing**.

- Two distributions, $\mathbb{P}$ and $\mathbb{Q}$, over $\mathbb{R}^N$.
- I draw $\boldsymbol{Y}$ from one of them secretly.
- You see $\boldsymbol{Y}$, and try to *infer* which one using a *test*:

$$f : \mathbb{R}^N \to \{\mathsf{p}, \mathsf{q}\}$$

What is asymptotic statistical inference?

For this talk, **statistical inference** = **hypothesis testing**.

- Two familes of distributions, $\mathbb{P}_n$ and $\mathbb{Q}_n$, over $\mathbb{R}^{N(n)}$.
- I draw $\boldsymbol{Y}$ from one of them secretly.
- You see $\boldsymbol{Y}$, and try to *infer* which one using a *test*:

$$f_n : \mathbb{R}^{N(n)} \to \{\mathrm{p}, \mathrm{q}\}$$

What is asymptotic statistical inference?

For this talk, **statistical inference** = **hypothesis testing**.

- Two familes of distributions, $\mathbb{P}_n$ and $\mathbb{Q}_n$, over $\mathbb{R}^{N(n)}$.
- I draw $\boldsymbol{Y}$ from one of them secretly.
- You see $\boldsymbol{Y}$, and try to *infer* which one using a *test*:

$$f_n : \mathbb{R}^{N(n)} \to \{\mathrm{p}, \mathrm{q}\}$$

This lets us define **asymptotic success** ("strong detection"):

$$\lim_{n\to\infty} \mathbb{P}_n[f_n(\boldsymbol{Y}) = \mathrm{p}] = 1,$$
$$\lim_{n\to\infty} \mathbb{Q}_n[f_n(\boldsymbol{Y}) = \mathrm{q}] = 1.$$

What kinds of distributions?

Think of $\mathbb{P}_n$ as **structured** ("planted") and $\mathbb{Q}_n$ as **null**.

[1] Just for optimal transport fans.

What kinds of distributions?

Think of $\mathbb{P}_n$ as **structured** ("planted") and $\mathbb{Q}_n$ as **null**.

- Principal component analysis
  - $\mathbb{Q}_n$: $(\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{\kappa n}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$
  - $\mathbb{P}_n$: $(\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{\kappa n}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n + \lambda \boldsymbol{x}\boldsymbol{x}^\top)$

[1] Just for optimal transport fans.

What kinds of distributions?

Think of $\mathbb{P}_n$ as **structured** (“planted”) and $\mathbb{Q}_n$ as **null**.

- Principal component analysis
  - $\mathbb{Q}_n$: $(\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{\kappa n}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$
  - $\mathbb{P}_n$: $(\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{\kappa n}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n + \lambda \boldsymbol{x}\boldsymbol{x}^\top)$
- Community detection
  - $\mathbb{Q}_n$: $G \sim$ Erdős-Rényi
  - $\mathbb{P}_n$: $G \sim$ Erdős-Rényi + clique

    $G \sim$ different edge prob. within/between blocks

[1] Just for optimal transport fans.

What kinds of distributions?

Think of $\mathbb{P}_n$ as **structured** ("planted") and $\mathbb{Q}_n$ as **null**.

- Principal component analysis
  - $\mathbb{Q}_n$: $(\boldsymbol{g}_1, \dots, \boldsymbol{g}_{\kappa n}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$
  - $\mathbb{P}_n$: $(\boldsymbol{g}_1, \dots, \boldsymbol{g}_{\kappa n}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n + \lambda \boldsymbol{x}\boldsymbol{x}^\top)$
- Community detection
  - $\mathbb{Q}_n$: $G \sim$ Erdős-Rényi
  - $\mathbb{P}_n$: $G \sim$ Erdős-Rényi + clique
    $G \sim$ different edge prob. within/between blocks
- Spiked transport model [Rigollet, Weed 2019][1]
  - $\mathbb{Q}_n$: $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_m), (\boldsymbol{y}_1, \dots, \boldsymbol{y}_m)$ i.i.d.
  - $\mathbb{P}_n$: $\boldsymbol{x}_i = \boldsymbol{a}_i^{(1)} + \boldsymbol{z}_i^{(1)}$, $\boldsymbol{y}_i = \boldsymbol{a}_i^{(2)} + \boldsymbol{z}_i^{(2)}$
    $\boldsymbol{a}^{(j)}$ different laws on low-dimensional subspace V, and
    $\boldsymbol{z}^{(j)}$ same law on $V^\perp$.

