

# Stat450 - 2025 Winter

## HW1

Tuorui Peng<sup>1</sup>

### Exercise 1

- Since  $f(x|\theta) \leq 1$  and  $\lambda(\theta) \rightarrow 0$  at the boundary of  $\Theta$ , we have

$$\int_{\Theta} \frac{d}{d\theta} f(x|\theta) \lambda(\theta) d\theta = f(x|\theta) \lambda(\theta) \Big|_{\text{boundary}} = 0$$

- Intergrade by parts, we have

$$\begin{aligned} \int_{\Theta} \psi(\theta) \frac{d}{d\theta} \{f(x|\theta) \lambda(\theta)\} d\theta &= \psi(\theta) f(x|\theta) \lambda(\theta) \Big|_{\text{boundary}} - \int_{\Theta} \frac{d}{d\theta} \psi(\theta) f(x|\theta) \lambda(\theta) d\theta \\ &= - \int_{\Theta} \frac{d}{d\theta} \{\psi(\theta)\} f(x|\theta) \lambda(\theta) d\theta \end{aligned}$$

- In the above equation, replace  $\psi(\theta) \mapsto \hat{\psi}(x) - \psi(\theta)$  and then add an intergral w.r.t.  $x$  on both sides, we have

$$\begin{aligned} \int_x \int_{\Theta} (\hat{\psi}(x) - \psi(\theta)) \frac{d}{d\theta} \{f(x|\theta) \lambda(\theta)\} d\theta dx &= \int_x \int_{\Theta} \frac{d}{d\theta} \{\psi(\theta)\} f(x|\theta) \lambda(\theta) d\theta dx \\ &= \int_{\Theta} \frac{d}{d\theta} \{\psi(\theta)\} \lambda(\theta) d\theta \end{aligned}$$

- By cauchy-schwarz inequality, we have

$$\begin{aligned} &\mathbb{E}_{\lambda} \left[ \mathbb{E}_{\theta} \left[ (\hat{\psi}(X) - \psi(\theta))^2 | \theta \right] \right] (\mathbb{E}_{\lambda} [\mathcal{I}(\theta)] + \mathcal{I}(\lambda)) \\ &= \mathbb{E}_{\lambda} \left[ \mathbb{E}_{\theta} \left[ (\hat{\psi}(X) - \psi(\theta))^2 | \theta \right] \right] \mathbb{E}_{\lambda} [\mathbb{E}_{\theta} [\mathcal{I}(\theta) + \mathcal{I}(\lambda) | \theta]] \\ &= \mathbb{E}_{\lambda} \left[ \mathbb{E}_{\theta} \left[ (\hat{\psi}(X) - \psi(\theta))^2 | \theta \right] \right] \mathbb{E}_{\lambda} \left[ \mathbb{E}_{\theta} \left[ \left( \frac{d}{d\theta} \log f(X|\theta) \right)^2 + \left( \frac{d}{d\theta} \log \lambda(\theta) \right)^2 | \theta \right] \right] \\ &\geq \mathbb{E}_{\lambda} \left[ \mathbb{E}_{\theta} \left[ (\hat{\psi}(X) - \psi(\theta)) \left( \frac{d}{d\theta} \log f(X|\theta) + \frac{d}{d\theta} \log \lambda(\theta) \right) | \theta \right] \right]^2 \\ &= \left( \int_x \int_{\Theta} (\hat{\psi}(x) - \psi(\theta)) \frac{d}{d\theta} \{f(x|\theta) \lambda(\theta)\} d\theta dx \right)^2 \\ &= \left( \int_{\Theta} \frac{d}{d\theta} \{\psi(\theta)\} \lambda(\theta) d\theta \right)^2 \end{aligned}$$

i.e. we have

$$\mathbb{E}_{\lambda} \left[ \mathbb{E}_{\theta} \left[ (\hat{\psi}(X) - \psi(\theta))^2 | \theta \right] \right] \geq \frac{(\mathbb{E}_{\lambda} [\frac{d}{d\theta} \psi(\theta)])^2}{\mathbb{E}_{\lambda} [\mathcal{I}(\theta)] + \mathcal{I}(\lambda)}$$

