1. A data Processing inequality

(a) Suppose $\theta \in \mathbb{R}^n$. Suppose $\forall \theta_i$,

the conditional density $p_{B|A}(y|x)$ satisfies

$$\frac{\partial}{\partial \theta_i} \int_{\mathcal{B}} p_{B|A}(y:\theta_i|x)\, dy = \int \frac{\partial p_{B|A}(y:\theta_i|x)}{\partial \theta_i}\, dy$$

Then this entails

$$\int_{\mathcal{B}} \nabla \ln_{p_{B|A}}(y:\underline{\theta}|x)\, dy = \int_{\mathcal{B}} \frac{\nabla p_{B|A}(y:\underline{\theta}|x)}{p_{B|A}(y:\underline{\theta}|x)}\, dy$$

$$= \nabla \int_{\mathcal{B}} \frac{p_{B|A}(y:\underline{\theta}|x)}{p_{B|A}(y:\underline{\theta}|x)}\, dy$$

$$= \nabla \cdot 1 = 0.$$

So

$$I_{A,B}(\underline{\theta}) = -\mathbb{E}_{A,B}\left[\nabla^2_{\underline{\theta}} \ln p_{A,B}(X,Y)\right]$$

$$= \mathbb{E}_{A,B}\left[\nabla_{\theta} \ln p_{A,B}(X,Y)\, \nabla_{\theta} \ln p_{A,B}(X,Y)^T\right]$$

$$= \int_{\mathcal{A}}\int_{\mathcal{B}} p_{A,B}(x,y)\, \nabla \ln p_{A,B}(x,y)\, \nabla \ln_{p_{A,B}}(x,y)^T \Big]\, dx\,dy$$

$$= \int_{\mathcal{A}}\int_{\mathcal{B}} p_{A,B}(x,y)\, \nabla \ln p_{B|A}(y|x)\, p_A(x)\, \nabla \ln_{p_{B|A}}(y|x)\, p_A(x)^T\, dx\,dy$$

$$= \int_{\mathcal{A}}\int_{\mathcal{B}} p_{A,B}(x,y)\left(\nabla \ln p_{B|A}(y|x) + \nabla \ln p_A(x)\right)\left(\nabla \ln p_{B|A}(y|x) + \nabla \ln p_A(x)\right)^T dx\,dy$$

$$= \underbrace{\mathbb{E}_{A,B}\left[\nabla \ln p_{B|A}(y|x)\, \nabla \ln p_{B|A}(y|x)^T\right]}_{(3)} + \underbrace{\mathbb{E}_{A,B}\left[\nabla \ln p_A(x)\, \nabla \ln p_A(x)^T\right]}_{(4)}$$

$$+\underbrace{\mathbb{E}_{A,B}\left[\nabla \ln p_{B|A}(y|x)\, \nabla \ln p_A(x)^T\right]}_{(1)} + \underbrace{\mathbb{E}_{A,B}\left[\nabla \ln p_A(x)\, \nabla \ln p_{B|A}(y|x)^T\right]}_{(2)}$$

(2) $$\int_A \int_B \nabla \ln p_A(x)\, \nabla \ln p_{B|A}(y|x)^T\, p_{B,A}(x,y)\, dx dy$$

$$= \int_A \nabla \ln p_A(x) \left(\int_B \nabla \ln p_{B|A}(y|x)^T\, p_{B|A}(y|x)\, p_A(x)\, dy \right) dx$$

$$= \int_A \nabla \ln p_A(x) \left(\nabla \int_B p_{B|A}(y|x) dy^T \right) p_A(x)\, dx \quad = 0$$

(1) $$\int_A \int_B \nabla \ln p_{B|A}(y|x)\, \nabla \ln p_A(x)^T\, p_{A,B}(x,y)\, dx dy$$

$$= \left(\int_A \int_B \nabla \ln p_A(x)\, \nabla \ln p_{B|A}(y|x)^T\, p_{A,B}(x,y)\, dx dy \right)^T = 0$$

(3) $$\mathbb{E}_{A,B}\left[\nabla \ln p_{B|A}(y|x)\, \nabla \ln p_{B|A}(y|x)^T\right]$$

$$= \int_A \left(\int_B \nabla \ln p_{B|A}(y|x)\, \nabla \ln p_{B|A}(y|x)^T\, p_{B|A}(y|x)\, dy \right) p_A(x)\, dx$$

$$= \mathbb{E}_A\left[I_{B|A}(\theta) \right]$$

(4) $$\int_A \left(\nabla \log p_A(x)\, \nabla \log p_A(x)^T \left(\int_B p_{A,B}(x,y)\, dy \right) dx \right.$$

$$= \int_A \nabla \log p_A(x)\, \nabla \log p_A(x)^T\, p_A(x)\, dx = I_A(\theta)$$

Hence $\underline{I_{A,B}(\theta) = I_{B|A}(\theta) + I_A(\theta)}$

(b) Y is sampled from $Q(Y|X)$, $X \sim P_\theta$.

Suppose $\log p_{X|Y}$ is concave.

① $I_{Y|X}(\underline{\theta}) \succeq 0$:

▸ $I_{Y|X}(\underline{\theta}) = -\mathbb{E}\left[\nabla^2 \ln p_{Y|X}(\underline{\theta})\right] \succeq 0$

Since $\nabla^2 \ln p_{Y|X}(\underline{\theta}) \preceq 0$

② $I_{X,Y}(\underline{\theta}) = I_{Y,X}(\underline{\theta})$:

▸ Using $p_{X,Y}(x,y) = p_{Y|X}\, p(y|x)\, p_X(x)$

$= p_{X|Y}\, p(x|y)\, p_Y(y)$,

$= p_{Y,X}(y,x)$

$$I_{X,Y}(\theta) = -\mathbb{E}_{X,Y}\left[\nabla^2 \ln p_{X,Y}\right]$$

$$= -\int_x \left(\int_y \nabla^2 \ln p_{X,Y}\, p_{Y|X}\, dy\right) p_X\, dx$$

$$= -\int_y \left(\int_x \nabla^2 \ln p_{Y,X}\, p_{X|Y}\, dx\right) p_Y\, dy$$

$$= -\mathbb{E}_{Y,X}\left[\nabla^2 \ln p_{Y,X}\right]$$

So $I_{X,Y}(\underline{\theta}) = I_{Y,X}(\underline{\theta})$

② $I_X(\underline{\theta}) \succeq I_Y(\underline{\theta})$

- From part (a), $I_{X,Y}(\underline{\theta}) = I_X(\underline{\theta}) + I_{Y|X}(\underline{\theta})$
$I_{Y,X}(\underline{\theta}) = I_Y(\underline{\theta}) + I_{X|Y}(\underline{\theta})$

$\Rightarrow I_X(\underline{\theta}) - I_Y(\underline{\theta}) = I_{X|Y}(\underline{\theta}) - I_{Y|X}(\underline{\theta})$

Note X is a sufficient statistic for Y

$Y|X \sim \dfrac{P_{X,Y}(X,Y)}{P_X(X)} = Q(Y|X)$, which does not depend on θ

Since $I_{Y|X}(\underline{\theta}) = \mathbb{E}\left[\frac{1}{(p_{Y|X})^2} \nabla_\theta p_{Y|X} \nabla_\theta p_{Y|X}^T\right]$,

$\frac{\partial}{\partial \theta_i} p_{Y|X} = 0 \; \forall i \Rightarrow I_{Y|X}(\underline{\theta}) = 0$

Hence $\underline{I_X(\theta) - I_Y(\theta) = I_{X|Y}(\underline{\theta}) \succeq 0}$.