[1] Just for optimal transport fans.

Testing without computational budget

If we could use **any** test f, which would we use?

Testing without computational budget

If we could use **any** test f, which would we use?

Heuristic: to build a function large on $\mathbb{P}$ but small on $\mathbb{Q}$,

$$\begin{array}{ll} \text{maximize} & \mathbb{E}_{\mathbb{P}}\, h(\boldsymbol{Y}) \\ \text{subject to} & \mathbb{E}_{\mathbb{Q}}\, h(\boldsymbol{Y})^2 \le 1 \end{array}$$

Testing without computational budget

If we could use **any** test f, which would we use?

Heuristic: to build a function large on $\mathbb{P}$ but small on $\mathbb{Q}$,

$$\begin{aligned} &\text{maximize} \quad \mathbb{E}_{\mathbb{P}}\, h(\boldsymbol{Y}) \\ &\text{subject to} \quad \mathbb{E}_{\mathbb{Q}}\, h(\boldsymbol{Y})^2 \leq 1 \end{aligned}$$

$$\updownarrow$$

$$\begin{aligned} &\text{maximize} \quad \left\langle h, \frac{d\mathbb{P}}{d\mathbb{Q}} \right\rangle \\ &\text{subject to} \quad \|h\|^2 \leq 1 \end{aligned}$$

Testing without computational budget

If we could use **any** test f, which would we use?

Heuristic: to build a function large on $\mathbb{P}$ but small on $\mathbb{Q}$,

$$\begin{aligned} &\text{maximize} && \mathbb{E}_{\mathbb{P}}\, h(\boldsymbol{Y}) \\ &\text{subject to} && \mathbb{E}_{\mathbb{Q}}\, h(\boldsymbol{Y})^2 \le 1 \end{aligned}$$

$$\updownarrow$$

$$\begin{aligned} &\text{maximize} && \left\langle h, \tfrac{d\mathbb{P}}{d\mathbb{Q}} \right\rangle \\ &\text{subject to} && \|h\|^2 \le 1 \end{aligned}$$

Optimizer: the (normalized) **likelihood ratio**

$$h^\star(\boldsymbol{Y}) = \frac{d\mathbb{P}}{d\mathbb{Q}}(\boldsymbol{Y}) \quad / \quad \underbrace{\left\| \frac{d\mathbb{P}}{d\mathbb{Q}} \right\|}_{\text{objective value}}$$

Justification 1: optimal error tradeoff

[Neyman, Pearson 1933] Of tests with $\mathbb{Q}[f(\boldsymbol{Y}) = \mathsf{p}] \leq \alpha$, the test that minimizes $\mathbb{P}[f(\boldsymbol{Y}) = \mathsf{q}]$ is

$$f_\xi(\boldsymbol{Y}) = \left\{ \begin{array}{ll} \mathsf{p} & \text{if } \frac{d\mathbb{P}}{d\mathbb{Q}}(\boldsymbol{Y}) \geq \xi \\ \mathsf{q} & \text{otherwise.} \end{array} \right\},$$

for suitable ξ.

Best tradeoff between "Type I" and "Type II" errors.

(And non-asymptotically!)

Justification 2: control of asymptotic success

[Le Cam, 1960's] Suppose $\| \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \| \leq K$ as $n \to \infty$. Then, $\mathbb{P}_n$ is *contiguous* to $\mathbb{Q}_n$:

$$\mathbb{Q}_n[A_n] \to 0 \;\Rightarrow\; \mathbb{P}_n[A_n] \to 0.$$

Justification 2: control of asymptotic success

[Le Cam, 1960's] Suppose $\|\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}\| \leq K$ as $n \to \infty$. Then, $\mathbb{P}_n$ is *contiguous* to $\mathbb{Q}_n$:

$$\mathbb{Q}_n[A_n] \to 0 \;\Rightarrow\; \mathbb{P}_n[A_n] \to 0.$$

Corollary: Set $A_n = \{f_n(\boldsymbol{Y}) = \mathrm{p}\}$. Then:

$$\underbrace{\mathbb{Q}_n[f_n(\boldsymbol{Y}) = \mathrm{p}] \to 0}_{\text{success under } \mathbb{Q}_n} \;\Rightarrow\; \underbrace{\mathbb{P}_n[f_n(\boldsymbol{Y}) = \mathrm{p}] \to 0}_{\text{failure under } \mathbb{P}_n}.$$

