STSCI 6940:

Lecture 13 Date: 03/25

§ Hypothesis testing.

"interesting" $Y = \sqrt{\frac{\lambda}{n}} xx^T + W$. $\mathbb{P}_\lambda$

"Dull" $Y = W$. $\mathbb{P}_0$

$x \in \mathbb{R}^n$, $x_i \overset{iid}{\sim} P_0$ zero mean, unit variance.

Interested in distinguishing $\mathbb{P}_\lambda$ from $\mathbb{P}_0$.

General theory:

A test $T : \mathbb{R}^{n \times n} \to \{0, 1\}$,

Type I: $\mathbb{P}_0(T = 1)$

Type II: $\mathbb{P}_\lambda(T = 0)$.

$$\text{err}(T) = \mathbb{P}_0(T=1) + \mathbb{P}_\lambda(T=0).$$

Interested in minimizing err(T).

- Neyman-Pearson Lemma: the test minimizing $T \to \text{err}(T)$ is

$$T^*(Y) = \begin{cases} 1 & \text{if } L(Y) \geq 1 \\ 0 & \text{o/w}. \end{cases}$$

likelihood ratio test

$$\mathcal{L}(Y) \equiv \frac{d\mathbb{P}_\lambda}{d\mathbb{P}_0}(Y) \quad : \text{ likelihood ratio}$$

$$\mathrm{err}(T) = \underset{\mathbb{Q}}{\mathbb{P}_0}(T=0) + \underset{\mathbb{P}}{\mathbb{P}_\lambda}(T=1).$$

$$= \int \left(\mathbb{1}\{T=0\} + L \cdot \mathbb{1}\{T=1\}\right) d\mathbb{Q}.$$

$$\leq \int \left(\mathbb{1}\{L<1\} + L \cdot \mathbb{1}\{L \geq 1\}\right) d\mathbb{Q}$$

$$T^* = \mathbb{1}\{L \geq 1\}$$

$$= 1 - d_{TV}(\mathbb{P}, \mathbb{Q}).$$

$$\to \quad d_{TV}(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} \int |L - 1| \, d\mathbb{Q}.$$

- strong detection: $\mathrm{err}(T^*) \xrightarrow[n \to +\infty]{} 0$.

  equivalently $d_{TV}(\mathbb{P}, \mathbb{Q}) \longrightarrow 1.$

- weak detection: $\overline{\lim}\, \mathrm{err}(T^*) < 1.$

- KL divergence:

$$d_{KL}(\mathbb{P}, \mathbb{Q}) = \mathbb{E}_{\mathbb{P}}\left[\log \frac{d\mathbb{P}}{d\mathbb{Q}}\right]$$

$$= \mathbb{E}_{\mathbb{P}}[\log L]$$

$$= \mathbb{E}_{\mathbb{Q}}[L \log L] \quad \geq 0.$$

$$\log x \leq x - 1$$

$$
\begin{aligned}
d_{KL}(\mathbb{P}, \mathbb{Q}) &= \mathbb{E}_{\mathbb{Q}}[L \log L] \\
&\leq \mathbb{E}_{\mathbb{Q}}[L(L-1)] \\
&= \mathbb{E}_{\mathbb{Q}}[L^2] - 1 \\
&= \mathbb{E}_{\mathbb{Q}}[(L-1)^2]. \\
&= d_{\chi^2}(\mathbb{P}, \mathbb{Q}).
\end{aligned}
$$

- when strong detection not possible?

- Def (Contiguity): $\mathbb{P}_n \triangleleft \mathbb{Q}_n$ if for every sequence of events (A_n) we have.

$$\mathbb{Q}_n(A_n) \longrightarrow 0 \Rightarrow \mathbb{P}_n(A_n) \longrightarrow 0.$$

Mutual contiguity $\mathbb{P}_n \triangleleft\triangleright \mathbb{Q}_n$ if $\mathbb{P}_n \triangleleft \mathbb{Q}_n$ and $\mathbb{Q}_n \triangleleft \mathbb{P}_n$.