(c) For a single X_i, $I_{X_i}(\theta) = \frac{1}{\theta(1-\theta)}$

$$I_{X_i}(\theta) = \mathbb{E}\left[\operatorname{cov}\left(\frac{\partial}{\partial\theta} \ln p_{X_i} x_i\right)\right.$$

$$= \operatorname{cov}\left[\frac{\partial}{\partial\theta}\left(x_i \log\theta + (1-x_i)\log(1-\theta)\right)\right]$$

$$= \operatorname{cov}\left[\frac{x_i}{\theta} - \frac{(1-x_i)}{1-\theta}\right]$$

$$= \mathbb{E}\left[\left(\frac{x_i(1-\theta)-\theta(1-x_i)}{\theta(1-\theta)} - \mathbb{E}\left[\frac{x_i(1-\theta)-\theta(1-x_i)}{\theta(1-\theta)}\right]\right)^2\right]$$

$$= \mathbb{E}\left[\left(\frac{x_i-\theta}{\theta(1-\theta)}\right)^2\right]$$

$$= \mathbb{E}\left[\frac{x_i^2 - 2x_i\theta + \theta^2}{\theta(1-\theta)}\right]$$

$$= \frac{1}{\theta(1-\theta)}\left(\theta(1-\theta) - \theta^2 + \theta^2\right) = \frac{1}{\theta(1-\theta)}$$

X_i's are sampled iid so $\underline{I_{\underline{X}}(\theta) = n\, I_{X_i}(\theta) = \frac{n}{\theta(1-\theta)}}$

For a sing Y_i, we have $p_{Y_i}(y_i) = \mathbb{E}_X\left[p_{Y_i|X_i}(y_i|x_i)\right]$

$$p_{Y_i}(y_i) = p_{X_i}(X_i=1)\, p_{Y_i|X_i=1} + p_{X_i}(X_i=0)\, p_{Y_i|X_i=0}$$

(*) $$= \theta\left(y_i\left(\frac{1+\varepsilon}{2}\right) + (1-y_i)\left(\frac{1-\varepsilon}{2}\right)\right) + (1-\theta)\left(y_i\left(\frac{1-\varepsilon}{2}\right) + (1-y_i)\left(\frac{1+\varepsilon}{2}\right)\right)$$

Hence

$$\log p_Y(y_i) = \begin{cases} \log\left(\theta\left(\frac{1+\varepsilon}{2}\right) + (1-\theta)\left(\frac{1-\varepsilon}{2}\right)\right) & \text{if } y_i = 1 \\ \log\left(\theta\left(\frac{1-\varepsilon}{2}\right) + (1-\theta)\left(\frac{1+\varepsilon}{2}\right)\right) & \text{if } y_i = 0 \end{cases}$$

$$= \begin{cases} \log\left(\varepsilon\theta + \frac{1-\varepsilon}{2}\right) & \text{if } y_i = 1 \\ \log\left(\frac{1+\varepsilon}{2} - \varepsilon\theta\right) & \text{if } y_i = 0 \end{cases}$$

$$\Rightarrow \log p_{Y_i}(y_i) = y_i \log\left(\varepsilon\theta + \frac{1-\varepsilon}{2}\right) + (1-y_i)\log\left(\frac{1+\varepsilon}{2} - \varepsilon\theta\right)$$

$$\Rightarrow \frac{\partial}{\partial\theta} \log p_{Y_i}(y_i) = y_i \frac{\varepsilon}{\varepsilon\theta + \frac{1-\varepsilon}{2}} - (1-y_i)\frac{\varepsilon}{\frac{1+\varepsilon}{2} - \varepsilon\theta}$$

$$\Rightarrow \left(\frac{\partial}{\partial\theta} \log p_{Y_i}(y_i)\right)^2 = \frac{y_i^2\varepsilon^2}{\left(\varepsilon\theta + \frac{1-\varepsilon}{2}\right)^2} + \frac{(1-y_i)\varepsilon^2}{\left(\frac{1+\varepsilon}{2} - \varepsilon\theta\right)^2}$$

=0 since

$y_i(1-y_i) = 0 \ \forall y_i \in \{0,1\}$

$$\left\{ -2\frac{\varepsilon^2 y_i(1-y_i)}{\left(\varepsilon\theta + \frac{1-\varepsilon}{2}\right)\left(\frac{1+\varepsilon}{2} - \varepsilon\theta\right)}\right.$$

$$\Rightarrow \left(\frac{\partial}{\partial\theta} \log p_{Y_i}(y_i)\right)^2 = \frac{y_i^2\varepsilon^2}{\left(\varepsilon\theta + \frac{1-\varepsilon}{2}\right)^2} + \frac{(1-y_i)\varepsilon^2}{\left(\frac{1+\varepsilon}{2} - \varepsilon\theta\right)^2}$$

$$\Rightarrow I_{Y_i}(\theta) = \mathbb{P}(Y_i = 1)\frac{\varepsilon^2}{\left(\varepsilon\theta + \frac{1-\varepsilon}{2}\right)^2} + \mathbb{P}(Y_i = 0)\frac{\varepsilon^2}{\left(\frac{1+\varepsilon}{2} - \varepsilon\theta\right)^2}$$

$$\mathbb{P}[Y_i=1 \mid X_i=1]\theta + \mathbb{P}[Y_i=1 \mid X_i=0)(1-\theta) = \theta\left(\frac{1-\varepsilon}{2}+\varepsilon\right) + (1-\theta)\left(\frac{1-\varepsilon}{2}\right)$$

$$= \theta\left(\frac{1+\varepsilon}{2}\right) + (1-\theta)\left(\frac{1-\varepsilon}{2}\right)$$

$$= \theta\varepsilon + \frac{1-\varepsilon}{2}$$

$$\mathbb{P}[Y_i=0] = 1-\left(\theta\varepsilon + \frac{1-\varepsilon}{2}\right) = 1-\left(\theta\varepsilon + \frac{1-\varepsilon}{2}\right) = \frac{1+\varepsilon}{2} - \theta\varepsilon$$

So $$I_{Y_i}(\theta) = \frac{\varepsilon^2}{\theta\varepsilon + \frac{1-\varepsilon}{2}} + \frac{\varepsilon^2}{\frac{1+\varepsilon}{2} - \theta\varepsilon}$$

$$= \varepsilon^2 \frac{\left(\theta\varepsilon + \frac{1-\varepsilon}{2}\right) + \left(\frac{1+\varepsilon}{2} - \theta\varepsilon\right)}{\left(\theta\varepsilon + \frac{1-\varepsilon}{2}\right)\left(\frac{1+\varepsilon}{2} - \theta\varepsilon\right)}$$

$$\left(\theta\varepsilon + \frac{1-\varepsilon}{2}\right)\left(\frac{1+\varepsilon}{2} - \theta\varepsilon\right) = \frac{\theta\varepsilon(1+\varepsilon)}{2} - \frac{(1-\varepsilon)\theta\varepsilon}{2} - \theta^2\varepsilon^2 + \left(\frac{1-\varepsilon}{2}\right)\left(\frac{1+\varepsilon}{2}\right)$$

$$= \theta\varepsilon^2 - \theta^2\varepsilon^2 + \frac{1}{4}(1-\varepsilon^2)$$

$$= \varepsilon^2\left(-\theta^2+\theta-\frac{1}{4}\right) + \frac{1}{4}$$

$$\Rightarrow I_{Y_i}(\theta) = \frac{\varepsilon^2}{\varepsilon^2\left(-\theta^2+\theta-\frac{1}{4}\right)+\frac{1}{4}} = \frac{4\varepsilon^2}{4\varepsilon^2\theta - 4\varepsilon^2\theta^2 + \varepsilon^2 + 1}$$

Hence since Y_i's are iid,

$$\boxed{I_Y(\theta) = n\varepsilon^2\left[\frac{1}{\varepsilon\theta + \frac{1-\varepsilon}{2}} + \frac{1}{\frac{1+\varepsilon}{2} - \varepsilon\theta}\right]}$$, i.e.

$$I_{\underline{Y}}(\underline{\theta}) = \frac{n4\varepsilon^2}{4\varepsilon^2\theta - 4\varepsilon^2\theta^2 + \varepsilon^2 + 1}$$

(d) Consider $\delta_n = \left(\frac{1}{n}\sum_{i=1}^{n} \frac{Y_i - \frac{1-\varepsilon}{2}}{\varepsilon}\right) = \frac{1}{\varepsilon}\left(\bar{Y}_n - \frac{1-\varepsilon}{2}\right)$

① Consistency

First note that $Y \sim Ber\left(\frac{1-\varepsilon}{2} + \varepsilon\theta\right)$, and $E[Y] = \frac{1-\varepsilon}{2} + \varepsilon\theta$.

