STSCI 6940.

<u>Lecture 2:</u> Date: 02/11.

Prior P_0 over $\mathbb{R}$ $(n=1)$

$P_0 = N(0,1)$.

$$F(\lambda) = \mathbb{E} \log \int e^{\sqrt{\lambda}\, y.x - \frac{\lambda}{2} x^2} dP_0(x).$$

$$= \mathbb{E} \log \mathbb{E}_{x \sim N(0,1)} \left[e^{\sqrt{\lambda} y x - \frac{\lambda}{2} x^2} \right].$$

$$= \mathbb{E} \log \int e^{\sqrt{\lambda} y x - \frac{(1+\lambda)}{2} x^2} \frac{dx}{\sqrt{2\pi}\,\frac{1}{1+\lambda}} \cdot \frac{1}{\sqrt{1+\lambda}}$$

$$= \mathbb{E} \log \mathbb{E}_g \left[e^{\sqrt{\lambda} y g} \right] - \frac{1}{2} \log(1+\lambda) \qquad g \sim N\Big(0, \underbrace{\tfrac{1}{1+\lambda}}_{\sigma^2}\Big).$$

$$= \mathbb{E} \log e^{\frac{\lambda y^2}{2} \sigma^2} - \frac{1}{2} \log(1+\lambda).$$

$y = \sqrt{\lambda} x + z$

$y \sim N(0, 1+\lambda)$

$$= \mathbb{E}\left[\frac{\lambda}{2(1+\lambda)} y^2 \right] - \frac{1}{2} \log(1+\lambda)$$

$$F(\lambda) = \frac{\lambda}{2} - \frac{1}{2} \log(1+\lambda).$$

The mutual information:

$$I(y;x) = \frac{\lambda}{2} \underbrace{\mathbb{E}[x^2]}_{=1} - F(\lambda).$$

$$I(y;x) = \frac{1}{2} \log(1+\lambda).$$

- MMSE ?

I – MMSE formula:

$$\frac{1}{2}\,\frac{1}{1+\lambda} = \frac{d}{d\lambda} I(y;x) = \frac{1}{2}\,\mathrm{MMSE}(\lambda)$$

$$\mathrm{MMSE}(\lambda) = \frac{1}{1+\lambda}.$$

- $\mathbb{E}[x \mid y]$?

- Remark: Let $(x_1, \dots, x_n)$ be a gaussian vector $N(0, \Sigma)$.

$$\mathbb{E}[x_1 \mid (x_j)_{j \geq 2}] = \sum_{j \geq 2} c_j x_j$$

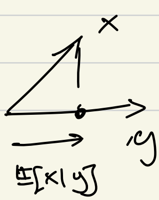
$(c_j)_j$ determined by a least squares problem.

More precisely $\mathbb{E}[x_1 \mid (x_j)_{j \geq 2}]$ is the orthogonal projection over $\mathrm{Span}\{(x_j)_{j \geq 2}\}$

$$\min_{(c_j)} \mathbb{E}\Big[\Big(\sum_{j \geq 2} c_j x_j - x_1\Big)^2\Big].$$

In our case (y, x) are jointly Gaussian

$$\mathbb{E}[x \mid y] = \text{orthogonal projection of } x \text{ on } y$$



$$= \frac{\mathbb{E}[xy]}{\mathbb{E}[y^2]}\, y$$

$$\boxed{\mathbb{E}[x \mid y] = \frac{\sqrt{\lambda}}{1+\lambda}\, y}$$

- $$\mathrm{MMSE}(\lambda) = \mathbb{E}\left[\left(x - \mathbb{E}[x|y]\right)^2\right].$$
$$= \mathbb{E}\left[\left(x - \frac{\sqrt{\lambda}}{1+\lambda} y\right)^2\right]$$
$$= \mathbb{E}\left[\left(x - \frac{\sqrt{\lambda}}{1+\lambda}\left(\sqrt{\lambda}x + z\right)\right)^2\right]$$
$$= \frac{1}{1+\lambda}.$$

- Case of general prior P_0 (over $\mathbb{R}$) ;

Consider : $\hat{x}(y) = \frac{\sqrt{\lambda}}{1+\lambda} y$

$$\mathrm{MSE}(\hat{x}) = \frac{1}{1+\lambda} \geq \mathrm{MMSE}_{P_0}(\lambda).$$

This means that
$$\sup_{\substack{P_0 \text{ with} \\ \text{unit variance}}} \mathrm{MMSE}_{P_0}(\lambda) = \frac{1}{1+\lambda}$$
and is acheived by the Gaussian $P_0 = \mathcal{N}(0,1)$.