---

<sup>1</sup>TuoruiPeng2028@u.northwestern.edu

1.(a)

For Normal distribution with known variance 1, we have

$$\mathcal{I}(\theta) = \frac{n}{\sigma^2} = n \quad \mathcal{I}(\lambda) = \mathcal{I}(\lambda_0)a^{-2}$$

then we have using Trees' inequality:

$$\begin{aligned} \mathbb{E}_\lambda \left[ \mathbb{E}_\theta \left[ (\hat{\psi}(X) - \theta^\alpha)^2 | \theta \right] \right] &\geq \frac{\left( \int a^{-1} \lambda_0(a^{-1}\theta) \alpha \theta^{\alpha-1} d\theta \right)^2}{n + \mathcal{I}(\lambda_0)a^{-2}} \\ &= \frac{Aa^{2(\alpha-1)}\alpha^2}{n + \mathcal{I}(\lambda_0)a^{-2}} := \text{R.H.S.} \end{aligned}$$

which is true for any  $a$  so we optimize it over  $a$ :

$$\begin{aligned} \text{R.H.S.} &= \exp \left[ \text{const} + 2(\alpha - 1) \log a - \log(n + \mathcal{I}(\lambda_0)a^{-2}) \right] \\ \frac{\partial 2(\alpha - 1) \log a - \log(n + \mathcal{I}(\lambda_0)a^{-2})}{\partial a} &= \frac{2(\alpha - 1)}{a} + \frac{\mathcal{I}(\lambda_0)2a^{-3}}{n + \mathcal{I}(\lambda_0)a^{-2}} = 0 \\ \Rightarrow a &= \sqrt{\frac{\alpha}{1 - \alpha} \frac{\mathcal{I}(\lambda_0)}{n}} \end{aligned}$$

thus we get the optimal  $a$  and the optimal bound:

$$\mathfrak{M} \geq \mathbb{E}_\lambda \left[ \mathbb{E}_\theta \left[ (\hat{\psi}(X) - \theta^\alpha)^2 | \theta \right] \right] \geq A(1 - \alpha)\alpha^{\alpha+2} \left( \frac{\mathcal{I}(\lambda_0)}{1 - \alpha} \right)^{\alpha-1} n^{-\alpha} \asymp n^{-\alpha}$$

1.(b)

By Le Cam's two point method, we have

$$\begin{aligned} \mathfrak{M} &\geq \sup_{\theta_0, \theta_1 \geq 0} \frac{\ell(\theta_0, \theta_1)}{8} (1 - d_{\text{TV}}(f(x|\theta_0), f(x|\theta_1))) \\ &\geq \sup_{\theta_0=0, \theta_1 \geq 0} \frac{\ell(\theta_0, \theta_1)}{8} (1 - d_{\text{TV}}(f(x|\theta_0), f(x|\theta_1))) \\ &\stackrel{(i)}{\geq} \sup_{\theta_0=0, \theta_1 \geq 0} \frac{\ell(\theta_0, \theta_1)}{8} (1 - \sqrt{\frac{1}{2} \text{KL}(f(x|\theta_0) \| f(x|\theta_1))}) \\ &\stackrel{(ii)}{=} \sup_{\theta_0=0, \theta_1 \geq 0} \frac{\ell(\theta_0, \theta_1 = \theta)}{8} (1 - \frac{\sqrt{n}}{2} \|\theta_1 - \theta_0\|) \\ &= \sup_{\theta \geq 0} \frac{\theta^{2\alpha}}{8} \left( 1 - \frac{\sqrt{n}}{2} \theta \right) \end{aligned}$$

in which (i) uses Pinsker's inequality and (ii) uses the fact that  $\text{KL}(f(x|\theta_0) \| f(x|\theta_1)) = \frac{1}{2} \|\theta_1 - \theta_0\|^2$  for normal distribution (with the same known variance 1). Optimize the above bound w.r.t.  $\theta \geq 0$  and we get optimal  $\theta = 2 \frac{2\alpha}{2\alpha + 1} \frac{1}{\sqrt{n}}$ , then we have the optimal bound for Le Cam's two point method:

$$\mathfrak{M} \gtrsim \frac{1}{2\alpha + 1} \left( \frac{2\alpha}{2\alpha + 1} \right)^{2\alpha} n^{-\alpha} \asymp n^{-\alpha}$$

which gives the same order of convergence as the optimal bound in the previous question.