Consider $\delta_n = \frac{1}{\varepsilon}\left(\bar{Y}_n - \frac{1-\varepsilon}{2}\right)$

By WLLN, $\bar{Y}_n \xrightarrow{P} \frac{1-\varepsilon}{2} + \varepsilon\theta$ (Y_i's iid)

By CMT, $\frac{1}{\varepsilon}\left(\bar{Y}_n - \frac{1-\varepsilon}{2}\right) \xrightarrow{P} \theta$

Hence $\delta_n \xrightarrow{P} \theta$ and δ_n is consistent.

② Efficient.

By CLT, $\sqrt{n}\left(\bar{Y}_n - \left(\frac{1-\varepsilon}{2} + \varepsilon\theta\right)\right) \xrightarrow{d} N(0, \Sigma)$

$\Sigma := V[Y_i] = \left(\frac{1-\varepsilon}{2} + \varepsilon\theta\right) - \left(\frac{1-\varepsilon}{2} + \varepsilon\theta\right)^2$ (Y_i is indicator)

$:= p - p^2$, with $p = \frac{1-\varepsilon}{2} + \varepsilon\theta$

But $\bar{Y}_n = \frac{1-\varepsilon}{2} + \varepsilon\delta_n$, so $\sqrt{n}\left(\frac{1-\varepsilon}{2} + \varepsilon\delta_n - \left(\frac{1-\varepsilon}{2} + \varepsilon\theta\right)\right) \xrightarrow{d} N(0, \Sigma)$

i.e. $\sqrt{n}(\delta_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{\varepsilon^2}\Sigma\right)$

But $I_y(\theta)^{-1} = \left(\varepsilon^2\left(\frac{1}{\beta} + \frac{1}{1-\beta}\right)\right)^{-1}$

$$= \left(\frac{\varepsilon^2}{\beta(1-\beta)}\right)^{-1}$$

So $\sqrt{n}(\delta_n - \theta) \xrightarrow{d} N(0, I_y(\theta)^{-1})$ so δ_n is efficient.

(e) By CLT, $\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{d} N(0, \theta - \theta^2)$

$$\sqrt{n}(\delta_n - \theta) \xrightarrow{d} N\left(0, \frac{\beta - \beta^2}{\varepsilon^2}\right)$$

$$\frac{\theta - \theta}{\frac{\beta - \beta^2}{\varepsilon^2}} = \frac{\varepsilon^2(\theta - \theta^2)}{\beta - \beta^2}$$

$$\beta - \beta^2 = \frac{1-\varepsilon}{2} + \varepsilon\theta - \left(\frac{1-\varepsilon}{2} + \varepsilon\theta\right)^2$$

$$= \frac{1-\varepsilon}{2} + \varepsilon\theta - \frac{(1-\varepsilon)^2}{4} + 2\frac{(\varepsilon\theta)(1-\varepsilon)}{2} + (\varepsilon\theta)^2$$

$$= \frac{1-\varepsilon}{2} + \varepsilon\theta - \frac{1}{4} + \frac{2\varepsilon}{4} - \frac{\varepsilon^2}{4} + \varepsilon\theta - \frac{2\varepsilon^2\theta}{2} + (\varepsilon\theta)^2$$

$$= \frac{1}{4} + \varepsilon\left(\frac{-1}{2} + \frac{1}{2}\right) + \varepsilon^2\left(\frac{-1}{4}\right) + \theta\left(\varepsilon + \varepsilon - \varepsilon^2\right) + \varepsilon^2\theta^2$$

$$= \frac{1}{4} - \frac{\varepsilon^2}{4} + \varepsilon^2(\theta^2 - \theta)$$

So ARE of δ_n wrt $\bar{X}_n$ is

$$\frac{\varepsilon^2(\theta-\theta^2)}{\frac{1-\varepsilon^2}{4}+\varepsilon^2(\theta^2-\theta)}$$

This is the "cost" because it is an estimator of how many more people need to be surveyed by the researcher to have an equally good confidence interval.

f) : $Y=1$

2. Potential Outcomes Causal Estimation

(a) $\mathbb{E}[Y_i(A_i)\mathbb{1}_{A_i=1}] = \mathbb{E}[Y_i(A_i)\mathbb{1}_{A_i=1} \mid A_i=1]\,\mathbb{P}[A_i=1]$

$= \mathbb{E}[Y_i(1)]\,\mathbb{P}[A_i=1]$ by independence

$= \mathbb{E}[Y_i(1)]\cdot\frac{1}{2} = \frac{1}{2}\mathbb{E}[Y(1)]$

Similarly, $\mathbb{E}[Y_i(A_i)\mathbb{1}_{A_i=0}] = \mathbb{E}[Y_i(0)]\cdot\frac{1}{2} = \frac{1}{2}\mathbb{E}[Y(0)]$

Hence $\tau = 2\left(\mathbb{E}[Y_i\mathbb{1}_{A_i=1}] - \mathbb{E}[Y_i\mathbb{1}_{A_i=0}]\right)$

(b) With $\hat{\tau}_n = \frac{2}{n}\sum_{i=1}^{n}\left(Y_i\cdot\mathbb{1}_{A_i=1} - Y_i\mathbb{1}_{A_i=0}\right)$,

$$\mathbb{E}[\hat{\tau}_n] = \frac{1}{n}\sum_{i=1}^{n} 2\,\mathbb{E}[Y_i\cdot\mathbb{1}_{A_i=1} - Y_i\mathbb{1}_{A_i=0}]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y(1)-Y(0)]$$

$$= \mathbb{E}[Y(1)-Y(0)]$$

and $\mathbb{E}[\hat{\tau}_n^2] = \mathbb{E}\left[\frac{4}{n^2}\left(\sum_i Y_i^2\left(\overbrace{\mathbb{1}_{A_i=1} + \mathbb{1}_{A_i=0}}^{1} - 2\,\overbrace{\mathbb{1}_{A_i=0}\mathbb{1}_{A_i=1}}^{0}\right)\right)\right]$

$$+\ \mathbb{E}\left[\frac{4}{n^2}\,2\sum_{i<j}\left(Y_i^2\mathbb{1}_{A_i=1} + Y_j^2\mathbb{1}_{A_j=0} - 2Y_iY_j\mathbb{1}_{A_i=1}\mathbb{1}_{A_j=0}\right)\right]$$

$$= \mathbb{E}\left[\frac{4}{n^2}\sum_{i=1}^{n} Y_i^2(A_i)\right] + \frac{8}{n^2}\sum_{i=1}^{n}\mathbb{E}\left[Y_i\mathbb{1}_{A_i=1} + Y_i\mathbb{1}_{A_i=0}\right]$$

$$- \frac{16}{n^2}\sum_{i<j}\mathbb{E}\left[Y_jY_i\mathbb{1}_{A_i=1}\mathbb{1}_{A_j=0}\right] \Big\}\ 0 \text{ if } A_j=1,\ 0\cdot Y_i \text{ if } A_j=0$$

$$= \frac{-4}{n}\mathbb{E}[Y_i(1)] + \frac{4}{n}\mathbb{E}[Y(1)+Y(0)] = \frac{4}{n}\mathbb{E}[Y(0)] < \infty$$