- Fact: If $\mathbb{P}_n \triangleleft \mathbb{Q}_n$ the strong detection is impossible

$$A_n = \{T = 1\}$$

suppose $\mathbb{Q}_n(A_n) \longrightarrow 0 \Rightarrow \mathbb{P}_n(A_n) \longrightarrow 0$

$$\Rightarrow \quad err_n(T_n) \longrightarrow 1.$$

- Proving contiguity: $\mathbb{Q}_n(A_n) \to 0$.

$$\mathbb{P}_n(A_n) = \mathbb{E}_{\mathbb{P}_n}[\mathbb{1}\{A_n\}].$$
$$= \mathbb{E}_{\mathbb{Q}_n}[L_n \cdot \mathbb{1}\{A_n\}].$$
$$\leq \mathbb{E}_{\mathbb{Q}_n}[L_n^2]^{\frac{1}{2}} \cdot \mathbb{E}_{\mathbb{Q}_n}[\mathbb{1}\{A_n\}]^{\frac{1}{2}}$$
$$= \mathbb{E}_{\mathbb{Q}_n}[L_n^2]^{\frac{1}{2}} \cdot \mathbb{Q}_n(A_n)^{\frac{1}{2}}.$$

If $\sup_n \mathbb{E}_{\mathbb{Q}_n}[L_n^2] < +\infty$ then

$$\mathbb{P}(A_n) \to 0 \ : \ \mathbb{P}_n \triangleleft \mathbb{Q}_n.$$

Remark: $d_{KL}(\mathbb{P}_n, \mathbb{Q}_n) \leq \mathbb{E}_{\mathbb{Q}_n}[L_n^2] - 1$

- Consider spiked wigner :

$$L = \frac{d\mathbb{P}_\lambda}{d\mathbb{P}_0} \qquad \mathbb{P}_n = \mathbb{P}_\lambda, \ \mathbb{Q}_n = \mathbb{P}_0.$$

$$\begin{cases} Y_{ij} = \sqrt{\frac{\lambda}{n}} x_i x_j + w_{ij} & i<j \\ Y_{ii} = \sqrt{\frac{\lambda}{n}} x_i^2 + w_{ii} & i. \end{cases}$$

$$d\mathbb{P}_\lambda(Y) = \left[\int \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}\sum_{i<j}\left(Y_{ij} - \sqrt{\frac{\lambda}{n}} x_i x_j\right)^2} dP_0^n(x).\right] \cdot dY.$$

$$L = \int e^{\sqrt{\frac{\lambda}{n}}\sum_{i<j} Y_{ij} x_i x_j - \frac{\lambda}{2n}\sum_{i<j} x_i^2 x_j^2} dP_0^n(x).$$

(Ignoring the diagonal)

Including the diagonal: $\quad W_{ii} \sim N(0, \sigma^2)$.

$$L = \int e^{\left(\sum_{i<j} \sqrt{\frac{\lambda}{n}} Y_{ij} x_i x_j - \frac{\lambda}{2n} x_i^2 x_j^2 + \frac{1}{\sigma^2} \sum_i \sqrt{\frac{\lambda}{n}} Y_{ii} x_i^2 - \frac{\lambda}{2n} x_i^4 \right)} dP_0^n(x).$$

Posterior measure:

$$\mathbb{P}(\cdot \mid Y) = \frac{e^{H_n(x)}}{Z_n}.$$

$$\Rightarrow \quad L = Z_n.$$

$$d_{KL}(\mathbb{P}_\lambda, \mathbb{P}_0) = \mathbb{E}_{\mathbb{P}_\lambda}[\log Z_n(Y)]$$

$$= F_n(\lambda).$$

we know:

$$\underset{\frac{1}{n}\mathbb{E}_{\mathbb{P}_\lambda} \log L}{\frac{1}{n} d_{KL}(\mathbb{P}_\lambda, \mathbb{P}_0)} \longrightarrow \phi_{RS}(\lambda) \geq 0$$