Justification 2: control of asymptotic success

[Le Cam, 1960's] Suppose $\| \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \| \leq K$ as $n \to \infty$. Then, $\mathbb{P}_n$ is *contiguous* to $\mathbb{Q}_n$:

$$\mathbb{Q}_n[A_n] \to 0 \;\Rightarrow\; \mathbb{P}_n[A_n] \to 0.$$

Corollary: Set $A_n = \{f_n(\boldsymbol{Y}) = \mathrm{p}\}$. Then:

$$\underbrace{\mathbb{Q}_n[f_n(\boldsymbol{Y}) = \mathrm{p}] \to 0}_{\text{success under } \mathbb{Q}_n} \;\Rightarrow\; \underbrace{\mathbb{P}_n[f_n(\boldsymbol{Y}) = \mathrm{p}] \to 0}_{\text{failure under } \mathbb{P}_n}.$$

"Information-theoretic" (no efficiency worries) limitations:

$\| \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \|$ **bounded** $\Rightarrow$ **no test succeeds**

Testing **with** computational budget

What if we want to restrict to **efficiently computable** $f_n(\boldsymbol{Y})$, e.g., in time poly(N)?

Testing **with** computational budget

What if we want to restrict to **efficiently computable** $f_n(\boldsymbol{Y})$, e.g., in time poly(N)?

Heuristic: suppose the relevant tests are **polynomials**: $p \in \mathbb{R}[y_1, \ldots, y_N]$ with $\deg(p) \leq D$ computable in $O(N^D)$.

$$\begin{aligned} \text{maximize} \quad & \mathbb{E}_{\mathbb{P}}\, h(\boldsymbol{Y}) \\ \text{subject to} \quad & \mathbb{E}_{\mathbb{Q}}\, h(\boldsymbol{Y})^2 \leq 1 \\ & h(\boldsymbol{Y}) \in \mathbb{R}[y_1, \ldots, y_N] \\ & \deg(h) \leq D \end{aligned}$$

Testing **with** computational budget

What if we want to restrict to **efficiently computable** $f_n(\boldsymbol{Y})$, e.g., in time poly(N)?

Heuristic: suppose the relevant tests are **polynomials**: $p \in \mathbb{R}[y_1, \ldots, y_N]$ with $\deg(p) \leq D$ computable in $O(N^D)$.

$$\begin{aligned} \text{maximize} \quad & \left\langle h, \frac{d\mathbb{P}}{d\mathbb{Q}} \right\rangle \\ \text{subject to} \quad & \|h\|^2 \leq 1 \\ & h \in V^{\leq D}, \text{ a } \textbf{subspace} \end{aligned}$$

Testing **with** computational budget

What if we want to restrict to **efficiently computable** $f_n(\boldsymbol{Y})$, e.g., in time poly(N)?

Heuristic: suppose the relevant tests are **polynomials**:
$p \in \mathbb{R}[y_1, \ldots, y_N]$ with $\deg(p) \leq D$ computable in $O(N^D)$.

$$\begin{array}{ll} \text{maximize} & \left\langle h, \frac{d\mathbb{P}}{d\mathbb{Q}} \right\rangle \\ \text{subject to} & \|h\|^2 \leq 1 \\ & h \in V^{\leq D}, \text{a } \textbf{subspace} \end{array}$$

Optimizer: the (normalized) **low-degree likelihood ratio**

$$h^\star(\boldsymbol{Y}) = P^{\leq D}\frac{d\mathbb{P}}{d\mathbb{Q}}(\boldsymbol{Y}) \ \Big/ \ \underbrace{\left\| P^{\leq D}\frac{d\mathbb{P}}{d\mathbb{Q}} \right\|}_{\text{objective value}}$$

The low-degree conjecture

One wrinkle: rather than $D = \omega(1)$, to include calculation of spectral norms of matrices $\rightsquigarrow$ $D = \omega(\log N)$.

The low-degree conjecture

One wrinkle: rather than $D = \omega(1)$, to include calculation of spectral norms of matrices $\rightsquigarrow$ $D = \omega(\log N)$.

Main conjecture:

$$\| P^{\leq (\log N)^{1+\epsilon}} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \| \text{ **bounded** } \Rightarrow \text{ **no efficient test succeeds** }$$

The low-degree conjecture

One wrinkle: rather than $D = \omega(1)$, to include calculation of spectral norms of matrices $\rightsquigarrow D = \omega(\log N)$.