By the CLT, $\sqrt{n}(\hat{\tau}_n - \tau) \xrightarrow{d} N(0, \mathbb{V}ar(\tau))$

Let $Z_i := 2(Y_i \mathbb{1}_{A_i=1} - Y_i \mathbb{1}_{A_i=0})$, $\sigma_0^2 = \mathbb{V}[Y(0)]$, $\sigma_1^2 = \mathbb{V}[Y(1)]$

law of total variance

$$\frac{1}{4}\mathbb{V}(Z_i) = \mathbb{E}_A \mathbb{V}_{Y|A}[Y_i \mathbb{1}_{A_i=1} - Y_i \mathbb{1}_{A_i=0}] + \mathbb{V}_A \mathbb{E}_{Y|A}[Y_i \mathbb{1}_{A_i=1} - Y_i \mathbb{1}_{A_i=0}]$$

$$= \frac{1}{2}(\mathbb{V}[Y(1)] + \mathbb{V}[Y(0)]) + \frac{1}{2}(E[Y(0)]^2 + E[Y(1)]^2) - \left(\frac{E[Y(1)] - E[Y(0)]}{2}\right)^2$$

$$= \frac{1}{2}(\mathbb{V}[Y(1)] + \mathbb{V}[Y(0)]) + \frac{1}{4}(E[Y(0)]^2 + E[Y(1)]^2 + \frac{E[Y(1)]\,E[Y(0)]}{2}$$

$$= \frac{1}{2}(\sigma_1^2 + \sigma_0^2) + \frac{1}{4}(E[Y(0)] + E[Y(1)])^2$$

So $\mathbb{V}[Z_i] = 2(\sigma_1^2 + \sigma_0^2) + (E[Y(0)] + E[Y(1)])^2$

$$\boxed{\text{Hence } \sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, 2(\sigma_1^2 + \sigma_0^2) + (E[Y(0)] + E[Y(1)])^2)}$$ by CLT

(c) $\hat{\tau}_n^{norm} = \frac{1}{|S_1|} \sum_{i \in S_1} Y_i - \frac{1}{|S_0|} \sum_{i \in S_0} Y_i$

$$= \frac{n}{|S_1|} \cdot \frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{1}_{\{A_i=1\}} - \frac{n}{|S_0|} \frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{1}_{\{A_i=0\}}$$

$$= \Big(\underbrace{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{A_i=1\}}}_{A}\Big)^{-1} \underbrace{\frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{1}_{\{A_i=1\}}}_{B} - \Big(\underbrace{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{A_i=0\}}}_{C}\Big)^{-1} \underbrace{\frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{1}_{\{A_i=0\}}}_{D}$$

A_i's iid hence Y_i's iid so by WLLN.

$$A \xrightarrow{P} \mathbb{E}[\mathbb{1}_{\{A_i=1\}}] = \mathbb{P}[A_i=1] = \frac{1}{2}$$

$$B \xrightarrow{P} \mathbb{E}[Y_i \mathbb{1}_{\{A_i=1\}}] = \frac{\mathbb{E}[Y(1)]}{2} \text{ (by (a))}$$

$$C \xrightarrow{P} \mathbb{E}[\mathbb{1}_{\{A_i=0\}}] = \mathbb{P}[A_i=0] = \frac{1}{2}$$

$$D \xrightarrow{P} \mathbb{E}[Y_i \mathbb{1}_{\{A_i=0\}}] = \frac{\mathbb{E}[Y(0)]}{2} \text{ by (a)}$$

- $\mathbb{V}[\mathbb{1}_{\{A_i=1\}}] = \mathbb{E}[\mathbb{1}_{A_i=1}] - \mathbb{E}^2[\mathbb{1}_{A_i=1}] = \frac{1}{4}$
- $\mathbb{V}[\mathbb{1}_{A_i=0}] = \mathbb{E}[\mathbb{1}_{A_i=0}] - \mathbb{E}^2[\mathbb{1}_{A_i=0}] = \frac{1}{4}$
- $\mathbb{V}[Y_i \mathbb{1}_{A_i=1}] = \frac{\sigma_1^2}{2} + \frac{1}{4}(\mathbb{E}[Y(0)])^2$ by (b)
- $\mathbb{V}[Y_i \mathbb{1}_{A_i=0}] = \frac{\sigma_0^2}{2} + \frac{1}{4}\mathbb{E}[Y(1)]^2$ by (b)

- Similarly, $\mathbb{V}\left[Y_i(A_i)\, \mathbb{1}_{A_i=0} \right] = \dfrac{\mathbb{V}[Y_i(0)]}{2}$

- $\mathrm{Cov}\left(\mathbb{1}_{A_j=1}, Y_i \mathbb{1}_{A_i=1}\right) = \mathbb{E}\left[\left(\mathbb{1}_{A_i=1} - \frac{1}{2}\right)\left(Y_i \mathbb{1}_{A_i=1} - \frac{E[Y(1)]}{2}\right)\right]$ $\qquad j \neq i$

$$= \mathbb{E}\left[\mathbb{1}_{A_i=1} Y \mathbb{1}_{A=1}\right] - \frac{1}{2} E\left[Y \mathbb{1}_{A=1}\right] - \frac{E[Y(1)]}{4} + \frac{E[Y(1)]}{4}$$

$$= \mathbb{E}\left[\mathbb{1}_{A_i=1} Y_i\right] - \frac{\mathbb{E}[Y(1)]}{4}$$

$$= \frac{E[Y(1)]}{4}$$

- $\mathrm{Cov}\left(\mathbb{1}_{A_i=1}, Y_i \mathbb{1}_{A_i=0}\right) = \mathbb{E}\left[\left(\mathbb{1}_{A_i=1} - \frac{1}{2}\right)\left(Y_i \mathbb{1}_{A_i=0} - \frac{E[Y(0)]}{2}\right)\right]$

independence $\quad = \mathbb{E}\left[Y_i \underbrace{\mathbb{1}_{A_i=0} \mathbb{1}_{A_i=1}}_{0}\right] - \mathbb{E}\left[\mathbb{1}_{A_i=1} \frac{E[Y(0)]}{2}\right] - \mathbb{E}\left[\frac{1}{2} Y_i \mathbb{1}_{A_i=0}\right] + \frac{E[Y(0)]}{4}$

$$= -\frac{E[Y(0)]}{2}\frac{1}{2} - E[Y_i \mathbb{1}_{A_i=0}] \cdot \frac{1}{2} + \frac{E[Y(0)]}{4}$$

$$= -\frac{E[Y(0)]}{4}$$

- $\mathrm{Cov}\left(\mathbb{1}_{A_i=1}, \mathbb{1}_{A_i=0}\right) = \mathbb{E}\left[\left(\mathbb{1}_{A_i=1} - \frac{1}{2}\right)\left(\mathbb{1}_{A_i=0} - \frac{1}{2}\right)\right]$

$$= \mathbb{E}\left[\overbrace{\mathbb{1}_{A_i=1} \mathbb{1}_{A_i=0}}^{0}\right] - \frac{1}{2}\left(\mathbb{E}[\mathbb{1}_{A_i=1}] + \mathbb{E}[\mathbb{1}_{A_i=0}]\right) + \frac{1}{4} = -1/4$$

