tuple $(X_1, \ldots, X_n) = X^n = X^{1:n}$ 400102204 [illegible] تمرین سری 3 ITSL

I : سؤال

Conditioning on $X^{1:n-1}$

$$d_{TV}(P,Q) = \frac{1}{2}\sum_{x^n \in \mathcal{X}^n} \left| P(x^n) - Q(x^n) \right| = \frac{1}{2}\sum_{x^n \in \mathcal{X}^n} \left| \sum_{x^{1:n-1} \in \mathcal{X}^{n-1}} P(x^{1:n} | x^{1:n-1}) P(x^{1:n-1}) - Q(x^{1:n}|x^{1:n-1}) Q(x^{1:n-1}) \right|$$

$$= \frac{1}{2}\sum_{x^n \in \mathcal{X}^n} \left| \sum_{x^{1:n-1} \in \mathcal{X}^{n-1}} P(x^{1:n}|x^{1:n-1}) P(x^{1:n-1}) - P(x^{1:n-1}) Q(x^n|x^{1:n-1}) + P(x^{1:n-1}) Q(x^n|x^{1:n-1}) - Q(x^{1:n}|x^{1:n-1}) Q(x^{1:n-1}) \right|$$

by Triangle Inequality

$$\le \frac{1}{2}\sum_{x^n \in \mathcal{X}^n} \left(\left| \sum_{x^{n-1} \in \mathcal{X}^{n-1}} P(x^{n-1}) \left(P(x^n|x^{n-1}) - Q(x^n|x^{n-1}) \right) \right| + \left| \sum_{x^{n-1} \in \mathcal{X}^{n-1}} Q(x^n|x^{n-1}) \left(P(x_{n-1}) - Q(x_{n-1}) \right) \right| \right)$$

$$= \frac{1}{2}\sum_{x^n \in \mathcal{X}^n} \sum_{x^{n-1} \in \mathcal{X}^{n-1}} P(x^{n-1}) \left| P(x^n|x^{n-1}) - Q(x^n|x^{n-1}) \right| + \frac{1}{2}\sum_{x^n \in \mathcal{X}^n} \sum_{x^{n-1} \in \mathcal{X}^{n-1}} Q(x^n|x^{n-1}) \left| P(x^{n-1}) - Q(x^{n-1}) \right|$$

$1 = \sum Q(\cdot|\cdot)$ [illegible]

$$= \mathbb{E}_{x^{n-1} \sim P} \left[\left\| P(\cdot|x^{n-1}) - Q(\cdot|x^{n-1}) \right\|_{TV} \right] + \frac{1}{2}\sum_{x^{n-1} \in \mathcal{X}^{n-1}} \left| P(x^{n-1}) - Q(x^{n-1}) \right|$$

$$= \mathbb{E}_{x^{n-1} \sim P} \left[\left\| P(\cdot|x^{n-1}) - Q(\cdot|x^{n-1}) \right\|_{TV} \right] + \Big\| \underbrace{P(x^{n-1}) - Q(x^{n-1})}_{\to P^{\otimes n-1} - Q^{\otimes n-1}} \Big\|_{TV}$$

So now we proved that: $\|P^n - Q^n\|_{TV} \le \|P^{n-1}_{(\cdot)} - Q^{n-1}_{(\cdot)}\|_{TV} + \mathbb{E}_{x^{n-1} \sim P^{n-1}} \left[\|P(\cdot|x^{n-1}) - Q(\cdot|x^{n-1})\|_{TV} \right]$

by applying this n times we can deduce that:

$$\|P - Q\|_{TV} = d_{TV}(P,Q) \le \sum_{i=1}^{n} \mathbb{E}_{x^{1:i-1} \sim P^{i-1}} \left[\| P_i(\cdot|X_{1:i-1}) - Q_i(\cdot|X_{1:i-1}) \|_{TV} \right] \quad \blacksquare \ \text{Q.E.D}$$

II $\quad \|P - Q\|_{TV} = \frac{1}{2}\sum_{x \in \mathcal{X}} |P(x) - Q(x)|$

$$\left.\begin{aligned} 1 + TV(P,Q) &= \sum_{x \in \mathcal{X}} \max(P(x), q(x)) \\ 1 - TV(P,Q) &= \sum_{x \in \mathcal{X}} \min(P(x), q(x)) \end{aligned}\right\} \; 1 - TV^2(P,Q) = \left(\sum_{x \in \mathcal{X}} \max\{P(x), q(x)\} \right) \left(\sum_{x \in \mathcal{X}} \min\{P(x), q(x)\} \right)$$

By Cauchy Schwartz $\Rightarrow$

$$1 - TV^2(P,Q) \ge \left(\sum_{x \in \mathcal{X}} \sqrt{\underbrace{\min(P_x, q_x) \cdot \max(P_x, q_x)}_{P_x \cdot q_x}} \right)^2 = BC^2(P,Q) \Rightarrow \begin{cases} 1 - TV^2 \ge BC^2 \\ TV \le \sqrt{1 - BC^2} \end{cases}$$

1

now using Jensen Inequality we obtain an upper bound for $BC(P,Q)$:

$$BC(P,Q) = \mathbb{E}_{X\sim Q}\left[\sqrt{\frac{P(x)}{q(x)}}\right] = \mathbb{E}\left[e^{\log\sqrt{\frac{P(x)}{Q(x)}}}\right] \geq e^{\frac{1}{2}\mathbb{E}_{x\sim Q}\log\left[\frac{P(x)}{Q(x)}\right]} = e^{-\frac{1}{2}D_{KL}(P,Q)}$$

$$\overset{(*)}{\Rightarrow} BC(P,Q) \geq e^{-1/2\, D_{KL}(P,Q)} \overset{(**)}{\Longrightarrow} TV \leq \sqrt{1-BC^2} \leq \sqrt{1-e^{-D_{KL}(P,Q)}}$$

Now for the last part we take advantage of convex functions properties:

let $f(x) = \sqrt{1-x} \xrightarrow{cvx} f(x) \leq f(y) + (x-y)^T \nabla f(x) \overset{y=0}{\Longrightarrow} f(x) \leq f(0) + x f'(0) = 1 - \frac{x}{2}$

$$x = e^{-D_{KL}(P,Q)} \Rightarrow \sqrt{1-e^{-D_{KL}(P,Q)}} \leq 1 - \frac{1}{2}D_{KL}(P,Q)$$

$\Rightarrow$ Now we proved all the inequalities. Q.E.D

$$TV \leq \sqrt{1-BC^2} \leq \sqrt{1-e^{-D_{KL}(P,Q)}} \leq 1 - \frac{1}{2}D_{KL}(P,Q)$$

(III) since $BC(P_n^{\otimes n}, Q_n^{\otimes n}) \underset{iid}{=} BC(P_n,Q_n)^n$ & $BC = 1 - H^2$ & the fact that

$1 - BC(P_n,Q_n)^n \leq TV(P_n^{\otimes n}, Q_n^{\otimes n}) \leq \sqrt{1 - BC(P_n,Q_n)^{2n}}$ (*). Let $a_n = n\, H^2(P_n,Q_n)$

, then $BC(P_n,Q_n) = 1 - \frac{a_n}{n}$ $\rightarrow$ plugging in (*) implies that:

$$1 - \left(1 - \frac{a_n}{n}\right)^n \leq TV(P_n^{\otimes n}, Q_n^{\otimes n}) \leq \sqrt{1 - \left(1-\frac{a_n}{n}\right)^{2n}}$$

since $\lim_{n\to\infty}\left(1+\frac{k}{n}\right)^n = e^{+k}$ we can deduce that:

$$\lim_{n\to\infty} 1 - \left(1-\frac{a_n}{n}\right)^n \leq \lim_{n\to\infty} TV(P_n^{\otimes n}, Q_n^{\otimes n}) \leq \lim_{n\to\infty}\sqrt{1-\left(1-\frac{a_n}{n}\right)^{2n}}$$

$$\lim_{n\to\infty} 1 - e^{-a_n} \leq \lim_{n\to\infty} TV(P_n^{\otimes n}, Q_n^{\otimes n}) \leq \lim_{n\to\infty}\sqrt{1-e^{-2a_n}}$$

now if $\begin{cases} \lim_{n\to\infty} a_n = 0, \text{ then } \lim_{n\to\infty} 1 - e^{-a_n} = 1 - 1 = 0 \\ \lim_{n\to\infty} a_n = \infty, \text{ then } \lim_{n\to\infty} 1 - e^{-a_n} = 1 - 0 = 1 \end{cases}$ (**)

if $\begin{cases} \lim_{n\to\infty} a_n = 0, \text{ then } \lim_{n\to\infty}\sqrt{1-e^{-2a_n}} = 0 \\ \lim_{n\to\infty} a_n = \infty, \text{ then } \lim_{n\to\infty}\sqrt{1-e^{-2a_n}} = 1 \end{cases}$ (***)

augmenting (***), (****) results in
طبق تعریف سامپتیک (اسم انگلیسیش را ندانستم) →

$$\begin{cases} TV\left(P_n^{\otimes n}, Q_n^{\otimes n}\right) = 0 & \text{if } a_n \to 0 \quad (\mathrm{I}) \\ TV\left(P_n^{a_n}, Q_n^{a_n}\right) = 1 & \text{if } a_n \to \infty \quad (\mathrm{II}) \end{cases}$$