Main conjecture:

$$\| P^{\leq (\log N)^{1+\epsilon}} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \| \textbf{ bounded } \Rightarrow \textbf{ no efficient test succeeds}$$

Originally from sum-of-squares optimization (fancy semidefinite programming) literature: controls whether a lower bound construction succeeds or not.

- [Barak, Hopkins, Kelner, Kothari, Moitra, Potechin 2016]
- [Hopkins, Steurer 2017]
- [Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer 2017]
- [Hopkins 2018] (PhD thesis)

$$\limsup_{n\to\infty} \left\| P^{\leq (\log N)^{1+\epsilon}} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\| = \begin{cases} +\infty & \rightsquigarrow \text{ maybe easy} \\ K & \rightsquigarrow \text{ hard} \end{cases}$$

$$\limsup_{n\to\infty} \left\| P^{\leq (\log N)^{1+\epsilon}} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\| = \begin{cases} +\infty & \rightsquigarrow \text{ maybe easy} \\ K & \rightsquigarrow \text{ hard} \end{cases}$$

Question 1:

How to project to low-degree polynomials?

Question 2:

How to evaluate asymptotics?

A simple gaussian model

Let's restrict to a special case to show how this works:

- $\mathcal{P}_n$ a "prior" over $\mathbb{R}^N$.
- $\mathbb{Q}_n$: $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_N)$.
- $\mathbb{P}_n$: draw $\boldsymbol{X} \sim \mathcal{P}_n$, then $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}, \boldsymbol{I}_N)$.

A simple gaussian model

Let's restrict to a special case to show how this works:

- $\mathcal{P}_n$ a "prior" over $\mathbb{R}^N$.
- $\mathbb{Q}_n$: $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$.
- $\mathbb{P}_n$: draw $\boldsymbol{X} \sim \mathcal{P}_n$, then $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}, \boldsymbol{I}_N)$.

A very special case: $N(n) = n^2$, $\mathcal{P}_n$ distribution over rank 1 matrices $\boldsymbol{X} = \sqrt{\frac{n}{2}}\lambda \boldsymbol{x}\boldsymbol{x}^\top$, e.g., $\boldsymbol{x} \sim \mathrm{Unif}(\mathbb{S}^{n-1})$. Symmetrizing,

$$\underbrace{\mathrm{GOE}(n)}_{\mathbb{Q}_n} \quad \text{vs.} \quad \underbrace{\mathrm{GOE}(n) + \sqrt{n} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top}_{\mathbb{P}_n}$$

A simple gaussian model

Let's restrict to a special case to show how this works:

- $\mathcal{P}_n$ a "prior" over $\mathbb{R}^N$.
- $\mathbb{Q}_n$: $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_N)$.
- $\mathbb{P}_n$: draw $\boldsymbol{X} \sim \mathcal{P}_n$, then $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}, \boldsymbol{I}_N)$.

A very special case: $N(n) = n^2$, $\mathcal{P}_n$ distribution over rank 1 matrices $\boldsymbol{X} = \sqrt{\frac{n}{2}}\lambda \boldsymbol{x}\boldsymbol{x}^\top$, e.g., $\boldsymbol{x} \sim \mathrm{Unif}(\mathbb{S}^{n-1})$. Symmetrizing,

$$\underbrace{\mathrm{GOE}(n)}_{\mathbb{Q}_n} \quad \text{vs.} \quad \underbrace{\mathrm{GOE}(n) + \sqrt{n} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top}_{\mathbb{P}_n}$$

[Féral, Péché 2007] Top eigenvalue test succeeds iff $\lambda > 1$.

Question: Is this optimal?

Step 1: computing the likelihood ratio

The model:

- $\mathbb{Q}_n$: $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_N)$.
- $\mathbb{P}_n$: draw $\boldsymbol{X} \sim \mathcal{P}_n$, then $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}, \boldsymbol{I}_N)$.

Step 1: computing the likelihood ratio

The model:

- $\mathbb{Q}_n$: $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_N)$.
- $\mathbb{P}_n$: draw $\boldsymbol{X} \sim \mathcal{P}_n$, then $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}, \boldsymbol{I}_N)$.

For likelihood ratio, just need gaussian densities:

$$\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}(\boldsymbol{Y}) = \mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mathcal{P}_n}\left[\frac{d\mathbb{P}_n[\bullet|\boldsymbol{X}]}{d\mathbb{Q}_n}(\boldsymbol{Y})\right]$$

Step 1: computing the likelihood ratio

The model:

- $\mathbb{Q}_n$: $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$.
- $\mathbb{P}_n$: draw $\boldsymbol{X} \sim \mathcal{P}_n$, then $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}, \boldsymbol{I}_N)$.