$\cdot\ \mathrm{Cov}(Y_i \mathbb{1}_{A_i=1}, \mathbb{1}_{A_i=0}) = \mathbb{E}\left[\left(Y_i \mathbb{1}_{A_i=1} - \frac{E[Y(1)]}{2} \right) \left(\mathbb{1}_{A_i=0} - \frac{1}{2} \right) \right] = \frac{-E[Y(1)]}{4}$

By same method as in $\mathrm{Cov}(\mathbb{1}_{A_i=1}, Y_i \mathbb{1}_{A_i=0})$

$$\cdot\ \mathrm{Cov}(Y_i \mathbb{1}_{A_i=1}, Y_i \mathbb{1}_{A_i=0}) = \mathbb{E}\left[\left(Y_i \mathbb{1}_{A_i=1} - \frac{E[Y(1)]}{2} \right) \left(Y_i \mathbb{1}_{A_i=0} - \frac{E[Y(0)]}{2} \right) \right]$$

$$= \mathbb{E}[\underbrace{Y_i \mathbb{1}_{A_i=1} Y_i \mathbb{1}_{A_i=0}}_{0}] - \mathbb{E}[Y_i \mathbb{1}_{A_i=1}] \frac{E[Y(0)]}{2} - \frac{E[Y(1)]}{2} \mathbb{E}[Y_i \mathbb{1}_{A_i=0}] + \frac{E[Y(1)]\,E[Y(0)]}{4}$$

$$= -\frac{E[Y(1)]\,E[Y(0)]}{4}$$

$$\cdot\ \mathrm{Cov}(Y_i \mathbb{1}_{A_i=0}, \mathbb{1}_{A_i=0}) = \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{1}_{A_i=0} - \frac{E[Y(0)]}{2} \right) \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{A_i=0} - \frac{1}{2} \right) \right] = \frac{E[Y(0)]}{4}$$

By same method as $\mathrm{Cov}(Y_i \mathbb{1}_{A_i=1}, \mathbb{1}_{A_i=1})$

The covariance matrix of $\begin{bmatrix} \mathbb{1}_{A_i=1} \\ Y_i \mathbb{1}_{A_i=1} \\ \mathbb{1}_{A_i=0} \\ Y_i \mathbb{1}_{A_i=0} \end{bmatrix}$ is

$$\Sigma = \begin{bmatrix} 1/4 & E[Y(1)]/4 & -1/4 & -E[Y(0)]/4 \\ E[Y(1)]/4 & \sigma_1^2/2 + \frac{E[Y(0)]^2}{4} & -E[Y(1)]/4 & x \\ -1/4 & -E[Y(1)]/4 & 1/4 & E[Y(0)]/4 \\ -E[Y(0)]/4 & x & E[Y(0)]/4 & \frac{\sigma_0^2}{2} + \frac{E[Y(1)]^2}{4} \end{bmatrix} \quad x = -\frac{E[Y(1)]\,E[Y(0)]}{4}$$

We now apply the delta method. Consider $f : \mathbb{R}\setminus\{0\} \times \mathbb{R} \times \mathbb{R}\setminus\{0\} \times \mathbb{R} \to \mathbb{R}$,

$$f(a,b,c,d) = a^{-1}b - c^{-1}d.$$

f is differentiable at $\underline{\theta} = \begin{bmatrix} 1/2 \\ E[Y(1)]/2 \\ 1/2 \\ E[Y(0)]/2 \end{bmatrix}$ and by the CLT

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} \mathbb{1}_{A_i=1} \\ Y_i \mathbb{1}_{A_i=1} \\ \mathbb{1}_{A_i=0} \\ Y_i \mathbb{1}_{A_i=0} \end{bmatrix} \right) \xrightarrow{d} N\left(\begin{bmatrix} 1/2 \\ E[Y(1)]/2 \\ 1/2 \\ E[Y(0)]/2 \end{bmatrix}, \Sigma \right)$$

$$\nabla f = \begin{bmatrix} -a^{-2}b \\ a^{-1} \\ c^{-2}d \\ -c^{-1} \end{bmatrix}, \text{ and hence } \nabla f_{\underline{\theta}} = \begin{bmatrix} -2E[Y(1)] \\ 2 \\ 2E[Y(0)] \\ -2 \end{bmatrix}$$

By Theorem 5.1,

with $T_n = \left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_i=1}, \ \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_i=1} Y_i, \ \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_i=0}, \ \frac{1}{n}\sum_{i=1}^{n} Y_i \mathbb{1}_{A_i=0} \right]^T$, we have

$$\sqrt{n}\left(f(T_n) - f(\underline{\theta}) \right) \xrightarrow{d} N\left(0, \nabla f_{\theta}^T \Sigma \nabla f_{\theta}\right)$$

i.e. $$\sqrt{n}\left(\frac{\sum_{i \in S_1} Y_i}{|S_1|} - \frac{\sum_{i \in S_0} Y_i}{|S_0|} - \left(E[Y(1)] - E[Y(0)] \right) \right)$$

$$\xrightarrow{d} N\left(\underline{0}, \nabla f_{\theta}^T \Sigma \nabla f_{\theta} \right)$$

i.e. $$\sqrt{n}\left(\hat{\tau}_n^{\,norm} - \left(E[Y(1)] - E[Y(0)] \right) \right) \xrightarrow{d} N\left(0, \nabla f_{\theta}^T \Sigma \nabla f_{\theta}\right).$$

$$\Sigma \nabla_\theta f = \begin{bmatrix} -1/4 \cdot 2E[Y(1)] + 2E[Y(1)]/4 - 1/4\, 2E[Y(0)] - E[Y(0)]/4 (-2) \\ -E[Y(1)]/4 \cdot 2E[Y(1)] + \sigma_1^2 + \frac{E[Y(0)]^2}{2} - E[Y(1)]/4 \cdot 2E[Y(0)] - 2x \\ 1/4\, 2E[Y(1)] - E[Y(1)]/4 \cdot 2 + 1/4\, 2E[Y(0)] - E[Y(0)]/4 \cdot 2 \\ E[Y(0)]/4 \cdot 2E[Y(1)] + 2x + 2E[Y(0)]\frac{E[Y(0)]}{4} - \sigma_0^2 - \frac{E[Y(1)]^2}{2} \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ -\frac{E[Y(1)]^2}{2} + \sigma_1^2 + \frac{E[Y(0)]^2}{2} \\ 0 \\ \frac{E[Y(0)]^2}{2} - \sigma_0^2 - \frac{E[Y(1)]^2}{2} \end{bmatrix}$$

$$\nabla_\theta f^\top \Sigma \nabla_\theta f = 2\left(-\frac{E[Y(1)]^2}{2} + \sigma_1^2 + \frac{E[Y(0)]^2}{2}\right) - 2\left(\frac{E[Y(0)]^2}{2} - \sigma_0^2 - \frac{E[Y(1)]^2}{2}\right))$$

$$= 2(\sigma_1^2 + \sigma_0^2)$$

Hence $\boxed{\sqrt{n}\,(\hat{\tau}_n^{norm} - \tau) \xrightarrow{d} N(0,\ 2(\sigma_1^2 + \sigma_0^2))}$

(d) Suppose $E[Y_i] \neq 0$.