ادامه سوال 1:

$(\mathrm{I})\ a_n \to 0 \Rightarrow H^2 = \frac{a_n}{n} \Rightarrow H^2 = o\left(\frac{1}{n}\right)$

$(\mathrm{II})\ a_n \to \infty \Rightarrow H^2 = \frac{a_n}{n} \Rightarrow H^2 = \Omega\left(\frac{1}{n}\right)$

So at last we can infer that:

$$\begin{cases} a_n \to 0 \Rightarrow \left(d_{TV}\left(P_n^{a_n}, Q_n^{a_n}\right) \to 0\right) \Longleftrightarrow \left(D_{H^2}(P_n, Q_n) = o\left(\frac{1}{n}\right)\right) \\ a_n \to \infty \Rightarrow \left(d_{TV}\left(P_n^{a_n}, Q_n^{a_n}\right) \to 1\right) \Longleftrightarrow \left(D_{H^2}(P_n, Q_n) = \Omega\left(\frac{1}{n}\right)\right) \end{cases}$$

IV Since $\int_{x \in \mathcal{X}} q(x)\,dx = 1 \Rightarrow \frac{1}{2}\sqrt{\int_{\mathcal{X}} \frac{(P_1(x) - P_2(x))^2}{q(x)}\,dx} = \frac{1}{2}\sqrt{\int_{\mathcal{X}} \frac{(P_1(x) - P_2(x))^2}{q(x)}\,dx}\sqrt{\int_{\mathcal{X}} q(x)\,dx}$

$\forall q \in \mathcal{P}$

$$\Rightarrow \frac{1}{2}\sqrt{\int_{\mathcal{X}} \frac{(P_1(x) - P_2(x))^2}{q(x)}\,dx} \underset{\text{Cauchy Shwartz}}{\geq} \frac{1}{2}\int_{\mathcal{X}} \sqrt{(P_1(x) - P_2(x))^2}\,dx = \int_{\mathcal{X}} \frac{1}{2}|P_1(x) - P_2(x)|\,dx = TV(P_1, P_2)$$

Now we will investigate the case of equality here, we now that equality holds in Cauchy-Shwartz when all of the terms have the same ration.

$$\overset{\forall x}{\Longrightarrow} \frac{|P_1(x) - P_2(x)|}{q(x)} = c \xrightarrow{\int q = 1} q(x) = \frac{|P_1(x) - P_2(x)|}{2\,TV(P_1, P_2)} \quad \blacksquare$$

So we can now deduce that

$$TV(P, Q) = \frac{1}{2} \inf_{q \in \mathcal{P}} \sqrt{\int_{\mathcal{X}} \frac{(P_1(x) - P_2(x))^2}{q(x)}\,dx}$$

3

V $\begin{matrix} 0 \searrow \\ 1 \nearrow \end{matrix} X \xrightarrow{P_{Y|X}} Y \begin{matrix} \nearrow P_0 : P_{Y|X=0} \\ \searrow P_1 : P_{Y|X=1} \end{matrix}$

let $\begin{cases} \mathbb{P}[X = 0] = p \\ P_Y = p P_0 + \bar{p} P_1 \end{cases}$ Bayes

$$I(X;Y) = \mathbb{E}_X\left[D_{KL}\left(P_{Y|X} \| P_Y\right)\right] = p\, D_{KL}(P_0 \| P_Y) + \bar{p}\, D_{KL}(P_1 \| P_Y) \quad (*)$$

now we'll use Pinsker to relate it to TV

$$\begin{cases} TV(P_0 \| P_Y) = TV(P_0 \| pP_0 + \bar{p}P_1) = \frac{1}{2}\sum_{x\in\mathcal{X}} \Big| \underbrace{P_0(x) - pP_0(x)}_{\bar{p}P_0(x)} - \bar{p}P_1(x) \Big| = \bar{p}\, TV(P_0 \| P_1) \\ TV(P_1 \| P_Y) = TV(P_1 \| pP_0 + \bar{p}P_1) = \frac{1}{2}\sum_{x\in\mathcal{X}} \Big| P_1(x) - pP_0(x) - \bar{p}P_1(x) \Big| = p\, TV(P_0 \| P_1) \end{cases} \quad (**)$$

now Pinsker tells us that. $D_{KL}(P, Q) \geq 2TV^2(P, Q)$, thus:

$$\xrightarrow{(*)} I(X;Y) = p\, D_{KL}(P_0 \| P_Y) + \bar{p}\, D_{KL}(P_1 \| P_Y) \geq 2p\, TV^2(P_0 \| P_Y) + 2\bar{p}\, TV^2(P_1 \| P_Y)$$

$$\xRightarrow{(**)} I(X;Y) \geq 2p\bar{p}^2 TV^2(P_0, P_1) + 2\bar{p}p^2 TV^2(P_0, P_1) = 2p\bar{p}\, TV^2(P_0, P_1)\, \underbrace{(p + \bar{p})}_{1}$$

now if we have a BSC $\rightarrow p = \frac{1}{2}$ $\Rightarrow$ $\boxed{I(X;Y) \geq \frac{1}{2} TV^2(P_0, P_1)}$ ■

4 Now for the other side of inequality

اولین سری IT

ادامه (V)

Now we'll prove the other side of inequality which is $I(X;Y) \leq d_{tv}(P_0 \| P_1)$

since $\delta_{KL} \leq \delta_{\chi^2} \rightarrow \chi^2(P\|Q) \geq D_{KL}(P\|Q)$, therefore:

$$\begin{cases} \chi^2(P_0\|P_Y) = \sum_{x\in\mathcal{X}} \dfrac{(P_0\bar{p} + P_1\bar{p})^2}{pP_0+\bar{p}P_1} = \bar{p}^2 \sum_{x\in\mathcal{X}} \dfrac{(P_0-P_1)^2}{P_1+(P_0-P_1)p} \\ \chi^2(P_1\|P_Y) = \sum_{x\in\mathcal{X}} \dfrac{(p\cdot P_0 - p\cdot P_1)^2}{p\cdot P_0 + \bar{p}P_1} = p^2 \sum_{x\in\mathcal{X}} \dfrac{(P_0-P_1)^2}{p\cdot P_0+\bar{p}P_1} \end{cases}$$

So $I(X;Y) \leq D_{KL}(P_0\|P_1) \leq p\cdot D_{KL}(P_0\|P_Y) + \bar{p}\, D_{KL}(P_1\|P_Y)$
$$\leq p\,\chi^2(P_0\|P_Y) + \bar{p}\,\chi^2(P_1\|P_Y)$$

which implies (since $p\bar{p}^2 + \bar{p}p^2 = p\bar{p}(p+\bar{p}) = p\bar{p}$):

$$I(X;Y) \leq p\bar{p} \sum_{x\in\mathcal{X}} \frac{(P_0-P_1)^2}{p\cdot P_0+\bar{p}P_1} = p\bar{p}\left[\sum_{\substack{x\in\mathcal{X} \\ P_0(x)>P_1(x)}} \cdots + \sum_{\substack{x\in\mathcal{X} \\ P_0(x)\leq P_1(x)}} \cdots \right]$$

$$(*) \quad \leq p\bar{p}\left[\frac{1}{p} \sum_{P_0>P_1} (P_0-P_1) + \frac{1}{\bar{p}} \sum_{P_1>P_0} (P_1-P_0) \right]$$