1.(c)

We prove that  $(\bar{X} + n^{-1/2})^\alpha$  is an estimator that achieves the optimal rate of convergence.

First for  $\bar{X}$  we have

$$\bar{X} \sim \mathcal{N}(\theta, n^{-1}) := \theta + \frac{1}{\sqrt{n}}\varepsilon$$

i.e.  $\varepsilon \sim \mathcal{N}(0, 1)$ .

We have the following using taylor expansion to the 2

$$\begin{aligned} \mathbb{E}_\theta \left[ ((\bar{X} + n^{-1/2})^\alpha - \theta^\alpha)^2 \right] &= \mathbb{E} \left[ \left( (\theta + n^{-1/2} + n^{-1/2}\varepsilon)^\alpha - \theta^\alpha \right)^2 \right] \\ &= \mathbb{E} \left[ \left( \theta^\alpha + \alpha\theta^{\alpha-1}(1+\varepsilon)\Theta(n^{-1/2}) + \alpha(\alpha-1)\theta^{\alpha-2}(1+\varepsilon)^2\Theta(n^{-1}) - \theta^\alpha \right)^2 \right] \\ &= \mathbb{E} \left[ \left( \alpha\theta^{\alpha-1}(1+\varepsilon)\Theta(n^{-1/2}) + \alpha(\alpha-1)\theta^{\alpha-2}(1+\varepsilon)^2\Theta(n^{-1}) \right)^2 \right] \\ &= \mathbb{E} \left[ \alpha^2\theta^{2\alpha-2}(1+\varepsilon)^2\Theta(n^{-1}) + 2\alpha^2(\alpha-1)\theta^{2\alpha-3}(1+\varepsilon)^3\Theta(n^{-3/2}) \right] \\ &\asymp \theta^{2\alpha-2}\Theta(n^{-1}) + \theta^{2\alpha-3}\Theta(n^{-3/2}) \end{aligned}$$

then we optimize the above bound w.r.t.  $\theta$  to get the optimal rate of  $\theta^* \asymp n^{-1/2}$ , which gives that

$$\begin{aligned} \sup_{\theta \geq 0} \mathbb{E}_\theta \left[ ((\bar{X} + n^{-1/2})^\alpha - \theta^\alpha)^2 \right] &= \mathbb{E}_{\theta^*} \left[ ((\bar{X} + n^{-1/2})^\alpha - (\theta^*)^\alpha)^2 \right] \\ &\asymp (\theta^*)^{2\alpha-2}n^{-1} + (\theta^*)^{2\alpha-3}n^{-3/2} \\ &= n^{-\alpha} \end{aligned}$$

Thus we have proved that  $(\bar{X} + n^{-1/2})^\alpha$  is an estimator that achieves the optimal rate of convergence  $n^{-\alpha}$ .

## Exercise 2

2.(a)

We have the following:

$$\begin{aligned} \mathbb{E}_\theta [Y] &= \int y \mathbb{P}_\theta(y) dy \\ &= \int y h(y) \exp \left[ \frac{y \langle x, \theta \rangle - \Phi(\langle x, \theta \rangle)}{s(\sigma)} \right] dy \\ &= \int h(y) s(\sigma) \left[ \frac{\partial}{\partial \langle x, \theta \rangle} \exp \left[ \frac{y \langle x, \theta \rangle - \Phi(\langle x, \theta \rangle)}{s(\sigma)} \right] + \frac{\Phi'(\langle x, \theta \rangle)}{s(\sigma)} \exp \left[ \frac{y \langle x, \theta \rangle - \Phi(\langle x, \theta \rangle)}{s(\sigma)} \right] \right] dy \\ &= s(\sigma) \frac{\partial}{\partial \langle x, \theta \rangle} 1 + \int h(y) \Phi'(\langle x, \theta \rangle) \exp \left[ \frac{y \langle x, \theta \rangle - \Phi(\langle x, \theta \rangle)}{s(\sigma)} \right] dy \\ &= \Phi'(\langle x, \theta \rangle) \end{aligned}$$