- I - MMSE relatiion :

$$\frac{d}{d\lambda} I(y;x) = \frac{1}{2} \mathrm{MMSE}_{P_0}(\lambda) \leq \frac{1}{2} \cdot \frac{1}{1+\lambda}.$$

$$I(y;x)\Big|_{\lambda} - \underbrace{I(y;x)\Big|_{\lambda=0}}_{=0} \leq \frac{1}{2}\int_0^{\lambda} \frac{dt}{1+t}$$
$$= \frac{1}{2}\log(1+\lambda).$$

$$\Rightarrow \quad I_{P_0}(y;x) \leq \frac{1}{2}\log(1+\lambda) = \underset{\text{under } P_0 = \mathcal{N}(0,1)}{I(y;x)}$$

- $$F'(\lambda) = \frac{1}{2} \mathbb{E}\left[\| \mathbb{E}[x|y] \|^2 \right] \geq 0 .$$

Properties of F :

1. F is non-decreasing.

2. $$F(0) = \mathbb{E} \log \int e^{0} \, dP_0(x)$$
$$= 0$$
$$\Rightarrow F(\lambda) \geq 0 \quad \forall \lambda .$$

3. F is convex, equivalently
$$\left(Q(\lambda) \equiv \mathbb{E}\left[\| \mathbb{E}[x|y] \|^2 \right] \right).$$
$\lambda \to Q(\lambda)$ is non-decreasing.

Proof: $$F''(\lambda) = \frac{1}{2} \mathbb{E}\left[\| \operatorname{cov}(x|y) \|_F^2 \right] \geq 0$$

- "Needle in a haystack" problem ;
  Gaussian mean location problem.

$$\sigma_0 \sim \text{Unif}(\{1, \ldots, 2^n\})$$

$$z_i \sim N(0,1) \qquad 1 \leq i \leq 2^n \quad \text{iid} .$$

$$y = \sqrt{\lambda n} \cdot e_{\sigma_0} + z = \begin{array}{c} \\ \sigma_0 \rightarrow \end{array} \underbrace{\begin{pmatrix} \vdots \\ \vdots \\ \vdots \end{pmatrix}}_{2^n}$$

$x = \sqrt{n}\, e_{\sigma_0}$. $\qquad \forall \varepsilon > 0$

$$\mathbb{P}\Big(\max_{1\le i\le n} z_i \in [\sqrt{2\log n} - \varepsilon, \sqrt{2\log n} + \varepsilon] \Big) \xrightarrow[n\to+\infty]{} 1$$

$n \to 2^n \qquad \sqrt{2(\log 2)\, n} \approx$ maximal coordinate of z

Compute:

$$F(\lambda) = \mathbb{E} \log\Big(\frac{1}{2^n} \sum_{\sigma=1}^{2^n} e^{\sqrt{\lambda n}\, y^T \cdot e_\sigma - \frac{\lambda n}{2} \|e_\sigma\|^2} \Big)$$

$$= \mathbb{E} \log\Big(\frac{1}{2^n} \sum_{\sigma=1}^{2^n} e^{\sqrt{\lambda n}\, y_\sigma - \frac{\lambda n}{2}} \Big)$$

upper bound.

Jensen's inequality

$$F(\lambda) \le \log\Big(\frac{1}{2^n} \sum_{\sigma=1}^{n} \mathbb{E}\big[e^{\sqrt{\lambda n}\, y_\sigma - \frac{\lambda n}{2}} \big] \Big).$$

$$= \log\Big(\frac{1}{2^n} \sum_{\sigma=1}^{n} \mathbb{E}\big[e^{\sqrt{\lambda n}\, z_\sigma + \lambda n \mathbb{1}\{\sigma=\sigma_0\} - \frac{\lambda n}{2}} \big] \Big)$$

$y_\sigma = \sqrt{\lambda n}\, \mathbb{1}\{\sigma=\sigma_0\} + z_\sigma$

$$= \log\Big(\frac{1}{2^n} \sum_{\sigma=1}^{2^n} \mathbb{E}\big[e^{\frac{\lambda n}{2} + \lambda n \mathbb{1}\{\sigma=\sigma_0\} - \frac{\lambda n}{2}} \big] \Big)$$

$$= \log\Big(\frac{1}{2^n} \sum_{\sigma=1}^{2^n} \frac{1}{2^n} \sum_{\sigma_0=1}^{2^n} e^{\lambda n \mathbb{1}\{\sigma=\sigma_0\}} \Big)$$

$$= \log\Big(\frac{1}{4^n} \sum_{\sigma,\sigma'=1}^{2^n} e^{\lambda n \mathbb{1}\{\sigma=\sigma'\}} \Big).$$

$$= \log\Big(\frac{1}{4^n} \cdot 2^n \cdot e^{\lambda n} + \frac{1}{4^n}\, 2^n(2^n-1) \cdot 1 \Big)$$

$$= \log\left(e^{(\lambda - \log 2)n} + 1 - \frac{1}{2^n} \right)$$

If:

$\lambda \leq \log 2$: $F(\lambda) \xrightarrow[n \to +\infty]{} 0$.