Since
$$\begin{cases} \dfrac{(P_0-P_1)^2}{P_1+(P_0-P_1)p} \overset{P_1\geq 0}{\leq} \dfrac{(P_0-P_1)^2}{p\cdot(P_0-P_1)} = \dfrac{P_0-P_1}{p} & (*) \\ \dfrac{(P_0-P_1)^2}{P_0+(P_1-P_0)\bar{p}} \leq \dfrac{(P_1-P_0)^2}{\bar{p}(P_1-P_0)} = \dfrac{P_1-P_0}{\bar{p}} & (*) \end{cases}$$

$$\Rightarrow I(X;Y) \leq p\sum_{P_0>P_1} |P_0-P_1| + \bar{p}\sum_{P_0<P_1} |P_1-P_0| = p\, d_{tv}(P_0\|P_1) + \bar{p}\, d_{tv}(P_0\|P_1)$$
$$= d_{tv}(P_0\|P_1)$$

So finally we got the proof:

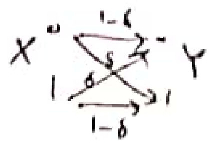
$$\frac{1}{2} d_{tv}^2(P_0\|P_1) \leq I(X;Y) \leq d_{tv}(P_0\|P_1) \qquad \text{Q.E.D} \blacksquare$$

$= 20\text{b}$

$G = (V, \mathcal{E})$
vertices $\leftarrow$ $\leftarrow$ edges

(i) $\{X_v ; v \in V\} \overset{iid}{\sim} \mathrm{Ber}(1/2)$
$\{Z_e ; e = (u,v) \in \mathcal{E}\} \overset{iid}{\sim} \mathrm{Ber}(\delta)$
$e = (u,v) : Y_e = X_u \oplus X_v \oplus Z_e$

(ii) [Contraction coefficient for BSC $\rightarrow \eta_{P_{Y|X}} = (1-2\delta)^2$



(iii) $U - X - Y \Rightarrow I(U;Y) \le I(U;X)$
Markov kernel} DPI

$\eta_{P_{Y|X}} = \sup_{P_{Y|X}} \frac{I(U;Y)}{I(U;X)}$. $I(U;Y) \le \eta_{P_{Y|X}} \cdot I(U;X)$

(I) $I(X_v ; Y_\mathcal{E}) = D_{KL}(P_{\mathcal{E}V} \| P_V P_\mathcal{E}) = \mathbb{E}_V\left[D_{KL}(P_{\mathcal{E}|V} \| P_\mathcal{E})\right] = 0$

$X_u \sim \mathrm{Ber}(1/2)$, $X_v \sim \mathrm{Ber}(1/2)$, $Z_e \sim \mathrm{Ber}(\delta)$ } $H_{u,e} = X_u \oplus Z_e \rightarrow \mathbb{P}[H_{u,e} = 1] = \delta \cdot \frac{1}{2} + \bar{\delta} \cdot \frac{1}{2} = \frac{1}{2}$

$H_{u,e,v} = H_{u,e} \oplus X_v \rightarrow \mathbb{P}[H_{u,e,v} = 1] = 1/2$

So we can see that regardless our choices of X_u, X_v; $Y_\mathcal{E} \sim \mathrm{Ber}(\frac{1}{2})$ & can't be determined by X_u, X_v (and generally by the set of $\{X_u ; u \in V\}$) so $Y_\mathcal{E}$ can't be determined given $\mathcal{E}$. ($Y_\mathcal{E}$ is independent of X_v) in other words.

(II) we will use induction on $|\mathcal{E}|$, we want to bound information by knowing X_S for some $S \subseteq V$ and $Y_\mathcal{E}$.

let us assume for the base case that $v \in S \subseteq V$; then $\mathbb{P}_G[v \to S] = 1$ & therefore $I(X_v ; X_S, Y_\mathcal{E}) \le H(X_v) \underbrace{\le}_{\text{Bernoulli}} \log 2 = \mathbb{P}_G[v \to S] \cdot \log 2$ ☑

from now on let $v \notin S$, $v \in V$. we can easily deduce that, the more v is closer to set S, the more information it'll provide and vise-versa. if there's no path from v to S then $\mathbb{P}[v \to S] = 0$ & $I(\cdot;\cdot) = 0$.

6

We start adding edges to see the fully-connected graph. It's obvious that adding an edge which connects two element of S will not provide us any info.
So let $\mathcal{E}^{(k)}$ be the set of all edges after adding the k^{th} edge. By conditioning on $\mathcal{E}^{(k)}$ for $\mathcal{E}^{(k+1)}$ we get: $\mathcal{E}^{(k+1)} = \{e\} \cup \mathcal{E}^{(k)}$

استدلال:

$$I(X_v; X_S, Y_{\mathcal{E}^{(k+1)}}) = I(X_v; X_S, Y_{\mathcal{E}^{(k)}}) + I(X_v; Y_{\mathcal{E}^{(k)}} | X_S, Y_{\mathcal{E}^{(k)}})$$

if $e \in S \times S \rightarrow I(X_v, Y_e | X_S, Y_{\mathcal{E}^{(k)}}) = 0$ since $Y_e | X_S \perp\!\!\!\perp X_v, Y_{\mathcal{E}^{(k)}}$

thus: $I(X_v, Y_e | X_S, Y_{\mathcal{E}^{(k)}}) = 0$

So ~~the~~ merely, in case where one of $e = (u,v)$ is at least outside of S, the information will leak, otherwise as we saw no information was ~~thought~~ get.

So by conditioning on $\mathcal{E}^{(k)}$ we get:

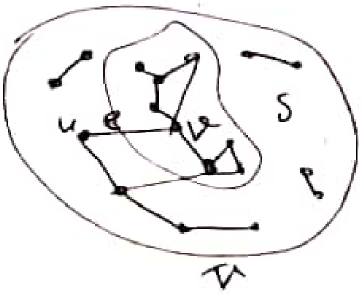
$$I(X_v, X_S, Y_{\mathcal{E}^{(k+1)}}) = I(X_v; X_S, Y_{\mathcal{E}^{(k)}}) + I(X_v; Y_e | X_S, Y_{\mathcal{E}^{(k)}})$$

& by contraction coefficient we can bound the second term:

$$I(X_v; Y_e | X_S, Y_{\mathcal{E}^{(k)}}) \leq \eta_k \cdot \underbrace{I(X_v; X_u | X_S, Y_{\mathcal{E}^{(k)}})}_{\leq H(X_v) = \log 2}$$

So by adding each node which is not currently in S,
(*1) we added $\eta \cdot \log 2$ information where $\eta = (1-2\delta)^2$

So now if we average over the set $\mathcal{E}$, we get



$$I(X_v; Y_e | X_S, Y_{\mathcal{E}^{(k)}}) = \mathbb{E}_{e \sim \mathcal{E}} \left[I(X_v, Y_e | X_S, Y_{\mathcal{E}^{(k)}}) \right] \qquad X_u \rightarrow Y_e \leftarrow X_S$$

7

$$= P[v \rightarrow u] \, I(X_v, Y_e | X_S, Y_{\mathcal{E}^k}, v \rightarrow u) + \mathbb{P}[v \not\rightarrow u] \underbrace{I(X_v; Y_e | X_S, Y_{\mathcal{E}^k}, v \not\rightarrow u)}_{0}$$

$$\leq \mathbb{P}[v \rightarrow u] \cdot \eta \log 2 = (\log 2)(1-2\delta)^2 \, \mathbb{P}[v \rightarrow u]$$

Now we'll take our inductive step:

(*)

$$\hookrightarrow I(X_v; X_S, Y_{\mathcal{E}^{(k+1)}}) \leq I(X_v; X_S, Y_{\mathcal{E}^{(k)}}) + \eta \, \mathbb{P}(v \rightarrow u) \log 2 \qquad (\text{***})$$

Scanned by CamScanner

since $\mathbb{P}_{G^{k+1},\eta}[v \to s] \geq \mathbb{P}\left[(v \to s)]_{G^k,\eta} \cup \left[(v \to u), (u \to s)\right)\right]$

the more general Case

$$= \mathbb{P}_{G^k,\eta}[v \to s] + \eta \, \mathbb{P}_{G^k,\eta}[v \to u] \qquad (**)$$

