Predicting computational difficulty of binary hypothesis testing.

Distinguish between two distributions $\mathcal{P}$ and $\mathcal{Q}$.

$H_1 : Y \sim \mathcal{P}_n$, $H_0 : Y \sim \mathcal{Q}_n$.

Test minimizes Type I + Type II error via

Likelihood ratio test $L_n = \frac{d \mathcal{P}_n}{d \mathcal{Q}_n}$ ($> 1 \text{ or } < 1$)

Restrict theory to the class of polynomial tests.

$\rightarrow p : R^n \rightarrow R$

$\rightarrow$ Apply test $p(Y)$ ($\approx$ $t$ or $< t$).

Another characterization the $t$ is:

Let's assume that $\mathcal{P}_n \subset \mathcal{Q}_n$, and

$\mathcal{P}_n$ has all moments finite.

$y_i \in \mathcal{L}^2(\mathbb{R}), \forall i, n$. 
\[
\begin{align*}
\text{max} & \quad \mathbb{E}_{f \in \mathcal{F}} \int f(y) \, dy = \langle L, f \rangle_{L^2(\mathcal{Q})} \\
\text{s.t.} & \quad \mathbb{E}_{f \in \mathcal{F}} \int f(y)^2 \, dy = 1 \\
& \quad \mathbb{E}_{f \in \mathcal{F}} \int L f(y) \, dy = 1
\end{align*}
\]

\[
\text{argmax} \quad \frac{\mathbb{E}_{f \in \mathcal{F}} \int f(y) \, dy}{\mathbb{E}_{f \in \mathcal{F}} \int f(y)^2 \, dy} = \frac{2}{\sqrt{2} \sqrt{2}}
\]

\[
\text{restrict optimization to bounded degree polynomials}
\]

Main idea (Hopkins, Steurer 16 (?)).
maximizer is $f^* = \frac{L \leq d}{\|L \leq d\|_2}$.

Can (in principle) compute $L_{sd}$ in $O(n^d)$ time.

Objective: $\|L \leq d\|_2$.

Prediction on computational difficulty:

whether $\|L \leq d\|_2$ is bounded or diverges.

$\xrightarrow{\text{under } F\text{-expander}} \xrightarrow{\text{Expander up}}$.

Analogy with contiguity via second moment:

$P_n \propto \Omega_n$ if $\lim_{n \to \infty} \mathbb{E}[L_n^2] < +\infty$

Application: Additive Gaussian channel.

$H_1: Y = X + Z$, $Z \sim N(0, I_n)$, $X \sim P_0$ on $\mathbb{R}^n$.

$H_0: Y = Z$.

$LR: L(x) = \max_{x \in \mathbb{R}^n} \sum_{x \in \mathbb{R}^n} \mathbb{E}_{x \sim P_0}[\sum_{x \sim P_0} \langle Y, x \rangle - \frac{1}{2}x'y'x']$.

How to compute $L_{sd}$?
Consider the basis of Hermite polynomials.

A Hilbert space is $L^2(\mathbb{R}^n)$, $\mathbb{R}^n = N(0, I_n)$ (univariate)

Hermite polynomials: $h_k(x) = (-1)^k e^{\frac{x^2}{2}} \frac{d^k}{dx^k} (e^{-x^2/2})$

$x \geq 0, x \in \mathbb{R}$

$$\int_{x \in \mathbb{R}^n} h_k(x) h_{k'}(x) \, dx = \delta_{k,k'}$$

$$h_k(x) = \frac{h_k(x)}{\sqrt{k!}} \text{ for } L^2(N(0,1))$$

Multivariate Hermite polynomials:

$x \in \mathbb{R}^n$

$\mathbb{R}^n$

$\sum_{i=1}^{n} h_{k_i}(x_i)$

$\alpha \in N^n$

$\sum_{i=1}^{n} h_{\alpha_i}(x_i)$

$\Rightarrow$ orthonormal basis for $L^2(\mathbb{R}^n)$.