For likelihood ratio, just need gaussian densities:

$$\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}(\boldsymbol{Y}) = \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mathcal{P}_n}\left[\frac{d\mathbb{P}_n[\bullet|\boldsymbol{X}]}{d\mathbb{Q}_n}(\boldsymbol{Y})\right] = \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mathcal{P}_n}\left[\frac{(2\pi)^{\text{something}}\exp(-\|\boldsymbol{Y}-\boldsymbol{X}\|^2/2)}{(2\pi)^{\text{something}}\exp(-\|\boldsymbol{Y}\|^2/2)}\right]$$

Step 1: computing the likelihood ratio

The model:

- $\mathbb{Q}_n$: $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_N)$.
- $\mathbb{P}_n$: draw $\boldsymbol{X} \sim \mathcal{P}_n$, then $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}, \boldsymbol{I}_N)$.

For likelihood ratio, just need gaussian densities:

$$\begin{aligned}
\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}(\boldsymbol{Y}) &= \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mathcal{P}_n}\left[\frac{d\mathbb{P}_n[\bullet|\boldsymbol{X}]}{d\mathbb{Q}_n}(\boldsymbol{Y})\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mathcal{P}_n}\left[\frac{(2\pi)^{\text{something}}\exp(-\|\boldsymbol{Y}-\boldsymbol{X}\|^2/2)}{(2\pi)^{\text{something}}\exp(-\|\boldsymbol{Y}\|^2/2)}\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mathcal{P}_n}\left[\exp\left(-\frac{1}{2}\|\boldsymbol{X}\|^2 + \langle\boldsymbol{X},\boldsymbol{Y}\rangle\right)\right]
\end{aligned}$$

Step 2: computing the low-degree projections

Use the orthogonal basis of **Hermite polynomials**,

$$h_k(y) \in \mathbb{R}[y]$$

$$H_{\boldsymbol{k}}(\boldsymbol{Y}) = \prod_{i=1}^{N} h_{k_i}(Y_i) \in \mathbb{R}[Y_1, \ldots, Y_N]$$

Step 2: computing the low-degree projections

Use the orthogonal basis of **Hermite polynomials**,

$$h_k(y) \in \mathbb{R}[y]$$

$$H_{\boldsymbol{k}}(\boldsymbol{Y}) = \prod_{i=1}^{N} h_{k_i}(Y_i) \in \mathbb{R}[Y_1, \ldots, Y_N]$$

Projections by **generalized gaussian integration by parts**:

$$\left\langle \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}, H_{\boldsymbol{k}} \right\rangle = \mathop{\mathbb{E}}_{\boldsymbol{Y} \sim \mathbb{Q}_n} \left[\frac{\partial^{\sum k_i}}{\partial Y_1^{k_1} \cdots \partial Y_N^{k_N}} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right]$$

Step 2: computing the low-degree projections

Use the orthogonal basis of **Hermite polynomials**,

$$h_k(y) \in \mathbb{R}[y]$$

$$H_{\boldsymbol{k}}(\boldsymbol{Y}) = \prod_{i=1}^{N} h_{k_i}(Y_i) \in \mathbb{R}[Y_1, \ldots, Y_N]$$

Projections by **generalized gaussian integration by parts**:

$$\begin{aligned}
\left\langle \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}, H_{\boldsymbol{k}} \right\rangle &= \mathop{\mathbb{E}}_{\boldsymbol{Y} \sim \mathbb{Q}_n} \left[\frac{\partial^{\sum k_i}}{\partial Y_1^{k_1} \cdots \partial Y_N^{k_N}} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right] \\
&= \mathop{\mathbb{E}}_{\substack{\boldsymbol{X} \sim \mathcal{P}_n \\ \boldsymbol{Y} \sim \mathbb{Q}_n}} \left[\prod X_i^{k_i} \exp\left(-\frac{1}{2} \|\boldsymbol{X}\|^2 + \langle \boldsymbol{X}, \boldsymbol{Y} \rangle \right) \right]
\end{aligned}$$

Step 2: computing the low-degree projections

Use the orthogonal basis of **Hermite polynomials**,

$$h_k(y) \in \mathbb{R}[y]$$

$$H_{\boldsymbol{k}}(\boldsymbol{Y}) = \prod_{i=1}^{N} h_{k_i}(Y_i) \in \mathbb{R}[Y_1, \ldots, Y_N]$$