(**)

$$\Rightarrow \quad I(X_v; X_S, Y_{\mathcal{E}(k+1)}) \leq \left[\mathbb{P}_{G^k,\eta}[v \to s] + \eta \, \mathbb{P}_{G^k,\eta}[v \to u] \right] \log 2$$

$$\leq \mathbb{P}_{G^{k+1},\eta}[v \to s] \log 2$$

$$\Rightarrow \boxed{I(X_v; X_S, Y_{\mathcal{E}}) \leq \mathbb{P}_{G,\eta}[v \to s] \log 2}$$

using induction into further steps

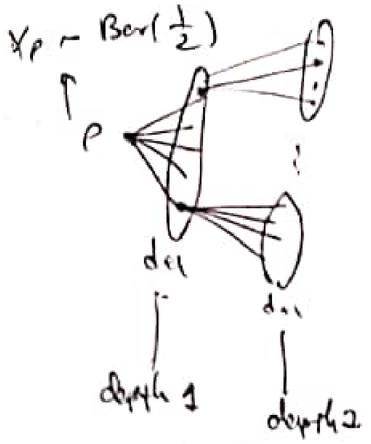
Q.E.D

III

ادامه‌ی سؤال 12
ادامه III

We know that in a tree $\mathcal{T}$ there's no loop & we have a root

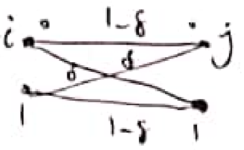
Since the tree is d-ary it looks like this:



$\forall i \in \mathcal{T} \rightarrow |N_i| = d+1$

$\forall i,j \in \mathcal{T}:\ N_i \cap N_j = \phi \rightarrow$ not to form any loop

each edge is a channel $BSC(\delta)$:



S_k is the set of nodes of $\mathcal{T}$ in depth k.

in this BSC channel we can see that for a connected $i \to j$ we have

$\mathbb{P}[X_i = X_j] = 1-\delta$ & $e = (i,j) \rightarrow X_i \oplus X_j = Z_e \sim Ber(\delta)$

$\Rightarrow Y_e = Z_e \oplus Z_e = 0 \quad \forall e \in \mathcal{E}$.

So in this case which we can see that Y_e is always zero so it has no uncertainty & we can drop it and nothing will happen

from previous sections $\rightarrow I(X_\rho ; X_{S_k}, \underbrace{Y_\mathcal{E}}_{\text{drop this}}) = I(X_\rho ; X_S) \le \underbrace{P_{(G,\eta)}[\rho \to S_k]}_{1} \log 2 = \log 2$ ↳ not useful

$|S_k| = (d+1)^k \longrightarrow d_{TV}(X_\rho, X_{S_k}) = \frac{1}{|S_k|} \sum_{i \in S_k} d_{TV}(X_\rho, X_i)$

let $V(\rho, i)$ be the set of all nodes connecting $\rho \to i$ in order. We have:

$d_{TV}(X_\rho, X_i) \le \sum_{j \in V(\rho, i)} d_{TV}(X_\rho, X_j)$

ideas: way to relate d_{TV} to I from both sides.

(I) let g to be a non-decreasing function, then

:3ذذ

$$\mathbb{P}\left[\frac{d\nu}{d\mu} > M\right] \le \mathbb{P}\left[g\left(\frac{d\nu}{d\mu}\right) \ge g(M)\right]$$

change of measure

$$\nu\left(\frac{d\nu(x)}{d\mu(x)} > M\right) = \mathbb{E}_\nu\left[\mathbb{1}\left[\nu(x) \ge M\mu(x)\right]\right] = \mathbb{E}_\mu\left[\frac{d\nu}{d\mu} \cdot \mathbb{1}\left\{\frac{d\nu}{d\mu} > M\right\}\right]$$

let $t > 0$

$$= \int_0^\infty \mu\left[\frac{d\nu(x)}{d\mu(x)} \cdot \mathbb{1}\left\{\frac{d\nu}{d\mu} > M\right\} > t\right] dt$$

using the lemma given in the group!

$$= \int_M^\infty \mu\left[\frac{d\nu}{d\mu}(x) > t\right] dt \le \int_M^\infty \mu\left[f\left(\frac{d\nu}{d\mu}(x)\right) > f(t)\right] dt$$

change of variable

$f(\infty) = \infty$

$$\Rightarrow \nu\left[\frac{d\nu}{d\mu}(x) > M\right] \le \int_M^\infty \mu\left[f\left(\frac{d\nu}{d\mu}\right) > f(t)\right] dt \le \int_{f(M)} \mu\left[f\left(\frac{d\nu}{d\mu}\right) > u\right] \frac{du}{f'(M)}$$

$u = f(M)$
$du = f'(M)\, dM$

since $f'(t) \ge f'(M)$

$D_f(\nu \| \mu)$

$$\nu\left[\frac{d\nu}{d\mu}(x) > M\right] = \frac{1}{f'(M)} \int_{f(M)}^\infty \mu\left[f\left(\frac{d\nu}{d\mu}(x)\right) > M\right] du = \boxed{\frac{\mathbb{E}_\mu\left[f\left(\frac{d\nu}{d\mu}\right)\right]}{f'(M)}}$$

■

Q.E.D

(II)

$\le \log\left(1 + \chi^2(P_{Y|\varepsilon} \| Q_Y)\right)$

$$D_{KL}(P_Y \| Q_Y) = D_{KL}(\bar{\delta} P_{Y|\varepsilon} + \delta P_{Y|\varepsilon^c} \| Q_Y) \underset{\text{Concave}}{\le} \bar{\delta} D_{KL}(P_{Y|\varepsilon}) + \delta D_{KL}(P_{Y|\varepsilon^c} \| Q_Y)$$

writing the KL divergence as a probability:

$\le \mathbb{E}_X\left[D_{KL}(P_{Y|X,\varepsilon^c} \| Q_Y)\right]$

$$P_{Y|X,\varepsilon^c}(y) = \frac{P_{Y,\varepsilon^c}(X|Y)}{\underset{\text{normalization}}{P(\varepsilon^c)}} = \frac{\mathbb{1}\{Y \in \varepsilon\}\, P_{Y|X}(y)}{\delta}$$

$$\rightarrow D_{KL}(P_{Y|X,\varepsilon^c} \| Q_Y) = \sum_{y \in \mathcal{Y}} \frac{\mathbb{1}(X \in \varepsilon)}{\delta} P_{Y|X}(y) \log\left(\frac{P_{Y|X}(y)}{\delta Q(y)}\right) = \mathbb{E}_{P_{Y|X}}\left[\frac{\mathbb{1}(y \in \varepsilon)}{\delta}(-\log\delta) + \dots\right]$$

$$\hookrightarrow D_{KL}(P_{Y|X,\varepsilon^c} \| Q_Y) \le \delta D_{KL}(P_{Y|\varepsilon^c} \| Q_Y) + \underbrace{\log\delta\, \mathbb{E}_{X,Y}\left[\frac{\mathbb{1}(y \in \varepsilon^c)}{\delta}\right]}_{\delta \log\frac{1}{\delta}} + \mathbb{E}_{XY}\left(\frac{\mathbb{1}(y \in \varepsilon^c)}{\log\left(\frac{P_{Y|X}}{Q_Y}\right)}\right)$$

Now we'll show that this second & third term below is $\sqrt{\delta \operatorname{Var}(\dots)}$

$$\mathbb{E}_{XY}\left[\mathbb{1}(y \in \varepsilon^c) \log\frac{P_{X|Y}(y)}{Q_Y(y)}\right] + \delta\, \mathbb{E}_X\left[D_{KL}(P_{Y|X} \| Q_Y)\right] - \mathbb{E}\left[\log\left(\frac{P_{Y|X}}{Q_Y}\right)^2\right]$$