$$L(Y) = \sum_{\alpha \in N^n} < L, \hat{h}_\alpha > \cdot \hat{h}_\alpha(Y).$$

$$L \circ \delta(Y) = \sum_{\alpha \in N^n} < L, \hat{h}_\alpha > \cdot \hat{h}_\alpha(Y).$$

$$|\alpha| = \sum_{i=1}^{n} \alpha_i.$$

hence $\prod_{x \in \mathbb{R}^n}$
Compute inner products: $< L, \mathbf{A}_2 >$

**Lemma 1**: $\| L \|^2_2 = \mathbb{E} \sum_{x_1, x_2} \frac{1}{2^d} \mathbb{E}_{x_1, x_2 \sim P} < x_1, x_2 >$

Parallel with $\| L \|^2_2$:

$\| L \|^2_2 = \mathbb{E} \sum_{x_1, x_2} L^2 < x_1, x_2 >$

Proof:

$\mathbb{E} \sum_{x_1, x_2} L^2 < x_1, x_2 > = \mathbb{E} \sum_{x_1, x_2} \left( \frac{1}{2^d} \mathbb{E}_{x_1, x_2 \sim P} < x_1, x_2 > \right)$

Proof of Lemma 1:

Gaussian integration by parts:

$\mathbb{E} \left\{ \sum_{k=0}^{+\infty} \frac{1}{k!} < x_1, x_2 > k \right\}$

$\iff$ weakly diff, poly growth at $+\infty$
A generalization of G.I.P.

\[ E \left[ \frac{1}{L} \sum_{x_1} \left( \frac{1}{L} \sum_{x_1} h(x_1) \right) \right] = E \left[ \frac{1}{L} \sum_{x_1} h(x_1) \right]. \]

\[ \text{for } f \text{ w-diff, poly growth.} \]

\[ L \leq d = \sum_{x \in \mathbb{N}^n} \left< L, \hat{A}_x \right> \cdot \hat{A}_x \]

\[ \left\| L \left( d \right) \right\|^2 \leq \sum_{x \in \mathbb{N}^n} \left< L, \hat{A}_x \right>^2 \cdot \left\| L \left( d \right) \right\|^2 \]

\[ \left\langle L, \hat{A}_x \right\rangle = E \sum_{L \cdot \hat{A}_x} L \cdot \hat{A}_x \]
\[
\langle L, H^\alpha \rangle = \frac{1}{\prod_{i=1}^{n} \alpha_i} \cdot \mathbb{E}_{\prod \alpha_i \leq p_0} \left[ \prod_{i=1}^{n} x_i ^\alpha \right] \exp \left\{ \gamma y_k - \frac{\|y\|^2}{2} \right\}
\]

\[
= \frac{1}{\prod_{i=1}^{n} \alpha_i} \cdot \mathbb{E}_{\prod \alpha_i \leq p_0} \left[ \prod_{i=1}^{n} x_i ^\alpha \right].
\]

\[
\Rightarrow \quad \prod_{1 \leq i \leq d} H^\alpha = \sum_{\prod \alpha_i \leq p_0} \frac{1}{\prod_{i=1}^{n} \alpha_i} \mathbb{E}_{\prod \alpha_i \leq p_0} \left[ \prod_{i=1}^{n} x_i ^\alpha \right] \exp \left\{ \sum_{i=1}^{n} (x_i x_i ^\gamma) \right\}.
\]

\[
= \sum_{k=0}^{d} \frac{1}{k!} \mathbb{E}_{\prod \alpha_i \leq p_0} \left[ \left( \sum_{i=1}^{n} x_i x_i ^\gamma \right)^k \right].
\]

\[
= \sum_{k=0}^{d} \frac{1}{k!} \mathbb{E}_{\prod \alpha_i \leq p_0} \left[ \langle x, x \rangle ^k \right].
\]