I - MMSE : $\frac{1}{2} Q_n(\lambda) = \frac{1}{2} \mathbb{E}\left[\| \mathbb{E}[x|y] \|^2 \right]$

$$= F_n'(\lambda).$$

$$\frac{1}{2} \int_0^\lambda \underbrace{Q_n(T)}_{\geq 0} \, dT = F_n(\lambda) \longrightarrow 0 \quad \text{if } \lambda \leq \log 2.$$

$\Rightarrow$ $Q_n(\lambda) \xrightarrow[n \to +\infty]{} 0$ for almost every $\lambda \leq \log 2$.

observation $\lambda \mapsto Q_n(\lambda)$ is non-decreasing.

$\Rightarrow$ $Q_n(\lambda) \xrightarrow[n \to +\infty]{} 0$ for all $\lambda < \log 2$.

It is impossible to estimate σ_0 (asymptotically) if $\lambda < \log 2$.

$$MMSE_{P_0}(\lambda) = \mathbb{E}[\|X\|^2] - Q_n(\lambda).$$

Is our upper tight? In particular, is the threshold $\log 2$ tight?

A refined analysis :

Jensen + conditionning.

$$F(\lambda) = \mathbb{E}\Big[\mathbb{E}[\log Z \mid \sigma_0, z_{\sigma_0}]\Big].$$

$$\leq \mathbb{E}\log\left(\mathbb{E}[Z \mid \sigma_0, z_{\sigma_0}]\right).$$

$$\mathbb{E}[Z \mid \sigma_0, z_{\sigma_0}] = \frac{1}{2^n}\sum_{\sigma=1}^{2^n}\mathbb{E}\Big[e^{\sqrt{\lambda n}\, z_\sigma + \lambda n \mathbb{1}\{\sigma=\sigma_0\} - \frac{\lambda n}{2}} \,\Big|\, \sigma_0, z_{\sigma_0}\Big]$$

$$= \frac{1}{2^n}\sum_{\sigma\neq\sigma_0}^{2^n} 1 + \frac{1}{2^n} e^{\sqrt{\lambda n}\, z_{\sigma_0} + \frac{\lambda n}{2}}.$$

$$= 1 - \frac{1}{2^n} + \frac{1}{2^n} e^{\sqrt{\lambda n}\, z_{\sigma_0} + \frac{\lambda n}{2}}.$$

$$\frac{1}{n}F_n(\lambda) \leq \frac{1}{n}\mathbb{E}_{\sigma_0, z_{\sigma_0}} \log\left(1 + \frac{1}{2^n} e^{\sqrt{\lambda n}\, z_{\sigma_0} + \frac{\lambda n}{2}}\right).$$

$$= \frac{1}{n}\mathbb{E}\log\left(1 + e^{\sqrt{\lambda n}\, g + \left(\frac{\lambda}{2} - \log 2\right)n}\right)$$

$g \sim N(0,1)$.

$$\xrightarrow[n\to+\infty]{} \begin{cases} \frac{\lambda}{2} - \log 2 & \text{if } \frac{\lambda}{2} - \log 2 > 0 \\ 0 & \text{if } \ldots \leq 0 \end{cases}$$

$$\frac{F_n}{n}(\lambda) \longrightarrow 0 \quad \text{if } \underline{\lambda \leq 2\log 2}$$

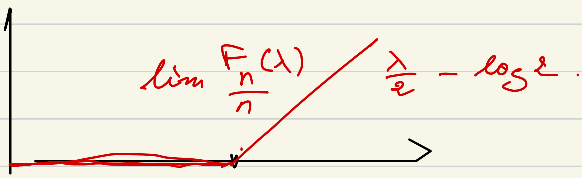
Lower bound:

$$F_n(\lambda) = \mathbb{E} \log\left(\frac{1}{2^n} \underbrace{\sum_{\sigma=1}^{2^n} e^{\sqrt{\lambda n}\, \xi_\sigma + \lambda n \mathbb{1}\{\sigma = \sigma_0\} - \frac{\lambda n}{2}}}_{\geqslant e^{\sqrt{\lambda n}\, \xi_{\sigma_0} + \frac{\lambda n}{2}}} \right)$$

$$\geqslant \mathbb{E} \log\left(\frac{1}{2^n} e^{\sqrt{\lambda n}\, \xi_{\sigma_0} + \frac{\lambda n}{2}} \right).$$

$$= \underbrace{\mathbb{E}\left[\sqrt{\lambda n}\, \xi_{\sigma_0} \right]}_{=0} + \left(\frac{\lambda}{2} - \log 2 \right) n.$$

$$\liminf_{n \to +\infty} \frac{F_n}{n}(\lambda) \geqslant \frac{\lambda}{2} - \log 2$$