9

Now using the Cauchy-Schwarz inequality we can deduce that the L.H.S. $\le \sqrt{\mathbb{E}\{\mathbb{1}(y \in \mathcal{E}^c)^2\}} \sqrt{\mathbb{E}\left[\log\left(\frac{P_{Y|X}}{Q_Y}\right) - \mathbb{E}\log\left(\frac{P_{Y|X}}{Q_Y}\right)\right]^2}$

$$= \sqrt{\delta} \cdot \sqrt{\mathrm{Var}\left(\log \frac{P_{Y|X}}{Q_Y}\right)} \quad (**)$$

now it implies that: (*+)

$$D_{KL}(P_Y \| Q_Y) \le \log(1+\chi^2(P_{Y|\mathcal{E}} \| Q_Y)) + \delta\log\delta + \delta\,\mathbb{E}_X\left[D_{KL}(P_{Y|X} \| Q_Y)\right] + \sqrt{\delta}\sqrt{\mathrm{Var}\left(\log\left(\frac{P_{Y|X}}{Q_Y}\right)\right)}$$

Q.E.D ■

(I) $X — Y = \sqrt{A}X + Z$, $Z \sim N(0,1)$

let $\ell(X,\hat{X}) = (X-\hat{X})^2$, $\hat{X} = f(y)$

It's obvious that $P_{Y|X} \sim N(\sqrt{A}X, 1)$

$$\ell(X,\hat{X}) = \mathbb{E}\left[(X-\hat{X})^2\right] = \mathbb{E}\left[(X - \mathbb{E}(X|Y) + \mathbb{E}(X|Y) - f(y))^2\right] =$$

$$= \mathbb{E}_{X,Y}\left[(X-\mathbb{E}[X|Y])^2\right] + \mathbb{E}_{X,Y}\left[(\mathbb{E}[X|Y] - f(Y))^2\right] + 2\mathbb{E}_{X,Y}\left[(X-\mathbb{E}[X|Y])(\mathbb{E}[X|Y] - f(Y))\right]$$

(1) $\mathbb{E}_{X,Y}\left[(X-\mathbb{E}X|Y)(\mathbb{E}X|Y - f(y))\right] = \mathbb{E}_Y\left[(\mathbb{E}(X|Y) - f(y))\,\underbrace{\mathbb{E}_{X|Y}(X - \mathbb{E}X|Y)}_{0}\right]$

$$\hookrightarrow \ell(X,\hat{X}) = \mathbb{E}\left[(X-\mathbb{E}X|Y)^2\right] + \mathbb{E}\left[(\mathbb{E}X|Y - f(y))^2\right]$$

$0 \leftarrow$ let to minimize

the first term is not up to us however, the second term can be optimized to zero. thus $f^*(y) = \mathbb{E}[X|Y] = \underset{f(y)}{\operatorname{argmin}}\ \ell(X, f(y))$.

(II) $Y = \sqrt{\delta}X + Z \rightarrow I(X;Y) = I(X;\sqrt{\delta}X+Z) = I(X; X+\sqrt{a}Z)$

$$I(X;Y) = \mathbb{E}_X\left[D_{KL}(P_{Y|X} \| P_Z)\right] - D_{KL}(P_Y \| P_Z)$$

$$= \mathbb{E}_X\left[D_{KL}\left(N(\sqrt{\delta}X,1), N(0,1)\right)\right] - D_{KL}(P_Y \| P_Z)$$

$\cdot\, O(\frac{\mu^2}{2})\delta$

$$= \mathbb{E}_X\left[\frac{\delta X^2}{2}\right] - D_{KL}(P_{\sqrt{\delta}X+Z} \| P_Z).$$

using taylor series at $\delta=0$

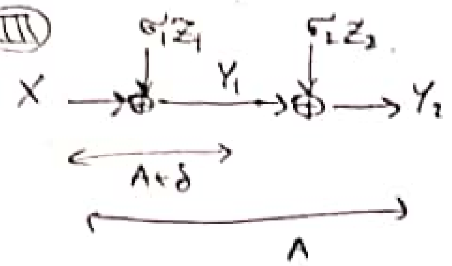
$$D_{KL}(P_{\sqrt{\delta}X+Z} \| P_Z) = \delta \frac{d}{d\delta} D_{KL}(P_{\sqrt{\delta}X+Z} \| P_Z)\Big|_{\delta=0} + 0 + o(\delta^2)$$

Since both $Z \sim N(0,1)$ & $\sqrt{\delta}X + Z \sim N(\sqrt{\delta}X, 1)$ are Normal distributions, we get

$$D_{KL}(P_{\sqrt{\delta}X+Z} \| P_Z) = \frac{1}{2}\mathbb{E}[\delta X^2] + o(\delta) = \frac{1}{2}\delta\,\mathbb{E}[X^2] + o(\delta)$$

$$\Rightarrow I(X;Y) = \frac{\delta}{2}\Big[\underbrace{\mathbb{E}[X^2] - \mathbb{E}[X]^2}_{Var(X) \longrightarrow \text{ equivalent to SNR}}\Big] + o(\delta) \qquad \text{Q.E.D} \ \blacksquare$$

(III)



$$M_E(A) = \mathbb{E}_{X,Y}\big[(X - \mathbb{E}[X|Y])^2\big]$$
$$= \mathbb{E}_{X,Z}\big[(X - \mathbb{E}[X|\sqrt{A}X + Z])^2\big]$$

Since the Markov kernel is of the form $X - Y_1 - Y_2$ & the fact that $X \perp\!\!\!\perp Y_2 | Y_1$ we can deduce that $I(X;Y_2|Y_1) = 0$, thus:

$$I(Y_1, Y_2; X) = I(X;Y_1) + \overbrace{I(X;Y_2|Y_1)}^{0} \xrightarrow{\text{conditioning on } Y_2} = I(X;Y_2) + I(X;Y_1|Y_2) \quad (*)$$

let $\left.\begin{array}{l} I(X;Y_1) = I(A+\delta) \\ I(X;Y_2) = I(A) \end{array}\right\} \underset{\text{assert}}{\Rightarrow} \frac{M_E(A)}{2} = \lim_{\delta \to 0} \frac{I(X;Y_1) - I(X;Y_2)}{\delta} = \lim_{\delta \to 0} \frac{I(A+\delta) - I(A)}{\delta}$

so it suffices to prove that ↑ to assume $I(X;Y_1) - I(X;Y_2) = \delta\left(\frac{1}{2}M_E(A) + o(1)\right)$

(IV) $Y_1 = X + \sigma_1 Z_1 = X + \dfrac{Z_1}{\sqrt{A+\delta}}$, $A = \dfrac{1}{\sigma_1^2 + \sigma_2^2}$ ← که دو گاوسی را یکی فرض می‌کند

$Y_2 = X + \sigma_1 Z_1 + \sigma_2 Z_2 = X + \dfrac{Z_3}{\sqrt{A}}$

$$(A+\delta)Y_1 = (A+\delta)X + \sqrt{A+\delta}\,Z_1 = \delta X + AX + \frac{Z_1}{\sigma_1} = \delta X + \frac{AY_2}{\sqrt{\sigma_1^2+\sigma_2^2}} \cdot \Bigg(\overset{\sim X}{\frac{\sigma_1 Z_1 + \sigma_2 Z_2}{\sqrt{\sigma_1^2+\sigma_2^2}}}\Bigg)$$

which results in $(A+\delta)Y_1 = AY_2 + \delta X + \dfrac{Z_1(\sigma_1^2+\sigma_2^2) - \sigma_1^2 Z_1 - \sigma_1\sigma_2 Z_2}{\sigma_1(\sigma_1^2+\sigma_2^2)} = AY_2 + \delta X + \underbrace{\dfrac{\sigma_2(\sigma_2 Z_1 - \sigma_1 Z_2)}{\sigma_1(\sigma_1^2+\sigma_2^2)}}_{\sim N(0,1)}$

so by $\left.\begin{array}{l} A+\delta = \frac{1}{\sigma_1^2} \\ A = \frac{1}{\sigma_1^2+\sigma_2^2} \end{array}\right\}$ δ will be $\Rightarrow \delta = \dfrac{\sigma_2^2}{\sigma_1^2(\sigma_1^2+\sigma_2^2)}$.