Projections by **generalized gaussian integration by parts**:

$$\begin{aligned}
\left\langle \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}, H_{\boldsymbol{k}} \right\rangle &= \mathop{\mathbb{E}}_{\boldsymbol{Y}\sim\mathbb{Q}_n} \left[\frac{\partial^{\sum k_i}}{\partial Y_1^{k_1} \cdots \partial Y_N^{k_N}} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right] \\
&= \mathop{\mathbb{E}}_{\substack{\boldsymbol{X}\sim\mathcal{P}_n \\ \boldsymbol{Y}\sim\mathbb{Q}_n}} \left[\prod X_i^{k_i} \exp\left(-\frac{1}{2}\|\boldsymbol{X}\|^2 + \langle \boldsymbol{X}, \boldsymbol{Y} \rangle \right) \right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mathcal{P}_n} \left[\prod X_i^{k_i} \right]
\end{aligned}$$

Step 3: computing the norm

Use (part of) the **replica trick** to handle squared $\mathbb{E}[\cdots]$:

$$\left\| P^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \sum_{\sum k_i \leq D} \frac{1}{\prod k_i!} \left\langle \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}, H_{\boldsymbol{k}} \right\rangle^2$$

$$= \sum_{\sum k_i \leq D} \frac{1}{\prod k_i!} \left(\mathbb{E}_{\boldsymbol{x} \sim \mathcal{P}_n} \left[\prod x_i^{k_i} \right] \right)^2$$

Step 3: computing the norm

Use (part of) the **replica trick** to handle squared $\mathbb{E}[\cdots]$:

$$\left\| P^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \sum_{\sum k_i \leq D} \frac{1}{\prod k_i!} \left\langle \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}, H_{\boldsymbol{k}} \right\rangle^2$$

$$= \sum_{\sum k_i \leq D} \frac{1}{\prod k_i!} \left(\mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mathcal{P}_n} \left[\prod X_i^{k_i} \right] \right)^2$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{X}, \boldsymbol{X}' \sim \mathcal{P}_n} \sum_{\sum k_i \leq D} \prod_i \frac{(X_i X_i')^{k_i}}{k_i!}$$

Step 3: computing the norm

Use (part of) the **replica trick** to handle squared $\mathbb{E}[\cdots]$:

$$\left\| P^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \sum_{\sum k_i \leq D} \frac{1}{\prod k_i!} \left\langle \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}, H_{\boldsymbol{k}} \right\rangle^2$$

$$= \sum_{\sum k_i \leq D} \frac{1}{\prod k_i!} \left(\mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mathcal{P}_n} \left[\prod X_i^{k_i} \right] \right)^2$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{X}, \boldsymbol{X}' \sim \mathcal{P}_n} \sum_{\sum k_i \leq D} \prod_i \frac{(X_i X_i')^{k_i}}{k_i!}$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{X}, \boldsymbol{X}' \sim \mathcal{P}_n} \sum_{d=0}^{D} \frac{1}{d!} \sum_{\sum k_i = d} \binom{d}{k_1 \cdots k_N} \prod_i (X_i X_i')^{k_i}$$

Step 3: computing the norm

Use (part of) the **replica trick** to handle squared $\mathbb{E}[\cdots]$:

$$\left\| P^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \sum_{\sum k_i \leq D} \frac{1}{\prod k_i!} \left\langle \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}, H_{\boldsymbol{k}} \right\rangle^2$$

$$= \sum_{\sum k_i \leq D} \frac{1}{\prod k_i!} \left(\mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mathcal{P}_n} \left[\prod X_i^{k_i} \right] \right)^2$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{X}, \boldsymbol{X}' \sim \mathcal{P}_n} \sum_{\sum k_i \leq D} \prod_i \frac{(X_i X_i')^{k_i}}{k_i!}$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{X}, \boldsymbol{X}' \sim \mathcal{P}_n} \sum_{d=0}^{D} \frac{1}{d!} \sum_{\sum k_i = d} \binom{d}{k_1 \cdots k_N} \prod_i (X_i X_i')^{k_i}$$

$$= \boxed{\mathop{\mathbb{E}}_{\boldsymbol{X}, \boldsymbol{X}' \sim \mathcal{P}_n} \sum_{d=0}^{D} \frac{1}{d!} \langle \boldsymbol{X}, \boldsymbol{X}' \rangle^d}$$

Step 4: evaluating the asymptotic

The special case: $\boldsymbol{X} = \sqrt{n/2} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top$, $\boldsymbol{x} \sim \mathrm{Unif}(\mathbb{S}^{n-1})$.