then we can compute the KL divergence:

$$\begin{aligned}
\text{KL}(\mathbb{P}_\theta \parallel \mathbb{P}_{\theta'}) &= \mathbb{E}_\theta \left[ \log \frac{\mathbb{P}_\theta}{\mathbb{P}_{\theta'}} \right] \\
&= \int \prod_{i=1}^n h(y_i) \exp \left[ \frac{y_i \langle x_i, \theta \rangle - \Phi(\langle x_i, \theta \rangle)}{s(\sigma)} \right] \sum_{i=1}^n \frac{y_i \langle x_i, \theta - \theta' \rangle - \Phi(\langle x_i, \theta \rangle) + \Phi(\langle x_i, \theta' \rangle)}{s(\sigma)} d^n y \\
&= \sum_{i=1}^n \left( \frac{\langle x_i, \theta - \theta' \rangle}{s(\sigma)} \Phi'(\langle x_i, \theta \rangle) - \frac{\Phi(\langle x_i, \theta \rangle) - \Phi(\langle x_i, \theta' \rangle)}{s(\sigma)} \right) \\
&= \frac{1}{s(\sigma)} \sum_{i=1}^n (\Phi(\langle x_i, \theta' \rangle) - \Phi(\langle x_i, \theta \rangle) - \Phi'(\langle x_i, \theta \rangle) \langle x_i, \theta' - \theta \rangle)
\end{aligned}$$

2.(b)

Note that we have

$$\begin{aligned}
\Phi(\langle x_i, \theta' \rangle) - \Phi(\langle x_i, \theta \rangle) &= \int_{\langle x_i, \theta \rangle}^{\langle x_i, \theta' \rangle} \Phi'(t) dt \\
&= \int_{\langle x_i, \theta \rangle}^{\langle x_i, \theta' \rangle} \left( \int_{\langle x_i, \theta \rangle}^t \Phi''(\tau) d\tau + \Phi'(\langle x_i, \theta \rangle) \right) dt \\
&\leq \int_{\langle x_i, \theta \rangle}^{\langle x_i, \theta' \rangle} (L \|t - \langle x, \theta \rangle\| + \Phi'(\langle x_i, \theta \rangle)) dt \\
&\leq \frac{L}{2} \|x'_i(\theta' - \theta)\|^2 + \Phi'(\langle x_i, \theta \rangle) \langle x_i, \theta' - \theta \rangle
\end{aligned}$$

Thus

$$\begin{aligned}
\text{KL}(\mathbb{P}_\theta \parallel \mathbb{P}_{\theta'}) &\leq \frac{1}{s(\sigma)} \sum_{i=1}^n (\Phi(\langle x_i, \theta' \rangle) - \Phi(\langle x_i, \theta \rangle) - \Phi'(\langle x_i, \theta \rangle) \langle x_i, \theta' - \theta \rangle) \\
&\leq \frac{L}{2s(\sigma)} \sum_{i=1}^n \|x'_i(\theta' - \theta)\|^2 \\
&= \frac{L}{2s(\sigma)} \|X'(\theta' - \theta)\|^2 \\
&\leq n \frac{L\eta_{\max}^2}{2s(\sigma)} \|\theta - \theta'\|_2^2
\end{aligned}$$

in which  $\eta_{\max}^2$  is the maximum singular value of  $X/\sqrt{n}$  (as define in question (c)).