11

So we get $(A+\delta)Y_1 = AY_2 + \delta X + \sqrt{\delta} Z$, $Z \perp\!\!\!\perp X$, $Z \sim N(0,1)) = \frac{\sigma_1 Z_2 - \sigma_2 Z_1}{\sqrt{\sigma_1^2+\sigma_2^2}}$

so directly $\begin{cases} (A+\delta)Y_1 = AY_2 + \delta X + \sqrt{\delta} Z \\ Z \perp\!\!\!\perp X \sim N(0,1) \end{cases}$, اثبات ٤:

Q.E.D ■

(V) Now directly, we will show that $I(X;Y_1|Y_2) = \delta\left(\frac{1}{2} m_E(A) + o(1)\right)$

to do so: we have to bound $I(X;Y_1) - I(X;Y_2)$

$$I(X;Y_1) - I(X;Y_2) = I(X;Y_1|Y_2) = \mathbb{E}_{Y_2,X}\left[D_{KL}\left(P_{Y_1|Y_2,X} \| P_Z\right)\right]$$

$$- \mathbb{E}_{Y_2}\left[D_{KL}\left(P_{Y_1|Y_2} \| P_{\hat{Z}}\right)\right]$$

$$\left.\begin{array}{l} P_Z \to \hat{Z} = \dfrac{\sqrt{\delta} Z + \delta\, \mathbb{E}X|Y_2 + AY_2}{A+\delta} \\ Y_1 = Y_2 - \sigma_2 Z_2 \\ \sigma_2^2 = \dfrac{1}{A} - \dfrac{1}{A+\delta} \end{array}\right\}$$

so we will get:

they differ in X and $\mathbb{E}X|Y_2$:

$$I(X;Y_1|Y_2) = \mathbb{E}_{Y_2}\left[I(X;Y_1|Y_2)\right] = \mathbb{E}_{Y_2X}\left[D_{KL}\left(P_{\frac{AY_2+\sqrt{\delta}Z+\delta X}{A+\delta}} \| P_{\frac{AY_2+\sqrt{\delta}Z+\delta\,\mathbb{E}X|Y_2}{A+\delta}}\right)\right]$$

$$- \mathbb{E}_{Y_2}\left[D_{KL}\left(P_{\frac{AY_2+\sqrt{\delta}Z+\delta X}{A+\delta}|Y_2} \| P_{\frac{AY_2+\sqrt{\delta}Z+\mathbb{E}X|Y_2}{A+\delta}|Y_2}\right)\right]$$

So the first term will be:

$$\mathbb{E}_{Y_2/X}\left[\underbrace{\left(N\left(\frac{AY_2+\delta X}{A+\delta}, \frac{\delta}{(A+\delta)^2}\right) \| N\left(\frac{AY_2+\delta\,\mathbb{E}X|Y_2}{A+\delta}, \frac{\delta}{(A+\delta)^2}\right)\right)}_{D_{KL}}\right]$$

$$= \frac{(\mu_2-\mu_1)^2}{2} = \frac{\delta}{2} \cdot \frac{\mathbb{E}\left[(X - \mathbb{E}X|Y_2)^2\right]}{A+\delta} = \frac{\delta}{2} m_E(A)$$

and the second term will be $o(1)$, so it will imply that:

$$I(X;Y_1|Y_2) = \frac{\delta}{2} m_E(A) + \delta\, o(1)$$ which proves the whole theorem.

Q.E.D. ■

سؤال 5:

(I) $P_\pi = \int P_{\theta\pi}(d\theta) \qquad W(\theta,\hat\theta) = \mathbb{E}_{Q}\left[\frac{P_\theta P_{\hat\theta}}{q^2}\right]$

from the definition of χ^2 we can say that $\chi^2(P_\pi \| Q) = \mathbb{E}_{\theta,\hat\theta\sim\pi}\left[W(\theta,\hat\theta)\right] - 1$ which we'll prove!

$$\to \mathbb{E}_{\theta,\hat\theta\sim\pi}\left[W(\theta,\hat\theta)\right] = \mathbb{E}_{\theta,\hat\theta\sim\pi}\left[\mathbb{E}_{Q}\left[\frac{P_\theta P_{\hat\theta}}{q^2}\right]\right] = \mathbb{E}_{Q}\left[\mathbb{E}_{\theta\sim\pi}\left[\mathbb{E}_{\hat\theta\sim\pi}\frac{P_\theta P_{\hat\theta}}{q^2}\right]\right]$$

change the order of expected $\mathbb{E}$

Since $\mathbb{E}_{\theta}\left[P_\theta\right] = P_\pi = \mathbb{E}_{\hat\theta}\left[P_{\hat\theta}\right]$ we can deduce that:

$$\mathbb{E}_{\theta,\hat\theta\sim\pi}\left[W(\theta,\hat\theta)\right] = \mathbb{E}_{Q}\left[\frac{P_\pi^2}{q^2}\right] = -1+1+\mathbb{E}_{Q}\left[\frac{P_\pi^2}{q^2}\right] = 1 + \underbrace{\mathbb{E}_{Q}\left[\frac{P_\pi^2 - q^2}{q^2}\right]}_{\chi^2(P_\pi, q)}$$

so therefore: $\chi^2(P_\pi \| Q) = \mathbb{E}_{\theta,\hat\theta}\left[\frac{P_\theta \cdot P_{\hat\theta}}{q^2}\right] - 1$

(II) $P = \{P_n\}_{n=1}^{\infty}$

$Q = \{Q_n\}_{n=1}^{\infty} \qquad \chi^2(P_n \| Q_n) = o(1)$ سؤال 5:

$$\infty > \|L_n\|^2 = \mathbb{E}_{X\sim Q_n}\left[\frac{P_n(X)^2}{Q_n(X)^2}\right] = 1 + \chi^2(P_n \| Q_n) < \infty$$

So as we saw, we can deduce that if $\forall n: \chi^2(P_n\|Q_n) < \infty$ then $\chi^2(P_n\|Q_n) = o(1)$ which leads to $P \triangleleft Q$. Q.E.D ■

(I) $\sigma \in \{1,-1\}^n$, $A_{ij} = \begin{cases} P & \sigma_i = \sigma_j \\ Q & \sigma_i \neq \sigma_j \end{cases}$ → the Adjacency Matrix سؤال 6:

$A = (A_{ij})_{n\times n}$ $\quad P \to Ber(p)$, $Q \to Ber(q)$ $\longrightarrow G = G(\sigma, P, Q)$

Hypotheses test $\to \begin{cases} H_0: & G \overset{iid}{\sim} R_0 = G\left(n, \frac{P+Q}{2}\right) \\ H_1: & G \overset{iid}{\sim} R_1 = G(P,Q) \end{cases}$ $\longleftrightarrow$ all from $\frac{P+Q}{2}$

13

$$P(A) = \prod_{1\le i<j\le n} P(A_{ij})^{\mathbb{1}(\sigma_i=\sigma_j)} \cdot q(A_{ij})^{\mathbb{1}(\sigma_i\neq\sigma_j)}$$

the distribution of A matrix (which is symmetric)

Since $\sigma_i \in \{\pm 1\}$ we can deduce that for each Rademacher variable we have $\begin{cases} \mathbb{1}\{\sigma_i = \sigma_j\} = \frac{\sigma_i\sigma_j + 1}{2} \in \{0,1\} \\ \mathbb{1}\{\sigma_i \neq \sigma_j\} = \frac{-\sigma_i\sigma_j + 1}{2} \in \{0,1\} \end{cases}$

ادامه سوال 6:

$$P_{\sigma}(A) = \prod_{1\le i<j\le n} \sqrt{P(A_{ij})\, q(A_{ij}) \left(\frac{P(A_{ij})}{q(A_{ij})} \right)^{\sigma_i \sigma_j}}$$

now let $R(A)$ be the distribution for $R = \frac{P+Q}{2}$ distribution we have

$$R(A) = \prod_{1\le i<j\le n} \frac{P(A_{ij}) + q(A_{ij})}{2}$$

Now with the idea of Problem 5 to generate $\frac{P_\sigma P_{\hat\sigma}}{q^2}$ we get:

$$\frac{P_\sigma(A) P_{\hat\sigma}(A)}{R(A)^2} = \prod_{1\le i<j\le n} \frac{4P(A_{ij}) q(A_{ij})}{(P(A_{ij}) + q(A_{ij}))^2} \cdot \left(\frac{P(A_{ij})}{Q(A_{ij})} \right)^{\frac{\sigma_i\sigma_j + \hat\sigma_i \hat\sigma_j}{2}}$$