Recalling $\sqrt{n}\,(\hat{\tau}_n - \tau) \xrightarrow{d} N(\underline{0},\ 2(\sigma_1^2 + \sigma_0^2) + E[Y(1)]^2 + E[Y(0)]^2]$

$\sqrt{n}\,(\hat{\tau}_n^{norm} - \tau) \xrightarrow{d} N(\underline{0},\ 2(\sigma_1^2 + \sigma_0^2))$

By proposition C.4. the ARE of $\hat{\tau}_n$ with respect to $\hat{\tau}_n^{norm}$ is given by

$$\frac{2\left(\sigma_1^2+\sigma_0^2\right)+E[Y(1)]^2+E[Y(0)]^2}{2\left(\sigma_1^2+\sigma_0^2\right)} = 1 + \frac{\left(E^2[Y(1)]+E^2[Y(0)]\right)}{\left(\sigma_1^2+\sigma_0^2\right)} > 1.$$

Since $E[Y(1)]^2 + E[Y(0)]^2 > 0$ since $E[Y_i] \neq 0 \ \forall i$.

3. The uniform MLE

(a) Let $X_{(n)} := \max_{i \in [n]} X_i$.

Recalling from problem set 2 that $F_{X_{(n)}}(x) = \left(\frac{x}{\theta^*}\right)^n \mathbb{1}_{0 \le x \le \theta^*}$

$$\mathbb{E}_{\theta^*}\left[\left|\hat{\theta}_n^{MLE} - \theta^*\right|\right] = \frac{1}{n+1}\theta^*$$

$$\mathbb{P}\left[n(\theta - \hat{\theta}_n) \le \alpha\right] = \mathbb{P}\left[X_{(n)} \ge \theta - \frac{\alpha}{n}\right]$$

$$= 1 - \mathbb{P}\left[X_{(n)} < \theta - \frac{\alpha}{n}\right]$$

$$= 1 - \mathbb{P}\left[\bigcap_{i=1}^{n} X_{(i)} \le \theta - \frac{\alpha}{n}\right]$$

$$= 1 - \prod_{i=1}^{n} \mathbb{P}\left[X_{(i)} \le \theta - \frac{\alpha}{n}\right]$$

$$= 1 - \left(\frac{\theta - \frac{\alpha}{n}}{\theta}\right)^n \mathbb{1}_{\{0 \le \theta - \frac{\alpha}{n} \le \theta\}} - \mathbb{1}_{\{\theta - \frac{\alpha}{n} > \theta\}}$$

$$= 1 - \left(1 - \frac{\alpha}{\theta n}\right)^n \mathbb{1}_{\{0 \le \theta - \frac{\alpha}{n} \le \theta\}} - \mathbb{1}_{\{\theta - \frac{\alpha}{n} > \theta\}}$$

$$\to 1 - e^{-\alpha/\theta} \cdot \mathbb{1}_{\{\alpha \ge 0\}}$$

$\lim_{n \to \infty} \mathbb{P}\left[n(\hat{\theta}_n - \theta) \le \alpha\right] = 1 - e^{-\alpha/\theta}$, hence $n(\hat{\theta}_n^{MLE} - \theta) \xrightarrow{d} Exp\left(\frac{1}{\theta}\right)$

$\hat{\theta}_n^{MLE}$ is not unbiased: from pb set 2,

$$\mathbb{E}\left[\left| \hat{\theta}_n^{MLE} - \theta^* \right| \right] = \frac{n}{n+1} \theta^* > 0$$

$\hat{\theta}_n^{MLE}$ is not asymptotically normal because the limiting distribution of $n\left(\hat{\theta}_n^{MLE} - \theta^*\right)$ is exponential.

(b) $\mathbb{E}\left[n^2 (\theta - \hat{\theta}_n)^2\right] \to 2\theta^2$ as $n \to \infty$

The pdf of $\hat{\theta}_n$ is given by $f(x) = n \dfrac{x^{n-1}}{\theta^n} \mathbb{1}_{0 \le x \le \theta}$

$$\mathbb{E}_{\theta^*}\left[(\theta - \hat{\theta}_n)^2 \right] = \int_0^\theta (\theta - \hat{\theta}_n)^2 \cdot n \frac{\hat{\theta}_n^{n-1}}{\theta^n} \, d\hat{\theta}_n$$

$$= \int_0^\theta \left(\theta^2 - 2\theta \hat{\theta}_n + \hat{\theta}_n^2\right) n \frac{\hat{\theta}_n^{n-1}}{\theta^n} \, d\hat{\theta}_n$$

$$= n\theta^{2-n} \int_0^\theta \hat{\theta}_n^{n-1} \, d\hat{\theta}_n - 2n\theta^{1-n} \int_0^\theta \hat{\theta}_n^{n} \, d\hat{\theta}_n + n\theta^{-n} \int_0^\theta \hat{\theta}_n^{n+1} \, d\hat{\theta}_n$$

$$= \theta^{2-n} n \left[\frac{\hat{\theta}_n^{n}}{n} \right]_0^\theta - 2n\theta^{1-n} \left[\frac{\hat{\theta}_n^{n+1}}{n+1} \right]_0^\theta + n\theta^{-n} \left[\frac{\hat{\theta}_n^{n+2}}{n+2} \right]_0^\theta$$

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

$$= \theta^2 \cdot \frac{2n^2(n+1)}{(n+1)(n+2)} = \theta^2 \frac{2}{(n+1)(n+2)}$$

Hence $\lim_{n\to\infty} \mathbb{E}\left[n^2(\hat{\theta}-\theta) \right] = \lim_{n\to\infty} \frac{2n^2\theta^2}{(n+1)(n+2)}$

L'Hopital's $= \lim_{n\to\infty} \frac{4\theta^2 n}{2n+3} = \lim_{n\to\infty} \frac{4\theta^2}{2} = 2\theta^2.$

(ii) $E_\theta\left[n(\theta-\delta_n) \right] = \int_0^\theta n\left(\theta - \frac{n+1}{n}\hat{\theta}_n\right) n \frac{\hat{\theta}_n^{n-1}}{\theta^n} \, d\hat{\theta}_n$

$= n^2 \theta^{1-n} \int_0^\theta \hat{\theta}_n^{n-1} \, d\hat{\theta}_n - n(n+1)\theta^{-n} \int_0^\theta \hat{\theta}_n^{n} \, d\hat{\theta}_n$

$= n^2 \theta^{1-n} \theta^n \cdot \frac{1}{n} - \frac{n(n+1)}{\theta^n} \theta^{n+1} \frac{1}{n+1} = n\theta - n\theta = 0$

$E_\theta\left[(\theta-\delta_n)^2 \right] = \int_0^\theta \left(\theta^2 - 2\frac{n+1}{n}\theta\hat{\theta}_n + \left(\frac{n+1}{n}\hat{\theta}_n\right)^2 \right) n \frac{\hat{\theta}_n^{n-1}}{\theta^n} \, d\hat{\theta}_n$

$= \theta^{2-n} n \int_0^\theta \hat{\theta}_n^{n-1} - 2(n+1)\theta^{1-n} \int_0^\theta \hat{\theta}_n^{n} \, d\hat{\theta}_n + \frac{(n+1)^2}{n} \theta^{-n} \int_0^\theta \hat{\theta}_n^{n+1} \, d\hat{\theta}_n$

$= n\theta^{2-n}\theta^n \frac{1}{n} - 2(n+1)\theta^{1-n} \frac{\theta^{n+1}}{n+1} + \frac{(n+1)^2}{n} \frac{\theta^{-n}\theta^{n+2}}{n+2}$

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

$\lim_{n\to\infty} \theta^2 \left(\frac{(n+1)^2}{n(n+2)} - 1 \right) = \lim_{n\to\infty} \theta^2 \frac{2n+2}{2n+2} = \theta^2.$

c) We have $P_{\hat{S}_n} = \text{Unif}[0, \hat{S}_n]$

Hence $F_{\hat{S}_n}(x) = \mathbb{P}[X \le x] = \frac{x}{\hat{S}_n} \mathbb{1}_{\{0 \le x \le \hat{S}_n\}} + \mathbb{1}_{\{x > \hat{S}_n\}}$

and $p_{\hat{S}_n}(x) = \frac{d}{dx} F_{\hat{S}_n}(x) = \frac{1}{\hat{S}_n} \mathbb{1}_{\{0 \le x \le \hat{S}_n\}}$