$$\left\| P^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{X}' \sim \mathcal{P}_n} \sum_{d=0}^{D} \frac{1}{d!} \langle \boldsymbol{X}, \boldsymbol{X}' \rangle^d$$

Step 4: evaluating the asymptotic

The special case: $\boldsymbol{X} = \sqrt{n/2} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top$, $\boldsymbol{x} \sim \mathrm{Unif}(\mathbb{S}^{n-1})$.

$$\left\| P^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{X}' \sim \mathcal{P}_n} \sum_{d=0}^{D} \frac{1}{d!} \langle \boldsymbol{X}, \boldsymbol{X}' \rangle^d$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{x}' \sim \mathrm{Unif}(\mathbb{S}^{n-1})} \sum_{d=0}^{D} \frac{1}{d!} \left(\frac{\lambda^2}{2} \cdot n \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^2 \right)^d$$

Step 4: evaluating the asymptotic

The special case: $\boldsymbol{X} = \sqrt{n/2} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top$, $\boldsymbol{x} \sim \text{Unif}(\mathbb{S}^{n-1})$.

$$\left\| P^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{X}' \sim \mathcal{P}_n} \sum_{d=0}^{D} \frac{1}{d!} \langle \boldsymbol{X}, \boldsymbol{X}' \rangle^d$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{x}' \sim \text{Unif}(\mathbb{S}^{n-1})} \sum_{d=0}^{D} \frac{1}{d!} \left(\frac{\lambda^2}{2} \cdot n \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^2 \right)^d$$

By CLT, $\sqrt{n} \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle \Rightarrow \mathcal{N}(0, 1)$, so...

Step 4: evaluating the asymptotic

The special case: $\boldsymbol{X} = \sqrt{n/2} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top$, $\boldsymbol{x} \sim \mathrm{Unif}(\mathbb{S}^{n-1})$.

$$\left\| P^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{X}' \sim \mathcal{P}_n} \sum_{d=0}^{D} \frac{1}{d!} \langle \boldsymbol{X}, \boldsymbol{X}' \rangle^d$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{x}' \sim \mathrm{Unif}(\mathbb{S}^{n-1})} \sum_{d=0}^{D} \frac{1}{d!} \left(\frac{\lambda^2}{2} \cdot n \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^2 \right)^d$$

By CLT, $\sqrt{n} \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle \Rightarrow \mathcal{N}(0,1)$, so...

$$\approx \mathop{\mathbb{E}}_{g \sim \mathcal{N}(0,1)} \sum_{d=0}^{D} \frac{1}{d!} \left(\frac{\lambda^2}{2} g^2 \right)^d \qquad (\text{if } D \ll n)$$

Step 4: evaluating the asymptotic

The special case: $\boldsymbol{X} = \sqrt{n/2} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top$, $\boldsymbol{x} \sim \mathrm{Unif}(\mathbb{S}^{n-1})$.

$$\left\| P^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{X}' \sim \mathcal{P}_n} \sum_{d=0}^{D} \frac{1}{d!} \langle \boldsymbol{X}, \boldsymbol{X}' \rangle^d$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{x}' \sim \mathrm{Unif}(\mathbb{S}^{n-1})} \sum_{d=0}^{D} \frac{1}{d!} \left(\frac{\lambda^2}{2} \cdot n \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^2 \right)^d$$

By CLT, $\sqrt{n} \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle \Rightarrow \mathcal{N}(0,1)$, so...

$$\approx \mathop{\mathbb{E}}_{g \sim \mathcal{N}(0,1)} \sum_{d=0}^{D} \frac{1}{d!} \left(\frac{\lambda^2}{2} g^2 \right)^d \qquad \text{(if } D \ll n\text{)}$$

$$\to \mathop{\mathbb{E}}_{g \sim \mathcal{N}(0,1)} \exp\left(\frac{\lambda^2}{2} g^2 \right).$$

Step 4: evaluating the asymptotic

The special case: $\boldsymbol{X} = \sqrt{n/2} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top$, $\boldsymbol{x} \sim \mathrm{Unif}(\mathbb{S}^{n-1})$.

$$\left\| P^{\leq D} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2 = \mathop{\mathbb{E}}_{\boldsymbol{X}, \boldsymbol{X}' \sim \mathcal{P}_n} \sum_{d=0}^{D} \frac{1}{d!} \langle \boldsymbol{X}, \boldsymbol{X}' \rangle^d$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{x}, \boldsymbol{x}' \sim \mathrm{Unif}(\mathbb{S}^{n-1})} \sum_{d=0}^{D} \frac{1}{d!} \left(\frac{\lambda^2}{2} \cdot n \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^2 \right)^d$$