Now let $W(\sigma, \hat\sigma) = \mathbb{E}_{R}\left[\frac{P_\sigma(A) P_{\hat\sigma}(A)}{R^2(A)} \right] = \prod_{1\le i<j\le n} \mathbb{E}_{A_{ij}\sim R = \frac{P+Q}{2}} \left[\frac{P_\sigma(A) P_{\hat\sigma}(A)}{R(A)^2} \right]$

finally

$$\longrightarrow W(\sigma, \hat\sigma) = \prod_{1\le i<j\le n} \int_x \frac{2P(x) q(x)}{[P(x) + q(x)]} \left(\frac{P(x)}{q(x)} \right)^{\frac{\sigma_i\sigma_j + \hat\sigma_i\hat\sigma_j}{2}} dx$$

$$= \prod_{1\le i<j\le n} \int_x \frac{1}{(P(x) + q(x))} \times P(x)^{\mathbb{1}(\sigma_i = \sigma_j)} \cdot q(x)^{\mathbb{1}(\sigma_i \neq \sigma_j)} \cdot P(x)^{\mathbb{1}(\hat\sigma_i = \hat\sigma_j)} \cdot q(x)^{\mathbb{1}(\hat\sigma_i \neq \hat\sigma_j)} dx$$

Now that we got here we can rewrite $p^{0} \times q^{0}$ as the sum of the phrases

$$\Rightarrow P(x)^{\mathbb{1}(\sigma_i = \sigma_j)} \cdot q(x)^{\mathbb{1}(\sigma_i \neq \sigma_j)} = \mathbb{1}(\sigma_i = \sigma_j) P(x) + \mathbb{1}(\sigma_i \neq \sigma_j) q(x)$$

$$= \frac{P(x) + q(x)}{2} + \sigma_i\sigma_j \frac{P(x) - q(x)}{2}$$

thus:

$$\Rightarrow W(\sigma, \hat\sigma) = \prod_{1\le i<j\le n} \int_x \frac{2}{P(x) + q(x)} \times \left[\frac{P(x) + q(x)}{2} + \frac{\sigma_i\sigma_j (P(x) - q(x))}{2} \right] \times \left[\frac{P(x) + q(x)}{2} + \hat\sigma_i\hat\sigma_j \frac{(P(x) - q(x))}{2} \right] dx$$

14

$$\Rightarrow W(\sigma, \hat\sigma) = \prod_{1\le i<j\le n} \int_x \frac{\left(\frac{2}{P(x) + q(x)} \right)}{4} \left[(P(x) + q(x))^2 + \sigma_i\sigma_j \hat\sigma_i \hat\sigma_j (P(x) - q(x))^2 + (\sigma_i\sigma_j + \hat\sigma_i\hat\sigma_j)(P(x) + q(x))^* \right] dx$$

& after simplification we will get

الطلب الثاني:

$$W(\underline{\sigma},\hat{\sigma}) = \prod_{1\le i<j\le n} \frac{1}{2}\int_{\mathcal{X}} \left[P(x)+q(x)\right] + \left(\sigma_i\sigma_j + \hat{\sigma}_i\hat{\sigma}_j\right)\left(P(x)-q(x)\right) + \sigma_i\sigma_j\hat{\sigma}_i\hat{\sigma}_j \frac{\left(P(x)-q(x)\right)^2}{P(x)+q(x)}\,dx$$

now since we are asked/given $\rho = \int_{\mathcal{X}} \frac{(P(x)-q(x))^2}{2(P(x)+q(x))}\,dx$ we'd get:

$$W(\underline{\sigma},\hat{\sigma}) = \prod_{1\le i<j\le n}\left[\frac{1+1}{2} + \underbrace{\left(\frac{\sigma_i\sigma_j+\hat{\sigma}_i\hat{\sigma}_j}{2}\right)(1-1)}_{0} + \sigma_i\sigma_j\hat{\sigma}_i\hat{\sigma}_j\,\rho\right]$$

$$= \prod_{1\le i<j\le n}\left[1+\sigma_i\sigma_j\hat{\sigma}_i\hat{\sigma}_j\,\rho\right] \le \prod_{i<j} e^{\sigma_i\sigma_j\hat{\sigma}_i\hat{\sigma}_j\rho}$$

$$\longrightarrow W(\underline{\sigma},\hat{\underline{\sigma}}) = \exp\Big(\sum_{1\le i<j\le n}\hat{\sigma}_i\hat{\sigma}_j\sigma_i\sigma_j\,\rho\Big) = \exp\Big(\frac{\rho}{2}\sum_{i,j=1}^{n}\left(\sigma_i\sigma_j\hat{\sigma}_i\hat{\sigma}_j\right)\Big)$$

$$= \exp\Big(\frac{\rho}{2}\sum_{i=1}^{n}\sigma_i\sigma_j\Big(\sum_{j=1}^{n}\hat{\sigma}_i\hat{\sigma}_j\Big)\Big) = \exp\Big(\frac{\rho}{2}\langle\sigma\hat{\sigma}\rangle^2\Big)$$

Q.E.D ∎

(II) Now in this part we'll investigate if $P=\frac{a}{n}, q=\frac{b}{n}, \tau = \frac{(a-b)^2}{2(a+b)}$ what will be $\rho = \frac{\tau+o(1)}{n}$. So we compute ρ:

$$\begin{cases} P(x)\sim Ber(P) \\ Q(x)\sim Ber(q)\end{cases}$$

$$\rho = \frac{1}{2}\int_{\mathcal{X}}\frac{(P(x)-q(x))^2}{P(x)+q(x)}\,dx = \frac{(P-q)^2}{2(P+q)} + \frac{(\bar{P}-\bar{q})^2}{2(\bar{P}+\bar{q})} = \frac{(P-q)^2}{(P+q)(\bar{P}+\bar{q})}$$

which finally implies that: $\rho = \frac{(P-q)^2}{(P+q)(2-P-q)} \xrightarrow[q=\frac{b}{n}]{P=\frac{a}{n}}$

$$\Rightarrow \rho = \frac{\left(\frac{a-b}{n}\right)^2}{2\left(\frac{a+b}{n}\right)-\left(\frac{a+b}{n}\right)^2} = \frac{1}{n}\cdot\frac{(a-b)^2}{2(a+b)}\times\frac{1}{1-\frac{a+b}{2n}}$$

$o(1)+1$ هنا تحليل

$$= \frac{\tau}{n}\times\left[o(1)+1\right] = \rho \quad \text{Q.E.D}$$ ∎

15

(III) since $\sigma_i, \hat{\sigma}_i \sim$ Radmacher $\rightarrow$ $\sigma_i \sigma_j \sim$ Radmacher so by symmetry we expect their inner product to be zero $\rightarrow$ $\mathbb{E}[\langle \sigma, \hat{\sigma} \rangle] = 0$

so we want to bound $\chi^2(R_0, R_1)$ to see when we can't distinguish the parameters:

$$\chi^2(R_0 \| R_1) \leq \mathbb{E}_{\sigma, \hat{\sigma} \sim \text{Radmacher}}\left(\lim_{n \to \infty} \exp\left(\frac{\langle \sigma, \hat{\sigma} \rangle^2}{n} \cdot \frac{\tau + o(1)}{2} \right) - 1 \right)$$

let $H = \langle \sigma, \hat{\sigma} \rangle$ since $\mathbb{E} H = 0 \rightarrow \mathbb{E}[H^2] = \text{Var}(H) = S^2 = \overline{H^2}$

so by CLT we get that

$$\chi^2(R_0 \| R_1) \leq \mathbb{E}_{S \sim N(0,1)}\left[\exp\left(\frac{\tau}{2} \underbrace{S^2}_{\mathbb{E}[H^2]}\right) \right] - 1 = -1 + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\frac{\tau}{2} S^2 - \frac{S^2}{2}} ds$$

$$\chi^2(R_0 \| R_1) \leq -1 + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{S^2}{2}(1-\tau)} ds$$

$\rightarrow$ now it can easily be seen that if $1 - \tau \leq 0$ then the coeff of $-\frac{S^2}{2}$ in the exponent will be greater equal to zero & that makes the integral to diverge.

so we'll get: $\chi^2(R_0 \| R_1) \leq \begin{cases} -1 + \frac{1}{\sqrt{1-\tau}} & \tau < 1 \\ -1 + \infty & \tau \geq 1 \end{cases}$

So we now assume that $\tau < 1$

so $\chi^2(R_0 \| R_1) = \frac{1}{\sqrt{1-\tau}} - 1 = O(1)$ and we can distinguish $\chi^2(R_0 \| R_1)$ if $\chi^2 < \infty$.