$$\begin{cases} \phi_{RS}(\lambda) > 0 & \text{if } \lambda > \lambda_c \\ \phi_{RS}(\lambda) = 0 & \text{if } \lambda \leq \lambda_c. \end{cases}$$

second moment bounded $\Rightarrow$ $d_{KL}(\mathbb{P}_\lambda, \mathbb{P}_0)$ bounded

$$\Rightarrow \phi_{RS}(\lambda) = 0$$

$\Rightarrow$ If λ is such that $\sup_n \mathbb{E}_{\mathbb{P}_0}[L(Y)^2] < +\infty$

then $\quad \lambda \leq \lambda_c$.

Conversely, there are scenarios where $\lambda < \lambda_c$ but

$$\mathbb{E}_{\mathbb{P}_0}\left[L(Y)^2 \right] \longrightarrow +\infty .$$

Conditioning the second moment computation:

Identify a sequence of good events $\mathcal{E}_n$ such that $\mathbb{P}_n(\mathcal{E}_n) \to 1$

[want to prove $\mathbb{P}_n \triangleleft \mathbb{Q}_n$].

Take a sequence of events (A_n) s.t. $\mathbb{Q}_n(A_n) \to 0$.

$$\mathbb{P}_n(A_n) = \mathbb{P}_n(A_n \mid \mathcal{E}_n)\, \mathbb{P}(\mathcal{E}_n) + \mathbb{P}_n(A_n \mid \mathcal{E}_n^c) \cdot \mathbb{P}(\mathcal{E}_n^c)$$

$$\leq \mathbb{P}_n(A_n \mid \mathcal{E}_n) + \underbrace{\mathbb{P}(\mathcal{E}_n^c)}_{\to 0} .$$

$$\leq \mathbb{E}_{\mathbb{Q}_n}\left[\tilde{L}_n^2 \right]^{\frac{1}{2}} . \mathbb{Q}_n(A_n)^{\frac{1}{2}} + \mathbb{P}(\mathcal{E}_n^c).$$

$$\tilde{L}_n = \frac{d\, \mathbb{P}_n(\cdot \mid \mathcal{E}_n)}{d\, \mathbb{Q}_n} .$$

For the spiked wigner model:

- Can we prove strong detection for all $\lambda > \lambda_c$
- Can we prove mutual contiguity for all $\lambda < \lambda_c$ and what is the error of the best test for $\lambda < \lambda_c$?

Remark: unclear what happens at $\lambda = \lambda_c$ ...

[Guess: detection is possible?]

$$L \sim n^{\alpha} \ ???$$

. Strong detection above λ_c :

we have $\frac{1}{n} \mathbb{E}_{\mathbb{P}_\lambda} \log L \to \phi_{RS}(\lambda)$.

$\phi_{RS}(\lambda) > 0$ if $\lambda > \lambda_c$.

$\mathbb{E}_{\mathbb{P}_0} \log L \leq \log \underbrace{\mathbb{E}_{\mathbb{P}_0}[L]}_{=1} = 0$.

$\overline{\lim} \frac{1}{n} \mathbb{E}_{\mathbb{P}_0} \log L \leq 0$.

Look at $\frac{1}{n} \log L$ and decide based on its sign.

. Concentration of measure ; (prior has bounded support)

$$\mathbb{P}_\lambda \left(\left| \frac{1}{n} \log L - \mathbb{E}_\lambda \log L \right| \geq t \right) \leq e^{-\frac{nt^2}{c}} \quad \forall t$$

$c > 0$.

$\forall \lambda \geq 0$.