2.(c)

Using Fano's bound we have

$$\inf_{\hat{\theta}} \sup_{\theta \in \mathbb{B}_d(1)} \mathbb{E} \left[ \|\hat{\theta} - \theta\|_2^2 \right] \geq \delta^2 \left( 1 - \frac{I(Z; J) + \log 2}{\log |\mathcal{M}(2\delta, \mathbb{B}_d(1), \|\cdot\|_2^2)|} \right), \quad \forall \delta.$$

We try the following:  $\delta^2 = c \frac{s(\sigma)}{L\eta_{\max}^2} \frac{d}{n}$  in which  $c$  t.b.d. and we hope  $c$  to satisfy:

$$\log |\mathcal{M}(2\delta, \mathbb{B}_d(1), \|\cdot\|_2^2)| \geq 2(I(Z; J) + \log 2)$$

For which we make the following argument:

- For mutual information we have bound

$$I(Z; J) \leq \frac{1}{M^2 \sum_{j,k=1}^M} \text{KL}(\mathbb{P}_{\theta_j} \parallel \mathbb{P}_{\theta_k}) \leq n \frac{L\eta_{\max}^2}{2s(\sigma)} (2\delta)^2 = \frac{2L\eta_{\max}^2}{s(\sigma)} n\delta^2$$

in which  $\theta_j, \theta_k \in \mathcal{M}(2\delta)$ .

- For  $2\delta$ -packing number we have by volume argument:

$$\log \left| \mathcal{M}(2\delta, \mathbb{B}_d(1), \|\cdot\|_2^2) \right| \gtrsim \log \left( \frac{1}{(2\delta)^d} \right) := Cd \log \frac{1}{\delta}$$

- Together we have the condition of such  $c$  being:

$$Cd \log \frac{1}{\delta} \geq 2 \left( \frac{2L\eta_{\max}^2}{s(\sigma)} n\delta^2 + \log 2 \right)$$

substitute  $\delta^2 = c \frac{s(\sigma)}{L\eta_{\max}^2} \frac{d}{n}$  we have

$$\begin{aligned} \frac{Cd}{2} \left( -\log \frac{s(\sigma)}{L\eta_{\max}^2} \frac{d}{n} - \log c \right) &\geq 2(2cd + \log 2) \\ \Leftrightarrow \left( \frac{C}{2} \log c + 4c \right) &\leq -\frac{C}{2} \log \frac{s(\sigma)}{L\eta_{\max}^2} \frac{d}{n} - \frac{2}{d} \log 2 \end{aligned}$$

in which we note that, if  $d \leq n$ , the right hand side is lower bounded by a universal constant, thus we can always find such  $c$ .

choose such  $c$  as above and we have for such  $c$  that:

$$\begin{aligned} \inf_{\hat{\theta}} \sup_{\theta \in \mathbb{B}_d(1)} \mathbb{E} \left[ \left\| \hat{\theta} - \theta \right\|_2^2 \right] &\geq \delta^2 \left( 1 - \frac{I(Z; J) + \log 2}{\log \left| \mathcal{M}(2\delta, \mathbb{B}_d(1), \|\cdot\|_2^2) \right|} \right) \\ &\geq \frac{1}{2} \delta^2 = \frac{c}{2} \frac{s(\sigma)}{L\eta_{\max}^2} \frac{d}{n} \end{aligned}$$

adding the trivial lower bound 1 for  $\hat{\theta} \equiv 0$ , and we have the desired result:

$$\inf_{\hat{\theta}} \sup_{\theta \in \mathbb{B}_d(1)} \mathbb{E} \left[ \left\| \hat{\theta} - \theta \right\|_2^2 \right] \geq \min \left\{ 1, c \frac{s(\sigma)}{L\eta_{\max}^2} \frac{d}{n} \right\}.$$

2.(d)

We have the following:

$$\begin{aligned} \inf_{\hat{\theta}} \sup_{\theta \in \mathbb{B}_d(1)} \mathbb{E} \left[ \left\| X(\hat{\theta} - \theta) \right\|_2^2 \right] &\geq \eta_{\min}^2 \inf_{\hat{\theta}} \sup_{\theta \in \mathbb{B}_d(1)} \mathbb{E} \left[ \left\| \hat{\theta} - \theta \right\|_2^2 \right] \\ &\geq c \frac{s(\sigma) \eta_{\min}^2}{L\eta_{\max}^2} \frac{d}{n} \\ &\gtrsim s(\sigma) \frac{d}{n} \end{aligned}$$

which gives the same rate as Linear regression  $s(\sigma) = \sigma^2$ , which is an example in MJW page 504.