However if $\tau \geq 1$ then $\chi^2(R_0 \| R_1)$ can be anything since $\chi^2 < \infty$ And we can't do anything in this case

$$t = \frac{(a-b)^2}{2(a+b)} \overset{!}{<} 1$$

$\underline{\underline{Q.E.D}}$

16

(I) As it was discussed in the class, all f-div's are convex. let =70/2
P_1, Q_1, P_2, Q_2 be some distributions when $\left(D_f(P_1,Q_1), D_g(P_1,Q_1)\right) \in \mathcal{R}$
$\left(D_f(P_2,Q_2), D_g(P_2,Q_2)\right) \in \mathcal{R}$: then: $\forall \lambda \in [0,1]$ any convex hull of
(P_1,P_2,Q_1,Q_2) such as : $(\lambda P_1 + \bar{\lambda} P_2)$ and $(\lambda Q_1 + \bar{\lambda} Q_2)$ are distributions;

thus:
$$\Rightarrow \left(D_f(\lambda P_1 + \bar{\lambda} P_2, \lambda Q_1 + \bar{\lambda} Q_2), D_g(\lambda P_1 + \bar{\lambda} P_2, \lambda Q_1 + \bar{\lambda} Q_2)\right) \in \mathcal{R}$$

we also have

$$D_f(\lambda P_1 + \bar{\lambda} P_2, \lambda Q_1 + \bar{\lambda} Q_2) \le \lambda D_f(P_1,Q_1) + \bar{\lambda} D_f(P_2,Q_2)$$
$$D_g(\lambda P_1 + \bar{\lambda} P_2, \lambda Q_1 + \bar{\lambda} Q_2) \le \lambda D_g(P_1,Q_1) + \bar{\lambda} D_g(P_2,Q_2)$$

so since between to points of distributions their convex hull will also be in the so called $\mathcal{R}$, we can prove that it's convex

(II) $$D_f(P\|Q) = \mathbb{E}_{x\sim Q}\left[f\left(\frac{dP}{dQ}\right)\right] = \mathbb{E}_{x\sim P}\left[\underbrace{\frac{dQ}{dP} f\left(\frac{dP}{dQ}\right)}_{\tilde{f}(t) = t f(\frac{1}{t})}\right] = \mathbb{E}_{x\sim P}\left[\tilde{f}\left(\frac{dQ}{dP}\right)\right] = D_{\tilde{f}}(Q\|P)$$

now let $Z\sim Q$, $M\sim R$, $H = \frac{P(Z)}{Q(Z)}$

$$D_f(P\|Q) = \sum_{x\in\mathcal{X}} q(x) f\left(\frac{P(x)}{q(x)}\right) = \sum_{x: q(x)=0} q(x) f\left(\frac{dP}{dQ}\right) + \sum_{x: q(x)\neq 0} q(x) f\left(\frac{dP}{dQ}\right)$$

$$= \underbrace{\sum_{q=0} \lim_{dq\to 0} dq\, f\left(\frac{dP}{dQ}\right)}_{} + \sum_{\substack{x\in\mathcal{X} \\ q(x)\neq 0}} q(x) f\left(\frac{dP}{dQ}\right)$$

by change of measure $f \to \tilde{f}$: $\tilde{f}(0) P(q(x)=0)$

$$= \tilde{f}(0)\left(1 - P(q(x)\neq 0)\right) = \tilde{f}(0)\left(1 - \sum_{x\in\mathcal{X}} q(x) \frac{P(x)}{q(x)}\right) = 1 - \underbrace{\mathbb{E}_{x\sim Q}\left[\frac{P(x)}{Q(x)}\right]}_{\mathbb{E}_R[H]}$$

$$\Rightarrow \text{As a result: } \sum_{\substack{x\in\mathcal{X} \\ q(x)\neq 0}} q(x) f\left(\frac{P(x)}{q(x)}\right) = \mathbb{E}_{X\sim R}\left[f(X)\right]$$

~~17~~

(II) الالم

Firstly, we can easily see that $\forall k \in \mathbb{N}$ $R_k \subseteq R_{k+1} \subseteq \mathbb{R}$ اداوي سوال 7:

So we can see that $\mathbb{P}(P \wedge Q) = \mathbb{E}_{X \sim R}\left[f(X)\right] + f(1,)\left[1 - \mathbb{E}_{X \sim R}[X]\right]$ holds for both R and R_4 since we didn't assume any binding assumptions.

$\longrightarrow$ Q.E.D ■

(III) Fenchel – Eggleston – Caratheodory

$S \subseteq \mathbb{R}^d$, $x \in Co(S)$, $\exists S' \in \mathbb{R}^{d+1}$, $S' = \{x_1, \ldots, x_{d+1}\}$, $S \rightarrow$ مترابط $x \in Co(S')$

$\hookrightarrow \exists\, \underline{X}, \underline{\alpha}$ such that $\begin{pmatrix} \mathbb{E}X \\ \mathbb{E}f(X) \\ \mathbb{E}g(X) \end{pmatrix} = \alpha^T\left(\underline{X}, f(\underline{X}), g(\underline{X})\right)$

$\alpha = (\alpha_1, \alpha_2, \alpha_3)^T$

$X = (X_1, X_2, X_3)^T$

since $\mathbb{E}X = \mathbb{E}X'$

$$\hookrightarrow \begin{cases} \mathbb{E}f(X) + (1 - \mathbb{E}X)\,\tilde{f}(1,) = \mathbb{E}f(X') + (-\mathbb{E}X' + 1)\,\tilde{f}(1,) \\ \mathbb{E}g(X) + (1 - \mathbb{E}X)\,\tilde{g}(1,) = \mathbb{E}g(X') + (-\mathbb{E}X' + 1)\,\tilde{g}(1,) \end{cases} \longrightarrow \in R_4$$

So

$\hookrightarrow \forall x \in R, \exists x' \in R_4 \rightarrow x = x' \rightarrow R = R_4 \longrightarrow$ Corollary $R_k \subset R_{k+1}$

IV since $\mathbb{E} f(X) = \sum_{x \in \mathcal{X}_k} P(x) f(x) = \sum_{i=1}^{k} P(x_i) f(x_i)$

اثبات سوال ۲:

Can be written as: $\sum_{i=1}^{k-1} P(x_i) f(x_i) + f(x_k) P(x_k)$

& let $q(x_i) = \frac{P(x_i)}{1-P(x_k)}$ for normalization, then:

$$\left[\sum_i q(x_i) = \frac{\sum P(x_i)}{1-P(x_k)} = \frac{1-P(x_k)}{1-P(x_k)} = 1 \right]$$

For each random variable X with its support vector with k elements:
$\exists X'$ a random variable with $(k-1)$ be its size of support vector & --

$$\xrightarrow{\&} \mathbb{E} f(X) = \alpha f(x_k) + (1-\alpha) \mathbb{E} f(X')$$

by convexity

so in this analogy X, X'' also take $(k-1)$ values:

$$\begin{pmatrix} \mathbb{E} f(X) + (1-\mathbb{E}X) \bar{f}(\cdot) \\ \mathbb{E} g(X) + (1-\mathbb{E}X) \bar{g}(\cdot) \end{pmatrix} = \alpha \begin{pmatrix} \mathbb{E} f(X') + (1-\mathbb{E}X') \bar{f}(\cdot) \\ \mathbb{E} g(X') + (1-\mathbb{E}X') \bar{g}(\cdot) \end{pmatrix}$$
$$+ \bar{\alpha} \begin{bmatrix} \mathbb{E} f(X'') + (1-\mathbb{E}X'') \bar{f}(\cdot) \\ \mathbb{E} g(X'') + (1-\mathbb{E}X'') \bar{g}(\cdot) \end{bmatrix}$$

$\rightarrow$ therefore each point in R_k was written by convex hull of points in R_{k-1} $\longrightarrow$ therefore $\forall k$: $R_k \subseteq R_{k-1}$

$R_k \subseteq \text{Convex hull}(R_{k-1})$

V So it was proven in the last part that $k=3$ will result in $R = R_4 = CH(R_3) = CH(CH(R_2)) = CH(R_1) = -$

Q.E.D ■

19