$\Rightarrow$ If $\lambda > \lambda_c$: $\frac{1}{n} \log L \geq \frac{\phi_{RS}(\lambda)}{2}$ w.p $\to 1$
under $\mathbb{P}_\lambda$

under $\mathbb{P}_0$: $\frac{1}{n} \log L \leq \varepsilon$ w.p $\to 1$

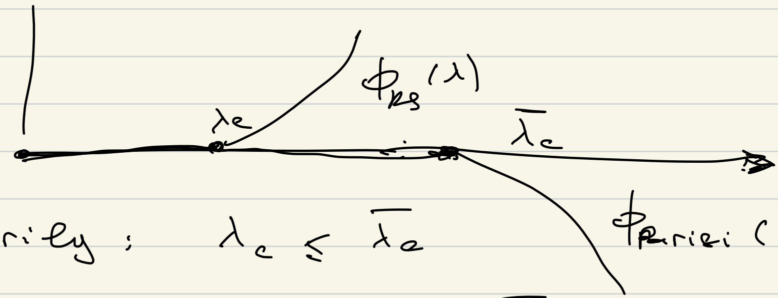
$\forall \varepsilon > 0$.

The test $T = \mathbb{1}\left\{ \frac{1}{n} \log L \geq \frac{1}{2} \phi_{RS}(\lambda) \right\}$

achieves strong detection.

• Remark: $\frac{1}{n} \mathbb{E}_{P_0} \log L \longrightarrow \phi_{\text{Parisi}}(\lambda) \quad \forall \lambda \geq 0$

[Talagrand (Rademacher prior) $\stackrel{\le}{2005}$].



necessarily ; $\lambda_c \leq \bar{\lambda}_c$

Question : Is $\lambda_c = \bar{\lambda}_c$ for all priors?

yes if $P_0 = \text{Rad}(\frac{1}{2})$ $\lambda_c = \bar{\lambda}_c = 1$.
(Gaussian ?)

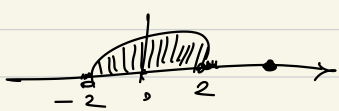
• Difficult to compute L in general.

what about efficient tests?

Spectral tests?

First eigenvalue of Y?

under $\mathbb{P}_0$: $Y = W$.



$$\lambda_1(Y) \longrightarrow 2 \quad \text{a.s.}$$

under $\underset{\lambda>0}{\mathbb{P}_\lambda}$: $Y = \sqrt{\frac{\lambda}{n}}\, x x^T + W$.

Feral - Péché 2007 :

$$\lambda_1(Y) \longrightarrow \begin{cases} 2 & \text{if } \lambda < 1 \\ \underbrace{\lambda + \frac{1}{\lambda}}_{>2} & \text{otherwise.} \end{cases}$$

The first eigenvalue acheives strong detection for $\lambda > 1$.

- what about $\lambda \in (\lambda_c, 1)$?
  unknown.

Proposition:

- In fact we can show that any test based on the eigenvalues of Y fails at strong detection for $\lambda < 1$.

- $$Y = \sqrt{\frac{\lambda}{n}}\, x x^T + W.$$

  Let R be orthognal matrix such wich sends x to the all ones vector

$$Qx = \rho_n \mathbb{1} \qquad \rho_n = \frac{\|x\|_2}{\sqrt{n}} \to 1$$
$$= 1 + o(1).$$

$$\tilde{Y} = Q Y Q^T$$
$$= \sqrt{\frac{\lambda}{n}}\, \rho_n^2 \cdot \mathbb{1}\mathbb{1}^T + Q W Q^T$$
$$\overset{d}{=} \underline{\qquad\qquad\qquad} + W.$$

The eigenvalues of $\tilde{Y}$ = The eigv. vals (Y).

$\lambda_c = 1$ for $\tilde{Y}$.

$\Rightarrow$ Law of $\underset{\tilde{\mathbb{P}}_\lambda}{\tilde{Y}}$ is contiguous to law of $\underset{\mathbb{P}_0}{W}$.

$\Rightarrow$ No test operating on the eigenvalues succeeds for $\lambda < 1$.

We will prove that $\mathbb{P}_\lambda \triangleleft \mathbb{P}_0$ for $\lambda < 1$.
for Rademacher. ∎

Easier to prove contiguity up to $\lambda = \frac{1}{\sigma^2}$.

where σ^2 is subgaussian parameter of the prior.

for Rademacher: $\sigma^2 = 1$. [via the second moment method].