Recalling also $\hat{S}_n = \hat{\theta}_n \cdot c$ so $\mathbb{P}[\hat{S}_n \le y] = \mathbb{P}[\hat{\theta}_n \le \frac{y}{c}]$,

Hence $p_{\hat{S}_n}(y) = n \frac{y^{n-1}}{(c\theta)^n} \mathbb{1}_{\{0 \le y \le \theta\}}$

$D_{KL}(P_{\hat{S}_n} \| P_\theta)$

$$= \int_{-\infty}^{\infty} p_{\hat{S}_n} \log\left(\frac{p_{\hat{S}_n}}{p_\theta}\right)$$

$$= \int_{-\infty}^{\infty} \left(\frac{1}{\hat{S}_n}\right) \mathbb{1}_{\{0 \le x \le \hat{S}_n\}} \left[\log\left(\frac{1}{\hat{S}_n} \mathbb{1}_{\{0 \le x \le \hat{S}_n\}}\right) - \log\left(1/\theta \, \mathbb{1}_{\{0 \le x \le \theta\}}\right)\right] dx$$

$$= \int_0^{\hat{S}_n} \frac{1}{\hat{S}_n} \left[\log\left(\frac{1}{\hat{S}_n}\right) - \log\left(\theta \mathbb{1}_{\{0 \le x \le \theta\}}\right)\right] dx$$

Hence taking expectation over $\hat{\theta}_n$,

$E_\theta D_{KL}(P_{\hat{S}_n} \| P_\theta)$

$$= \int_{-\infty}^{\infty} \int_0^{\hat{S}_n} \frac{1}{\hat{S}_n} \log\left(\frac{1}{\hat{S}_n}\right) - \log\left(\frac{1}{\theta} \mathbb{1}_{\{0 \le x \le \theta\}}\right)\Big] \, dx \, dP_{\hat{S}_n}$$

$$= \int_{-\infty}^{\infty} \int_0^{y} \frac{1}{y} \log\left(\frac{1}{y}\right) - \log\left(\frac{1}{\theta} \mathbb{1}_{\{0 \le x \le \theta\}}\right) dx \, \frac{n y^{n-1}}{(c\theta)^n} \mathbb{1}_{\{0 \le y \le c\theta\}} dy$$

$$= \int_0^{c\theta} \int_0^{y} \frac{1}{y} \log\left(\frac{c}{y}\right) - \log\left(\frac{1}{\theta} \mathbb{1}_{\{0 \le x \le \theta\}}\right) \frac{n y^{n-1}}{(c\theta)^n} \, dx \, dy$$

If $c > 1$, then for $x \in [\theta, c\theta]$ $-\log\left(\frac{1}{\theta} \mathbb{1}_{\{0 \le x \le \theta\}}\right)$ will take value $+\infty$, so the integral diverges and the K-L divergence will be infinite.

We proceed with $c \le 1$:

$$\mathbb{E}_\theta\left[D_{KL}\left(P_{\hat{S}_n} \| P_\theta\right)\right] = \int_0^{\theta c} \int_0^{y} \frac{1}{y}\left(\log\left(\frac{1}{y}\right) - \log\left(\frac{1}{\theta}\right)\right) \frac{n y^{n-1}}{(c\theta)^n} \, dx \, dy$$

$$= \int_0^{\theta c} y \cdot \frac{1}{y} \left(\log\left(\frac{1}{y}\right) - \log\left(\frac{1}{\theta}\right)\right) \frac{n y^{n-1}}{(c\theta)^n} \, dy$$

$$= \frac{n}{(c\theta)^n} \int_0^{c\theta} (\log\theta - \log y) \, y^{n-1} \, dy$$

$$= \frac{n}{(c\theta)^n}\left(\left[\log\theta \, (c\theta)^n \frac{1}{n}\right] - \int_0^{c\theta} (\log y) \, y^{n-1} \, dy\right)$$

$$= \log\theta - \frac{n}{(c\theta)^n}\left[\left[\log y \cdot \frac{y^n}{n}\right]_0^{c\theta} - \int_0^{c\theta} \frac{y^{n-1}}{n} \, dy\right]$$

$$= \log\theta - \frac{n}{(c\theta)^n}\left[\log c\theta \cdot \frac{(c\theta)^n}{n} - \frac{(c\theta)^n}{n^2}\right]$$

$$= \log\theta - \log(c\theta) + \frac{1}{n}$$

$$= \log\left(\frac{1}{c}\right) + \frac{1}{n}$$

$-\log c$ is non-increasing in c, so for $c \leq 1$, the minimum value is achieved at $c = 1$.

$$\boxed{\text{Hence for } c=1, \quad \mathbb{E}_\theta\left[D_{KL}\left(P_{\hat{S}_n} \,\|\, P_\theta\right)\right] = \frac{1}{n}}$$

Which is the minimum of expected $D_{KL}(P_{\hat{S}_n} \| P_\theta)$ for $c \leq 1$.

Relative Entropy favors $\hat{S}_n = c\hat{\theta}_n = \hat{\theta}_n$ rather than $\hat{S}_n = \frac{n+1}{n}\hat{\theta}_n$ because if $\hat{S}_n$ "overshoots", relative entropy can be infinity with non-zero probability.

Heuristically, $\forall c > 1$, there is a non-zero probability that $\frac{n+1}{n}\hat{\theta}_n$ overshoots theta and relative entropy become $+\infty$.

$\hat{\theta}_n \leq \theta$ a.s. for $c < 1$, we will shrink the estimator, but that is necessary as we just get further away from θ and decrease entropy which is already $+\infty$, which explains why $c = 1$ is the best for relative entropy loss.

4. The cost of super-efficiency

(a) We see that $\hat{T}_n$ is a better estimator than $\bar{X}_n$ when θ is very close to 0, but behaves much worse as $|\theta_n^h|$ grows, while $|\theta_n^h| < n^{-1/4}$.
As $|\theta_n^h|$ increases, $|\bar{X}_n|$, the tails of $\hat{L}(\hat{T}_n)$ go to $\hat{L}(\hat{X}_n)$