### Exercise 3

We use the Heillinger distance, which has the following properties:

$$\begin{aligned} H^2(P, Q) &:= \int (\sqrt{p} - \sqrt{q})^2 d\mu \\ 1 - \frac{1}{2}H^2(P, Q) &\stackrel{(i)}{=} \int \sqrt{pq} d\mu \\ d_{\text{TV}}(P, Q) &\leq H(P, Q) \sqrt{1 - \frac{H^2(P, Q)}{4}} \end{aligned}$$

then we construct the following:

1.

$$\begin{aligned} \max_{j=0,1} P_j(\psi \neq j) &\geq \frac{1}{2} (P_0(\psi \neq 0) + P_1(\psi \neq 1)) \\ &\geq \frac{1}{2} (1 - d_{\text{TV}}(P_0, P_1)) \\ &\geq \frac{1}{2} \left( 1 - H(P_0, P_1) \sqrt{1 - \frac{H^2(P_0, P_1)}{4}} \right) \end{aligned}$$

2.

$$\begin{aligned} \left( \int \sqrt{pq} d\mu \right)^2 &= \exp 2 \log \left( \int \sqrt{pq} d\mu \right) \\ &= \exp 2 \log \int \frac{\sqrt{p}}{\sqrt{q}} q d\mu \\ &= \exp 2 \log \mathbb{E}_Q [\sqrt{p}/\sqrt{q}] \\ &\geq \exp 2 \mathbb{E}_Q [\log \sqrt{p}/\sqrt{q}] \\ &= \exp [-\mathbb{E}_Q [\log p/q]] \\ &= \exp [-\text{KL}(Q, P)] \end{aligned}$$

and with  $Q \mapsto P_0$ ,  $P \mapsto P_1$  we have

$$\left( \int \sqrt{dP_0 dP_1} d\mu \right)^2 \stackrel{(ii)}{\geq} \exp [-\text{KL}(P_1, P_0)]$$

3. Using the relation  $1 - \frac{1}{2}H^2(P, Q) = \int \sqrt{pq} d\mu$ , it suffices to show the following  $\stackrel{(?)}{\geq}$ :

$$2 - 2H(P_0, P_1) \sqrt{1 - \frac{H^2(P_0, P_1)}{4}} \stackrel{(?)}{\geq} \left( 1 - \frac{1}{2}H^2(P_0, P_1) \right)^2 \stackrel{(i)}{=} \left( \int \sqrt{dP_0 dP_1} d\mu \right)^2 \stackrel{(ii)}{\geq} \exp [-\text{KL}(P_1, P_0)]$$

in which denote  $H(P_0, P_1) = 2 \cos \theta$ , then we have

$$\begin{aligned} 2 - 2H(P_0, P_1) \sqrt{1 - \frac{H^2(P_0, P_1)}{4}} &\stackrel{(?)}{\geq} \left( 1 - \frac{1}{2}H^2(P_0, P_1) \right)^2 \\ \Leftrightarrow 2 - 4 \cos \theta \sin \theta &\geq (1 - 2 \cos \theta)^2 \\ \Leftrightarrow 2 - 2 \sin 2\theta &\geq 1 - \sin^2 2\theta \\ \Leftrightarrow \sin^2 2\theta - 2 \sin 2\theta + 1 &\geq 0 \\ \Leftrightarrow (\sin 2\theta - 1)^2 &\geq 0 \end{aligned}$$

which is always true, thus we prove that

$$\max_{j=0,1} P_j(\psi \neq j) \geq \exp[-\text{KL}(P_1, P_0)]$$