Another view: Fix \( x \in \mathbb{R} \)

\[
y \mapsto e^{xy} = x^{3/2} \text{ expanded on the basis of } (h_k)_{k=0}^{\infty} ?
\]

\[
e^{xy} = x^{3/2} = \sum_{k=0}^{\infty} \frac{1}{k!} h_k (y), \quad \text{by}(1).
\]
\[ L = \mathbb{E}_{x \sim \mathcal{P}_0} \left[ \sum_{i=1}^{\infty} \left( \sum_{k=0}^{\infty} \frac{1}{k!} x_i^k h_k(y_i) \right) \right] \]

\[ = \sum_{x \in \mathbb{N}^n} \mathbb{E}_{x \sim \mathcal{P}_0} \left[ \prod_{i=1}^{\infty} x_i^{\lambda_i} h_{\lambda_i}(y_i) \right] \]

\[ = \sum_{x \in \mathbb{N}^n} \mathbb{E}_{x \sim \mathcal{P}_0} \left[ \prod_{i=1}^{\infty} x_i^{\lambda_i} \right] \mathbb{E}_{x \sim \mathcal{P}_0} \left[ H_\lambda(y) \right] \]

\[ \Rightarrow L \leq d = \sum_{x \in \mathbb{N}^n} \mathbb{E}_{x \sim \mathcal{P}_0} \left[ \prod_{i=1}^{\infty} x_i^{\lambda_i} \right] \mathbb{E}_{x \sim \mathcal{P}_0} \left[ H_\lambda(y) \right] \]

Special case: spiked Wigner model:

\[ Y_{ij} = \sum_{i,j} x_i x_j + \epsilon_{ij} \quad \epsilon_{ij} \sim \mathcal{N}(0,1) \]

\[ x_i \sim \mathcal{P}_0, \text{ zero mean, unit variance} \]

\[ \lambda_\epsilon : \text{ see Threshold } (\lambda_\epsilon < 1) \]

Look at the LDLR:

\[ \| L \leq \lambda_\epsilon \|_2^2 = \sum_{k=0}^{d} \left( \frac{\lambda_\epsilon}{2\pi} \right)^{\frac{k}{2}} \frac{1}{k!} \mathbb{E}_{x \sim \mathcal{P}_0} \left[ \frac{1}{\sqrt{n}} \left( x, x' \right)^2 \right] \]

\[ = \sum_{k=0}^{d} \frac{1}{2^k \lambda_\epsilon^k} \mathbb{E}_{x \sim \mathcal{P}_0} \left[ \frac{1}{\sqrt{n}} \left( x, x' \right)^2 \right] \]
By the CLT:

\[
\lim_{n \to +\infty} \frac{1}{2} \sum_{k=0}^{2n} \frac{X_k}{k!} = \mathbb{E} \left[ \sum_{k=0}^{2n} \frac{X_k}{k!} \right].
\]

Fact: If \( \left| \frac{X_{k+1}}{X_k} \right| \to R > 1 \) then \( \sum |a_k| = +\infty \).

\[
\mathbb{E} \left[ \sum \frac{X_k}{k!} \right] = (\ell_k - 1)(\ell_k - 3) - 1.
\]

\[
a_k = \frac{1}{2k} \frac{X_k}{k!} \in \mathbb{E} \left[ \sum \frac{X_k}{k!} \right].
\]

\[
\frac{q_{k+1}}{a_k} = \frac{1}{2k+1} \frac{(2k+1)(3k-1)(5k-3)}{(2k-1)(3k-3)} - \frac{1}{2} \frac{X_k}{k!} \to \lambda.
\]

If \( \lambda < 1 \):

\[
\lim_{n \to +\infty} \frac{1}{2} \sum_{k=0}^{2n} \frac{X_k}{k!} < +\infty.
\]

\[
\Rightarrow \text{LDLR predicts testing is computationally difficult for all } \lambda < 1.
\]