$\lambda_c = 2\log 2$

$x = \sqrt{n}\, e_{\sigma_0}.$

$$Q_n(\lambda) = \mathbb{E}\left[\| \mathbb{E}[x \mid y] \|^2 \right]$$

$$= n\, \mathbb{E}\left[\| \mathbb{E}[e_\sigma \mid y] \|_2^2 \right].$$

$$\boxed{\text{Consequence: } \quad \frac{Q_n(\lambda)}{n} \xrightarrow[n \to +\infty]{} \begin{cases} 0 & \text{if } \lambda < 2\log 2. \\ 1 & \text{if } \lambda > 2\log 2 \end{cases}}$$

- Lemma: if $f_n \to f$, f_n is convex and differentiable, the f is convex and $f_n'(t) \to f'(t)$ for all t at which f is differentiable.

I. MMSE: $\frac{1}{n} Q_n(\lambda) = \frac{2}{n} F_n'(\lambda)$. $F(\lambda) = \lim \frac{F_n(\lambda)}{n}$.

$\Rightarrow \quad \frac{1}{n} Q_n(\lambda) \longrightarrow 2F'(\lambda)$

$$F(\lambda) = \begin{cases} \frac{\lambda}{2} - \log 2 & \text{if } \lambda \geq 2\log 2 \\ 0 & \text{otherwise.} \end{cases}$$

$\frac{1}{n} Q_n(\lambda) \longrightarrow 1$ if $\lambda \geq 2\log 2$.

$\Rightarrow \quad \frac{1}{n} MMSE_{P_0}(\lambda) \longrightarrow 0$ if $\lambda > 2\log 2$.

Conclusion: • one can estimate σ_0 almost perfectly if $\lambda > 2\log 2$

• one cannot estimate σ_0 better chance if $\lambda \leq 2\log 2$.

$\Rightarrow$ sharp threshold for estimation.

Information-theoretic analysis

$\Rightarrow$ precise characterization.

Maximum likelihood estimator:

$$\hat{\sigma} = \underset{\sigma \in [2^n]}{\operatorname{argmax}} \; y_\sigma$$

$$y_\sigma = \sqrt{\lambda n} \; \mathbb{1}\{\sigma = \sigma_0\} + z_\sigma .$$

$$\max_{\sigma \in [2^n]} z_\sigma \simeq \sqrt{2 \log 2^n} \quad \text{w.h.p.}$$

$$= \sqrt{2(\log 2) n} .$$

$$\sqrt{\lambda n} \quad \text{v.s.} \quad \sqrt{(2\log 2) n}$$

This means $\lambda > 2\log 2 \Rightarrow y_{\sigma_0} > \max_{\sigma \neq \sigma_0} y_\sigma$ w.h.p.

$\Rightarrow$ MLE will succeed above the threshold $2\log 2$.

- why did conditioning help?

$\mathbb{E}[I(y;\lambda)]$ is dominated by rare events.

$\Rightarrow \mathbb{E}[I(y;\lambda)] \gg$ "typical value of Z".

- $\frac{1}{n} \log Z$ concentrates.

$$\Rightarrow \quad \frac{1}{n} \log Z \simeq \frac{1}{n} \mathbb{E} \log Z = \frac{F_n}{n}(\lambda)$$

$$\longrightarrow F(\lambda) = \left\{ \begin{array}{l} \cdot \\ \cdot \end{array} \right.$$

w.h.p $\quad Z_n \asymp e^{n F(\lambda)} . \quad = e^{n\left(\frac{\lambda}{2} - \log 2\right)}$

$\lambda >$ ___

$$\mathbb{E}[Z_n] = e^{n(\lambda - \log 2)}$$

$$\gg e^{n F(\lambda)}$$

$$\mathcal{E}_T = \{ 3_{\sigma_0} > \sqrt{Tn} \} \text{ for some } T > 0$$

$$\mathbb{P}(\mathcal{E}_T) \sim e^{-\frac{Tn}{2}} .$$

$$\mathbb{E}[Z_n] \geq \mathbb{E}[Z_n / \mathcal{E}_T] . \ \mathbb{P}(\mathcal{E}_T) .$$

$$\gg e^{n F(\lambda)} .$$