By CLT, $\sqrt{n} \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle \Rightarrow \mathcal{N}(0, 1)$, so...

$$\approx \mathop{\mathbb{E}}_{g \sim \mathcal{N}(0,1)} \sum_{d=0}^{D} \frac{1}{d!} \left(\frac{\lambda^2}{2} g^2 \right)^d \qquad \text{(if } D \ll n\text{)}$$

$$\to \mathop{\mathbb{E}}_{g \sim \mathcal{N}(0,1)} \exp\left(\frac{\lambda^2}{2} g^2 \right).$$

Key: $D(n) \ll n$, so CLT “kicks in” in time for moments.

Summary of spiked matrix model

$$\mathrm{GOE}(n) \quad \text{vs.} \quad \mathrm{GOE}(n) + \sqrt{n} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top$$

Summary of spiked matrix model

$$\mathrm{GOE}(n) \quad \text{vs.} \quad \mathrm{GOE}(n) + \sqrt{n} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top$$

$$\downarrow$$

$$\limsup_{n \to \infty} \left\| P^{\leq D(n)} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2$$

Summary of spiked matrix model

$$\mathrm{GOE}(n) \quad \text{vs.} \quad \mathrm{GOE}(n) + \sqrt{n} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top$$

$$\downarrow$$

$$\limsup_{n \to \infty} \left\| P^{\leq D(n)} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2$$

$$\downarrow$$

$$\limsup_{n \to \infty} \mathop{\mathbb{E}}_{\boldsymbol{x}, \boldsymbol{x}'} \sum_{d=0}^{D(n)} \frac{1}{d!} \left(\frac{\lambda^2}{2} \cdot n \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^2 \right)^d$$

Summary of spiked matrix model

$$\mathrm{GOE}(n) \quad \text{vs.} \quad \mathrm{GOE}(n) + \sqrt{n} \cdot \lambda \boldsymbol{x}\boldsymbol{x}^\top$$

$$\downarrow$$

$$\limsup_{n\to\infty} \left\| P^{\le D(n)} \frac{d\mathbb{P}_n}{d\mathbb{Q}_n} \right\|^2$$

$$\downarrow$$

$$\limsup_{n\to\infty} \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{x}'} \sum_{d=0}^{D(n)} \frac{1}{d!} \left(\frac{\lambda^2}{2} \cdot n \cdot \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^2 \right)^d$$

$$\downarrow$$

$$\underbrace{\mathop{\mathbb{E}}_{g\sim\mathcal{N}(0,1)} \exp\left(\frac{\lambda^2}{2} g^2 \right)}_{\text{natural, \textbf{scalar} expectation!}}$$

Review

1. The *low-degree conjecture* connects hardness of statistical testing with the norm of the *low-degree likelihood ratio*.

Review

1. The *low-degree conjecture* connects hardness of statistical testing with the norm of the *low-degree likelihood ratio*.

2. To analyze a problem, we proceed as follows:
   2.1 Compute the likelihood ratio
   2.2 Find the orthogonal polynomials of the null model ($\mathbb{Q}$)
   2.3 Project (using **special distributional properties**)
   2.4 Compute the norm (using “**baby replica trick**”)
   2.5 Reduce to scalar expectation (**limit theorem** heuristic)

Other frameworks for hardness predictions

1. Conjecturally optimal algorithms
   1.1 BP / AMP $\sim$ cavity and replica methods of stat. physics
   1.2 Sum-of-squares hierarchy (semidefinite programming)
   1.3 Monte Carlo sampling from posterior
   1.4 Local algorithms
   1.5 Problem-specific algorithms (e.g. PCA)
2. Structure of solution space ("shattering" & co.)
3. Geometric analysis of optimization landscapes
4. Average-case reductions

The bright side

The low degree method is...

- Easy
- Uniform across problems
- Broadly applicable (to nice “toy-ish” setups)
- Intuitively plausible
- Always correct (so far)

The other hand

The low degree method is...

- Coarse-grained in runtimes
- Hard to handle correlated models with
- **Dependent on orthogonal polynomial magic**
- Dependent on good control of signal priors
- Not a great way to design actual algorithms

So...give it a try when you are wearing your theorist hat, and want to make a **quick, painless prediction of thresholds for a nice model.**

Thank you!