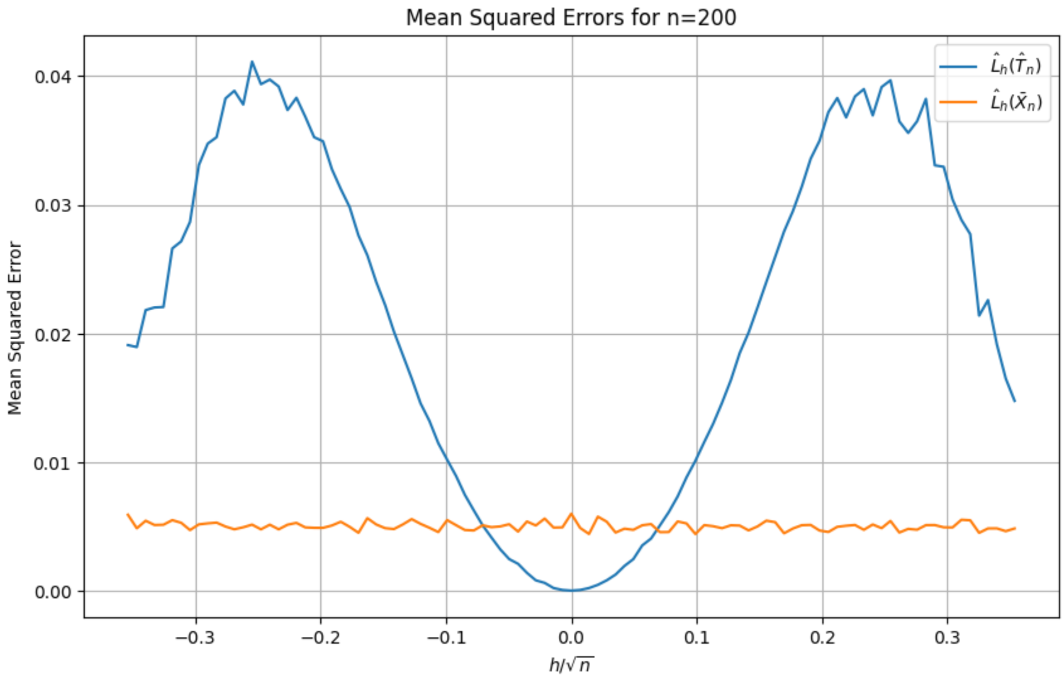
The shape of $\hat{L}_n(\hat{T}_n)$ is due to the fact that if $|\theta_n^h| \approx n^{-1/4}$, then $\hat{T}_n$ is volatile, alternating between 0 and $\bar{X}_n$, and hence $\frac{1}{N}\sum_{n=1}^{N}(\hat{T}_n - \theta_n^h)^2$ will be large due to $(\hat{T}_n - \theta_n^h)^2 \Big|_{\hat{T}_n = 0} = n^{-1/4}$
As θ_1^h increases, $\hat{T}_n$ will be equal to $\bar{X}_n$ more often and hence $\hat{L}_h(\hat{T}_n)$ will go to $\hat{L}(\hat{X}_n)$.

(b) For larger n, we see convergence of the tails of $\hat{L}_n(\hat{T}_n)$ to $\hat{L}_n(\bar{X}_n)$ faster since $200^{-1/4} < 100^{-1/4} < 50^{-1/4}$, and $\hat{T}_n \to \bar{X}_n$ as $\theta_n^n >> n^{-1/4}$.
However, for larger n, the error is larger around $\theta_n^h = n^{-1/4}$ since $(\hat{T}_n - \theta_n^h)^2 \Big|_{\hat{T}_n = 0} \approx n^{-1/4}$.

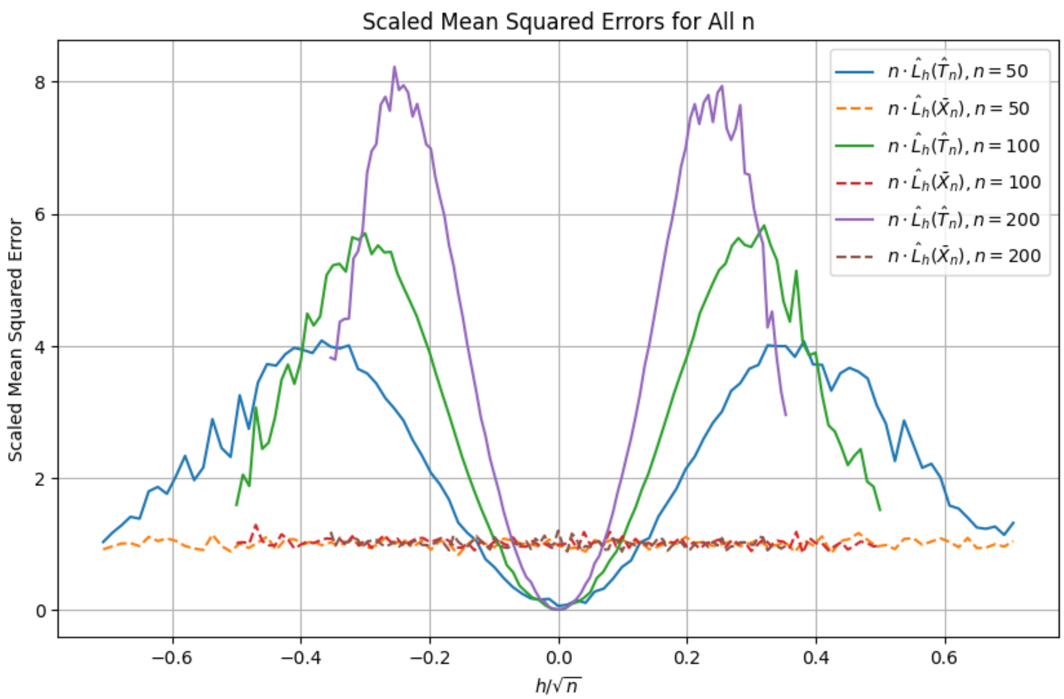
If we care about a large range of possible values of θ_n^h, then it is better to use $\bar{X}_n$ which is more robust. However, if we a priori know some qualitative bound on $|\theta_n^h|$ and know it should be much smaller than $n^{-1/4}$, then we should use $\hat{T}_n$.

```python
import numpy as np
import matplotlib.pyplot as plt

n_values = [50, 100, 200]
h_values = np.arange(-5.0, 5.1, 0.1)
N = 500
results = {}

for n in n_values:
    mse_Tn = []
    mse_Xn = []
    threshold = n ** (-1 / 4)

    for h in h_values:
        theta_h = h / np.sqrt(n)
        errors_Tn = []
        errors_Xn = []

        for _ in range(N):
            X = np.random.normal(theta_h, 1, n)
            bar_Xn = np.mean(X)

            T_n = bar_Xn if abs(bar_Xn) >= threshold else 0

            errors_Tn.append((T_n - theta_h) ** 2)
            errors_Xn.append((bar_Xn - theta_h) ** 2)

        mse_Tn.append(np.mean(errors_Tn))
        mse_Xn.append(np.mean(errors_Xn))

    results[n] = {
        "h": h_values / np.sqrt(n),
        "mse_Tn": mse_Tn,
        "mse_Xn": mse_Xn,
    }

n = 200

plt.figure(figsize=(10, 6))
plt.plot(results[n]["h"], results[n]["mse_Tn"], label=r"$\hat{L}_h(\hat{T}_n)$")
plt.plot(results[n]["h"], results[n]["mse_Xn"], label=r"$\hat{L}_h(\bar{X}_n)$")
plt.xlabel(r"$h / \sqrt{n}$")
plt.ylabel("Mean Squared Error")
plt.title(f"Mean Squared Errors for n={n}")
plt.legend()
plt.grid()
plt.show()
```

```python
plt.figure(figsize=(10, 6))
for n in n_values:
    plt.plot(
        results[n]["h"],
        np.array(results[n]["mse_Tn"]) * n,
        label=r"$n \cdot \hat{{L}}_h(\hat{{T}}_n)"+f", n={n}$",
    )
    plt.plot(
        results[n]["h"],
```




```
        np.array(results[n]["mse_Xn"]) * n,
        label=r"$n \cdot \hat{{L}}_h(\bar{{X}}_n)"+f", n={n}$",
        linestyle="dashed",
    )

plt.xlabel(r"$h / \sqrt{n}$")
plt.ylabel("Scaled Mean Squared Error")
plt.title("Scaled Mean Squared Errors for All n")
plt.legend()
plt.grid()
plt.show()
```



Scaled Mean Squared Errors for All n
Scaled Mean Squared Error
$h/\sqrt{n}$
$n \cdot \hat{L}_h(\hat{T}_n), n = 50$
$n \cdot \hat{L}_h(\bar{X}_n), n = 50$
$n \cdot \hat{L}_h(\hat{T}_n), n = 100$
$n \cdot \hat{L}_h(\bar{X}_n), n = 100$
$n \cdot \hat{L}_h(\hat{T}_n), n = 200$
$n \cdot \hat{L}_h(\bar{X}_n), n = 200$
0
2
4
6
8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6